Bottleneck congestion and residential location of heterogeneous commuters

Takayama, Yuki and Kuwahara, Masao

1 March 2017

Bottleneck congestion and residential location of heterogeneous commuters*

Yuki Takayama† and Masao Kuwahara‡

March 1, 2017

Abstract

This study examines the effects of bottleneck congestion and an optimal time-varying congestion toll on the spatial structure of cities. We develop a model in which heterogeneous commuters choose departure times from home and residential locations in a monocentric city with a bottleneck located between a central downtown and an adjacent suburb. We then demonstrate that commuters sort themselves temporally and spatially on the basis of their value of travel time and their flexibility. Furthermore, we reveal that introducing an optimal congestion toll alters the urban spatial structure. We further demonstrate using an example that congestion tolling can lead to urban sprawl, which helps rich commuters but hurts poor commuters.

JEL classification: D62; R21; R41; R48

Keywords: bottleneck congestion; residential location; congestion toll; urban sprawl

*We are grateful to Daisuke Fukuda, Tomoru Hiramatsu, Yukihiro Kidokoro, and Tatsuhito Kono for helpful comments and discussions. This research was supported by JSPS Grant-in-Aid for Young Scientists (B) 15K18136.

†Corresponding author. Institute of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. Phone/Fax: +81-76-234-4915, E-mail: ytakayama@se.kanazawa-u.ac.jp

‡Graduate School of Information Sciences, Tohoku University.

1 Introduction

Traditional residential location models describe the spatial structures of cities and their evolution based on the trade-off between land rent and commuting costs (Alonso, 1964; Mills, 1967; Muth, 1969). Traditional and subsequent studies have successfully predicted the empirically observed patterns of residential location (e.g., spatial distribution of rich and poor) and the effects of assorted urban policies.[1] However, most previous studies describe traffic congestion using static flow congestion models. Their use renders these models inappropriate for dealing with peak-period traffic congestion and for examining the effects of measures intended to alleviate it (e.g., time-varying congestion tolls, flextime, staggered work hours).

The bottleneck model most successfully describes peak-period congestion and how commuters choose their departure times from home (Vickrey, 1969; Hendrickson and Kocur, 1981; Arnott et al., 1990b, 1993). Its simple, effective framework for studying the efficacy of various measures to alleviate peak-period congestion has inspired numerous extensions and modifications. However, only Arnott (1998), Gubins and Verhoef (2014), and Fosgerau et al. (2016) developed models to describe how commuters choose where they live and when they depart from home.

Arnott (1998) considered a (discrete space) monocentric closed city comprising two areas—a downtown and a suburb—connected by a single road with a bottleneck. He showed that imposing an optimal congestion toll without redistributing its revenues affects neither commuting costs nor commuters' residential locations. Gubins and Verhoef (2014) considered a (continuous space) monocentric closed city with a bottleneck at the entrance to its central business district (CBD). Their model introduced an incentive for commuters to spend time at home, which the standard bottleneck model disregards,[2] and assumed that a commuters' house size affects their marginal utility of spending time at home. They demonstrated that congestion tolling causes commuters to spend more time at home and to have larger houses, leading to urban sprawl. Fosgerau et al. (2016) developed a model similar to that of Gubins and Verhoef (2014). Unlike Gubins and Verhoef (2014), Fosgerau et al. (2016) did not introduce the assumption that the marginal utility of spending time at home depends on house size and considered a monocentric open city. They defined the social optimum as the global maximizer of total revenue from congestion tolling and land rents and showed that the optimal policy induces lower density in the center and higher density farther out.

The results obtained by Arnott (1998), Gubins and Verhoef (2014), and Fosgerau et al. (2016) differ fundamentally from the results given by traditional models with static flow congestion, which predict that cities become denser with congestion pricing (Kanemoto, 1980; Wheaton, 1998; Anas et al., 1998). Their models, however, assumed that commuters are homogeneous, although it has been established that optimal congestion tolling changes commuting costs in bottleneck models with heterogeneous commuters (Arnott et al., 1992, 1994; van den Berg and Verhoef, 2011). That is, the effects of congestion tolling in the bottleneck model with heterogeneous commuters can fundamentally differ from those in models with homogeneous commuters.

[1] Fujita (1989), Glaeser et al. (2008), Fujita and Thisse (2013), and Duranton and Puga (2015) provided detailed overviews of traditional residential location models. For the effects of congestion pricing, see Kanemoto (1980), Wheaton (1998), and Anas et al. (1998). Recently, Brueckner (2007), Anas and Rhee (2007), Joshi and Kono (2009), Kono et al. (2008); Kono and Joshi (2012), and Pines and Kono (2012) showed the efficiency of urban policies to substitute for congestion pricing (urban growth boundary, floor-to-area ratio regulations).

[2] Vickrey (1973), Tseng and Verhoef (2008), Fosgerau and Engelson (2011), Fosgerau and de Palma (2012), Fosgerau and Lindsey (2013), and Fosgerau and Small (2014) considered the utility of spending time at home.

This study extends the model developed by Arnott (1998) to consider commuter heterogeneity and a continuous space monocentric city with a bottleneck located between a central downtown and an adjacent suburb.[3] We systematically analyzed our model using the properties of complementarity problems that define the equilibrium.[4] Our analysis shows that commuters sort themselves both temporally and spatially on the basis of their value of time and their flexibility: commuters with a higher time-based cost per unit schedule delay (marginal schedule delay cost divided by marginal travel time cost) arrive at work earlier; commuters with a higher value of travel time and a higher marginal schedule delay cost live closer to their workplace.

This study also investigates the effects of an optimal time-varying congestion toll on the spatial distribution of commuters. We show that introducing a congestion toll (with and without redistributing its revenues) changes commuters' commuting costs, thereby altering their spatial distribution. To demonstrate the effects of congestion tolling concretely, we also analyzed the model for a case in which commuters with a high value of travel time have a higher time-based cost per unit schedule delay.[5] This analysis indicates that congestion tolling causes urban sprawl and induces higher density and land rent at suburban locations and lower density and land rent at downtown locations. This finding is not merely inconsistent with the standard results of traditional location models, but it also contradicts those of Arnott (1998). This implies that interactions among heterogeneous commuters change the effects of congestion tolling on urban spatial structure. We further show that, although the optimal congestion toll (without toll-revenue redistribution) generates a Pareto improvement in this case if commuters do not relocate (Arnott et al., 1994; Hall, 2015), it leads to an unbalanced distribution of benefits: commuters with a high value of time (rich commuters) gain, whereas those with a low value of time (poor commuters) lose from tolling.

This study proceeds as follows. Section 2 presents a model in which heterogeneous commuters choose their departure times from home and residential locations in a monocentric city. Sections 3 and 4 characterize equilibrium in our model without and with tolling using the properties of complementarity problems, respectively. Section 5 presents the effects of the optimal time-varying congestion toll. Section 6 concludes the study.

2 The model

2.1 Assumptions

We consider a long narrow city with a spaceless CBD, where all job opportunities are located. The CBD is located at the edge of the city, and a residential location is indexed by distance x from the CBD (Figure 1). Land is uniformly distributed with unit density along a road. As is common in the literature, the land is owned by absentee landlords.[6] The road has a single bottleneck with capacity μ at location $d > 0$. If arrival rates at the bottleneck exceed its capacity, a queue develops.

[3] We do not introduce the utility of spending time at home.

[4] As we show in Appendix A, the equilibrium defined by the complementarity problems coincides with that obtained by the traditional bit-rent approach (Alonso, 1964; Fujita, 1989; Duranton and Puga, 2015).

[5] This case includes situations analyzed in the literature employing bottleneck models and commuter heterogeneity (Arnott et al., 1992, 1994; van den Berg and Verhoef, 2011).

[6] We can make the alternative assumption that land is publicly owned and that the aggregate land rent is equally redistributed to all commuters. The results under this assumption are essentially identical to those obtained with absentee landlords since we assume that the utility function u is quasi-linear (i.e., income elasticity of demand for land is zero).

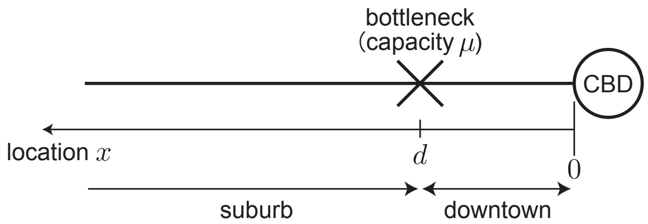

Figure 1: Long narrow city

To model queuing congestion, we employ first-in-first-out (FIFO) and a point queue, in which vehicles have no physical length as in standard bottleneck models (Vickrey, 1969; Arnott et al., 1993). Free-flow travel time per unit distance is assumed to be constant at $\tau > 0$ (i.e., free-flow speed is $1/\tau$).

There are I types of commuters, each of whom must travel from home to the CBD and who have the same preferred arrival time t^* at work. The number of commuters of type $i \in \mathcal{I} \equiv \{1, 2, \cdots, I\}$, whom we call "commuters i," is fixed and denoted by N_i. Since the bottleneck is located at d, only commuters who reside at $x > d$ pass through it, whereas those who reside at $x \in [0, d]$ do not. Following Arnott (1998), we denote locations $\mathcal{X}^s = \{x \in \mathbb{R}_+ \mid x > d\}$ as "suburb" and locations $\mathcal{X}^d = \{x \in \mathbb{R}_+ \mid x \in [0, d]\}$ as "downtown." We denote the number of commuters i in the suburb and downtown by N_i^s and $N_i^d (= N_i - N_i^s)$, respectively. If d is sufficiently large, all commuters reside downtown and no commuter traverses the bottleneck. Because we are uninterested in that case, d is assumed to be small, such that $\sum_{k \in \mathcal{I}} N_k^s > 0$.

Commuting cost $c_i(x, t)$ of commuters i who reside at x and arrive at work at time t is expressed as the sum of travel time cost $m_i(x, t)$ and schedule delay cost $s_i(t - t^*)$:

$$c_i(x, t) = m_i(x, t) + s_i(t - t^*), \tag{1a}$$

$$m_i(x, t) = \begin{cases} \alpha_i \tau x & \text{if} \quad x \in \mathcal{X}^d, \\ \alpha_i \{q(t) + \tau x\} & \text{if} \quad x \in \mathcal{X}^s, \end{cases} \tag{1b}$$

$$s_i(t - t^*) = \begin{cases} \beta_i (t^* - t) & \text{if} \quad t \leq t^*, \\ \gamma_i (t - t^*) & \text{if} \quad t \geq t^*, \end{cases} \tag{1c}$$

where $q(t)$ denotes the queuing time of commuters arriving at work at time t, τx represents free-flow travel time of commuters residing at x, and $\alpha_i > 0$ is the value of travel time of commuters i. $\beta_i > 0$ and $\gamma_i > 0$ are early and late delay costs per unit time, respectively. We assume $\alpha_i > \beta_i$ for all $i \in \mathcal{I}$ so that an equilibrium in our model satisfies the FIFO property (i.e., vehicles must leave the bottleneck in the same order as their arrival at the bottleneck). We also assume the value of travel time α_i of commuters i is positively correlated to their income w_i.

The utility of commuters i who reside at x and arrive at work at time t is given by the following quasi-linear utility function[7]:

$$u(z_i(x, t), a_i(x)) = z_i(x, t) + f(a_i(x)), \tag{2}$$

where $z_i(x, t)$ denotes consumption of the numéraire goods, $a_i(x)$ is the lot size at x, and $f(a_i(x))$

[7] As Arnott (1998) proved, if commuters are homogeneous, congestion tolling does not affect their spatial distribution under a quasi-linear utility function (2).

is the utility from land consumption. We assume $f(x)$ is a strictly increasing, concave, and twice differentiable function for $x > 0$. We also assume $\lim_{x\to 0+} f'(x) = \infty$. The logarithmic ($f(x) = \kappa \ln[a]$) and the hyperbolic ($f(x) = -\frac{\kappa}{2x}$) utility functions are examples satisfying these assumptions.[8] The budget constraint is expressed as

$$w_i = z_i(x,t) + \{r(x) + r_{\mathrm{A}}\}\, a_i(x) + c_i(x,t), \tag{3}$$

where $r_{\mathrm{A}} > 0$ is exogenous agricultural rent and $r(x) + r_{\mathrm{A}}$ denotes land rent at x.

The first-order condition of the utility maximization problem gives

$$\begin{cases} f'(a_i(x)) = r(x) + r_{\mathrm{A}} & \text{if} \quad a_i(x) > 0, \\ f'(a_i(x)) \leq r(x) + r_{\mathrm{A}} & \text{if} \quad a_i(x) = 0, \end{cases} \tag{4}$$

where the prime denotes differentiation. Since the marginal utility of land consumption is infinity at $a_i(x) = 0$, we must have $a_i(x) > 0$ and

$$a_i(x) = g(r(x) + r_{\mathrm{A}}), \tag{5}$$

where $g(\cdot)$ is the inverse function of $f'(\cdot)$. This implies that lot size $a_i(x)$ is independent of commuters' type i as well as commuting cost (and congestion toll levels) $c_i(x,t)$. Therefore, we denote lot size at x by $a(x)$.

From (2), (3), and (5), we obtain the indirect utility $v_i(x,t)$ as follows:

$$v_i(x,t) = w_i - c_i(x,t) + H(r(x) + r_{\mathrm{A}}), \tag{6}$$

where $H(r) = f(g(r)) - rg(r)$. Because $H(r(x) + r_A)$ can be rewritten as $f(a(x)) - \{r(x) + r_A\}a(x)$, this represents net utility from land consumption at x. Furthermore, since $H'(r(x)+r_A) = -g(r(x)+r_A) < 0$, $H(\cdot)$ is a strictly decreasing function.

2.2 Equilibrium conditions

Similar to models in Peer and Verhoef (2013), Gubins and Verhoef (2014), and Takayama (2015), we assume commuters make long-run decisions about residential location and short-run decisions about day-specific trip timing. In the short run, commuters i minimize commuting cost $c_i(x,t)$ by selecting their arrival time t at work taking their residential location x as given. In the long run, each commuter i chooses a residential location x so as to maximize his/her utility. We therefore formalize the short- and long-run equilibrium conditions.

2.2.1 Short-run equilibrium conditions

Commuters in the short run determine only their day-specific arrival time t at work, which implies that the number $N_i(x)$ of commuters i residing at x (i.e., spatial distribution of commuters) is assumed to be a given. Since commuting costs are given by (1), short-run equilibrium conditions differ according to commuters' residential locations. We first consider commuters residing in the suburb (suburban commuters), who must traverse the bottleneck. The commuting cost $c_i^s(x,t)$ of

[8] The same utility function has been introduced by, e.g., Blanchet et al. (2016) and Akamatsu et al. (2017).

suburban commuters i consists of a cost $\alpha_i \tau x$ of free-flow travel time depending only on residential location x and a bottleneck cost $c_i^b(t)$ owing to queuing congestion and a schedule delay depending only on arrival time t:

$$c_i^s(x,t) = \alpha_i \tau x + c_i^b(t), \tag{7a}$$

$$c_i^b(t) = \alpha_i q(t) + s_i(t - t^*). \tag{7b}$$

This implies that each suburban commuter chooses arrival time t so as to minimize bottleneck cost $c_i^b(t)$. Therefore, short-run equilibrium conditions coincide with those in the standard bottleneck model, which are given by three conditions:

$$\begin{cases} n_i^s(t)\left\{c_i^b(t) - c_i^{b*}\right\} = 0 \\ n_i^s(t) \geq 0,\ c_i^b(t) - c_i^{b*} \geq 0 \end{cases} \quad \forall i \in \mathcal{I}, \tag{8a}$$

$$\begin{cases} q(t)\left\{\mu - \sum_{k\in\mathcal{I}} n_k^s(t)\right\} = 0 \\ q(t) \geq 0,\ \mu - \sum_{k\in\mathcal{I}} n_k^s(t) \geq 0 \end{cases} \quad \forall t \in \mathbb{R}_+, \tag{8b}$$

$$\int n_i^s(t)\mathrm{d}t = N_i^s \qquad \forall i \in \mathcal{I}, \tag{8c}$$

where $n_i^s(t)$ denotes the number of suburban commuters i who arrive at work at time t (i.e., arrival rate of suburban commuters i at the CBD) and c_i^{b*} is the short-run equilibrium bottleneck cost of suburban commuters i.

Condition (8a) represents the no-arbitrage condition for the choice of arrival time. This condition means that, at the short-run equilibrium, no commuter can reduce the bottleneck cost by altering arrival time unilaterally. Condition (8b) is the capacity constraint of the bottleneck, which requires that the total departure rate $\sum_{k\in\mathcal{I}} n_k^s(t)$ at the bottleneck[9] equals capacity μ if there is a queue; otherwise, the total departure rate is (weakly) lower than μ. Condition (8c) is flow conservation for commuting demand. These conditions give $n_i^s(t)$, $q(t)$, and c_i^{b*} at the short-run equilibrium as functions of the number $\boldsymbol{N}^s = [N_i^s]$ of suburban commuters $i \in \mathcal{I}$. This implies that, at the short-run equilibrium, the bottleneck cost of suburban commuters i depends on $\boldsymbol{N}^s$ but not on $N(x)$.

We next consider commuters who reside downtown (downtown commuters). Since these commuters do not traverse the bottleneck, their commuting cost $c_i^d(x,t)$ is expressed as

$$c_i^d(x,t) = \alpha_i \tau x + s_i(t - t^*). \tag{9}$$

Thus, all downtown commuters will arrive at $t = t^*$, and their commuting cost at the short-run equilibrium is given by $\alpha_i \tau x$.

2.2.2 Long-run equilibrium conditions

In the long run, each commuter i chooses a residential location x so as to maximize indirect utility $v_i(x)$, which is expressed as

$$v_i(x) = y_i(x) + H(r(x) + r_A), \tag{10a}$$

[9]Note that the departure rate from the bottleneck coincides with the arrival rate of suburban commuters at the CBD.

$$y_i(x) = \begin{cases} w_i - \alpha_i \tau x - c_i^{b*}(\boldsymbol{N}^s) & \text{if} \quad x \in \mathcal{X}^s, \\ w_i - \alpha_i \tau x & \text{if} \quad x \in \mathcal{X}^d, \end{cases} \tag{10b}$$

where $y_i(x)$ denotes the income net of commuting cost earned by commuters i residing at x. Thus, long-run equilibrium conditions are given by

$$\begin{cases} N_i(x)\left\{v_i^* - v_i(x)\right\} = 0 \\ N_i(x) \geq 0,\ v_i^* - v_i(x) \geq 0 \end{cases} \qquad \forall x \in \mathbb{R}_+, \quad \forall i \in \mathcal{I}, \tag{11a}$$

$$\begin{cases} r(x)\left\{1 - \sum_{k\in\mathcal{I}} a(x)N_k(x)\right\} = 0 \\ r(x) \geq 0,\ 1 - \sum_{k\in\mathcal{I}} a(x)N_k(x) \geq 0 \end{cases} \qquad \forall x \in \mathbb{R}_+, \tag{11b}$$

$$\int_0^\infty N_i(x)\,\mathrm{d}x = N_i \qquad \forall i \in \mathcal{I}, \tag{11c}$$

where v_i^* denotes the long-run equilibrium utility of commuters i.

Condition (11a) is the equilibrium condition for commuters' choice of residential location. This condition implies that, at the long-run equilibrium, no commuter has incentive to change residential location unilaterally. Condition (11b) is the land market clearing condition. This condition requires that, if total land demand $\sum_{k\in\mathcal{I}} a(x)N_k(x)$ for housing at x equals supply 1, land rent $r(x) + r^A$ is (weakly) larger than agricultural rent r_A. Condition (11c) expresses the population constraint.

Note that *the traditional bid-rent approach (Alonso, 1964; Fujita, 1989; Duranton and Puga, 2015) is equivalent to our approach using complementarity problems*, as shown in Appendix A. More precisely, long-run equilibrium conditions (11) coincide with those of the bid-rent approach.[10] Therefore, even if we used the traditional bid-rent approach, we would obtain the same results as those presented in this study.

Substituting (5) into (11b), we obtain $r(x)$ as follows:

$$r(x) + r_\mathrm{A} = \begin{cases} f'(\frac{1}{N(x)}) & \text{if} \quad f'(\frac{1}{N(x)}) \geq r_\mathrm{A}, \\ r_\mathrm{A} & \text{if} \quad f'(\frac{1}{N(x)}) \leq r_\mathrm{A}, \end{cases} \tag{12}$$

where $N(x) = \sum_{k\in\mathcal{I}} N_k(x)$ represents the total number of commuters residing at x. It follows from (10a) and (12) that the equilibrium conditions in (11) are rewritten as (11a) and (11c) with (10a) and

$$v_i(x) = \begin{cases} y_i(x) + h(N(x)) & \text{if} \quad f'(\frac{1}{N(x)}) \geq r_\mathrm{A}, \\ y_i(x) + H(r_A) & \text{if} \quad f'(\frac{1}{N(x)}) \leq r_\mathrm{A}, \end{cases} \tag{13a}$$

$$h(N(x)) = H(f'(\tfrac{1}{N(x)})) = f(\tfrac{1}{N(x)}) - \tfrac{1}{N(x)} f'(\tfrac{1}{N(x)}), \tag{13b}$$

where $h(N(x)) = H(r(x)+r_A)$. Since $h'(N(x)) = \frac{1}{N(x)^3} f''(\frac{1}{N(x)}) < 0$, $h(\cdot)$ is a strictly decreasing function; that is, the net utility from land consumption decreases as the number of residents increases.

To study the spatial distribution of commuters, it is useful to rewrite the equilibrium conditions

[10] The equivalence between the bid-rent and our approaches has been shown in Fujita (1989, Chapter 2).

(11a) and (11c) as follows:

$$\begin{cases} N_i(x)\left\{v_i^{s*}(\boldsymbol{N}^s) - v_i^s(x)\right\} = 0 \\ N_i(x) \geq 0,\ v_i^{s*}(\boldsymbol{N}^s) - v_i^s(x) \geq 0 \end{cases} \qquad \forall x \in \mathcal{X}^s, \quad \forall i \in \mathcal{I}, \tag{14a}$$

$$\int_d^\infty N_i(x)\,\mathrm{d}x = N_i^s \qquad \forall i \in \mathcal{I}, \tag{14b}$$

$$\begin{cases} N_i(x)\left\{v_i^{d*}(\boldsymbol{N}^d) - v_i^d(x)\right\} = 0 \\ N_i(x) \geq 0,\ v_i^{d*}(\boldsymbol{N}^d) - v_i^d(x) \geq 0 \end{cases} \qquad \forall x \in \mathcal{X}^d, \quad \forall i \in \mathcal{I}, \tag{14c}$$

$$\int_0^d N_i(x)\,\mathrm{d}x = N_i^d \qquad \forall i \in \mathcal{I}, \tag{14d}$$

$$\begin{cases} v_i^{d*}(\boldsymbol{N}^d) \geq v_i^{s*}(\boldsymbol{N}^s) \quad \text{if} \quad N_i^d \geq 0 \\ v_i^{d*}(\boldsymbol{N}^d) \leq v_i^{s*}(\boldsymbol{N}^s) \quad \text{if} \quad N_i^s \geq 0 \end{cases} \qquad \forall i \in \mathcal{I}, \tag{14e}$$

$$N_i^d + N_i^s = N_i \qquad \forall i \in \mathcal{I}, \tag{14f}$$

where $v_i^{s*}(\boldsymbol{N}^s)$ and $v_i^{d*}(\boldsymbol{N}^d)$ denote the utilities that commuters i receive from residing in the suburb and downtown, respectively.

Conditions (14a) and (14b) are equilibrium conditions for suburban commuters' choice of residential location x. Similarly, conditions (14c) and (14d) are those for downtown commuters' choice of residential location x. Conditions (14e) and (14f) are equilibrium conditions for commuters' choice between residing in the suburb and downtown. We use these conditions to characterize the equilibrium spatial distribution of commuters in Section 3.

3 Equilibrium

3.1 Short-run equilibrium

The short-run equilibrium conditions (8) of suburban commuters coincide with those in the standard bottleneck model, as shown above. Therefore, we can invoke the results of studies utilizing the bottleneck model to characterize the short-run equilibrium (Arnott et al., 1994; Lindsey, 2004; Iryo and Yoshii, 2007; Liu et al., 2015). In particular, the following properties of the short-run equilibrium are useful for investigating the properties of our model.

Lemma 1 (Lindsey, 2004; Iryo and Yoshii, 2007)**.**

(i) The short-run equilibrium bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$ is uniquely determined.

(ii) The short-run equilibrium number $[n_i^{s*}(t)]$ of suburban commuters arriving at time t coincides with the solution of the following linear programming problem:

$$\min_{[n_i^s(t)]} \int \frac{s_i(t-t^*)}{\alpha_i} n_i^s(t)\,\mathrm{d}t \qquad s.t. \quad \mu - \sum_{k \in \mathcal{I}} n_k^s(t) \geq 0, \quad \int n_i^s(t)\mathrm{d}t = N_i^s. \tag{15}$$

Let us define (travel) time-based cost as the cost converted into equivalent travel time. Since that cost for commuters i is given by dividing the cost by α_i, we say that $s_i(t-t^*)/\alpha_i$ represents the time-based schedule delay cost of commuters i. Therefore, Lemma 1 (ii) shows that, at the

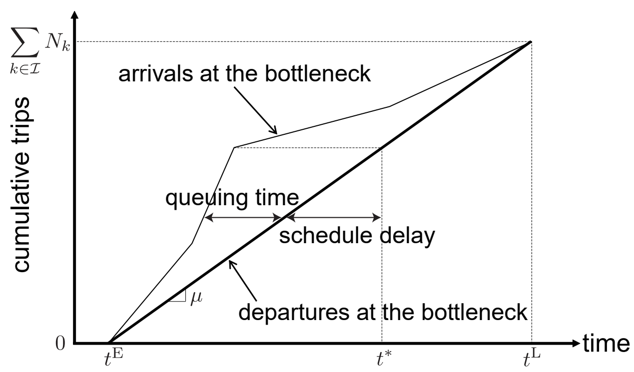

Figure 2: An example of cumulative arrival and departure curves at the short-run equilibrium

short-run equilibrium, the total *time-based* schedule delay cost is minimized, but the total schedule delay cost is not necessarily minimized.

We let $\mathrm{supp}\,(n_i^{s*}) = \{t \in \mathbb{R}_+ \mid n_i^{s*}(t) > 0\}$ be the support of the short-run equilibrium number $n_i^{s*}(t)$ of suburban commuters i who arrive at work at t. From Lemma 1 (ii), we then have

$$\mathrm{supp}\,(\textstyle\sum_{i\in\mathcal{I}} n_i^{s*}) = [t^{\mathrm{E}}, t^{\mathrm{L}}], \tag{16}$$

where t^{E} and t^{L} denote the earliest and latest arrival times of commuters, which satisfy

$$t^{\mathrm{L}} = t^{\mathrm{E}} + \frac{\sum_{i\in\mathcal{I}} N_i}{\mu}. \tag{17}$$

This indicates that, at the short-run equilibrium, a rush hour in which queuing congestion occurs must be a single time interval (Figure 2).

Furthermore, by using short-run equilibrium condition (8a), we obtain

$$c_i^b(t_i) + c_j^b(t_j) \leq c_i^b(t_j) + c_j^b(t_i) \qquad \forall t_i \in \mathrm{supp}\,(n_i^{s*}),\ t_j \in \mathrm{supp}\,(n_j^{s*}). \tag{18}$$

Substituting (7) into this, we have the following conditions as given in Arnott et al. (1994) and Liu et al. (2015): for any $t_i \in \mathrm{supp}\,(n_i^{s*})$, $t_j \in \mathrm{supp}\,(n_j^{s*})$, and $i, j \in \mathcal{I}$,

$$\left(\frac{\beta_i}{\alpha_i} - \frac{\beta_j}{\alpha_j}\right)\left(t_i - t_j\right) \geq 0 \qquad \text{if} \quad \max\{t_i, t_j\} \leq t^*, \tag{19a}$$

$$\left(\frac{\gamma_i}{\alpha_i} - \frac{\gamma_j}{\alpha_j}\right)\left(t_i - t_j\right) \leq 0 \qquad \text{if} \quad \min\{t_i, t_j\} \geq t^*. \tag{19b}$$

These conditions indicate that the short-run equilibrium has the following properties: if marginal time-based early delay cost of commuters i is lower than that of commuters j (i.e., $\beta_i/\alpha_i < \beta_j/\alpha_j$), early-arriving commuters i (commuters i arriving at the CBD before the preferred arrival time t^*) arrive at the CBD earlier than early-arriving commuters j; if the marginal time-based late delay cost of commuters i is lower than that of commuters j (i.e., $\gamma_i/\alpha_i < \gamma_j/\alpha_j$), late-arriving commuters i (commuters i arriving after t^*) arrive later than late-arriving commuters j. This occurs because commuters with a lower time-based schedule delay cost avoid queuing time rather than a schedule delay. This result shows that, *at the short-run equilibrium, commuters sort* *themselves temporally on the basis of their marginal time-based schedule delay cost.*

In the following analysis, we make the following assumption about marginal early and late delay costs, which is common to the literature employing a bottleneck model with commuter heterogeneity (Vickrey, 1973; Arnott et al., 1992, 1994; van den Berg and Verhoef, 2011; Hall, 2015).

Assumption 1. $\frac{\gamma_i}{\beta_i} = \eta$ *for all* $i \in \mathcal{I}$.

This assumption implies that commuters with a high early delay cost also have a higher late delay cost. Therefore, commuters are heterogeneous in two dimensions: the value of travel time α_i and marginal schedule delay cost β_i (γ_i).

For convenience, we call commuters with a high (low) value of travel time "*rich (poor)* commuters." We call commuters with a low (high) marginal schedule delay cost "*flexible (inflexible)* commuters." We can then say from (19) that richer or more flexible commuters prefer to arrive further from their preferred arrival time t^* to avoid queuing.

Under Assumption 1, we can explicitly obtain the short-run equilibrium bottleneck cost as a function of the number $\boldsymbol{N}^s = [N_i^s]$ of suburban commuters i. For the moment, we assume, without loss of generality, that commuters with small i have a (weakly) higher time-based schedule delay cost:

Assumption 2. $\frac{\beta_{i-1}}{\alpha_{i-1}} \geq \frac{\beta_i}{\alpha_i}$ *for any* $i \in \mathcal{I}\backslash\{1\}$.

Then, from (19), early-arriving (late-arriving) commuters with smaller i arrive later (earlier) at the short-run equilibrium. Under Assumptions 1 and 2, therefore, the short-run equilibrium bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$ of suburban commuters i is derived by following the procedure employed in literature featuring bottleneck models with commuter heterogeneity (see, e.g., van den Berg and Verhoef, 2011):

$$c_i^{b*}(\boldsymbol{N}^s) = \frac{\eta}{1+\eta}\left\{\beta_i \frac{\sum_{k=1}^{i} N_k^s}{\mu} + \alpha_i \sum_{k=i+1}^{I} \frac{\beta_k}{\alpha_k}\frac{N_k^s}{\mu}\right\} \qquad \forall i \in \mathcal{I}. \tag{20}$$

This clearly shows that richer or more inflexible commuters incur higher bottleneck costs at the short-run equilibrium.

Properties of the short-run equilibrium obtained above can be summarized as follows.

Proposition 1. *The short-run equilibrium has the following properties.*

(i) Total time-based schedule delay cost is minimized.

(ii) Early-arriving commuters arrive at work in order of increasing marginal time-based early delay cost (β_i/α_i). Late-arriving commuters arrive at work in order of decreasing marginal time-based late delay cost (γ_i/α_i).

(iii) The short-run equilibrium bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$ of commuters i is uniquely determined. Furthermore, if Assumptions 1 and 2 hold, $c_i^{b*}(\boldsymbol{N}^s)$ is given by (20).

3.2 Long-run equilibrium

3.2.1 Suburban and downtown spatial structures

We first examine the properties of suburban and downtown spatial structures at the long-run equilibrium using the properties of equilibrium conditions (14a), (14b), (14c), and (14d). We therefore consider in this subsection that the suburban and downtown populations are given. From equilibrium conditions (14a) and (14c), we see there is no vacant location between any two populated locations, as shown in Lemma 2.

Lemma 2. *The long-run equilibrium number $N^{*}(x)$ of commuters residing at x has the following properties:*

(i) the support of $N^{*}(x)$ is given by

$$\mathrm{supp}\,(N^{*}) = [0, X^{\mathrm{B}}], \tag{21}$$

where X^{B} denotes the residential location for commuters farthest from the CBD (city boundary).

(ii) $N^{*}(x)$ satisfies

$$f'(\tfrac{1}{N^{*}(x)}) > r^{\mathrm{A}} \qquad \forall x \in \mathrm{supp}\,(N^{*})\backslash\{X^{\mathrm{B}}\}, \tag{22a}$$

$$f'(\tfrac{1}{N^{*}(X^{\mathrm{B}})}) = r^{\mathrm{A}}. \tag{22b}$$

Proof. See Appendix B. □

Let $N_i^{s*}(x)$ and $N_i^{d*}(x)$ be the respective long-run equilibrium number of suburban and downtown commuters i residing at x. Then, it follows from Lemma 2 that, for any $x_i^s \in \mathrm{supp}\,(N_i^{s*})$ and $x_i^d \in \mathrm{supp}\,(N_i^{d*})$, indirect utilities $v_i^s(x_i^s)$ and $v_i^d(x_i^d)$ are expressed as

$$v_i^s(x_i^s) = w_i - c_i^{b*}(\boldsymbol{N}^s) - \alpha_i \tau x_i^s + h(N(x_i^s)), \tag{23a}$$

$$v_i^d(x_i^d) = w_i - \alpha_i \tau x_i^d + h(N(x_i^d)). \tag{23b}$$

In addition, equilibrium conditions (14a) and (14c) give the following conditions for $N_i^{s*}(x)$ and $N_i^{d*}(x)$:

$$v_i^s(x_i^s) + v_j^s(x_j^s) \geq v_i^s(x_j^s) + v_j^s(x_i^s) \qquad \forall x_i^s \in \mathrm{supp}\,(N_i^{s*}),\ \forall x_j^s \in \mathrm{supp}\,(N_j^{s*}),\ \forall i, j \in \mathcal{I}, \tag{24a}$$

$$v_i^d(x_i^d) + v_j^d(x_j^d) \geq v_i^d(x_j^d) + v_j^d(x_i^d) \qquad \forall x_i^d \in \mathrm{supp}\,(N_i^{d*}),\ \forall x_j^d \in \mathrm{supp}\,(N_j^{d*}),\ \forall i, j \in \mathcal{I}. \tag{24b}$$

Substituting (23) into (24) yields the following conditions: for any $x_i \in \mathrm{supp}\,(N_i^{s*})$, $x_j \in \mathrm{supp}\,(N_j^{s*})$, and $i, j \in \mathcal{I}$,

$$\left(\alpha_i - \alpha_j\right)\left(x_i - x_j\right) \leq 0. \tag{25}$$

This condition also holds for any $x_i \in \mathrm{supp}\,(N_i^{d*})$, $x_j \in \mathrm{supp}\,(N_j^{d*})$, and $i, j \in \mathcal{I}$.

This condition states that richer suburban commuters reside closer to the CBD to reduce their free-flow travel time cost. This property also holds for downtown commuters. This implies that *suburban and downtown commuters sort themselves spatially on the basis of their value of travel time.*

Furthermore, spatial distribution of suburban commuters and that of downtown commuters are unaffected by the short-run equilibrium bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$.

In our model, the free-flow travel time cost is more income elastic than the demand for land since we assume that the income elasticity of demand for land is zero. Therefore, this result is in accordance with the standard result of traditional location models, which show that rich commuters reside closer to the CBD if the income elasticity of commuting costs is larger than the income elasticity of demand for land.

By using condition (25), we can obtain the spatial distribution of commuters $N(x)$, land rent $r(x)$, and lot size $a(x)$. For this, we introduce the following assumption.[11]

Assumption 3. $\alpha_{i-1} > \alpha_i$ *for all* $i \in \mathcal{I} \backslash \{1\}$.

This means that commuters with small i are richer than those with large i.

Let X_i^s and X_i^d denote the respective locations for suburban and downtown commuters i residing nearest the CBD. It follows from (25) and Assumption 3 that suburban and downtown commuters i reside in $[X_i^s, X_{i+1}^s]$ and $[X_i^d, X_{i+1}^d]$, respectively (i.e., $\text{supp}\,(N_i^{s*}) = [X_i^s, X_{i+1}^s]$ and $\text{supp}\,(N_i^{d*}) = [X_i^d, X_{i+1}^d]$ for all $i \in \mathcal{I}$). Therefore, we have

$$v_i^s(x) = v_i^s(X_i^s) \qquad \forall x \in [X_i^s, X_{i+1}^s], \tag{26a}$$

$$v_i^d(x) = v_i^d(X_i^d) \qquad \forall x \in [X_i^d, X_{i+1}^d]. \tag{26b}$$

These, together with the population constraints (14b) and (14d), lead to the following lemma.

Lemma 3. *Suppose Assumption 3. Then the long-run equilibrium land rents at locations X_i^s and X_i^d are given by*

$$r(X_i^s) + r_A = r_A + \sum_{k=i}^{I} \alpha_k \tau N_k^s, \qquad r(X_i^d) + r_A = r(d) + r_A + \sum_{k=i}^{I} \alpha_k \tau N_k^d, \tag{27}$$

where $r(d) + r_A$ is the land rent at location d.

Proof. See Appendix C. □

Substituting (27) into (26), we obtain X_i^s and X_i^d as follows.

$$X_i^s = d + \sum_{j=1}^{i-1} \frac{1}{\alpha_j \tau} \left\{ H(r_A + \textstyle\sum_{k=j+1}^{I} \alpha_k \tau N_k^s) - H(r_A + \sum_{k=j}^{I} \alpha_k \tau N_k^s) \right\}, \tag{28a}$$

$$X_i^d = \sum_{j=1}^{i-1} \frac{1}{\alpha_j \tau} \left\{ H(r(d) + r_A + \textstyle\sum_{k=j+1}^{I} \alpha_k \tau N_k^d) - H(r(d) + r_A + \sum_{k=j}^{I} \alpha_k \tau N_k^d) \right\}. \tag{28b}$$

Recall that $H(r) = f(g(r)) - rg(r)$, and thus $h(N(X_i^s)) = H(r(X_i^s) + r_A)$. Land rent $r(d) + r_A$ at location d is obtained from the following condition:

$$\begin{cases} r(d)\left\{d - X_{I+1}^d\right\} = 0, \\ r(d) \geq 0, \quad d - X_{I+1}^d \geq 0. \end{cases} \tag{29}$$

[11] This assumption ensures that "no pair of bit rent functions intersects more than once" in the suburb and downtown when we use the bit-rent approach. As discussed in Fujita (1989, Chapter 4), this is a necessary condition for the uniqueness of the equilibrium in the traditional residential location model.

This condition means that, if $X^d_{I+1} < d$ for any $r(d) \in \mathbb{R}_+$, land rent at d equals the agricultural rent r_A (i.e., $r(d) = 0$); otherwise, $r(d)$ is determined such that $X^d_{I+1} = d$. $r(d)$ is uniquely determined by condition (29), because the following conditions hold:

$$\frac{\mathrm{d}X^d_{I+1}}{\mathrm{d}r(d)} = \sum_{k\in\mathcal{I}} \left\{ g(r(X^d_k) + r(d) + r_A) - g(r(X^d_{k+1}) + r(d) + r_A) \right\} < 0 \qquad \text{if} \quad \sum_{k\in\mathcal{I}} N^d_k > 0, \tag{30a}$$

$$\lim_{r(d)\to\infty} X^d_{I+1} = 0, \tag{30b}$$

where the second condition follows from $H'(\cdot) = -g(\cdot)$ and $\lim_{r\to\infty} g(r) = 0$ ($\lim_{a\to 0+} f'(a) = \infty$).

By using (26), (27), and (28), we obtain Lemma 4.

Lemma 4. *Suppose Assumption 3. Then at the long-run equilibrium,*

(i) the city boundary X^B is given by

$$X^B = d + \frac{1}{\tau}\left[\frac{1}{\alpha_I} H(r_A) - \sum_{k\in\mathcal{I}\setminus\{1\}} \left\{\frac{1}{\alpha_k} - \frac{1}{\alpha_{k-1}}\right\} H^s_k - \frac{1}{\alpha_1} H^s_1\right], \tag{31}$$

where $H^s_i = H(r_A + \sum_{k=i}^{I} \alpha_k \tau N^s_k)$;

(ii) the number $N(x)$ of commuters, lot size $a(x)$, and land rent $r(x) + r_A$ are given by

$$N(x) = \begin{cases} h^{-1}\left(\alpha_i \tau (x - d) + \alpha_i \sum_{k=2}^{i} \left\{\frac{1}{\alpha_k} - \frac{1}{\alpha_{k-1}}\right\} H^s_k + \frac{\alpha_i}{\alpha_1} H^s_1\right) & \text{if} \quad x \in [X^s_i, X^s_{i+1}], \\ h^{-1}\left(\alpha_i \tau x + \alpha_i \sum_{k=2}^{i} \left\{\frac{1}{\alpha_k} - \frac{1}{\alpha_{k-1}}\right\} H^d_k + \frac{\alpha_i}{\alpha_1} H^d_1\right) & \text{if} \quad x \in [X^d_i, X^d_{i+1}], \end{cases} \tag{32a}$$

$$a(x) = \frac{1}{N(x)} \qquad \forall x \in [0, X^B], \tag{32b}$$

$$r(x) + r_A = f'(a(x)) \qquad \forall x \in [0, X^B], \tag{32c}$$

where $h^{-1}(\cdot)$ is the inverse function of $h(\cdot)$ and $H^d_i = H(r(d) + r_A + \sum_{k=i}^{I} \alpha_k \tau N^d_k)$.

Proof. Since $X^B = X^s_{I+1}$ and $H^s_i = H(r(X^s_i) + r_A)$, we have (31). (32) is obtained from the straightforward calculation of (4), (12), and (26). □

It follows from this lemma that, for any $i, j \in \mathcal{I}$,

$$\frac{\partial X^B}{\partial N^s_i} > 0, \qquad \frac{\partial N(x)}{\partial N^s_i} \begin{cases} > 0 & \text{if} \quad x \in (X^s_j, X^s_{j+1}), \\ < 0 & \text{if} \quad x \in (X^d_j, X^d_{j+1}). \end{cases} \tag{33}$$

This shows that the city boundary moves outward as the suburban population increases. That is, a population increase in the suburb leads to urban sprawl. Furthermore, it induces higher density and land rent at any populated suburban location and lower density and land rent at any populated downtown location.

We see from this lemma that the long-run equilibrium spatial distribution of suburban commuters and that of downtown commuters are uniquely determined if X^s_i and X^d_i are finite. Therefore, we make the following assumption to ensure the uniqueness of the long-run equilibrium in the suburb and downtown.

Assumption 4. $H(r) < \infty$ *for any* $r \geq r_A$.

Indeed, since X_i^s and X_i^d are given by (28), they are finite under Assumption 4.

In addition, the long-run equilibrium conditions (14a), (14b), (14c), and (14d) are equivalent to Karush–Kuhn–Tucker (KKT) conditions of the following optimization problem:

Lemma 5. *The spatial distribution* $[N_i^*(x)]$ $(x \in \mathcal{X}^s)$ *of suburban commuters is a long-run equilibrium if and only if it is a KKT point of the following optimization problem:*

$$\max_{[N_i(x)]} \int_d^\infty \left[\sum_{k \in \mathcal{I}} \left\{ w_k - \alpha_k \tau x - c_k^{s*}(\boldsymbol{N}^s) + f(\tfrac{1}{N(x)}) \right\} N_k(x) - r_A \right] \mathrm{d}x \tag{34a}$$

$$\text{s.t.} \quad \int_d^\infty N_i(x) \mathrm{d}x = N_i^s \quad \forall i \in \mathcal{I}, \quad N_i(x) \geq 0 \quad \forall i \in \mathcal{I}, \forall x \in \mathbb{R}_+. \tag{34b}$$

Furthermore, the spatial distribution $[N_i^*(x)]$ $(x \in \mathcal{X}^d)$ *of downtown commuters is a long-run equilibrium if and only if it is a KKT point of the following optimization problem:*

$$\max_{[N_i(x)]} \int_0^d \left[\sum_{k \in \mathcal{I}} \left\{ w_k - \alpha_k \tau x + f(\tfrac{1}{N(x)}) \right\} N_k(x) - r_A \right] \mathrm{d}x \tag{35a}$$

$$\text{s.t.} \quad \int_0^d N_i(x) \mathrm{d}x = N_i^d \quad \forall i \in \mathcal{I}, \quad N_i(x) \geq 0 \quad \forall i \in \mathcal{I}, \forall x \in \mathbb{R}_+. \tag{35b}$$

Since $\frac{1}{N(x)}$ equals the lot size $a(x)$ and $f'(\frac{1}{N(x)})$ equals the market land rent at location x, the objective functions of (34) and (35) represent the total surplus $\int \sum_{k \in \mathcal{I}} v_k(x) N_k(x) \mathrm{d}x + \int r(x) a(x) N(x) \mathrm{d}x$ of the suburb and the downtown, respectively. Hence, this lemma demonstrates that the land market is efficient in both the suburb and downtown, as in the traditional residential location model (Fujita, 1989). Note that since the number $\boldsymbol{N}^s$ of suburban commuters is taken as given, Lemma 5 does not indicate that the long-run equilibrium is efficient but instead shows that market failures in our model are caused only by traffic (bottleneck) congestion.

The results obtained above can be summarized as Proposition 2.

Proposition 2. *Suppose Assumptions 3 and 4. Then, given the number* $\boldsymbol{N}^s$ *of suburban commuters, the long-run equilibrium suburban and downtown spatial structures have the following properties.*

(i) The long-run equilibrium spatial distribution of suburban commuters and that of downtown commuters are uniquely determined.

(ii) Among commuters residing in the suburb, those with a high value of travel time reside closer to the CBD. Among commuters residing downtown, those with a high value of travel time reside closer to the CBD.

(iii) Population increase in the suburb leads to urban sprawl. Furthermore, it induces higher density and land rents at any populated suburban location, and lower density and land rents at any populated downtown location.

(iv) The total surplus of the suburb and that of the downtown are maximized.

3.2.2 Population of suburban and downtown commuters

We next characterize the long-run equilibrium number $\boldsymbol{N}^{s*} = [N_i^{s*}]$ and $\boldsymbol{N}^{d*} = [N_i^{d*}]$ of suburban and downtown commuters i under Assumptions 3 and 4 by using equilibrium conditions (14e) and (14f). From (28) and Lemma 4, utilities that commuters receive from residing in the suburb and downtown are given by

$$v_i^{s*}(\boldsymbol{N}^s) = y_i - c_i^{b*}(\boldsymbol{N}^s) - \alpha_i \tau X_i^s + H(r_A) - \sum_{k=i}^{I} \alpha_k \tau (X_{k+1}^s - X_k^s), \tag{36a}$$

$$v_i^{d*}(\boldsymbol{N}^d) = y_i - \alpha_i \tau X_i^d + H(r(d) + r_A) - \sum_{k=i}^{I} \alpha_k \tau (X_{k+1}^d - X_k^d), \tag{36b}$$

where X_i^s and X_i^d are represented as (28). The utility difference $v_i^{d*}(\boldsymbol{N}^d) - v_i^{s*}(\boldsymbol{N}^s)$ is thus given by

$$v_i^{d*}(\boldsymbol{N}^d) - v_i^{s*}(\boldsymbol{N}^s) = c_i^{b*}(\boldsymbol{N}^s) + \sum_{k=i+1}^{I} (\alpha_{k-1} - \alpha_k)\tau(X_k^s - X_k^d) + \alpha_I \tau (X^B - d) + H(r(d) + r_A) - H(r_A). \tag{37}$$

The difference in utility of commuters with high bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$ or small i grows. More specifically, because the utility difference satisfies

$$\begin{aligned} &\left\{v_i^{d*}(\boldsymbol{N}^d) - v_i^{s*}(\boldsymbol{N}^s)\right\} - \left\{v_{i-1}^{d*}(\boldsymbol{N}^d) - v_{i-1}^{s*}(\boldsymbol{N}^s)\right\} \\ &\qquad = \left\{c_i^{b*}(\boldsymbol{N}^s) + \alpha_i \tau \left(X_i^s - X_i^d\right)\right\} - \left\{c_{i-1}^{b*}(\boldsymbol{N}^s) + \alpha_{i-1} \tau \left(X_i^s - X_i^d\right)\right\}, \end{aligned} \tag{38}$$

where $c_i^{b*}(\boldsymbol{N}^s) + \alpha_i \tau \left(X_i^s - X_i^d\right)$ denotes the commuting cost difference between suburban and downtown commuters i, commuters with a large commuting cost difference have a greater utility difference. Since the short-run equilibrium bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$ is given by (20), this implies that richer or more inflexible commuters prefer to reside downtown under Assumptions 1, 3, and 4. Indeed, if rich commuters are inflexible, they reside downtown. To see this, we consider the case in which Assumption 3 and the following assumption hold.

Assumption 5. $\beta_{i-1} > \beta_i$ *for all* $i \in \mathcal{I}\backslash\{1\}$.

Under Assumptions 1 and 3–5, $c_i^{b*}(\boldsymbol{N}^s) < c_{i-1}^{b*}(\boldsymbol{N}^s)$ for any $\boldsymbol{N}^s$ and $i \in \mathcal{I}\backslash\{1\}$; that is, the income elasticity of commuting cost differences is positive. Thus, we have

$$v_i^{d*}(\boldsymbol{N}^d) - v_i^{s*}(\boldsymbol{N}^s) < v_{i-1}^{d*}(\boldsymbol{N}^d) - v_{i-1}^{s*}(\boldsymbol{N}^s) \tag{39}$$

for any $\boldsymbol{N}^s$ and $i \in \mathcal{I}\backslash\{1\}$. This implies that there exists $i^* \in \mathcal{I}$ such that

$$\begin{cases} v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) > v_i^{s*}(\boldsymbol{N}^s) & \text{for any} \quad i < i^*, \\ v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) \leq v_i^{s*}(\boldsymbol{N}^s) & \text{if} \quad i = i^*, \\ v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) < v_i^{s*}(\boldsymbol{N}^s) & \text{for any} \quad i > i^*, \end{cases} \tag{40}$$

where $\boldsymbol{N} = [N_i]$ denotes the total number of commuters i. Because $v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{s*}(\boldsymbol{N}^s)$ increases with an increase in N_i^s, this condition indicates that a long-run equilibrium number $\boldsymbol{N}^{s*}$ of suburban commuters exists uniquely and is given by

$$N_i^{s*} = 0 \quad \text{if} \quad i < i^*, \tag{41a}$$

$$N_i^{s*} = N_i \quad \text{if} \quad i > i^*, \tag{41b}$$

$$N_{i^*}^{s*} = \begin{cases} N_{i^*} & \text{if} \quad v_{i^*}^{d*}([N_1, \cdots, N_{i^*-1}, 0, 0, \cdots, 0]) < v_{i^*}^{s*}([0, \cdots, 0, N_{i^*}, N_{i^*+1}, \cdots, N_I]), \\ \phi N_{i^*} & \text{otherwise}, \end{cases} \tag{41c}$$

where $\phi \in [0, 1]$ is uniquely determined from

$$v_{i^*}^{d*}([N_1, \cdots, N_{i^*-1}, (1-\phi)N_{i^*}, 0, \cdots, 0]) = v_{i^*}^{s*}([0, \cdots, 0, \phi N_{i^*}, N_{i^*+1}, \cdots, N_I]). \tag{42}$$

Therefore, we have the following proposition.

Proposition 3. *Suppose Assumptions 1 and 3–5. Then the long-run equilibrium number of suburban and downtown commuters is uniquely determined. Furthermore, at the long-run equilibrium, commuters with a high value of travel time reside downtown and commuters with a low value of travel time reside in the suburb.*

Propositions 2 and 3 show that, under Assumptions 1 and 3–5, *rich and inflexible commuters reside closer to the CBD* since rich commuters have a higher commuting cost. This result is consistent with empirical observations in cities with heavy traffic congestion (see, e.g., McCann, 2013, p.126). Furthermore, introducing Assumptions 3 and 5 implies that the income elasticity of commuting costs is positive and larger than the income elasticity of demand for land. Therefore, this result is consistent with the standard result given by the traditional location model.

Although we mainly focus on this type of heterogeneity to clearly demonstrate the properties of our model, in other cases, we can have different equilibrium spatial distributions of commuters. As an example, we consider the opposite case: poor commuters are highly inflexible and rich commuters are highly flexible. In this case, the bottleneck cost of poor commuters can be much higher than that of rich commuters, such that $c_i^{b*}(\boldsymbol{N}^s) + \alpha_i \tau(X_i^s - X_i^d) > c_{i-1}^{b*}(\boldsymbol{N}^s) + \alpha_{i-1}\tau(X_i^s - X_i^d)$ for some $\boldsymbol{N}^s$ and $i \in \mathcal{I}$ under Assumption 3 (i.e., income elasticity of commuting cost differences is negative). This, together with (38), implies that *poor commuters can reside downtown at the long-run equilibrium.* This result occurs because commuters who are relatively rich but highly flexible are able to traverse the bottleneck early or late to avoid queuing congestion at a low cost. This will save them a significant amount on their lot size, with a fairly small bottleneck cost. Thus, they prefer to do so. Likewise, commuters who are relatively poor but highly inflexible can avoid their incredibly costly traversing of the bottleneck by paying to reside downtown. Therefore, we can say that, in our model, *commuters sort into residing in the suburb or downtown based on their value of travel time and their flexibility.*

4 Optimal congestion toll

Studies utilizing the standard bottleneck model show that queuing time is a pure deadweight loss. Hence, in our model, there is no queue at the social optimum, and the social optimum can be achieved by imposing an optimal time-varying congestion toll that eliminates queuing congestion, as shown later. This section considers the introduction of an optimal congestion toll $p(t)$ and characterizes equilibrium under this pricing policy.

4.1 Short-run equilibrium

Congestion toll $p(t)$ eliminates queuing congestion.[12] Thus, the commuting cost $c_i^t(x,t)$ of commuters i is given by

$$c_i^t(x,t) = \begin{cases} s_i(t-t^*) + \alpha_i \tau x & \text{if} \quad x \in \mathcal{X}^d, \\ c_i^{bt}(t) + \alpha_i \tau x & \text{if} \quad x \in \mathcal{X}^s, \end{cases} \tag{43a}$$

$$c_i^{bt}(t) = p(t) + s_i(t-t^*). \tag{43b}$$

Superscript t describes variables under the congestion toll. Since we consider heterogeneous commuters, congestion toll $p(t)$ does not equal queuing time cost $\alpha_i q(t)$ at the no-toll equilibrium and is set so that travel demand $\sum_{i\in\mathcal{I}} n_i^{st}(t)$ at the bottleneck equals supply (i.e., capacity) μ. Therefore, short-run equilibrium conditions for suburban commuters are expressed as

$$\begin{cases} c_i^{bt}(t) = c_i^{bt*} & \text{if} \quad n_i^{st}(t) > 0 \\ c_i^{bt}(t) \geq c_i^{bt*} & \text{if} \quad n_i^{st}(t) = 0 \end{cases} \quad \forall i \in \mathcal{I},\ \forall t \in \mathbb{R}, \tag{44a}$$

$$\begin{cases} \sum_{i\in\mathcal{I}} n_i^{st}(t) = \mu & \text{if} \quad p(t) > 0 \\ \sum_{i\in\mathcal{I}} n_i^{st}(t) \leq \mu & \text{if} \quad p(t) = 0 \end{cases} \quad \forall t \in \mathbb{R}, \tag{44b}$$

$$\int n_i^{st}(t)\, \mathrm{d}t = N_i^{st} \quad \forall i \in \mathcal{I}. \tag{44c}$$

Condition (44a) is the no-arbitrage condition for suburban commuters' arrival time choices. Condition (44b) denotes the bottleneck's capacity constraints, which assure that queuing congestion is eliminated at the equilibrium. Condition (44c) provides the flow conservation for commuting demand. From these conditions, we have $n_i^{st}(t)$, $p(t)$, and c_i^{bt*} at the short-run equilibrium as functions of the number $\boldsymbol{N}^{st}$ of suburban commuters $i \in \mathcal{I}$.

As in the case without the congestion toll, by invoking the results of studies employing the bottleneck model, we have Lemma 6

Lemma 6 (Lindsey, 2004; Iryo and Yoshii, 2007)**.**

(i) The short-run equilibrium bottleneck cost $c_i^{bt*}(\boldsymbol{N}^{st})$ under the congestion toll is uniquely determined.

(ii) The short-run equilibrium number $[n_i^{st*}(t)]$ of suburban commuters arriving at time t under the congestion toll coincides with the solution of the following linear programming problem:

$$\min_{[n_i^{st}(t)]} \sum_{i\in\mathcal{I}} \int s_i(t-t^*)\, n_i^{st}(t)\, \mathrm{d}t \tag{45a}$$

$$s.t. \quad \sum_{i\in\mathcal{I}} n_i^{st}(t) \leq \mu \quad \forall t \in \mathbb{R}, \quad \int n_i^{st}(t)\, \mathrm{d}t = N_i^{st} \quad \forall i \in \mathcal{I}, \quad n_i^{st}(t) \geq 0 \quad \forall i \in \mathcal{I},\ \forall t \in \mathbb{R}. \tag{45b}$$

Lemma 6 (ii) suggests that total schedule delay cost is minimized at the short-run equilibrium under the congestion toll. Note that total schedule delay cost equals total commuting cost minus

[12] The tradable network permit scheme proposed by Akamatsu (2007) and Wada and Akamatsu (2013) has the same effect as the optimal congestion toll. Similar schemes have been proposed by, e.g., Verhoef et al. (1997), Yang and Wang (2011), Nie (2012), He et al. (2013), and Nie and Yin (2013).

total toll revenue. Hence, Lemma 6 (ii) indicates that, in the short run, the optimal congestion toll minimizes the social cost of commuting.

From equilibrium condition (44a), we have

$$c_i^{bt*}(t_i) + c_j^{st*}(t_j) \leq c_i^{bt*}(t_j) + c_j^{st*}(t_i) \quad \forall t_i \in \text{supp}\,(n_i^{st*}),\ \forall t_j \in \text{supp}\,(n_j^{st*}),\ \forall i, j \in \mathcal{I}. \tag{46}$$

Thus, the following condition is obtained by substituting (43b) into (46): for any $t_i \in \text{supp}\,(n_i^{st*})$, $t_j \in \text{supp}\,(n_j^{st*})$, and $i, j \in \mathcal{I}$,

$$\left(\beta_i - \beta_j\right)\left(t_i - t_j\right) \geq 0 \qquad \text{if} \quad \max\{t_i, t_j\} \leq t^*, \tag{47a}$$

$$\left(\gamma_i - \gamma_j\right)\left(t_i - t_j\right) \leq 0 \qquad \text{if} \quad \min\{t_i, t_j\} \geq t^*. \tag{47b}$$

This condition indicates that early-arriving commuters arrive at the CBD in order of increasing β_i and that late-arriving commuters arrive in order of decreasing γ_i under the congestion toll. Since commuters with a high marginal time-based schedule delay cost arrive closer to their preferred arrival time at the no-toll equilibrium, *imposing the congestion toll alters the arrival order of commuters.*

This result and Lemmas 1 and 6 reveal that *the equilibrium bottleneck cost under the congestion toll $c_i^{bt*}(\boldsymbol{N}^{st})$ generally differs from the no-toll equilibrium bottleneck cost $c_i^{b*}(\boldsymbol{N}^{st})$ when we consider commuter heterogeneity in the value of travel time.* Indeed, we can see that the bottleneck cost at equilibrium with tolling differs from that at the no-toll equilibrium. For this, we suppose Assumptions 1 and 5. Then equilibrium bottleneck cost $c_i^{bt*}(\boldsymbol{N}^{st})$ under the toll is obtained in the same manner as in (20).

$$c_i^{bt*}(\boldsymbol{N}^{st}) = \frac{\eta}{1+\eta}\left\{\beta_i \frac{\sum_{k=1}^{i} N_k^{st}}{\mu} + \sum_{k=i+1}^{I} \beta_k \frac{N_k^{st}}{\mu}\right\} \qquad \forall i \in \mathcal{I}. \tag{48}$$

This shows that inflexible commuters have higher bottleneck costs at the equilibrium under the congestion toll, which is fundamentally different from the properties of the no-toll equilibrium bottleneck cost.

We summarize the properties of the equilibrium in Proposition 4.

Proposition 4. *The short-run equilibrium under the congestion toll has the following properties.*

(i) Total schedule delay cost is minimized.

(ii) Early-arriving commuters arrive at work in order of increasing marginal early delay cost (β_i). Late-arriving commuters arrive at work in order of decreasing marginal late delay cost (γ_i).

(iii) Equilibrium bottleneck cost $c_i^{bt*}(\boldsymbol{N}^{st})$ of commuters i is uniquely determined. Furthermore, under Assumptions 1 and 5, $c_i^{bt*}(\boldsymbol{N}^{st})$ is given by (48).

As in the case without tolling, downtown commuters arrive at $t = t^*$, as they need not traverse the bottleneck. That is, the commuting cost of downtown commuters i at the short-run equilibrium under the congestion toll is given by $\alpha_i \tau x$.

4.2 Long-run equilibrium

We characterize the long-run equilibrium spatial distribution of commuters by using the short-run equilibrium bottleneck cost. In the long-run, the difference between cases with and without tolling appears only in the indirect utility of suburban commuters. Specifically, under the congestion toll, the indirect utility $v_i^{st}(x)$ of suburban commuters i is expressed as

$$v_i^{st}(x) = w_i - \alpha_i \tau x - c_i^{bt*}(\boldsymbol{N}^{st}) + H(r^t(x) + r_A), \tag{49}$$

where $r^t(x) + r_A$ denotes the land rent at x under the congestion toll. The long-run equilibrium conditions are thus represented as (11) with the use of (49).

Following the same procedure as in Section 3.2 reveals that the urban spatial structure at the long-run equilibrium under the congestion toll has the same properties as those without tolling (Propositions 2 and 3).

Proposition 5. *Suppose Assumptions 1 and 3–5. Then under the congestion toll, the long-run equilibrium has the following properties.*

(i) Spatial distributions of commuters are uniquely determined.

(ii) Commuters with a high value of travel time reside closer to the CBD.

(iii) The city boundary X^{Bt}, spatial distribution $N^t(x)$ of commuters, lot size $a^t(x)$, and land rent $r^t(x)$ in the suburb and downtown have the same functional form as in the no-toll equilibrium (i.e., (31) *and* (32)*).*

Note that this proposition does not suggest that the urban structure at the long-run equilibrium with tolling coincides with that at the no-toll long-run equilibrium. Indeed, imposing a congestion toll can change the number of suburban and downtown commuters since it changes the short-run equilibrium bottleneck cost. Therefore, this proposition demonstrates that *differences between the long-run equilibria with and without tolling arise only when the long-run equilibrium number ($\boldsymbol{N}^{s*}$ and $\boldsymbol{N}^{st*}$) of suburban commuters changes by tolling.* In the next section, we will show the effects of tolling on the urban spatial structure by examining differences between $\boldsymbol{N}^{s*}$ and $\boldsymbol{N}^{st*}$.

Before studying the effects of tolling, we show that the social optimum is achieved by imposing the congestion toll. We define the social optimum as the global maximizer of commuters' total utility subject to budget, land, bottleneck capacity, and population constraints:

$$\max_{[n_i(x,t)],[z_i(x,t)],[a_i(x,t)]} \sum_{i \in \mathcal{I}} \iint u(z_i(x,t), a_i(x,t))\, n_i(x,t)\, \mathrm{d}t \mathrm{d}x, \tag{50a}$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{I}} \iint \left\{ w_i - z_i(x,t) - r_{\mathrm{A}} a_i(x,t) - c_i(x,t) \right\} n_i(x,t)\, \mathrm{d}t \mathrm{d}x \geq 0, \tag{50b}$$

$$1 - \sum_{i \in \mathcal{I}} \int a_i(x,t) n_i(x,t) \mathrm{d}t \geq 0 \qquad \forall x \in \mathbb{R}_+, \tag{50c}$$

$$\mu - \sum_{i \in \mathcal{I}} \int n_i(x,t) \mathrm{d}x \geq 0 \qquad \forall t \in \mathbb{R}, \tag{50d}$$

$$N_i - \iint n_i(x,t) \mathrm{d}x \mathrm{d}t = 0 \qquad \forall i \in \mathcal{I}, \tag{50e}$$

$$n_i(x,t) \geq 0,\ z_i(x,t) \geq 0,\ a_i(x,t) \geq 0 \qquad \forall i \in \mathcal{I},\ \forall x \in \mathbb{R}_+,\ \forall t \in \mathbb{R}. \tag{50f}$$

Then we have Proposition 6.

Proposition 6. *The KKT conditions of problem* (50) *coincide with the short-run and long-run equilibrium conditions under the congestion toll.*

Proof. See Appendix D □

This proposition shows that the social optimum is a long-run equilibrium under the congestion toll, which indicates that market failures are caused only by bottleneck congestion in our model.

5 Effects of congestion toll on urban spatial structure

5.1 Long-run equilibria with and without tolling

This section demonstrates that imposing the congestion toll alters urban spatial structure. As discussed in the previous section, the difference between the long-run equilibria with and without tolling arises when the number of suburban and downtown commuters ($\boldsymbol{N}^s$ and $\boldsymbol{N}^d$) changes by imposing the congestion toll. Therefore, to examine effects of the congestion toll, we compare the number of suburban commuters at the long-run equilibria with and without tolling.

We denote the utilities of commuters i residing in the suburb and downtown under the congestion toll by $v_i^{st*}(\boldsymbol{N}^s)$ and $v_i^{dt*}(\boldsymbol{N}^d)$, respectively, which are derived from (14a)–(14d), with the use of (49). Then under Assumptions 3 and 4, $v_i^{st*}(\boldsymbol{N}^s)$ and $v_i^{dt*}(\boldsymbol{N}^d)$ are obtained in the manner of (36):

$$v_i^{st*}(\boldsymbol{N}^s) = y_i - c_i^{bt*}(\boldsymbol{N}^s) - \alpha_i \tau X_i^{st} + H(r_A) - \sum_{k=i}^{I} \alpha_k \tau (X_{k+1}^{st} - X_k^{st}), \tag{51a}$$

$$v_i^{dt*}(\boldsymbol{N}^d) = y_i - \alpha_i \tau X_i^{dt} + H(r^t(d) + r_A) - \sum_{k=i}^{I} \alpha_k \tau (X_{k+1}^{dt} - X_k^{dt}), \tag{51b}$$

where X_k^{st} and X_k^{dt} are, respectively, residential locations for suburban and downtown commuters i closest to the CBD, which are given by the same functional form as in the no-toll equilibrium (i.e., (28)). Thus, $v_i^{dt*}(\boldsymbol{N}^d) - v_i^{ds*}(\boldsymbol{N}^s)$ is represented as

$$v_i^{dt*}(\boldsymbol{N}^d) - v_i^{st*}(\boldsymbol{N}^s) = c_i^{bt*}(\boldsymbol{N}^s) + \sum_{k=i+1}^{I} (\alpha_{k-1} - \alpha_k) \tau (X_k^{st} - X_k^{dt}) + \alpha_I \tau (X^{Bt} - d) + H(r^t(d) + r_A) - H(r_A). \tag{52}$$

It follows from this and (37) that

$$\left(v_i^{dt*}(\boldsymbol{N}^d) - v_i^{st*}(\boldsymbol{N}^s)\right) - \left(v_i^{d*}(\boldsymbol{N}^d) - v_i^{s*}(\boldsymbol{N}^s)\right) = c_i^{bt*}(\boldsymbol{N}^s) - c_i^{b*}(\boldsymbol{N}^s) \qquad \forall i \in \mathcal{I}. \tag{53}$$

This leads to Proposition 7.

Proposition 7. *Suppose Assumptions 3 and 4. Then for any* $\boldsymbol{N}^s$ *and* $i \in \mathcal{I} \backslash \{1\}$,

$$\begin{aligned} &\left(v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{st*}(\boldsymbol{N}^s)\right) - \left(v_{i-1}^{dt*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_{i-1}^{st*}(\boldsymbol{N}^s)\right) \\ &\qquad = \left(v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{s*}(\boldsymbol{N}^s)\right) - \left(v_{i-1}^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_{i-1}^{s*}(\boldsymbol{N}^s)\right), \end{aligned} \tag{54}$$

if and only if there exists δ such that $c_i^{bt*}(\boldsymbol{N}^s) = c_i^{b*}(\boldsymbol{N}^s) + \delta$ for all $i \in \mathcal{I}$.

This proposition implies that the urban spatial structure does not change by imposing the congestion toll if $c_i^{bt*}(\boldsymbol{N}^s) = c_i^{b*}(\boldsymbol{N}^s) + \delta$ for all $i \in \mathcal{I}\backslash\{1\}$. However, in general, this condition does not hold when commuters are heterogeneous in their value of travel time, as discussed in Section 4.1. This means that imposing the optimal congestion toll changes the short-run equilibrium bottleneck cost and creates incentives for commuters to relocate. *Unlike Arnott (1998), therefore, congestion tolling does alter the urban spatial structure in our model.* It is, however, difficult to examine how the urban spatial structure changes by imposing the congestion toll. Therefore, the following subsection analyzes our model in a simple setting to elucidate the effects of tolling.

Note that the results presented thus far were obtained under the assumption that toll revenues are not redistributed. Since the optimal congestion toll minimizes the short-run social cost of traversing the bottleneck, bottleneck costs for all suburban commuters can be reduced. More specifically, if policymakers can observe the type of commuters, they can redistribute toll revenue to suburban commuters such that

$$c_i^{bt*}(\boldsymbol{N}^s) - \rho_i(\boldsymbol{N}^s) < c_i^{b*}(\boldsymbol{N}^s), \tag{55}$$

where $\rho_i(\boldsymbol{N}^s)$ denotes the toll-revenue redistribution for each suburban commuter i (type-specific lump-sum rebate). Thus, under this toll-revenue redistribution $\rho_i(\boldsymbol{N}^s)$, the following condition is satisfied for any $\boldsymbol{N}^s$ and $i \in \mathcal{I}$:

$$v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{st*}(\boldsymbol{N}^s) - \rho_i(\boldsymbol{N}^s) < v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{s*}(\boldsymbol{N}^s). \tag{56}$$

This indicates that *if every commuter does not relocate (i.e., in the short-run), imposing congestion toll with this toll-revenue redistribution helps all suburban commuters and hence makes residing in the suburb more desirable.* Furthermore, this and (40) indicates that this toll-revenue redistribution leads to urban sprawl under Assumptions 1 and 3–5.

5.2 A simple example

5.2.1 Theoretical analysis

We consider a simple setting to show concretely the effects of the congestion toll on urban spatial structure. Specifically, we suppose that Assumptions 1–4 hold.[13] That is, rich commuters are assumed to have a higher marginal time-based schedule delay cost. This implies that rich commuters tend to avoid a schedule delay rather than queuing time and paying the toll.

As Hall (2015) shows, this is a situation wherein congestion tolling does not alter the arrival order of commuters and generates a Pareto improvement if every commuter does not relocate. Indeed, it follows from (20) and (48) that the difference between *short-run* equilibrium bottleneck costs with and without tolling is non-positive for any $\boldsymbol{N}^s$:

$$c_i^{bt*}(\boldsymbol{N}^s) - c_i^{b*}(\boldsymbol{N}^s) = \frac{\eta}{1+\eta}\alpha_i \sum_{k=i+1}^{I} \left\{\frac{1}{\alpha_i} - \frac{1}{\alpha_k}\right\} \beta_k \frac{N_k^s}{\mu} \begin{cases} = 0 & \forall i \geq \max\{\mathrm{supp}\,(\boldsymbol{N}^s)\}, \\ < 0 & \forall i < \max\{\mathrm{supp}\,(\boldsymbol{N}^s)\}. \end{cases} \tag{57}$$

[13]Note that if Assumptions 1–4 hold, the condition in Assumption 5 is also satisfied.

This shows that the short-run equilibrium bottleneck cost incurred by the poorest commuters does not change. This reflects the fact that the poorest commuters are the most inflexible and have the highest time-based marginal schedule delay cost. That is, commuters who face no queuing cost at the equilibrium without tolling and face no toll at the equilibrium with tolling are the poorest ones.

We see from (57) that congestion tolling weakly decreases short-run equilibrium bottleneck costs of all commuters. However, as we see later, congestion tolling cannot lead to a Pareto improvement if we consider commuters' relocation. Moreover, rich commuters gain and poor commuters lose from imposing the congestion toll.[14]

We first examine the effects of tolling on urban spatial structure. It follows from (53) and (57) that, for any $\boldsymbol{N}^s$ and $i \in \mathcal{I}$,

$$\begin{cases} v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{st*}(\boldsymbol{N}^s) = v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{s*}(\boldsymbol{N}^s) & \forall i \geq \max\{\mathrm{supp}\,(\boldsymbol{N}^s)\}, \\ v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{st*}(\boldsymbol{N}^s) < v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^s) - v_i^{s*}(\boldsymbol{N}^s) & \forall i < \max\{\mathrm{supp}\,(\boldsymbol{N}^s)\}. \end{cases} \tag{58}$$

This clearly indicates that imposing the congestion toll can create incentives for commuters to reside in the suburb. Since the long-run equilibria with and without tolling are uniquely determined, this result leads to

$$N_i^{st*} \geq N_i^{s*} \qquad \forall i \in \mathcal{I}. \tag{59}$$

This implies that *imposing the congestion toll can increase the suburban population.* Furthermore, it follows from Propositions 2 and 4 that if there exists $i \in \mathcal{I}$ such that $N_i^{st*} > N_i^{s*}$,

$$X^{B*} < X^{Bt*}, \tag{60a}$$

$$N^*(x) \begin{cases} > N^{t*}(x) & \text{if} \quad x \in \mathcal{X}^d, \\ < N^{t*}(x) & \text{if} \quad x \in \mathcal{X}^s \cap \mathrm{supp}\,(N^{t*}), \end{cases} \tag{60b}$$

$$r^*(x) \begin{cases} > r^{t*}(x) & \text{if} \quad x \in \mathcal{X}^d, \\ < r^{t*}(x) & \text{if} \quad x \in \mathcal{X}^s \cap \mathrm{supp}\,(N^{t*}), \end{cases} \tag{60c}$$

where superscripts $*$ and $t*$ describe variables at the long-run equilibria without and with tolling, respectively. This indicates that *the population increase in the suburb leads to urban sprawl and induces higher (lower) density and land rents at any populated suburban (downtown) location.*

This finding is in contrast to the standard results of traditional location models, which consider static flow congestion (Kanemoto, 1980; Wheaton, 1998; Anas et al., 1998). It also differs from the results obtained by Arnott (1998), who considers homogeneous commuters. This thereby demonstrates that *interactions among heterogeneous commuters may cause urban sprawl resulting from imposition of the optimal congestion toll.*

We next examine changes in commuters' utility due to the population increase in the suburb. There exist commuters $i \in \mathcal{I}^R \equiv \{i \in \mathcal{I} \mid N_i^{dt*} > 0 \text{ and } N_i^{d*} > 0\}$ who reside downtown at the long-run equilibria with and without tolling, and their utility changes by tolling from $v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^{s*})$ to

[14] As Takayama and Kuwahara (2016) shows, essentially the same conclusion is obtained if we introduce the assumption "$\beta_{i-1}/\alpha_{i-1} \leq \beta_i/\alpha_i$ for any $i \in \mathcal{I}\backslash\{1\}$" instead of Assumption 2.

$v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^{st*})$. Their difference is obtained from (28), (36), and (51) as

$$v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^{dt*}) - v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^{d*}) = \left\{H(r^{t*}(0) + r_A) - H(r^*(0) + r_A)\right\} + \sum_{k=2}^{i} (\alpha_{k-1} - \alpha_k)\, \tau \left(X_k^{dt*} - X_k^{d*}\right). \tag{61}$$

Because $X_i^{dt*} \geq X_i^{d*}$ for all $i \in \mathcal{I}^R$ and $r^{t*}(0) < r^*(0)$ hold if there exists $i \in \mathcal{I}$ such that $N_i^{st*} > N_i^{s*}$, we have $v_i^{dt*}(\boldsymbol{N} - \boldsymbol{N}^{dt*}) > v_i^{d*}(\boldsymbol{N} - \boldsymbol{N}^{d*})$ for all $i \in \mathcal{I}^R$. This shows that *rich commuters $i \in \mathcal{I}^R$ gain from the population increase in the suburb.*

The poorest commuters I reside farthest from the CBD at the long-run equilibria with and without tolling. Therefore, their utility difference is equal to the commuting cost difference:

$$v_I^{st*}(\boldsymbol{N}^{st*}) - v_I^{s*}(\boldsymbol{N}^{s*}) = \left\{c_I^{b*}(\boldsymbol{N}^{s*}) + \alpha_I \tau X^{B*}\right\} - \left\{c_I^{bt*}(\boldsymbol{N}^{st*}) + \alpha_I \tau X^{Bt*}\right\}. \tag{62}$$

Furthermore, (20) and (48) yield

$$c_I^{bt*}(\boldsymbol{N}^{st*}) - c_I^{b*}(\boldsymbol{N}^{s*}) = \frac{\eta}{1+\eta}\frac{\beta_i}{\mu}\left\{\sum_{k\in\mathcal{I}} N_k^{st*} - \sum_{k\in\mathcal{I}} N_k^{s*}\right\}. \tag{63}$$

Thus, we obtain $v_I^{st*}(\boldsymbol{N}^{st*}) < v_I^{s*}(\boldsymbol{N}^{s*})$ if there exists $i \in \mathcal{I}$ such that $N_i^{st*} > N_i^{s*}$; that is, *the population increase in the suburb harms the poorest commuters.*

These results establish the following proposition.

Proposition 8. *Suppose Assumptions 1–4. Then*

(i) congestion tolling weakly decreases bottleneck costs of all commuters in the short-run and can increase the suburban population in the long-run;

(ii) the population increase in the suburb leads to urban sprawl and induces higher (lower) density and land rents at any populated suburban (downtown) location;

(iii) rich commuters $i \in \mathcal{I}^R$ gain and the poorest commuters I lose from the population increase in the suburb.

The results obtained in this subsection can be summarized as follows. In the short-run, congestion tolling reduces bottleneck costs of all commuters except those incurred by the poorest commuters, creating incentives for downtown commuters to reside in the suburb. Furthermore, commuters' relocation from downtown to the suburb causes downtown rents to fall and utilities to rise. Since rich commuters reside downtown in this case, they are made better off by congestion tolling if it leads to relocation. Since the poorest commuters reside farthest from the CBD, downtown commuters' relocation pushes them farther out from the CBD. This, together with increased demand for traversing the bottleneck, exacerbates commuting costs (free-flow travel time cost and schedule delay cost) of the poorest commuters.

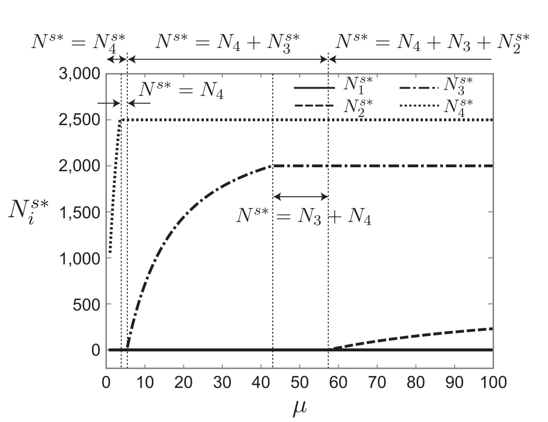

Figure 3: Number N_i^{s*} of suburban commuters i at no-toll equilibrium

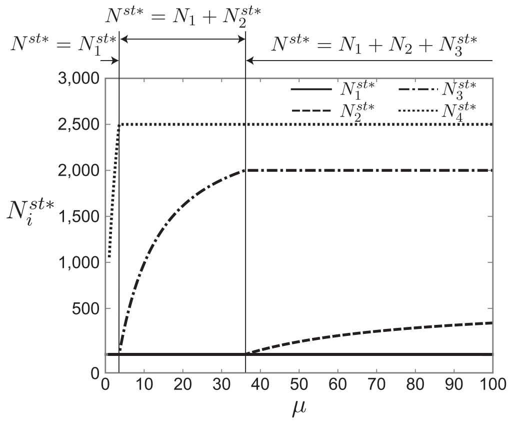

Figure 4: Number N_i^{st*} of suburban commuters i under the optimal congestion toll

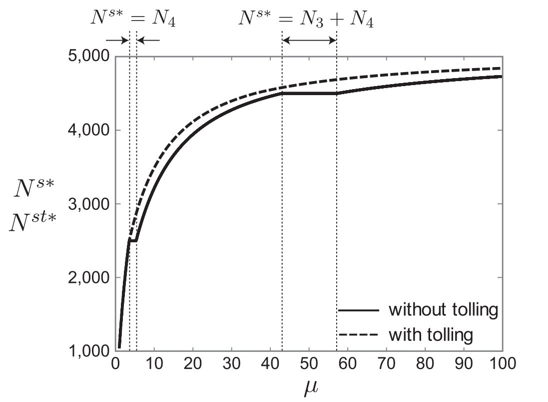

Figure 5: Total number N^s of suburban commuters

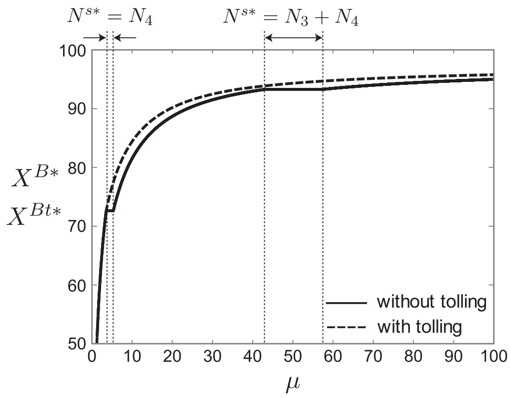

Figure 6: City boundary X^{B}

5.2.2 Numerical analysis

We numerically analyze our model and show effects of the optimal congestion toll. In this analysis, we assume $f(a) = \kappa \ln[a]$ and use the following parameter values:

$$I = 4, \quad d = 10 \text{ (km)}, \quad \tau = 2 \text{ (min/km)}, \quad [N_i] = [1000, 1500, 2000, 2500], \tag{64a}$$

$$[y_i] = [300, 200, 150, 100], \quad \kappa = 10, \quad r_{\mathrm{A}} = 200. \tag{64b}$$

The values of α_i, β_i, and η are set to be consistent with the empirical result (Small, 1982) and Assumptions 1–4.

$$[\alpha_i] = [0.3, 0.2, 0.15, 0.1], \quad [\beta_i] = [0.15, 0.09, 0.06, 0.03], \quad \eta = 4. \tag{64c}$$

We conduct comparative statics with respect to bottleneck capacity μ. The no-toll equilibrium number of commuters $i \in \mathcal{I}$ is presented in Figure 3. This figure shows that downtown commuters relocate to the suburb in order of decreasing i with increases in bottleneck capacity. They do so because increasing μ reduces bottleneck cost $c_i^{b*}(\boldsymbol{N}^s)$ of all commuters, creating incentives for downtown commuters to relocate to the suburb. This is consistent with the results presented in Section 5.2.1.

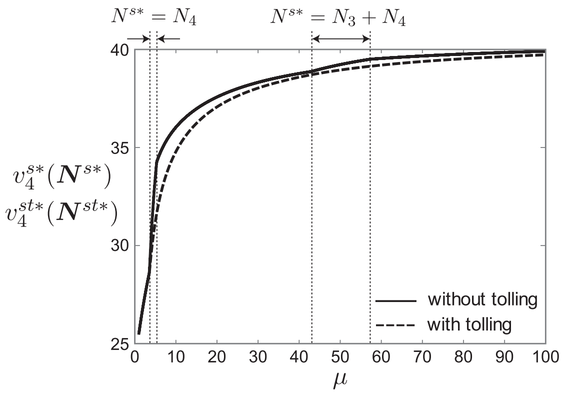

Figure 7: Utility v_4^* of commuters 4

The effects of the optimal congestion toll appear in Figures 4–7. Figure 4 presents the long-run equilibrium number N_i^{st*} of suburban commuters i under the optimal congestion toll. Although this result is qualitatively the same as that at the no-toll equilibrium (Figure 3), congestion tolling changes the total number $N^s = \sum_{i \in \mathcal{I}} N_i^s$ of suburban commuters, as illustrated in Figure 5. Note that when μ is small, imposition of the congestion toll does not alter N^s. This occurs because for small μ, only commuters 4 reside in the suburb (i.e., commuters traversing the bottleneck are homogeneous). Thus, congestion tolling does not affect the commuting costs of suburban commuters, as shown in Arnott (1998). Furthermore, a suburban population increase attributable to congestion tolling leads to expansion of the city boundary X^{B}, as illustrated in Figure 6. That is, imposing the optimal congestion toll causes urban sprawl. Figure 7 indicates that congestion tolling reduces the utility of commuters 4 (i.e., commuters with the lowest value of time). That is, the poorest commuters lose from congestion tolling. We also see from Figures 5–7 that the difference between equilibria without and with tolling increases in two ranges of μ where N^{s*} is constant. This occurs because, in these ranges, increases in bottleneck capacity do not alter the suburban population at the no-toll equilibrium but increase its population at the equilibrium with tolling. These results are also consistent with those presented in Section 5.2.1.

6 Conclusion

This study has developed a model in which heterogeneous commuters choose their departure time from home and residential locations in a monocentric city with a single bottleneck. By using the properties of the complementarity problem, we systematically examined the spatial distribution of commuters and the effects of time-varying congestion tolling. The results indicate that commuters sort themselves temporally and spatially on the basis of their value of time and their flexibility. Furthermore, imposing an optimal congestion toll alters the urban spatial structure. This finding differs fundamentally from the results obtained by Arnott (1998), who considered homogeneous commuters. Our finding thus also suggests that interactions among heterogeneous commuters change the effects of congestion tolling.

In addition, we used a simple example to demonstrate that imposing a congestion toll without redistributing toll revenues causes urban sprawl, which is opposite to the standard results of traditional location models considering static traffic flow congestion. This difference arises from the following reasons: in the traditional residential location model, imposing a congestion toll makes commuting more expensive; in our model, however, tolling eliminates the queuing congestion; hence, it can make commuting less expensive. We further show that, although congestion tolling generates a Pareto improvement in this example when commuters do not relocate, it leads to an unbalanced distribution of benefits among commuters: rich commuters gain and poor commuters lose from tolling. These results suggest that considering commuter heterogeneity and commuters' residential location choice is important when we examine the efficacy of transportation policies intended to alleviate peak-period congestion.

This study made simplifying assumptions that each commuter traverses only one bottleneck and that rich commuters are more inflexible than poor commuters. Furthermore, although we considered the quasi-linear utility function, it is well known that the income elasticity of demand for land is positive. Therefore, it is important to examine the robustness of our results by analyzing a model with multiple bottlenecks,[15] general heterogeneity,[16] and other utility functions such as Cobb-Douglas utility. In addition, it would be valuable for future research to investigate effects of policies other than optimal congestion tolling, such as step tolls (Arnott et al., 1990a; Laih, 1994, 2004; Lindsey et al., 2012) and transportation demand management measures for alleviating traffic congestion (Mun and Yonekawa, 2006; Takayama, 2015).

A Equivalence between the bid-rent and complementarity approaches

We show that long-run equilibrium conditions (11) coincide with those of the bid-rent approach. The condition (11a) can be rewritten as

$$\begin{cases} N_i(x)\left\{r(x) + r_A - \Psi_i(x, v_i^*)\right\} = 0 \\ N_i(x) \geq 0,\ r(x) + r_A - \Psi_i(x, v_i^*) \geq 0 \end{cases} \quad \forall x \in \mathbb{R}_+, \quad \forall i \in \mathcal{I}, \tag{65a}$$

or equivalently,

$$\begin{cases} r(x) + r_A = \Psi_i(x, v_i^*) & \text{if} \quad N_i(x) > 0 \\ r(x) + r_A \geq \Psi_i(x, v_i^*) & \text{if} \quad N_i(x) = 0 \end{cases} \quad \forall x \in \mathbb{R}_+, \quad \forall i \in \mathcal{I}. \tag{65b}$$

$\Psi_i(x, v_i^*)$ is given by

$$\Psi_i(x, v_i^*) = H^{-1}(v_i^* - y_i(x)), \tag{66}$$

where $H^{-1}(\cdot)$ is the inverse function of $H(\cdot)$. Furthermore, since $\max_{a(x)}\{y_i(x) + f(a(x)) - v_i^*\}/a(x) = \Psi_i(x, v_i^*)$,[17] $\Psi_i(x, v_i^*)$ can be interpreted as the bid-rent function of commuters i. This shows that conditions in (11b), (11c), and (65a) are the equilibrium conditions of the bid-rent approach (see, e.g., Fujita, 1989, Definition 4.2).

[15] Kuwahara (1990) and Akamatsu et al. (2015) have shown the properties of a bottleneck model with multiple bottlenecks.

[16] Liu et al. (2015) proposes a semi-analytical approach for solving an equilibrium of a bottleneck model with general heterogeneous users, which is applicable to our model.

[17] As shown in, e.g., Fujita (1989), this maximization problem defines the bid-rent function.

B Proof of Lemma 2

We can show that, for any $x^a, x^b \in \mathrm{supp}\,(N^*)$, there is no $x^c \in (x^a, x^b)$ such that $N^*(x^c) = 0$, because the indirect utilities of suburban and downtown commuters i are given by (13). Thus, we obtain Lemma 2 (i).

Differentiating the indirect utilities $v_i^s(x)$ and $v_i^d(x)$ with respect to location x, we have

$$\frac{\mathrm{d}v_i^s(x)}{\mathrm{d}x} = \begin{cases} -\alpha_i\tau + h'(N^*(x))\dfrac{\mathrm{d}N^*(x)}{\mathrm{d}x} & \text{if} \quad f'(\frac{1}{N^*(x)}) \geq r_{\mathrm{A}}, \\ -\alpha_i\tau & \text{if} \quad f'(\frac{1}{N^*(x)}) \leq r_{\mathrm{A}}, \end{cases} \tag{67a}$$

$$\frac{\mathrm{d}v_i^d(x)}{\mathrm{d}x} = \begin{cases} -\alpha_i\tau + h'(N^*(x))\dfrac{\mathrm{d}N^*(x)}{\mathrm{d}x} & \text{if} \quad f'(\frac{1}{N^*(x)}) \geq r_{\mathrm{A}}, \\ -\alpha_i\tau & \text{if} \quad f'(\frac{1}{N^*(x)}) \leq r_{\mathrm{A}}. \end{cases} \tag{67b}$$

Therefore, the long-run equilibrium number $N^*(x)$ of commuters residing at x satisfies

$$f'(\tfrac{1}{N^*(x)}) \geq r_{\mathrm{A}} \quad \forall x \in \mathrm{supp}\,(N^*). \tag{68}$$

Furthermore, it follows from long-run equilibrium conditions (14a) and (14c) that $N^*(x)$ also satisfies

$$\begin{cases} f'(\frac{1}{N^*(x)}) > r_{\mathrm{A}} \quad \forall x \in \mathrm{supp}\,(N^*)\backslash\{X^{\mathrm{B}}\}, \\ f'(\frac{1}{N^*(X^{\mathrm{B}})}) = r_{\mathrm{A}}. \end{cases} \tag{69}$$

This completes the proof.

C Proof of Lemma 3

It follows from (26) that $N(x)$ and $\frac{\mathrm{d}N(x)}{\mathrm{d}x}$ are given by

$$N(x) = \begin{cases} h^{-1}(h(N(X_i^s)) + \alpha_i\tau(x - X_i^s)) & \text{if} \quad x \in [X_i^s, X_{i+1}^s], \\ h^{-1}(h(N(X_i^d)) + \alpha_i\tau(x - X_i^d)) & \text{if} \quad x \in [X_i^d, X_{i+1}^d], \end{cases} \tag{70a}$$

$$\frac{\mathrm{d}N(x)}{\mathrm{d}x} = \frac{\alpha_i\tau}{h'(N(x))} \qquad \forall x \in \mathrm{supp}\,(N), \tag{70b}$$

where $h^{-1}(\cdot)$ is the inverse function of $h(\cdot)$. Hence, the population constraints (14b) and (14d) can be rewritten as

$$\begin{aligned} N_i^s &= \int_{X_i^s}^{X_{i+1}^s} N(x)\mathrm{d}x = \int_{N(X_i^s)}^{N(X_{i+1}^s)} N(x)\frac{\mathrm{d}x}{\mathrm{d}N(x)}\mathrm{d}N(x) \\ &= \int_{N(X_i^s)}^{N(X_{i+1}^s)} N(x)\frac{h'(N(x))}{\alpha_i\tau}\mathrm{d}N(x) = \frac{1}{\alpha_i\tau}\int_{N(X_i^s)}^{N(X_{i+1}^s)} \left\{\frac{1}{N(x)^2}f''(\tfrac{1}{N(x)})\right\}\mathrm{d}N(x) \\ &= \frac{1}{\alpha_i\tau}\left\{-r(X_{i+1}^s) + r(X_i^s)\right\}, \end{aligned} \tag{71a}$$

$$N_i^d = \int_{X_i^d}^{X_{i+1}^d} N(x)\mathrm{d}x = \frac{1}{\alpha_i\tau}\left\{-r(X_{i+1}^d) + r(X_i^d)\right\}. \tag{71b}$$

Since $r(X^s_{I+1}) = r_A$ and $r(X^d_{I+1}) = r(d) + r_A$, we have Lemma 3.

D Proof of Proposition 6

The KKT conditions of problem (50) are given by

$$\begin{cases} n_i(x,t)\Big[u(z_i(x,t),a_i(x,t)) + \lambda\left\{w_i - z_i(x,t) - r_{\mathrm{A}}a_i(x,t) - c_i(x,t)\right\} - \eta(x)a_i(x,t) - \rho(t) - \nu_i\Big] = 0 \\ n_i(x,t) \geq 0, \quad u(z_i(x,t),a_i(x,t)) + \lambda\left\{w_i - z_i(x,t) - r_{\mathrm{A}}a_i(x,t) - c_i(x,t)\right\} - \eta(x)a_i(x,t) - \rho(t) - \nu_i \leq 0, \end{cases} \tag{72a}$$

$$\begin{cases} z_i(x,t)n_i(x,t)(1-\lambda) = 0 \\ z_i(x,t) \geq 0, \quad n_i(x,t)(1-\lambda) \leq 0, \end{cases} \tag{72b}$$

$$\begin{cases} a_i(x,t)n_i(x,t)\left\{f'(a_i(x,t)) - \eta(x) - r_{\mathrm{A}}\right\} = 0, \\ a_i(x,t) \geq 0, \quad n_i(x,t)\left\{f'(a_i(x,t)) - \eta(x) - r_{\mathrm{A}}\right\} \leq 0, \end{cases} \tag{72c}$$

$$\begin{cases} \lambda \sum_{i\in\mathcal{I}} \iint \Big\{w_i - z_i(x,t) - r_{\mathrm{A}}a_i(x,t) - c_i(x,t)\Big\} n_i(x,t)\,\mathrm{d}t\mathrm{d}x = 0 \\ \lambda \geq 0, \quad \sum_{i\in\mathcal{I}} \iint \Big\{w_i - z_i(x,t) - r_{\mathrm{A}}a_i(x,t) - c_i(x,t)\Big\} n_i(x,t)\,\mathrm{d}t\mathrm{d}x \geq 0, \end{cases} \tag{72d}$$

$$\begin{cases} \eta(x)\left[1 - \sum_{i\in\mathcal{I}} \int a_i(x,t)n_i(x,t)\mathrm{d}t\right] = 0 \\ \eta(x) \geq 0, \quad 1 - \sum_{i\in\mathcal{I}} \int a_i(x,t)n_i(x,t)\mathrm{d}t \geq 0, \end{cases} \tag{72e}$$

$$\begin{cases} \rho(t)\left[\mu - \sum_{i\in\mathcal{I}} \int n_i(x,t)\mathrm{d}x\right] = 0 \\ \rho(t) \geq 0, \quad \mu - \sum_{i\in\mathcal{I}} \int n_i(x,t)\mathrm{d}x \geq 0, \end{cases} \tag{72f}$$

$$N_i - \iint n_i(x,t)\mathrm{d}x\mathrm{d}t = 0, \quad \nu_i \geq 0, \tag{72g}$$

where λ, $\eta(x)$, $\rho(t)$, and ν_i are Lagrange multipliers.

These conditions lead to $\lambda = 1$. It follows from this and $\lim_{a\to 0} f'(a) = \infty$ that condition (72c) can be rewritten as

$$\begin{cases} n_i(x,t)\left\{f'(a_i(x,t)) - \eta(x) - r_{\mathrm{A}}\right\} = 0, \\ n_i(x,t) \geq 0, \quad f'(a_i(x,t)) - \eta(x) - r_{\mathrm{A}} \leq 0. \end{cases} \tag{73}$$

This condition is equivalent to

$$\begin{cases} a_i(x,t) = g(\eta(x) + r_{\mathrm{A}}) & \text{if} \quad n_i(x,t) \geq 0, \\ a_i(x,t) \geq g(\eta(x) + r_{\mathrm{A}}) & \text{if} \quad n_i(x,t) = 0. \end{cases} \tag{74}$$

Because $f(a) - af'(a)$ is monotonically increasing with increases in a, we can rewrite condition (72a) as

$$\begin{cases} w_i + f(g(\eta(x) + r_{\mathrm{A}})) - \{\eta(x) + r_{\mathrm{A}}\}g(\eta(x) + r_{\mathrm{A}}) - \alpha_i\tau x - s_i(t - t^*) - \rho(t) - \nu_i = 0 & \text{if} \quad n_i(x,t) \geq 0, \\ w_i + f(g(\eta(x) + r_{\mathrm{A}})) - \{\eta(x) + r_{\mathrm{A}}\}g(\eta(x) + r_{\mathrm{A}}) - \alpha_i\tau x - s_i(t - t^*) - \rho(t) - \nu_i \leq 0 & \text{if} \quad n_i(x,t) = 0. \end{cases} \tag{75}$$

Since this condition is separable with respect to t and x, we have

$$\begin{cases} c_i^* - s_i(t - t^*) - \rho(t) = 0 & \text{if} \quad \int n_i(x,t)\mathrm{d}x \geq 0, \\ c_i^* - s_i(t - t^*) - \rho(t) \leq 0 & \text{if} \quad \int n_i(x,t)\mathrm{d}x = 0, \end{cases} \tag{76}$$

$$\begin{cases} w_i + f(g(\eta(x) + r_\mathrm{A})) - \{\eta(x) + r_\mathrm{A}\}g(\eta(x) + r_\mathrm{A}) - \alpha_i \tau x - c_i^* - \nu_i = 0 & \text{if} \quad \int n_i(x,t)\mathrm{d}t \geq 0, \\ w_i + f(g(\eta(x) + r_\mathrm{A})) - \{\eta(x) + r_\mathrm{A}\}g(\eta(x) + r_\mathrm{A}) - \alpha_i \tau x - c_i^* - \nu_i \leq 0 & \text{if} \quad \int n_i(x,t)\mathrm{d}t = 0. \end{cases} \tag{77}$$

Furthermore, conditions (72e) and (72g) can be represented as

$$\begin{cases} \eta \left[1 - g(\eta(x) + r_\mathrm{A}) \sum_{i \in \mathcal{I}} \int n_i(x,t)\mathrm{d}t\right] = 0, \\ \eta(x) \geq 0, \quad 1 - g(\eta(x) + r_\mathrm{A}) \sum_{i \in \mathcal{I}} \int n_i(x,t)\mathrm{d}t \geq 0, \end{cases} \tag{78}$$

$$N_i - \int \left\{ \int n_i(x,t)\mathrm{d}t \right\} \mathrm{d}x = 0, \quad \nu_i \geq 0. \tag{79}$$

Therefore, KKT conditions in (72) can be rewritten as

$$\begin{cases} c_i^* - s_i(t - t^*) - \rho(t) = 0 & \text{if} \quad n_i(t) \geq 0, \\ c_i^* - s_i(t - t^*) - \rho(t) \leq 0 & \text{if} \quad n_i(t) = 0, \end{cases} \tag{80a}$$

$$\begin{cases} \rho(t) \left[\mu - \sum_{i \in \mathcal{I}} n_i(t)\right] = 0 \\ \rho(t) \geq 0, \quad \mu - \sum_{i \in \mathcal{I}} n_i(t) \geq 0, \end{cases} \tag{80b}$$

$$\int n_i(t)\mathrm{d}t = \int_d^\infty N_i(x)\mathrm{d}x, \tag{80c}$$

$$\begin{cases} w_i + f(g(\eta(x) + r_\mathrm{A})) - \{\eta(x) + r_\mathrm{A}\}g(\eta(x) + r_\mathrm{A}) - \alpha_i \tau x - c_i^* - \nu_i = 0 & \text{if} \quad N_i(x) \geq 0, \\ w_i + f(g(\eta(x) + r_\mathrm{A})) - \{\eta(x) + r_\mathrm{A}\}g(\eta(x) + r_\mathrm{A}) - \alpha_i \tau x - c_i^* - \nu_i \leq 0 & \text{if} \quad N_i(x) = 0. \end{cases} \tag{80d}$$

$$\begin{cases} \eta \left[1 - g(\eta(x) + r_\mathrm{A}) \sum_{i \in \mathcal{I}} N_i(x)\right] = 0, \\ \eta(x) \geq 0, \quad 1 - g(\eta(x) + r_\mathrm{A}) \sum_{i \in \mathcal{I}} N_i(x) \geq 0, \end{cases} \tag{80e}$$

$$N_i = \int N_i(x)\mathrm{d}x. \tag{80f}$$

This proves Proposition 6.

References

Akamatsu, T. (2007) "Tradable network permits: A new scheme for the most efficient use of network capacity," *mimeograph*.

Akamatsu, T., Wada, K., and Hayashi, S. (2015) "The corridor problem with discrete multiple bottlenecks," *Transportation Research Part B: Methodological*, Vol. 81, No. 3, pp. 808–829.

Akamatsu, T., Fujishima, S., and Takayama, Y. (2017) "Discrete-space agglomeration models with social interactions: Multiplicity, stability, and continuous limit of equilibria," *Journal of Mathematical Economics*, Vol. 69, pp. 22–37.

Alonso, W. (1964) *Location and Land Use*: Harvard University Press.

Anas, A., Arnott, R., and Small, K. A. (1998) "Urban spatial structure," *Journal of Economic Literature*, Vol. 36, No. 3, pp. 1426–1464.

Anas, A. and Rhee, H.-J. (2007) "When are urban growth boundaries not second-best policies to congestion toll?," *Journal of Urban Economics*, Vol. 61, No. 2, pp. 263–286.

Arnott, R. (1998) "Congestion tolling and urban spatial structure," *Journal of Regional Science*, Vol. 38, No. 3, pp. 495–504.

Arnott, R., de Palma, A., and Lindsey, R. (1990a) "Departure time and route choice for the morning commute," *Transportation Research Part B: Methodological*, Vol. 248, No. 3, pp. 209–228.

Arnott, R., de Palma, A., and Lindsey, R. (1990b) "Economics of a bottleneck," *Journal of Urban Economics*, Vol. 27, No. 1, pp. 111–130.

Arnott, R., de Palma, A., and Lindsey, R. (1992) "Route choice with heterogeneous drivers and group-specific congestion costs," *Regional Science and Urban Economics*, Vol. 22, No. 1, pp. 71–102.

Arnott, R., de Palma, A., and Lindsey, R. (1993) "A structural model of peak-period congestion: A traffic bottleneck with elastic demand," *American Economic Review*, Vol. 83, No. 1, pp. 161–179.

Arnott, R., de Palma, A., and Lindsey, R. (1994) "The welfare effects of congestion tolls with heterogeneous commuters," *Journal of Transport Economics and Policy*, Vol. 28, No. 2, pp. 139–161.

van den Berg, V. A. C. and Verhoef, E. T. (2011) "Winning or losing from dynamic bottleneck congestion pricing? The distributional effects of road pricing with heterogeneity in values of time and schedule delay," *Journal of Public Economics*, Vol. 95, No. 7-8, pp. 983–992.

Blanchet, A., Mossay, P., and Santambrogio, F. (2016) "Existence and uniqueness of equilibrium for a spatial model of social interactions," *International Economic Review*, Vol. 57, No. 1, pp. 31–60.

Brueckner, J. K. (2007) "Urban growth boundaries: An effective second-best remedy for unpriced traffic congestion?," *Journal of Housing Economics*, Vol. 16, No. 3-4, pp. 263–273.

Duranton, G. and Puga, D. (2015) "Urban land use," in Duranton, G., Henderson, J. V., and Strange, W. C. eds. *Handbook of Regional and Urban Economics*, Vol. 5: North Holland, pp. 405–465.

Fosgerau, M. and Engelson, L. (2011) “The value of travel time variance,” *Transportation Research Part B: Methodological*, Vol. 45, No. 1, pp. 1–8.

Fosgerau, M., Kim, J., and Ranjan, A. (2016) “Vickrey meets Alonso: Commute scheduling and congestion in a monocentric city,” *mimeograph*.

Fosgerau, M. and Lindsey, R. (2013) “Trip-timing decisions with traffic incidents,” *Regional Science and Urban Economics*, Vol. 43, No. 5, pp. 764–782.

Fosgerau, M. and de Palma, A. (2012) “Congestion in a city with a central bottleneck,” *Journal of Urban Economics*, Vol. 71, No. 3, pp. 269–277.

Fosgerau, M. and Small, K. A. (2014) “Endogenous scheduling preferences and congestion,” *Working Paper, University of California at Irvine*.

Fujita, M. (1989) *Urban Economic Theory: Land Use and City Size*: Cambridge University Press.

Fujita, M. and Thisse, J.-F. (2013) *Economics of Agglomeration: Cities, Industrial Location, and Globalization*: Cambridge University Press, 2nd edition.

Glaeser, E. L., Kahn, M. E., and Rappaport, J. (2008) “Why do the poor live in cities? The role of public transportation,” *Journal of Urban Economics*, Vol. 63, No. 1, pp. 1–24.

Gubins, S. and Verhoef, E. T. (2014) “Dynamic bottleneck congestion and residential land use in the monocentric city,” *Journal of Urban Economics*, Vol. 80, pp. 51–61.

Hall, J. D. (2015) “Pareto improvements from lexus lanes: the effects of pricing a portion of the lanes on congested highways,” *University of Tronto Working Paper 548*.

He, F., Yin, Y., Shirmohammadi, N., and Nie, Y. M. (2013) “Tradable credit schemes on networks with mixed equilibrium behaviors,” *Transportation Research Part B: Methodological*, Vol. 57, pp. 47–65.

Hendrickson, C. and Kocur, G. (1981) “Schedule delay and departure time decisions in a deterministic model,” *Transportation Science*, Vol. 15, No. 1, pp. 62–77.

Iryo, T. and Yoshii, T. (2007) “Equivalent optimization problem for finding equilibrium in the bottleneck model with departure time choices,” in Heydecker, B. G. ed. *Mathematics in Transport*: Elsevier, pp. 231–244.

Joshi, K. K. and Kono, T. (2009) “Optimization of floor area ratio regulation in a growing city,” *Regional Science and Urban Economics*, Vol. 39, No. 4, pp. 502–511.

Kanemoto, Y. (1980) *Theories of Urban Externalities*: North-Holland.

Kono, T., Kaneko, T., and Morisugi, H. (2008) “Necessity of minimum floor area ratio regulation: A second-best policy,” *Annals of Regional Science*, Vol. 44, No. 3, pp. 523–539.

Kono, T. and Joshi, K. K. (2012) “A new interpretation on the optimal density regulations: Closed and open city,” *Journal of Housing Economics*, Vol. 21, No. 3, pp. 223–234.

Kuwahara, M. (1990) “Equilibrium queueing patterns at a two-tandem bottleneck during the morning peak,” *Transportation Science*, Vol. 24, No. 3, pp. 217–229.

Laih, C.-H. (1994) "Queueing at a bottleneck with single- and multi-step tolls," *Transportation Research Part A: Policy and Practice*, Vol. 28, No. 3, pp. 197–208.

Laih, C.-H. (2004) "Effects of the optimal step toll scheme on equilibrium commuter behaviour," *Applied Economics*, Vol. 36, No. 1, pp. 59–81.

Lindsey, R. (2004) "Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes," *Transportation Science*, Vol. 38, No. 3, pp. 293–314.

Lindsey, R., van den Berg, V. A. C., and Verhoef, E. T. (2012) "Step tolling with bottleneck queuing congestion," *Journal of Urban Economics*, Vol. 72, No. 1, pp. 46–59.

Liu, Y., Nie, Y. M., and Hall, J. (2015) "A semi-analytical approach for solving the bottleneck model with general user heterogeneity," *Transportation Research Part B: Methodological*, Vol. 71, pp. 56–70.

McCann, P. (2013) *Modern Urban and Regional Economics*: Oxford University Press, 2nd edition.

Mills, E. S. (1967) "An aggregative model of resource allocation in a metropolitan area," *American Economic Review*, Vol. 57, No. 2, pp. 197–210.

Mun, S.-i. and Yonekawa, M. (2006) "Flextime, traffic congestion and urban productivity," *Journal of Transport Economics and Policy*, Vol. 40, No. 3, pp. 329–358.

Muth, R. F. (1969) *Cities and Housing*: University of Chicago Press.

Nie, Y. M. (2012) "Transaction costs and tradable mobility credits," *Transportation Research Part B: Methodological*, Vol. 46, No. 1, pp. 189–203.

Nie, Y. M. and Yin, Y. (2013) "Managing rush hour travel choices with tradable credit scheme," *Transportation Research Part B: Methodological*, Vol. 50, pp. 1–19.

Peer, S. and Verhoef, E. T. (2013) "Equilibrium at a bottleneck when long-run and short-run scheduling preferences diverge," *Transportation Research Part B: Methodological*, Vol. 57, pp. 12–27.

Pines, D. and Kono, T. (2012) "FAR regulations and unpriced transport congestion," *Regional Science and Urban Economics*, Vol. 42, No. 6, pp. 931–937.

Small, K. A. (1982) "The scheduling of consumer activities: work trips," *American Economic Review*, Vol. 72, No. 3, pp. 467–479.

Takayama, Y. (2015) "Bottleneck congestion and distribution of work start times: The economics of staggered work hours revisited," *Transportation Research Part B: Methodological*, Vol. 81, No. 3, pp. 830–847.

Takayama, Y. and Kuwahara, M. (2016) "Bottleneck congestion and residential location of heterogeneous commuters," *MPRA Paper, University Library of Munich, Germany*, No. 68940.

Tseng, Y. Y. and Verhoef, E. T. (2008) "Value of time by time of day: A stated-preference study," *Transportation Research Part B: Methodological*, Vol. 42, pp. 607–618.

Verhoef, E. T., Nijkamp, P., and Rietveld, P. (1997) "Tradeable permits: Their potential in the regulation of road transport externalities," *Environment and Planning B*, Vol. 24, No. 4, pp. 527–548.

Vickrey, W. S. (1969) "Congestion theory and transport investment," *American Economic Review*, Vol. 59, No. 2, pp. 251–260.

Vickrey, W. S. (1973) "Pricing, metering, and efficiently using urban transportation facilities," *Highway Research Record*, Vol. 476, pp. 36–48.

Wada, K. and Akamatsu, T. (2013) "A hybrid implementation mechanism of tradable network permits system which obviates path enumeration: An auction mechanism with day-to-day capacity control," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 60, pp. 94–112.

Wheaton, W. C. (1998) "Land use and density in cities with congestion," *Journal of Urban Economics*, Vol. 43, No. 2, pp. 258–272.

Yang, H. and Wang, X. (2011) "Managing network mobility with tradable credits," *Transportation Research Part B: Methodological*, Vol. 45, No. 3, pp. 580–594.