

Downtown tolls and the distribution of trip lengths

Lewis Lehe^a,

^a*Lecturer, UC Berkeley Department of Civil Engineering
lewis500@berkeley.edu
<http://lewislehe.com>*

Abstract

Currently, all downtown tolling systems are what we call “access tolls,” meaning they charge for gross access to the zone, but tolls levied on distance-traveled are on the horizon. This paper shows how such tolls affect the distribution of trip lengths and what the consequences are for social welfare. A model is presented in which travelers with different trip lengths make a probabilistic choice of whether to drive into a downtown zone governed by a Macroscopic Fundamental Diagram, with the choice probability declining as tolls and travel time rise. Analysis proves and numerical simulation demonstrates that access tolls inefficiently shift the distribution of car trip lengths toward long trips, whereas a distance toll is able to achieve the welfare-maximizing set of car trips.

1. Introduction

Economists usually recommend road tolls on the principle of marginal cost pricing—the idea that drivers should pay for their delay externalities. On a downtown network, what a rigorous application of this principle would mean is a web of differentiated tolls on every link (Beckmann et al., 1956; Yang and Huang, 1998). Obviously, such precision is impossible in practice. Instead, Singapore (Santos et al., 2004), London (Santos, 2008; Leape, 2006), Stockholm (Eliasson, 2014), Milan (Gibson and Carnovale, 2015) and Gothenburg (Börjesson and Kristoffersson, 2015) have all established downtown tolling systems that charge drivers, not on a link-by-link basis, but for the right to access a downtown zone. Hereafter, such systems will be classified as “access tolls,” because gross “access” to the downtown is what the driver buys from the toll authority. Meanwhile, Singapore has plans in place to switch to a satellite-based system that would charge drivers what will be called a “distance toll,” a charge per km travelled (Tan, 2016). Recently, Sadiq Khan, the Mayor of London, proposed replacing the current London Congestion Charge with some form of a distance toll (Topham, 2017).

This paper was inspired by an observation about how these systems are discussed: often the systems’ initial design and later evaluation emphasizes changes in *arrival* flows—the gross number of vehicles passing the tolled zone or crossing a cordon around it. As an example, the Stockholm Congestion Tax

was designed with the goal of reducing traffic across the cordon by 10-15 percent (Eliasson, 2008, p.396). As another, the designers of Singapore’s Area License Scheme set the toll to reduce morning inbound traffic by 25-30 percent (Watson and Holland, 1978, p. 12).

The problem with this emphasis is that essential characteristics of traffic—its circulation, density and speed—depend, not only on the arrival flow’s *magnitude*, but also its *distribution* by trip length. Consider a zone where half of all arriving vehicles park within a few blocks after crossing the boundary, while the other half are ridesharing vehicles that circulate endlessly. That zone could exhibit mild arrival flow alongside severe congestion, while a zone with higher arrival flow might have mild congestion if fewer ridesharing vehicles enter.

Of course, if enacting or raising a toll preserves the arrival flow’s distribution—that is, if it changes the number of trips of all lengths by the same proportion—then change in arrival flow might be a fine proxy of how much the toll change has eased the burden on the street system. But what if toll changes *do* change the trip length distribution? A little thought shows tolls might reasonably do so, since the rise in speed consequent to a toll change saves longer trips more time than short trips. Of course, with a distance toll, the inequality of travel time savings across trip lengths is attenuated by a money penalty on longer trips, but not with an access toll. To investigate such issues, this paper proposes an original static traffic model designed to show how downtown tolls might alter the trip length distribution and to uncover the consequences of that redistribution for social welfare and measurable traffic variables.

In designing such a model, a principal question is how best to represent traffic physics. One option is to have the downtown network be a graph of origins and destinations connected by links, which might have different capacities and wind up with different congestion levels. This is the approach taken by Meng et al. (2012), Liu et al. (2013), Verhoef (2002), Lawphongpanich and Yin (2012) and other studies in the “second-best” literature on feasible downtown tolls; this literature is devoted in large part to solution heuristics and algorithms, owing to the difficulty of finding optimal tolls and equilibrium route flows simultaneously. Similarly, May and Milne (2000) experiments with different toll designs on a simulation of Cambridge. But while potentially useful for engineering real systems, the particularity of network modeling can bury insights that generalize across networks and populations.¹ In fact, for a network model it is not even exactly meaningful to speak of how something affects “long trips” or “short trips,” per se; if there is congestion on one side of the network but not the other, then two trips with the same length might have very different durations. Therefore, to avoid network modelling, we assume that traffic is spread sufficiently evenly across the network that a trip’s duration depends only on its distance and on the network’s space-averaged traffic conditions—not on the trip’s origin,

¹Similarly, Mun et al. (2003) notes, “Although network models are useful for practical applications, they are not suitable for investigating the general properties of the [pricing] problem, since the results depend on the network structures specified for calculations.”

destination or route. Under this assumption, the network exhibits a stable relationship among space-averaged flow, density and speed called the Macroscopic Fundamental Diagram (MFD). Traffic readings have testified to the existence of MFD’s in Yokohama (Geroliminis and Daganzo, 2008) and Toulouse (Buisson and Ladier, 2009) among other places; and the MFD can even be derived analytically from network parameters in some cases (Daganzo and Geroliminis, 2008; Daganzo and Lehe, 2016).

The MFD approach is a popular one for studies of downtown tolling, and models in this vein—e.g., Small and Chu (2003), Arnott (2013), Geroliminis and Levinson (2009)—are sometimes called “bathtub” or “isotropic” models. What most distinguishes the model here from other isotropic models is (i) the static time character, (ii) variable trip lengths, and (iii) elastic demand arising from a random utility model. The static time character was chosen because dynamic models involve schedule preferences not essential to the topic (the question is *whether* certain trips are driven, not *when*) and require certain strong assumptions for tractability². One similar study is Arnott and Inci (2010), which also features a static, isotropic model with elastic demand; if all trip were of the same length, then our model would be a special case of Arnott and Inci’s in which parking is sufficient to preclude cruising. Another similar contribution is the elastic demand (Sec. 3) extension in Fosgerau (2015). In addition to the dynamic character of that model, a major difference here is that our travelers vary not only in their trip lengths but also in the benefit each derives from a car trip. This heterogeneous benefit makes the choice of a traveler with any given trip length probabilistic, so that the aggregate arrival flow of such trips rises smoothly with the trip’s expected utility. As we shall see, this smoothly-changing demand draws out certain inefficiencies of an access toll that are suppressed when either all l -length trips are driven or none are.

The paper is organized as follows: Section 2 sets up the physics and demand mechanics of the model. Section 3 derives the untolled user equilibrium and shows how to frame the model in terms of both density congestion as well as flow congestion. Section 4 derives the model’s social optimum. Sections 5 and 6 show how to insert access and distance tolls, respectively, into the model and derives several useful insights—the most important being that the distance toll can achieve the social optimum but the access toll cannot. Section 7 provides a numerical simulation, while Section 8 concludes.

2. Model setup

This section describes the details of the paper’s main model as composed of two sides: a “physics” side, describing how congestion occurs; and a “demand” side, describing how travelers choose whether to drive. The two sides are then linked by an application of Little’s Law.

²See discussions of tractability in Fosgerau (2015), Daganzo and Lehe (2015), Arnott et al. (2016) and Mariotte et al. (2017).

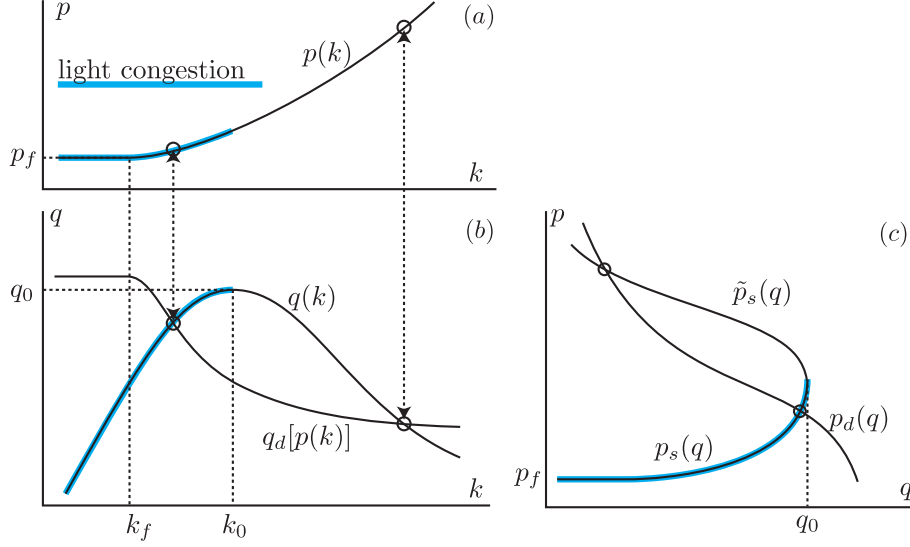


Figure 1: User equilibria along the MFD.

2.1. Physics

The setting is a downtown zone governed by a Macroscopic Fundamental Diagram³. A non-increasing function $v(k)$ (km/min) gives the space-averaged traffic speed v (km/min) for a space-averaged density k (veh/lane-km), but it will be convenient to use the inverse $p(k) := 1/v(k)$ (min/km), which engineers call *pace*. Minutes are the time-unit, rather than hours, to avoid having pace and circulation be small fractions in simulations. As Fig. 1(a) shows, $p(k)$ maintains a free-flow pace, p_f , for $k < k_f$, then rises monotonically.

Let q (veh/min/lane) give the space-averaged “circulation.” By the traffic identity $q = kv$, there is a function

$$q(k) := \frac{k}{p(k)} \quad (1)$$

with the humped shape in Fig. 1(b); it reaches a maximum q_0 at $k = k_0$ ($> k_f$). For densities $k < k_0$, blue highlighting denotes what is sometimes called “light congestion,” where q rises as k and p rise. In light congestion, the addition of cars to the road may diminish speed, but not by so much as to lower the circulation. For $k > k_0$, traffic is “hypercongested,” meaning q falls as p and k rise. In hypercongestion, the addition of cars lowers pace enough that circulation falls.

³For further background on the aggregate definitions involved in MFD modelling, see Daganzo (2007)

2.2. Demand

There is a continuum of “travelers” who might potentially drive every minute. A traveler i is characterized by a tuple (ε_i, l_i) , where l_i (km) is i ’s trip length and ε_i (min) is i ’s “gross benefit” from a car trip, measured in minutes of travel time. The utility of a car trip to traveler i , when there are no tolls and the pace of traffic is p , is

$$U_i := \varepsilon_i - pl_i. \quad (2)$$

A traveler who chooses not to drive earns utility 0. We will call the choice not to drive the “outside option,” which could ostensibly represent traveling by another mode, choosing a route that does not pass through the downtown, or cancelling the trip.

The population is distributed with joint density $f(\varepsilon, l)$. The densities of l and ε are potentially independent, but, arguably, the two properties are likely to be positively correlated: travelers with long trips might be engaged in ridesharing, deliveries or trips to the airport, or they might not have good transit alternatives. The total mass of travelers is λ (veh/min/lane-km), so that

$$\lambda \int_{l_0}^{l_1} \int_{\varepsilon_0}^{\varepsilon_1} f(\varepsilon, l) d\varepsilon dl$$

gives the mass of travelers with gross benefits in $[\varepsilon_0, \varepsilon_1]$ and trip lengths in $[l_0, l_1]$.

Before passing to the derivations, we will clarify two points about interpretation. First, the trip length is limited to *intrazonal* travel only; any travel accomplished en route to the zone or after exiting the zone is outside the scope of the model. Second, due to the static time character, the “population” of travelers cannot be thought of as a single, fixed crowd of people; rather, travelers stand in for a flow of constantly-generated trip opportunities, so the “population” is constantly refreshed with a continuum of travelers from an identical distribution.

2.3. Little’s Law

“Little’s Law” is a classic result from queueing theory that states “under steady state conditions, the average number of items in a queueing system equals the average rate at which items arrive multiplied by the average time that an item spends in the system” (Little and Graves, 2008, p. 82). The relationship turns out to be useful in linking arrival flows—as determined by demand—to traffic conditions.

Let L be the lane-km of the network (the network being equivalent to the “queueing system” of Little’s Law), k^l the traffic density of travelers with l -length trips, and a^l the rate at which such travelers arrive per lane-km. It follows that the number of such travelers on the road at once is Lk^l , and the rate at which they arrive in the whole network is La^l . The time such a traveler

spends in the network, during a steady state with speed v , is l/v , so Little's Law can be stated

$$Lk^l = (La^l) \left(\frac{l}{v} \right). \quad (3)$$

Multiplying by v/L yields

$$q^l = a^l l, \quad (4)$$

where q^l is the circulation of travelers with l -length trips (derived from $q^l = k^l v$). Thus, given knowledge of arriving traffic in the steady state, we can learn the total circulation—even without measuring speed or density directly—by calculating and summing the circulations that correspond to each trip length.

3. User equilibrium (UE)

3.1. Derivation

We now derive the UE that prevails without tolls. Since the outside option grants utility 0, i drives if and only if

$$\varepsilon_i \geq pl_i. \quad (5)$$

Figure 2 shows the resulting division of travelers between drivers and not-drivers, with shading that denotes the population density in l/ε space. The line $D(l) := lp$ divides traffic, such that every traveler above D drives and none below does.

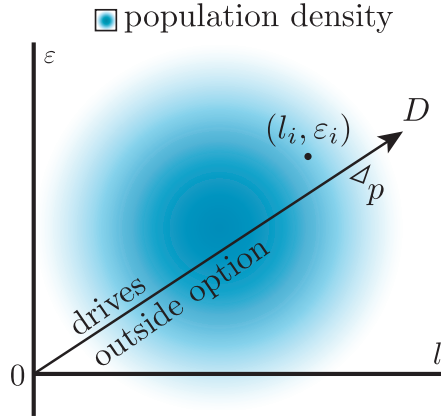


Figure 2: distribution of population in untolled UE

If every traveler satisfying (5) does drive, then, by Little's Law, circulation will be

$$q_d(p) := \lambda \int_0^\infty l \cdot \left(\int_{lp}^\infty f(\varepsilon, l) d\varepsilon \right) dl \text{ (veh/lane/min)} \quad (6)$$

This is the “circulation demanded” and is equal to the total mass above D in the figure, weighted by the l ’s. Note that even though the demand of an individual traveler—from the perspective of a toll authority who cannot measure ε_i —is probabilistic, the *aggregate* circulation demanded, q_d , is deterministic. Also, distinguish q_d from the *arrival* flow demanded:

$$a_d(p) := \lambda \int_0^\infty \left(\int_{l_p}^\infty f(\varepsilon, l) d\varepsilon \right) dl \text{ (veh/lane-km/min)}, \quad (7)$$

which is not weighted by trip lengths.

A UE occurs when the choices expressed in (6) coincide with the physics that give rise to them—i.e., when $k = k_e$ such that

$$q_d[p(k^e)] = q(k^e). \quad (8)$$

The two sides of this equation appear in Fig. 1(b), where their intersections are UE. Note that there do not have to be exactly two user equilibria nor that one occur in light congestion.

So far we have proceeded in terms of k , p and q together, but a more traditional way to frame a static traffic model is in terms of p and q alone—as in Fig. 1(c). This framing, often called “flow congestion” technology, comes from Walters (1961) and invites comparisons to microeconomic concepts such as marginal cost. Let $p_s(q)$ map $q(k) \rightarrow p(k)$ on $[0, k_0]$ (light congestion), let $\tilde{p}_s(q)$ do so on $(k_0, \infty]$ (hypercongestion), and let $p_d(q) := q_d^{-1}$ give the pace inducing q . Figure 1(c) shows $p_s(q)$, $\tilde{p}_s(q)$ and $p_d(q)$. The first two curves, taken together, form a “supply” and “average cost” curve, while the $p_d(q)$ is the inverse demand curve. Intersections are, as before, the model’s user equilibria.

3.2. Choice of equilibria

A question naturally arises as to which equilibria matter, since $q_d(p)$ and $q(k)$ might intersect many times. To answer, we make two assumptions.

Assumption 1. *If a lightly-congested equilibrium exists, then it will happen.*

This assumption makes the most sense if one thinks of the equilibrium as having been reached by the zone filling up from a near-empty state. The lightly-congestion equilibrium would be the first equilibrium reached as k rises, and so traffic might reasonably be expected to stay there.

On the other hand, suppose demand is so high that there are only hypercongested equilibria. There are two types of hypercongested equilibria possible, shown in Fig. 3. Those like E_1 , where $q_d[p(k)]$ cuts $q(k)$ from above (i.e., where $q_d[p(k)]$ is steeper than $q(k)$), will be called *Type 1*; those like E_2 , where the opposite holds, will be called *Type 2*. There is much discussion in the literature as to which type satisfies various stability criteria for static models of traffic on highways; see summaries in Verhoef (1999), Verhoef (2005) and Small and Chu (2003). Our own physics builds on the following assumption:

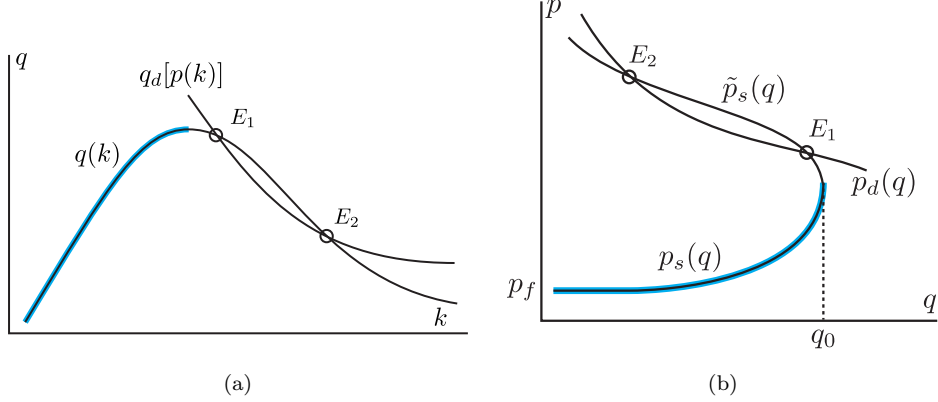


Figure 3: Types of hypercongested equilibria.

Assumption 2. *If a lightly-congested user equilibrium does not exist, the network achieves the Type 1 equilibrium with the lowest density/highest circulation.*

This assumption is justified by the analysis in Arnott and Inci (2010), which shows that Type 1 equilibria satisfy all the stability criteria that the authors consider for a static isotropic model with elastic demand, whereas Type 2 equilibria are stable only in a restricted sense. Their arguments are deep, but a simple mental exercise can make the point intuitive: Reasonably, q_d is a quantity related to instantaneous arrival flow; it tells the circulation one would expect to measure if the magnitude and composition of arrival flow induced by a pace p were sustained. Meanwhile, $q(k)$ measures instantaneous vehicle motion. Thus, at any density k where $q_d[p(k)]$ exceeds $q(k)$, the network is receiving more “work” than it can handle at density k , and so k must rise. Likewise, when $q_d[p(k)] < q(k)$, demand is insufficient to sustain the observed circulation, and so k will decline. Type 1 equilibria, then, are stable, because if k were slightly higher (lower) than their associated densities, then k would decline (rise). Type 2 equilibria are *unstable* by the converse logic.

The proviso that the network settles on the *lowest-density* Type 1 Equilibria is justified by the same logic as Assumption 1: as the network fills up with cars, it settles down at the first stable equilibrium it reaches.

4. Social optimum (SO)

The set of realized car trips in the untolled UE derived above is not, in general, optimal in the sense that it maximizes social welfare. There is almost certainly another set that does better. To derive the social optimum (SO), consider the impact on social welfare of a particular traveler i ’s decision to drive in light congestion. (We only consider light congestion, because hypercongestion cannot be socially optimal; the flow in any hypercongested state could be

achieved in light congestion regime at a lower pace.) Another l_i -length entry every minute raises circulation by l_i and thus changes pace by $l_i p'_s(q)$. Consequently, the cost of i 's entry to other drivers is

$$l_i c(q),$$

where

$$c(q) := p'_s(q)q \quad (9)$$

is the externality of driving one kilometer in light congestion. Thus, it is socially efficient for i to drive if and only if

$$U_i = \varepsilon_i - l_i p_s(q) \geq l_i c(q), \quad (10)$$

which means

$$\varepsilon_i \geq l_i m(q), \quad (11)$$

where

$$m(q) := p_s(q) + c(q) \quad (12)$$

is the marginal social cost of a km-driven. At the SO circulation, q^* , travelers drive if and only if their car trip satisfies (11). At the aggregate level, physical consistency requires that

$$q^* = q_d[m(q^*)]. \quad (13)$$

As in Walters (1961), this q^* can be found graphically by way of the flow-congestion framework. First, insert both sides of (13) to $p_d(q)$ to obtain

$$p_d(q^*) = m(q^*). \quad (14)$$

Then, plot the two sides in the (q, p) plane. The circulation where $m(q)$ and $p_d(q)$ intersect is q^* .

5. Distance tolls

The first toll considered is a charge per km-traveled, like the one that should soon come to Singapore and is being talked about in London. Since the toll is assumed to be denominated in minutes of travel time, the utility of a trip with a distance toll τ_d (min/km) is $\varepsilon_i - l_i(p + \tau_d)$. Hence, i will drive if and only if

$$\varepsilon_i \geq l_i(p + \tau_d). \quad (15)$$

Note the distance toll enters the RHS (the disutility of a car trip) just like a rise in pace would. Thus, increasing the distance toll by $\Delta\tau_d$ translates the curve $p_d(q)$, of Fig. 1(c), vertically downward by $\Delta\tau_d$ at all points.

Proposition 1. *A distance toll equal to $c(q^*)$ (min/km) decentralizes the social optimum.*

Proof. If $\tau_d = c(q^*)$, then (15) is the same condition as (11) for all i who drive when pace is p^* . By the construction of q^* , when only such travelers drive, pace is p^* . Finally, by Assumption 1, if the social optimum has been made a user equilibrium then it will happen. \square

6. Access tolls

Let τ_a give the size of an access toll, denominated in units of travel time. The utility of i 's car trip with an access toll is $\varepsilon_i - l_i p - \tau_a$, so i drives if and only if

$$\varepsilon_i > \tau_a + pl. \quad (16)$$

Since raising the access toll lowers the utility of every traveler's car trip, given p , such a rise shifts $p_d(q)$ downward at all points. But unlike the distance tolls, the access toll does not vertically shift $p_d(q)$ downward by the same amount everywhere. We will now discuss two propositions of practical interest about access tolls. The first is made under the following assumption:

Assumption 3. *The density of travelers is positive everywhere in the (ε, l) plane.*

This assumption makes the proposition succinct. Later we discuss the consequences of relaxing it.

Proposition 2. *When the access toll is increased at an equilibrium with $k > k_f$, there is a trip length \tilde{l} such that the number of travelers with $l > \tilde{l}$ rises and the number with $l < \tilde{l}$ falls.*

Proof. For lightly-congested equilibria with $k > k_f$ or Type 1 equilibria, the downward shift in $p_d(q)$ resulting from a rise in τ_a lowers p (raises speed). Let p^0 be the pace before the increase, p^1 be the pace after and

$$\tilde{l} := -(\tau_a^1 - \tau_a^0)/(p^1 - p^0). \quad (17)$$

If $l_i > \tilde{l}$, then

$$\tau_a^1 + l_i p^1 < \tau_a^0 + l_i p^0, \quad (18)$$

meaning the cost of car travel is strictly lower after the change for i . If ε_i falls between two bounds, then i 's trip will have be induced by the toll increase. Conversely, if $l_i < \tilde{l}$, then the toll increase raises the net cost of car travel, and so any i with ε_i between the (reversed) bounds will be discouraged. By Assumption 3, such travelers do exist. \square

The mechanics of Proposition 2 resemble the “adverse selection” phenomenon that troubles health insurance markets (Akerlof, 1970). Consider what happens when an insurer raises premiums and offers better coverage. If customers know something about their own health well and the insurer has to price everyone the same, then the healthiest (and most profitable) customers will tend to switch to a less generous, cheaper plan, while any new customers tend to be likely to fall ill. The non-discriminating insurance premium resembles the access toll, and travel distances resemble the customers' likelihood of needing care.

Next, we prove that, due to such “adverse selection,” the access toll is almost always inadequate for maximizing social welfare. But to avoid the edge case where the untolled UE is so low that everyone travels at free-flow speed, in which case the UE and the SO are the same, we preface the proposition with the following assumption:

Assumption 4. *In the untolled user equilibrium, $k > k_f$.*

Proposition 3. *An access toll cannot achieve the social optimum. Even if it achieves the social optimum circulation q^* and pace p^* , then there is a critical length l_c such that there are travelers with trips longer than l_c who drive but should not and travelers with trips shorter than l_c who should drive but do not.*

Proof. Suppose $\tau_a > 0$ has attained p^* and q^* . In this case, i 's car trip improves social welfare if and only if it satisfies (11), but i will drive if and only if $\varepsilon_i \geq \tau_a + p^*l_i$. Let

$$l_c = \tau_a / c(q^*). \quad (19)$$

If $l_i > l_c$, then

$$\tau_a + p^*l_i < l_i[p^* + c(q^*)] = l_i m(q^*), \quad (20)$$

If ε_i falls between these bounds, then i will drive and in doing so damage social welfare. Conversely, if $l < l_c$, then the inequalities are reversed, and if ε_i falls between the bounds she will not drive even though her trip would have raised social welfare. By Assumption 3, such travelers do exist, and so the equilibrium with τ_a is inefficient. \square

This proposition shows that the access toll is not only inefficient, but inefficient in a specific way: it allows too many long car trips to be realized and too few short trips. This shortcoming should be intuitive, given that the access toll fails to charge travelers their marginal external costs. The situation is helpfully visualized in Fig. 4, where the lines

$$D^*(l) := lm(q^*) \quad (21)$$

$$D^{\tau_a}(l) := \tau_a + lp \quad (22)$$

express the bounds in (20). In the shaded region R_1 lie the short-distance travelers who have been needlessly discouraged from driving. In R_2 lie the long-distance travelers whose driving ought to have been prevented. The more travelers fall into these regions, the farther the access toll is from the social optimum.

Relaxing Assumption 3 opens the door to degenerate cases of $f(\varepsilon, l)$ for which the access toll actually *can* decentralize the SO. For instance, if all travelers have the same trip length \bar{l} , then the population will lie along a vertical line in Fig. 4. Then it is efficient for traffic to be split such that only travelers with $\varepsilon_i > \bar{l}c(q^*)$ drive, whereupon $\tau_a = \bar{l}c(q^*)$ (equal to every traveler's externality) enforces the SO. Alternatively, consider the deterministic demand case where $\varepsilon_i = \bar{\varepsilon}$ for every traveler. In this case, only those travelers with trips shorter than $\bar{\varepsilon}/m(q^*)$ should drive, and a toll $\tau_a = \bar{\varepsilon}c(q^*)/m(q^*)$ (equal to the externality of the longest trip in the SO) ensures that these and only these travelers do drive.

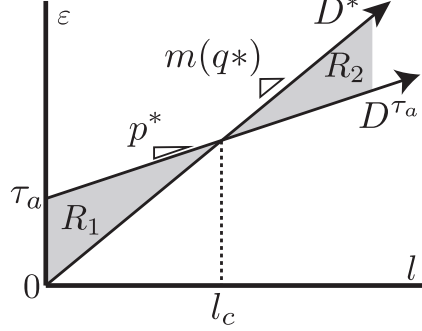


Figure 4: Inefficiency of population division with access toll

7. Numerical simulation

This section will apply the model in simulations. The purpose of simulation is to confirm that the model is computable and to demonstrate the existence and magnitudes of some of the effects about above. Section 7.1 explains the functional forms and parameter values chosen, Sec. 7.2 shows how to calculate relevant aggregates and Sec. 7.3 presents results for several parameter combinations.

7.1. Functional forms

To build the simulation, there are two functional forms to choose: the MFD and the population distribution f . The MFD is in the family given by Drake et al. (1967) for freeway traffic:

$$p(k) = p_f \exp(-1/2(k/k_0)^2), \quad (23)$$

so that $q(k)$ reaches a maximum $q_0 = k_0/(p_f\sqrt{e})$ at $k = k_0$. This function was chosen for its tractability. While Daganzo et al. (2012) fit a more complex model to data from Yokohama, Olszewski et al. (1995) and Williams et al. (1987) show the exponential to fit other network data well.

The population distribution $f(\varepsilon, l)$ is bivariate lognormal with means (μ_ε, μ_l) and covariance matrix

$$\Sigma := \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon,l}^2 \\ \sigma_{\varepsilon,l}^2 & \sigma_l^2 \end{pmatrix},$$

where σ_ε^2 gives the variance of gross benefit, σ_l^2 the variance of trip length and $\sigma_{\varepsilon,l}^2$ the covariance of gross benefit and trip length.

7.2. Aggregates

To evaluate the simulation requires functions for various aggregates. The three most important are TSS (total social surplus), TCS (total consumer surplus) and TR (total revenue). Let

$$X(p, l, \tau_a, \tau_d) = (p + \tau_d)l + \tau_a \quad (24)$$

toll	τ	k	\bar{l}	p	q	a	TCS	TR	TSS
none	0.0	70.5	2.3	5.0	14.1	6.0	20.4	0.0	20.4
distance	2.3	37.1	2.3	2.8	13.4	5.8	19.4	31.0	50.5
access	6.3	39.9	3.1	2.9	13.9	4.6	17.3	28.7	46.0

Table 1: TSS-maximization results

give the disutility of a car trip of length l , given pace p and tolls τ_a and τ_d . Depending on the tolling regime, either $\tau_a = 0$ or $\tau_d = 0$, since we do not consider a mixed regime. It follows that

$$TCS = \lambda \int_0^\infty \int_{X(p,l,\tau_a,\tau_d)}^\infty f(\varepsilon, l)(\varepsilon - X(p, l, \tau_a, \tau_d)) d\varepsilon dl \quad (25)$$

gives total consumer surplus and

$$TR = \lambda \int_0^\infty \int_{X(p,l,\tau_a,\tau_d)}^\infty f(\varepsilon, l)(l\tau_d + \tau_a) d\varepsilon dl \quad (26)$$

total revenue. Adding TCS and TR gives TSS:

$$TSS = \lambda \int_0^\infty \int_{X(p,l,\tau_a,\tau_d)}^\infty f(\varepsilon, l)(\varepsilon - lp) d\varepsilon dl. \quad (27)$$

Finally, the average trip length of travelers who drive (not of the generating population), \bar{l} can be found by the ratio

$$\bar{l} = \frac{q}{a}, \quad (28)$$

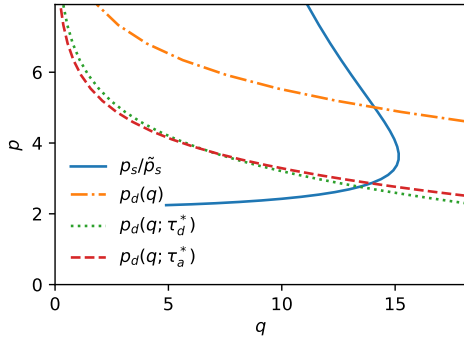
where a is the arrival flow.

7.3. Results

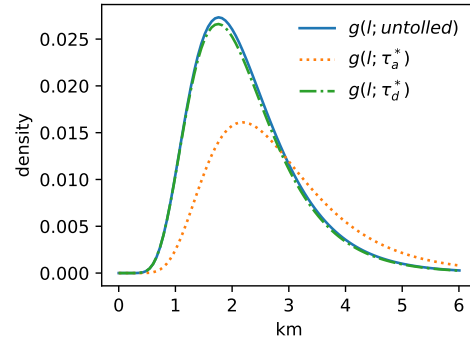
A computer program was used to numerically calculate (i) the untolled UE; (ii) the equilibrium with TSS-maximizing access tolls; and (iii) the social optimum along with the distance toll that decentralizes it. The MFD parameters were chosen to resemble the famous Yokohama experiment that produced the original MFD result in Geroliminis and Daganzo (2008): $k_0 = 55$ veh/lane-km and $p_f = 2.2$ min/km (27.7 km/hr).

First, we investigate the model under demand parameters chosen so the untolled UE exhibits moderate hypercongestion and the average trip length, 2.3 km, observed in Yokohama: $\lambda = 20$, $\mu_\varepsilon = 2.4$, $\mu_l = 1.0$, $\sigma_\varepsilon^2 = \sigma_l^2 = 0.2$, $\sigma_{\varepsilon,l}^2 = 0.12$. The population's mean trip length is thus 3 km, and the mean ε is 12.2 min. For these parameters, Fig. 5a shows the curves p_s/\tilde{p}_s and alongside the inverse demand curves for each toll regime—with intersections being equilibria.

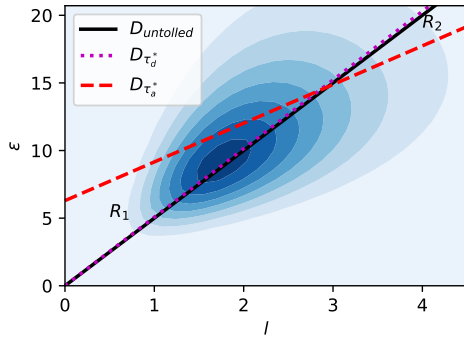
Statistics for each toll regime appear in Table 1. The untolled UE is moderately hypercongested, with a $k > k_0 = 55$. All three regimes produce a similar circulation, although the tolls are able to do so with much lower paces. The



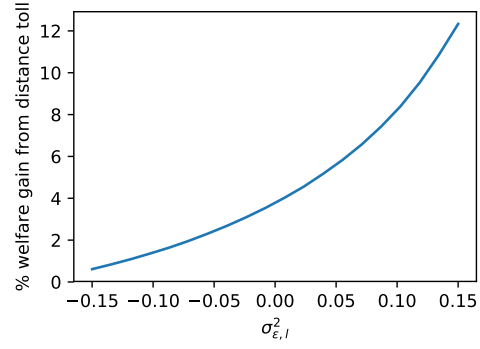
(a) equilibria in flow congestion framework



(b) trip length distributions



(c) division of the traveler population



(d) welfare advantage of distance over access toll

Figure 5: simulation results

distance toll achieves the highest TSS, yielding a 148% improvement over the untolled UE and a 10% improvement over the best-possible access toll. For both tolls, however, consumer surplus declines somewhat, although the effect is stronger for the access toll.

Turning now to the distribution of trip lengths, Figure 5b shows, for each toll regime, the function

$$g(l; \tau_a, \tau_d) = \int_{X(p, l, \tau_a, \tau_d)}^{\infty} f(\varepsilon, l) d\varepsilon \quad (29)$$

giving the density of travelers by trip length. Observe that the distribution is similar without tolls and with the distance toll, whereas the access toll permits longer trips and discourages many short trips. The mass of trips longer than about 3km actually rises with the access toll (evidenced by the τ_a^* line being higher after this length). In consequence, the access toll causes \bar{l} to rise 35%, whereas the distance toll maintains approximately the same \bar{l} as in the untolled UE. Note that, since any change in the density for a given trip length is caused by a change of equal sign in the utility for that trip, Fig. 5b also indicates who wins and loses from each toll change. Note moreover that, due to the higher \bar{l} , the arrival flow a is significantly lower with the access toll even though circulation is not. Thus, an analyst focused on changes in arrival flows might rate the access toll chosen as being even more potent than the distance toll, although it reduces pace and raises welfare by a smaller amount.

Figure 5c shows the population division between drivers and not-drivers under each regime, as per Fig. 4; travelers to the left of each D line drive, and shading denotes the population density as in Fig. 2. $D_{\tau_d^*}$ and $D_{untolled}$ are almost identical, whereas $D_{\tau_a^*}$ has the shallower slope and higher intercept expected. Clearly, a substantial number of trips fall into the R_1 and R_2 regions.

In this example, the access toll has been handicapped by the positive correlation between ε and l , which puts more mass in R_2 and R_2 . As a final experiment, we change the covariance parameter $\sigma_{\varepsilon, l}^2$ while keeping the other demand parameters constant and plot the welfare improvement gained by switching from the best-possible access toll to the best-possible distance toll. Results are plotted in Fig. 5d over the range $-.15$ to $.15$. When trip length and gross benefit are very negatively correlated, the distance toll's advantage is minor; when the two are very positively correlated, the advantage is substantial.

8. Conclusion

This paper has presented a static model of traffic into a downtown with an MFD, in which commuters with varying trip lengths and benefits from driving decide between driving and a zero-utility outside option. An application of Little's Law allows the user equilibrium to be derived from curves in pace/circulation space. Theory shows that an access toll cannot raise total social surplus as much as a distance toll (except in the special cases where demand is deterministic or trip lengths invariant), because it allows and even invites some

long, low-value car trips while discouraging certain short, high-value car trips. The distance toll, by contrast, is able to decentralize the socially-optimal split of traffic between drivers and non-drivers.

It is easy to imagine promising extensions to the model. One would involve letting the value of time vary among travelers. While the random benefit goes part of the way towards realistic heterogeneity, variation in value of travel time savings would go even further. Meng et al. (2012) have already solved for optimal distance tolls in a network model where the value of time varies continuously over the population.

Another extension could involve time-varying conditions. It seems there are two main ways to do so. One approach, epitomized by Gonzales (2015), is to say that trip opportunities are generated over a finite interval according to a fixed distribution of the traveler’s schedules—e.g., their work start or end times. In such a model, travelers cannot reschedule their trips, but traffic conditions can change over time; the zone might fill up with cars during the beginning of the rush and empty out at the end. Thus, as in the static model, the potential gains from tolling consist of reductions in delay and in realizing higher-valued car trips. A second approach might involve a realistic model in which travelers have schedule preferences and can reschedule their trips. Such a model is extremely complex, but Lamotte and Geroliminis (2017) show that, even though the untolled equilibrium in such a model cannot generally be derived analytically, some of its properties can be (a claim confirmed by the authors’ simulations). An open question, then, is what, if any, results for distance and access tolls hold when travelers can reschedule, and if such insights resemble those gleaned from our static setting.

Another rich avenue of investigation is to lose the static time character and instead adopt time-varying conditions. It seems there are two ways main approaches to incorporating time-varying conditions into an isotropic model. One approach, epitomized by Gonzales (2015), is to say that trip opportunities are constantly generated over an interval with a specific start and end time. That is, travelers cannot reschedule their trips, but traffic conditions change over time as the zone fills up with cars during the beginning of the rush and empties out at the end. A second approach might involve a realistic model in which travelers have schedule preferences and can reschedule their trips. Such a model is extremely complex, but Lamotte and Geroliminis (2017) show that, even though the untolled equilibrium in such a model cannot generally be derived analytically, some of its properties can be (a claim confirmed by the authors’ simulations). An open question, then, is what, if any, results for distance and access tolls hold when travelers can reschedule, and if such insights resemble those gleaned from our static setting.

Another rich avenue of investigation is to lose the static time character and instead adopt time-varying conditions. It seems there are two ways main approaches to incorporating time-varying conditions into an isotropic model. One approach, epitomized by Gonzales (2015), is to say that trip opportunities are constantly generated over an interval with a specific start and end time. That is, travelers cannot reschedule their trips, but traffic conditions change over time as the zone fills up with cars during the beginning of the rush and empties out at the end. A second approach might involve a realistic model in which travelers have schedule preferences and can reschedule their trips. Such a model is extremely complex, but Lamotte and Geroliminis (2017) show that, even though the untolled equilibrium in such

a model cannot generally be derived analytically, some of its properties can be (a claim confirmed by the authors’ simulations). An open question, then, is what, if any, results for distance and access tolls hold when travelers can reschedule, and if such insights resemble those gleaned from our static setting.

As for policy, a tempting conclusion is that cities should switch from access to distance tolls. Distance tolls raise social welfare more than access tolls, and they allow more trips to be made, by tolling off long trips that create significant congestion—likely a political advantage. However, some caution is in order. In the first place, whereas access tolls can rely on the proven number-plate recognition and RFID technologies already used widely for bridge and highway tolls, administering a distance toll might pose challenges that are hard to foresee until such schemes have been tried. Access tolls themselves, for instance, have proven at times significantly more expensive to administer than originally planned—see the London experience (Santos, 2008). More broadly, while time and experience will assuage the technical problems, it may be inevitable that distance tolls must by their nature invite delicate privacy concerns; they require authorities to be able to somehow track not only the vehicle’s entry to a zone—as with an access toll—but the overall motion of the vehicle, too. As in the domains of security and taxation, how to balance the gains of rational policy against the dangers of a prying Leviathan—perhaps by cryptography or the creation of new institutions at a remove from the state—will remain a puzzle for some time.

Acknowledgements

The author is indebted to Carlos Daganzo, Richard Arnott, Mogens Fosgerau and two anonymous reviewers for comments that have greatly improved the paper.

References

- Akerlof, G. A., 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics* 84 (3), 488–500.
URL <https://doi.org/10.2307/1879431>
- Arnott, R., 2013. A bathtub model of downtown traffic congestion. *Journal of Urban Economics* 76, 110–121.
URL <https://doi.org/10.1016/j.jue.2013.01.001>
- Arnott, R., Inci, E., 2010. The stability of downtown parking and traffic congestion. *Journal of Urban Economics* 68 (3), 260–276.
URL <http://dx.doi.org/10.1016/j.jue.2010.05.001>
- Arnott, R., Kokoza, A., Naji, M., 2016. Equilibrium traffic dynamics in a bathtub model: A special case. *Economics of Transportation* 7, 38–52.
URL <https://doi.org/10.1016/j.ecotra.2016.11.001>
- Beckmann, M., McGuire, C., Winsten, C. B., 1956. *Studies in the economics of transportation*. Published for the Cowles Commission for Research in Economics by Yale University Press, New Haven.
- Börjesson, M., Kristoffersson, I., 2015. The Gothenburg congestion charge. Effects, design and politics. *Transportation Research Part A: Policy and Practice* 75, 134–146.
URL <https://doi.org/10.1016/j.tra.2015.03.011>
- Buisson, C., Ladier, C., 2009. Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. *Transportation Research Record: Journal of the Transportation Research Board* 2124, 127–136.
URL <http://trrjournalonline.trb.org/doi/10.3141/2124-12>
- Daganzo, C. F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41, 49–62.
URL <https://doi.org/10.1016/j.trb.2006.03.001>
- Daganzo, C. F., Gayah, V. V., Gonzales, E. J., 2012. The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *EURO Journal on Transportation and Logistics* 1 (1-2), 47–65.
URL <https://doi.org/10.1007/s13676-012-0003-z>
- Daganzo, C. F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological* 42 (9), 771–781.
URL <https://doi.org/10.1016/j.trb.2008.06.008>
- Daganzo, C. F., Lehe, L. J., 2015. Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B: Methodological* 75, 89–99.
URL <https://doi.org/10.1016/j.trb.2015.02.010>

- Daganzo, C. F., Lehe, L. J., 2016. Traffic flow on signalized streets. *Transportation Research: Part B: Methodological* 90, 56–69.
URL <https://doi.org/10.1016/j.trb.2016.03.010>
- Drake, J. S., Schofer, J. L., May Jr, A. D., 1967. A Statistical Analysis of Speed-Density Hypotheses. *Highway Research Record* (154), 53–87.
URL <https://trid.trb.org/view.aspx?id=693312>
- Eliasson, J., 2008. Lessons from the Stockholm congestion charging trial. *Transport Policy* 15 (6), 395–404.
URL <https://doi.org/10.1016/j.tranpol.2008.12.004>
- Eliasson, J., 2014. The stockholm congestion charges: an Overview. CTS Working Paper 14.
URL <http://www.transportportal.se/swopec/CTS2014-7.pdf>
- Fosgerau, M., 2015. Congestion in the bathtub. *Economics of Transportation* 4 (4), 241–255.
URL <http://dx.doi.org/10.1016/j.ecotra.2015.08.001>
- Geroliminis, N., Daganzo, C. F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42 (9), 759–770.
URL <https://doi.org/10.1016/j.trb.2008.02.002>
- Geroliminis, N., Levinson, D. M., 2009. Cordon Pricing Consistent with the Physics of Overcrowding. In: Lam, W., Wong, S., Lo, H. (Eds.), *Transportation and Traffic Theory 2009: Golden Jubilee*. Springer, pp. 219–240.
URL https://doi.org/10.1007/978-1-4419-0820-9_11
- Gibson, M., Carnovale, M., 2015. The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics* 89, 62–73.
URL <http://dx.doi.org/10.1016/j.jue.2015.06.005>
- Gonzales, E. J., 2015. Coordinated pricing for cars and transit in cities with hypercongestion. *Economics of Transportation* 4 (1-2), 64–81.
URL <http://dx.doi.org/10.1016/j.ecotra.2015.04.003>
- Lamotte, R., Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. *Transportation Research Procedia* 23, 591–611.
URL <http://dx.doi.org/10.1016/j.trpro.2017.05.033>
- Lawphongpanich, S., Yin, Y., 2012. Nonlinear pricing on transportation networks. *Transportation Research Part C: Emerging Technologies* 20 (1), 218–235.
URL <http://dx.doi.org/10.1016/j.trc.2011.05.010>
- Leape, J., 2006. The London Congestion Charge. *Journal of Economic Perspectives* 20 (4), 157–176.
URL <http://www.jstor.org/stable/30033688>

- Little, J. D. C., Graves, S. C., 2008. Little's Law. In: Chhajed, D., Lowe, T. (Eds.), *Building Intuition* (International Series in Operations Research Management Science). Vol. 115. Springer, Boston, Ch. 5, pp. 81–100.
URL <http://www.springerlink.com/content/n7h5375373hh8kh1/>
- Liu, Z., Meng, Q., Wang, S., 2013. Speed-based toll design for cordon-based congestion pricing scheme. *Transportation Research Part C: Emerging Technologies* 31, 83–98.
URL <http://dx.doi.org/10.1016/j.trc.2013.02.012>
- Mariotte, G., Leclercq, L., Laval, J. A., 2017. Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. *Transportation Research Part B: Methodological* 101, 245–267.
URL <https://doi.org/10.1016/j.trb.2017.04.002>
- May, A. D., Milne, D. S., 2000. Effects of alternative road pricing systems on network performance. *Transportation Research Part A: Policy and Practice* 34 (6), 407–436.
- Meng, Q., Liu, Z., Wang, S., 2012. Optimal distance tolls under congestion pricing and continuously distributed value of time. *Transportation Research Part E: Logistics and Transportation Review* 48 (5), 937–957.
URL <http://dx.doi.org/10.1016/j.tre.2012.04.004>
- Mun, S.-i., Konishi, K.-j., Yoshikawa, K., 2003. Optimal cordon pricing. *Journal of Urban Economics* 54 (1), 21–38.
URL [https://doi.org/10.1016/S0094-1190\(03\)00022-6](https://doi.org/10.1016/S0094-1190(03)00022-6)
- Olszewski, P., Fan, H. S. L., Tan, Y. W., 1995. Area-wide traffic speed-flow model for the Singapore CBD. *Transportation Research Part A* 29 (4), 273–281.
- Santos, G., 2008. London Congestion Charging. *Brookings-Wharton Papers on Urban Affairs* (2008), 177–234.
- Santos, G., Li, W. W., Koh, W. T. H., 2004. Transport Policies in Singapore. *Research in Transportation Economics* 9 (February), 209–235.
- Small, K. A., Chu, X., 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37 (3), 319–352.
URL <http://www.jstor.org/stable/20053940>
- Tan, C., feb 2016. 10 things to know about the next-generation ERP system.
URL <http://www.straitstimes.com/singapore/transport/lta-to-roll-out-next-generation-erp-s>
- Topham, G., jun 2017. London mayor considers pay-per-mile road pricing and ban on new parking.
URL <https://www.theguardian.com/uk-news/2017/jun/20/london-mayor-considers-pay-per-mile-r>

- Verhoef, E. T., 1999. Time, speeds, flows and densities in static models of road traffic congestion and congestion pricing. *Regional Science and Urban Economics* 29 (3), 341–369.
- Verhoef, E. T., 2002. Second-best congestion pricing in general static transportation networks with elastic demands. *Regional Science and Urban Economics* 32, 281–310.
- Verhoef, E. T., 2005. Second-best congestion pricing schemes in the monocentric city. *Journal of Urban Economics* 58 (3), 367–388.
- Walters, A. A., 1961. The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica* 29 (4), 676–699.
- Watson, P. L., Holland, E. P., 1978. Relieving Traffic Congestion: the Singapore Area License Scheme.
URL <http://documents.worldbank.org/curated/en/1978/06/1555142/relieving-traffic-congestion>
- Williams, J. C., Mahmassani, H. S., Herman, R., 1987. Urban Traffic Network Flow Models. *Transportation Research Record* (1112), p. 78–88.
- Yang, H., Huang, H.-J., 1998. Principle of marginal-cost pricing: how does it work in a general road network? *Transportation Research Part A: Policy and Practice* 32 (1), 45–54.
URL <http://linkinghub.elsevier.com/retrieve/pii/S0965856497000189>