

# CONGESTION TOLLING AND URBAN SPATIAL STRUCTURE

Richard Arnott

Department of Economics

Boston College

Chestnut Hill, MA 02167

U.S.A.

*Acknowledgments:* The author would like to thank the editors and referees for helpful comments on an earlier version of the paper, and Alexander Kalenik for excellent research assistance.

*Abstract:* According to the standard model of urban traffic congestion and urban spatial structure, congestion tolling results in a more concentrated city. In recent years, a new model of rush hour urban auto congestion has been developed which incorporates trip-timing decisions — the bottleneck model. In the simplest bottleneck model, optimal congestion tolling without toll revenue redistribution has no effect on trip price since the efficiency gains exactly equal the toll revenue collected. Optimal congestion tolling then has *no* effect on urban spatial structure. This paper formalizes this result and extends it somewhat.

JEL code: R14, R41

## Congestion Tolling and Urban Spatial Structure

According to the standard model of urban traffic congestion *cum* urban spatial structure (e.g. Kanemoto (1980)), the first-round effect of congestion tolling is to increase the transport costs of living at less accessible locations, which steepens the rent and (population) density functions and results in a denser and more concentrated urban spatial structure. Subsequent effects depend on whether the city's population is fixed (the "closed city") or variable (the "open city") and the extent to which urban land rents and toll revenues are redistributed to the population. But the overall effect typically remains a denser urban spatial structure.

In this standard model, congestion tolling influences only two margins of consumer choice — lot size and residential location. In recent years, a new model of urban traffic congestion, which was introduced by Vickrey (1969), has been elaborated (e.g., Arnott, de Palma, and Lindsey (1993a)) — the bottleneck model. The bottleneck model focuses on another margin of consumer choice — the time at which to travel.<sup>1</sup> The bottleneck model has not yet been incorporated into a model of urban spatial structure.

This paper has two related objectives. The more general aim is to incorporate the bottleneck model into a model of urban spatial structure, albeit a very simple one. The more specific aim is to show that, in contrast to the standard model, congestion tolling in the bottleneck model can actually cause urban spatial structure to become *less* concentrated. The policy insight is that, when travel timing decisions are considered, congestion tolling may have less pronounced effects on urban spatial structure than was previously thought.

### 1. THE STANDARD MODEL

This paper employs a highly simplified monocentric city model of urban spatial structure. There are two islands, downtown and suburbia, connected by a causeway which is subject to traffic congestion, as shown in Figure 1. The areas of downtown and suburbia are  $\bar{A}_1$  and  $\bar{A}_2$  respectively. The city's population comprises  $N$  identical individuals, all of whom work downtown. The  $N_1$  who live downtown incur no commuting costs. The  $N_2$  who live in suburbia have to incur the costs of traveling on the causeway. Tastes are given by the utility function  $U(X, T)$  where  $X$  is a composite numéraire commodity and  $T$  lot size.

INSERT FIGURE 1 HERE

Eight cases could be considered, according to whether the city is open or closed, whether land is owned internally or by absentee landlords, and whether the revenue from congestion tolling is redistributed. But since the redistribution of land rents and toll revenues generates only income effects, the main points of the paper can be made by considering only the open and closed city cases, and for each identifying the direction of the income effect.

Thus far the model description applies to both the standard and bottleneck models. The two models differ in their treatment of the congestion technology. For the moment, only the congestion technology of the standard model shall be described. The bottleneck congestion technology, which is more complex, shall be described in the next section.

The congestion technology for the standard model is very simple. It is assumed that the cost of travel between downtown and suburbia is a function of only the number of individuals who travel between the two locations,  $c(N_2)$ ,  $c' > 0$ . Then the total cost of travel is  $N_2c(N_2)$ , the marginal social cost of travel is  $d(N_2c(N_2))/dN_2 = c(N_2) + N_2(dc(N_2)/dN_2)$ , and the congestion externality is  $N_2(dc(N_2)/dN_2)$ . The optimal Pigouvian congestion toll,  $\tau^*$ , equals the congestion externality, evaluated at the optimum, and is increasing in the number of suburban residents.

The open-city model is characterized by the following four equations:

$$(1) \quad N_1T(R_1, \bar{U}) - \bar{A}_1 = 0$$

$$(2) \quad N_2T(R_2, \bar{U}) - \bar{A}_2 = 0$$

$$(3) \quad V(R_1, I) - \bar{U} = 0$$

$$(4) \quad V(R_2, I - c(N_2) - \theta\tau^*(N_2)) - \bar{U} = 0$$

where  $R_i$  is land rent on island  $i$ ,  $I$  lump-sum income,  $\theta$  a shift parameter with  $\theta = 0$  characterizing no congestion tolling and  $\theta = 1$  full congestion tolling,  $\bar{U}$  the exogenous utility level for the open city,  $T(\cdot)$  the compensated demand for land, and  $V(\cdot)$  the indirect utility function. Equations (1) and (2) are the land market clearing conditions, and (3) and (4) the migration equilibrium conditions.

Of central concern is the effect of an increase in  $\theta$  on the steepness of the density function, which is crudely measured in the paper as  $T_2/T_1$  or  $N_1/N_2$ . The  $\theta$ -row in Table 1 gives the direction of changes caused by an increase in  $\theta$ , holding  $I$  constant and thereby ignoring income effects; the  $I$ -row gives the direction of changes caused by an increase in

lump-sum income. Consider first the case where toll revenues and land rents are not redistributed, so that lump-sum income is unaffected by  $\theta$ . The increase in  $\theta$  has no effect on downtown residents, so that  $R_1$ ,  $N_1$ , and hence  $T_1$  are unchanged. But the higher congestion toll decreases the disposable (after transport costs plus the toll) income of suburban residents. To restore utility to its exogenous level, suburban rent falls which induces larger suburban lots and a reduced suburban population. Hence, there is an unambiguous steepening of the density function.

INSERT TABLE 1 HERE

Now consider the case where the increase in  $\theta$  has an income effect, in addition to the substitution effect of the previous case. The increase in  $\theta$  causes aggregate land rents and toll revenues to change, and, in contrast to the previous case, some portion of land rents and/or toll revenues accrue to the city's residents. Where  $\chi$  denotes one of  $R_1$ ,  $N_1$ ,  $R_2$ , and  $N_2$ ,

$$\frac{d\chi}{d\theta} = \frac{\partial\chi}{\partial\theta} + \frac{\partial\chi}{\partial I} \frac{dI}{d\theta}$$

The first term on the RHS is the substitution effect given in the  $\theta$ -row of Table 1, and the second term is the income effect. From Table 1,  $\frac{\partial\chi}{\partial I} > 0$  for  $R_1$ ,  $N_1$ ,  $R_2$ , and  $N_2$ . Thus, the sign of the income effect is the same as the sign of  $\frac{dI}{d\theta}$ , which is ambiguous, depending on tastes, the congestion technology, the pattern of land ownership, and policy with respect to the redistribution of toll revenues. Even when the income effect is negative, it is unlikely to dominate the substitution effect. Thus, the result that heavier tolls steepen the density function should normally continue to hold when income effects are accounted for.

The closed-city model is characterized by the following five equations:

$$N_1 T(R_1, U) - \bar{A}_1 = 0$$

$$N_2 T(R_2, U) - \bar{A}_2 = 0$$

$$V(R_1, I) - U = 0$$

$$V(R_2, I - c(N_2) - \theta\tau^*(N_2)) - U = 0$$

$$N_1 + N_2 - \bar{N} = 0$$

where  $\bar{N}$  is the city's fixed population. It is assumed that land is a normal good in the sense that  $\partial T(\cdot)/\partial U > 0$ . The case where  $\theta$  is increased, holding  $I$  constant, is considered to begin.

Suppose that the increase in  $\theta$  had no effect on utility. From the results for the open-city case, this would imply no change in  $T_1$  and an increase in  $T_2$ . Since the city's population is fixed, there would be an excess demand for land. Utility must fall to reestablish equilibrium in the land market. Since downtown residents experience a fall in utility with no change in disposable income,  $R_1$  rises. The fall in  $U$  combined with the rise in  $R_1$  causes  $T_1$  to fall and  $N_1$  to rise. The suburban population must therefore fall. Since the disposable income of suburban residents falls, this is consistent with a fall in utility only if  $R_2$  falls. Thus, the density function steepens. The story is more complicated when income effects are present, but, as in the open-city case, the presumption is that the overall effect is qualitatively the same as the substitution effect.

INSERT TABLE 2 HERE

In conclusion: In both the open-city and closed-city variants of the standard model, the substitution effect of heavier congestion tolls causes the density function to steepen. In both cases, though for different reasons, the income effect of heavier tolls on the slope of the density function is ambiguous in sign, and is unlikely to dominate the substitution effect. Hence, the standard model leads to the presumption that heavier congestion tolls result in greater concentration of economic activity within a metropolitan area. And thus runs the conventional wisdom.

## 2. THE BOTTLENECK MODEL

The standard model is static. It has been variously interpreted as describing a stationary state in which traffic flow is uniform over the day or at least over the rush hour (of presumably fixed length) or as being a reduced form which implicitly captures the dynamic equilibrium of traffic flow over the rush hour. The bottleneck model, in contrast, is explicitly dynamic, deriving the time pattern of congestion over the rush hour.

Here only the simplest variant of the bottleneck model shall be presented. In the morning rush hour, each of the  $N_2$  identical individuals living in suburbia would like to arrive at work downtown at the same time  $t^*$ . But since the causeway has a bottleneck of fixed flow capacity  $s$ , this is physically impossible. As a result, some individuals arrive at work early and others late, which entails a "schedule delay" cost. Individuals also incur travel costs that include a fixed component, which is set equal to zero to simplify the algebra, and a variable component — the cost of time spent queuing behind the bottleneck. The essential innovation of the bottleneck model is to treat individuals' decisions concerning when to depart from home.

And the essential insight is that equilibrium requires that trip price, which includes schedule delay cost, travel cost and the toll, be constant over the departure interval. Thus, with a zero or uniform toll for example, since schedule delay cost is higher for individuals who arrive at work more early or more late, travel cost must be higher for individuals arriving at work closer to the common desired arrival time. Generally, the queuing pattern over the rush hour must be such that trip price is equalized over the departure time interval. And this queuing pattern implies a particular time pattern of departures over the rush hour. To simplify the analysis, it is assumed that trip price is linear in its components:

$$p_t = \alpha(\text{queuing time})_t + \beta(\text{time early})_t^+ + \gamma(\text{time late})_t^+ + (\text{toll})_t$$

where  $p_t$  is the trip price of an individual who arrives at time  $t$ ;  $\alpha$ ,  $\beta$ , and  $\gamma$  are the shadow values of queuing time, time early, and time late respectively, with  $\alpha > \beta$ ; and  $(x)^+ \equiv (x \text{ if } x > 0, 0 \text{ otherwise})$ .

The no-toll equilibrium is now solved for. Let  $t_0$  denote the time at which the first individual arrives at work,  $t_L$  the time of the last arrival, and  $T(t)$  variable travel time. The bottleneck must be used to capacity throughout the rush hour; otherwise, an individual could depart in the interior of the rush hour, incurring zero queuing costs and lower schedule delay costs than either the first or last person to arrive, which is inconsistent with equilibrium. This implies that  $t_L - t_0 = N_2/s$ . The first person to arrive incurs only schedule delay costs of  $\beta(t^* - t_0)$ ; the last person to arrive also faces no queue, and incurs schedule delay costs of  $\gamma(t_L - t^*)$ . Since the trip prices of the first and last person to arrive are equal in equilibrium,  $t^* - t_0 = (\gamma/(\gamma + \beta))(N_2/s)$ , and hence the equilibrium price is  $\bar{p} = \beta\gamma N_2/((\gamma + \beta)s)$ . Thus, for early arrivals,  $\bar{p} = p_t = \alpha T(t) + \beta(t^* - t)$ , while for late arrivals  $\bar{p} = p_t = \alpha T(t) + \gamma(t - t^*)$ , from which the evolution of the queue and the time pattern of departures over the rush hour can be solved straightforwardly. Since there is no toll, travel cost equals trip price. Thus, the no-toll travel cost function is  $c(N_2) = \beta\gamma N_2/((\gamma + \beta)s)$ .

Now consider imposing a time-varying toll which, as a function of arrival time, equals the queuing cost in the no-toll equilibrium. Such a toll eliminates the queue, without altering the departure interval. *Each individual then faces the same trip price as in the no-toll equilibrium*, and since these trip prices are equalized no individual has an incentive to change her behavior. Furthermore, such a toll decentralizes the optimum. Queuing costs are eliminated and total schedule delay costs are minimized since the first and last person to arrive face the same schedule delay cost and the bottleneck is used to capacity over the rush hour. The social cost of travel is reduced, without any reduction in the amount of travel, by

reallocating traffic over the rush hour. When toll revenues are not redistributed, the optimal time-varying toll has no effect on trip price or individuals' income, and hence has no effect on urban spatial structure and land rents, whether the city is closed or open. When toll revenues are redistributed in whole or in part, imposition of the toll has only an income effect (a change in  $I$ ) the comparative static properties of which are similar to those in the standard model.

The same line of argument applies more generally. If the switch from one toll to another has no effect on trip price, the only effect of the toll on spatial structure is due to the income effect resulting from the possible redistribution of toll revenue. To keep the analysis comparable to that of the standard model, attention is focused on families of toll functions that vary monotonically from the no-toll situation ( $\theta = 0$ ) to the optimal time-varying toll ( $\theta = 1$ ), and do not alter the equilibrium trip price. Refer to a family in the set as  $f \in \mathcal{F}$ . Then the analog to Table 1 for the open city and to Table 2 for the closed city are as shown below.

INSERT TABLE 3 HERE

INSERT TABLE 4 HERE

The central result is recorded in the following:

*Proposition:* In the above simple urban model with bottleneck congestion, for any family of toll functions  $f \in \mathcal{F}$ :

- a) In the absence of income effects, an increase in the toll has no effect on urban spatial structure.
- b) With income effects, an increase in the toll affects urban spatial structure only via these income effects which are ambiguous in sign.

Thus, the bottleneck model suggests that congestion tolling need not cause spatial concentration of economic activity and may actually cause dispersion.

### 3. AN EXAMPLE

The example starts out with a base case without a congestion toll and without redistribution of land rents, that is the same for the standard model and for the bottleneck model. Also, the no-toll travel cost function,  $\hat{c}(N_2)$ , is the same for the standard and bottleneck models. The respective optimal congestion tolls are then applied in both models, and the standard-model and bottleneck-model spatial structures ( $T_2/T_1$ ) are compared for four

cases: open city, no redistribution of toll revenue; open city, full redistribution of toll revenue; closed city, no redistribution of toll revenue; closed city, full redistribution of toll revenue.

The parameters and functions are chosen to be broadly realistic. For example, individuals spend 10% of their disposable incomes on land rent, and individual income is \$20000.

- Parameters and functions

Units are miles, dollars, and years.

$$\begin{array}{lll}
 U = X^{.90}T^{.10} & \bar{A}_1 = 20 & \hat{c}(N_2) = .057102 N_2 \\
 I = 20000 & \bar{A}_2 = 80 &
 \end{array}$$

Open city:  $\bar{U} = 3089.95$

Closed city:  $\bar{N} = 100,000$

Base case:  $\theta = 0$

- Results

INSERT TABLE 5 HERE

INSERT TABLE 6 HERE

INSERT TABLE 7 HERE

The base case, in which there is no toll, is displayed in Table 5. The parameters and functions were chosen such that in this base case the city's population is 100,000, evenly divided between downtown and suburbia. Since suburbia has four times the land area of downtown, suburban lot sizes are four times larger than those downtown, and the steepness of the density function,  $T_2/T_1$ , equals 4.0.

Tables 6 and 7 show how urban spatial structure is altered by imposition of the respective optimal congestion tolls. Note that, with standard congestion, imposition of the optimal toll does not alter the travel cost function, so that  $\hat{p}(N_2) = \hat{c}(N_2) + N_2(\partial\hat{c}(N_2)/\partial N_2) = 2\hat{c}(N_2)$ , while with bottleneck congestion  $\hat{p}(N_2) = \hat{c}(N_2)$  — as explained above.

Table 6 gives the results for an open city. The left side of the Table treats the case where toll revenues are not redistributed, so that only the substitution effect of the toll is



operative. In accord with Table 1, with the standard congestion technology the toll causes population downtown to remain the same, population in suburbia to fall, and the density function to steepen. And in accord with Table 3, with bottleneck congestion the toll has no effect on urban spatial structure or population. The right side of the Table treats the case where toll revenues are redistributed, and captures both substitution and income effects of the toll. Toll redistribution increases lump-sum income. Relative to the no-redistribution case: i) since utility is fixed, rents are higher, lot sizes smaller, and populations larger, both downtown and in suburbia and for both congestion technologies; and ii)  $T_2/T_1$  increases with both forms of congestion. Note finally that the toll with redistribution causes the city's population to increase significantly more with bottleneck congestion than with the standard form of congestion. This reflects the substantially greater efficiency gains from tolling under bottleneck congestion that derive from the rescheduling of trips, which the standard model ignores.

Table 7 is the same as Table 6 but for a closed city. In accord with Table 2, with the standard congestion technology imposition of the optimal toll without redistribution of revenues causes population downtown to rise and population in suburbia to fall, implying steepening of the density function. And in accord with Table 4, with bottleneck congestion and without redistribution of revenues the toll has no effect on urban spatial structure. With both forms of congestion, the income effect due to the redistribution of toll revenues flattens the density function.<sup>2</sup> Interestingly, with bottleneck congestion, imposition of the optimal toll along with redistribution of the toll revenue causes urban spatial structure to become *more* dispersed. The substitution effect is inoperative and the income effect flattens the density function.

#### 4. CONCLUSION

The conventional wisdom is that congestion tolling would lead to the spatial concentration of economic activity. The intuition is that travel would be more expensive with the congestion toll in place, which would discourage travel and encourage building at higher density. This intuition has been formalized by urban economists using a combination of the monocentric city model and the standard model of congestion which assumes that travel costs are simply a function of mean traffic flow.

In recent years, a new model of traffic congestion — the bottleneck model — has been developed that focuses on individuals' decisions concerning when to travel and on the equilibrium time pattern of congestion over the rush hour. According to this model, the

imposition of an optimal congestion toll does not make travel more expensive. By inducing a reallocation of traffic over the rush hour, the toll reduces travel costs by exactly the amount of the toll revenue collected. Thus, imposition of the toll generates efficiency gains without increasing trip price. Application of the toll therefore affects the spatial concentration of economic activity only via the income effect resulting from the possible redistribution of toll revenue.

This paper contrasted the effects of congestion tolling on urban spatial structure under the two alternative models of traffic congestion, employing a very simple characterization of urban spatial structure — two islands, downtown and suburbia, connected by a congestible causeway. The congestion technology employed in the bottleneck model — queuing behind a bottleneck — implies unrealistically that, even with inelastic travel demand, imposition of the optimal toll completely eliminates congestion. The standard model, meanwhile, implies that with fixed lot size and inelastic trip demand, imposition of the optimal toll has no effect on congestion, which is also unrealistic. The true situation probably lies between these two extremes. Accordingly, the best guess is that congestion tolling does cause urban spatial structure to become more concentrated, though not by as much as conventional wisdom suggests. And turning this result around, the *absence* of congestion tolling has probably led to less excessive decentralization of economic activity than was hitherto believed.

There are four obvious directions for future research on the effects of congestion tolling on urban spatial structure. The first is to employ more realistic models of congestion which take into account at the same time that i) congestion is not constant over the rush hour and its time pattern can be influenced by tolling, and ii) flow and queuing congestion are both important.<sup>3</sup> The second is to adopt more realistic non-monocentric models of urban spatial structure, perhaps along the lines pioneered by Fujita and Ogawa (1982) but also accounting for the durability of structures. This is important since congestion tolling could alter urban spatial structure qualitatively e.g., by reducing the CBD orientation of cities and inducing more subcentering — a possible effect which existing analyses completely ignore. The third is to undertake a thorough analysis of the income effects from congestion tolling on urban spatial structure. They are complex, which is probably why the previous literature has ignored them. The fourth is to take into account population heterogeneity, which is likely to be quantitatively important since different income-demographic groups tend to live at different locations and to travel at different times.<sup>4</sup>

The main policy insight provided by this paper is that, when account is taken of the ability of congestion tolls to reallocate traffic over the rush hour (which the standard model

does not), the effects of tolling on urban spatial structure are probably less pronounced than was previously thought.

## FOOTNOTES

<sup>1</sup> Since the model has been developed for the morning rush hour, the margin of choice is, more specifically, departure time from home in the morning commute.

<sup>2</sup> This result is not general, but is due to the functions and parameters assumed in the example.

<sup>3</sup> This is proving a hard nut to crack. Some work has been done (e.g., Small and Chu (1997)) which provides alternative state-variable representations of traffic congestion to the bottleneck model, i.e., which captures the effects of traffic congestion earlier in the rush hour on current congestion via a state variable. But no one has yet succeeded in solving a model which respects the “physics” of traffic congestion, in particular the equation of continuity.

<sup>4</sup> Arnott, de Palma, and Lindsey (1989, 1994) examine the bottleneck model with heterogeneous commuters on a road with a single origin and a single destination. Different commuter groups order themselves in a systematic way over the rush hour. And Arnott, de Palma, and Lindsey (1993b) examine the bottleneck model with homogeneous commuters on a road with more than one entry point and a single destination. By combining the two and making endogenous the location of different groups, one could analyze the effects of bottleneck congestion in a monocentric city model with population heterogeneity. Since the analytics would be cumbersome, numerical solution would probably be advisable.

## Reference List

- Arnott, Richard, André de Palma, and Robin Lindsey. 1989. "Schedule Delay and Departure Time Decision with Heterogeneous Commuters," *Transportation Research Record*, 1197, 56-67.
- Arnott, Richard, André de Palma, and Robin Lindsey. 1993a. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand", *American Economic Review*, 83, 161-179.
- Arnott, Richard, André de Palma, and Robin Lindsey. 1993b. "Properties of Dynamic Traffic Equilibrium Involving Bottlenecks, Including a Paradox and Metering", *Transportation Science*, 27, 148-160.
- Arnott, Richard, André de Palma, and Robin Lindsey. 1994. "The Welfare Effects of Congestion Tolls with Heterogeneous Commuters", *Journal of Transport Economics and Policy*, 139-161.
- Fujita, Masahisa and Hidemitsu Ogawa. 1982. "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations", *Regional Science and Urban Economics*, 12, 161-196.
- Kanemoto, Yoshitsugu. 1980. *Theories of Urban Externalities*. Amsterdam: North-Holland.
- Small, Kenneth and Xuehao Chu. 1997. "Hypercongestion", Working Paper No. 96-97-11, Department of Economics, University of California, Irvine.
- Vickrey, William. 1969. "Congestion Theory and Transport Investment", *American Economic Review, Papers and Proceedings*, 59, 251-260.

FIGURE 1: The Urban Spatial Structure.

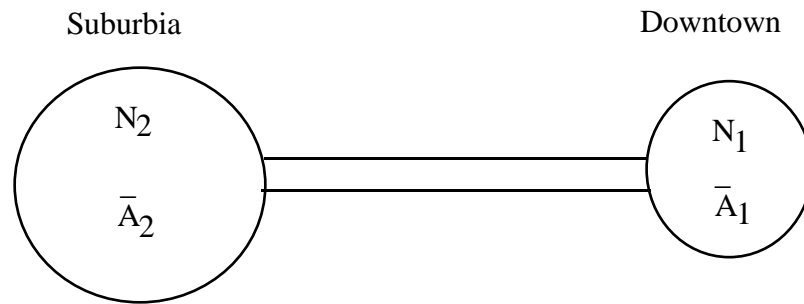


TABLE 1: Comparative Statics, Standard Model, Open City

	$R_1$	$N_1$	$R_2$	$N_2$
$\theta$	0	0	-	-
$I$	+	+	+	+

TABLE 2: Comparative Statics, Standard Model, Closed City

	$R_1$	$N_1$	$R_2$	$N_2$	$U$
$\theta$	+	+	-	-	-
$I$	?	?	?	?	+



TABLE 3: Comparative Statics, Bottleneck Model, Open City

	$R_1$	$N_1$	$R_2$	$N_2$
$\theta$	0	0	0	0
$I$	+	+	+	+

TABLE 4: Comparative Statics, Bottleneck Model, Closed City

	$R_1$	$N_1$	$R_2$	$N_2$	$U$
$\theta$	0	0	0	0	0
$I$	?	?	?	?	+

TABLE 5: Base Case, No Toll

	$N_1$	$N_2$	$T_2/T_1$
Standard	50000	50000	4.0
Bottleneck	50000	50000	4.0

TABLE 6: Open City, Optimal Toll ( $\theta = 1$ )

	No redistribution			Redistribution		
	$N_1$	$N_2$	$T_2/T_1$	$N_1$	$N_2$	$T_2/T_1$
Standard	50000	32176	6.2158	68800	36487	7.5424
Bottleneck	50000	50000	4.0	69776	58138	4.8007

TABLE 7: Closed City, Optimal Toll ( $\theta = 1$ )

	No redistribution			Redistribution		
	$N_1$	$N_2$	$T_2/T_1$	$N_1$	$N_2$	$T_2/T_1$
Standard	65012	34987	7.4327	64266	35735	7.1936
Bottleneck	50000	50000	4.0	49229	50771	3.8785