Recreation

ISSN 0316-5078

INSTITUTE FOR ECONOMIC RESEARCH

QUEEN'S UNIVERSITY

Kingston, Ontario, Canada K7L 3N6

OPTIMAL TOLLS WITH HIGH-PEAK TRAVEL DEMAND

Richard J. Arnott

and

Ronald E. Grieson

DISCUSSION PAPER NO. 391

Richard J. Arnott
Department of Economics
Queen's University
Kingston, Canada K7L 3N6

Ronald E. Grieson
Woodrow Wilson School
Princeton University
Princeton, N.J. 08540

June 1980

Optimal Tolls with High-Peak Travel Demand*

I. Introduction

Corresponding to any level of stationary-state traffic flow are two velocities. Zero flow, for instance, corresponds to travel both on an empty road and in a stationary traffic jam. We refer to travel at the lower velocity associated with a given flow as hyper-congestion, and periods during which hyper-congestion is possible as being characterized by high-peak travel demand. Our aim is to determine optimal congestion tolls with high-peak demand. Our note builds on Vickrey's discussion of this problem.[1]

Before proceeding with the analysis, we must clarify our terminology. The technology of stationary-state travel is given by a function which relates stationary-state traffic flow to each vehicle's travel time. The seminal papers on optimal congestion tolls, Walters [1961] and Vickrey [1963, 1969], framed the analysis in terms of average and marginal cost. To facilitate comparison between their analyses and ours, we treat travel cost as functionally related to travel time; the technology of stationary-state travel may then be characterized by a function relating stationary-state traffic flow to each driver's time cost of travel, which is termed either average social cost or marginal private cost. This relationship is shown in Figure 1. We define the flat portion of this curve, on which travel is so light that a few extra cars on the road do not increase a driver's travel costs, as the region of uncongested travel;[2] the upward-sloping portion as the region of congested travel; and the downward-sloping portion as the region of hyper-congested travel. Stationary-state travel in the

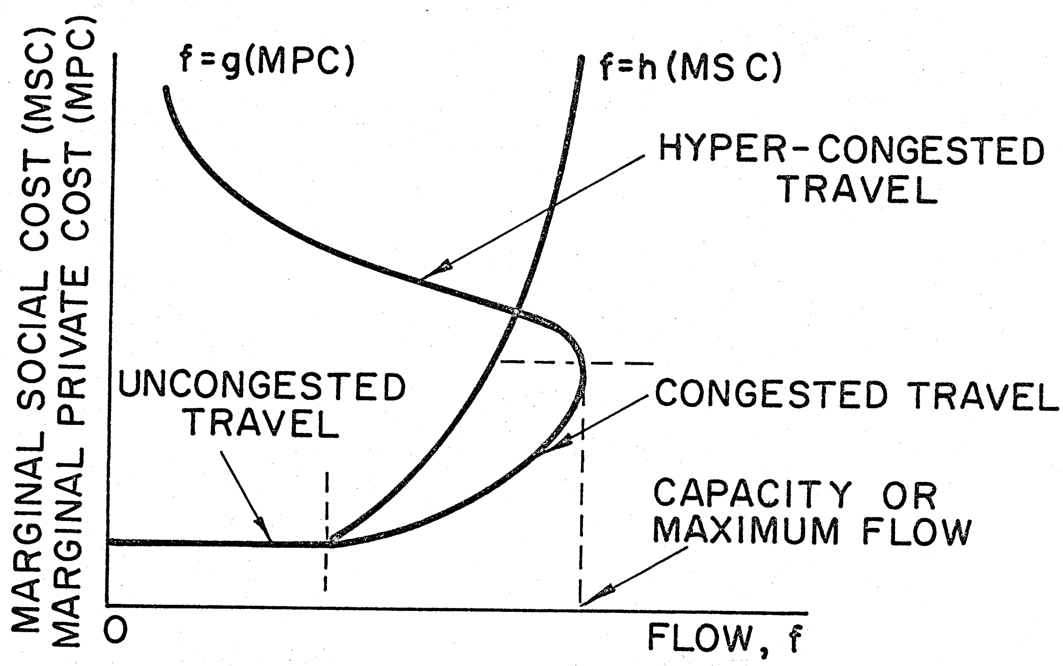

Figure 1: The technology of road travel

region of hyper-congestion is demonstrably inefficient, since corresponding to any point in this region there is another point with the same flow and lower travel costs, which may lie in either the region of congested or that of uncongested travel.

Previous published analyses of the determination of optimal congestion tolls have treated marginal private cost as a uni-valued function of flow. However, when hyper-congestion is possible there are two marginal private costs associated with each level of flow, one corresponding to a point in the region of hyper-congested travel and another to a point in one of the other two regions. To circumvent this difficulty, we treat flow as a function of marginal private cost,[3] and denote this function as $f = g(MPC)$.

In a stationary state, total travel costs per unit time equal each driver's travel costs times flow. We define the increase in total travel costs from a unit increase in flow as marginal social cost.[4] From $f = g(MPC)$, one can calculate the relationship between marginal social cost and flow, which is denoted $f = h(MSC)$. This function is plotted in Figure 1 for congested and uncongested, but not for hyper-congested, travel.[5]

We now characterize the demand side of the travel "market". The number of cars entering the road per unit time depends on the cost to the driver of travelling on the road. This cost comprises the monetized value of travel time, MPC, plus the congestion toll, τ. We denote the stationary demand for travel function as $d = d(MPC + \tau)$. We say that there is <u>high-peak demand</u> if the demand curve intersects the hyper-congested region of $g(MPC)$, $d = d^2(MPC + \tau)$ in Figure 2; <u>low-peak demand</u> if the demand curve intersects the congested, but not the hyper-congested, region of $g(MPC)$, $d = d^1(MPC + \tau)$ in Figure 2; and <u>off-peak demand</u> if the demand curve cuts neither the congested nor hyper-congested regions of $g(MPC)$, $d = d^0(MPC + \tau)$ in Figure 2. In our terminology, uncongested, congested, and hyper-congested refer to the characteristics of traffic flow, while high-peak, low-peak, and off-peak refer

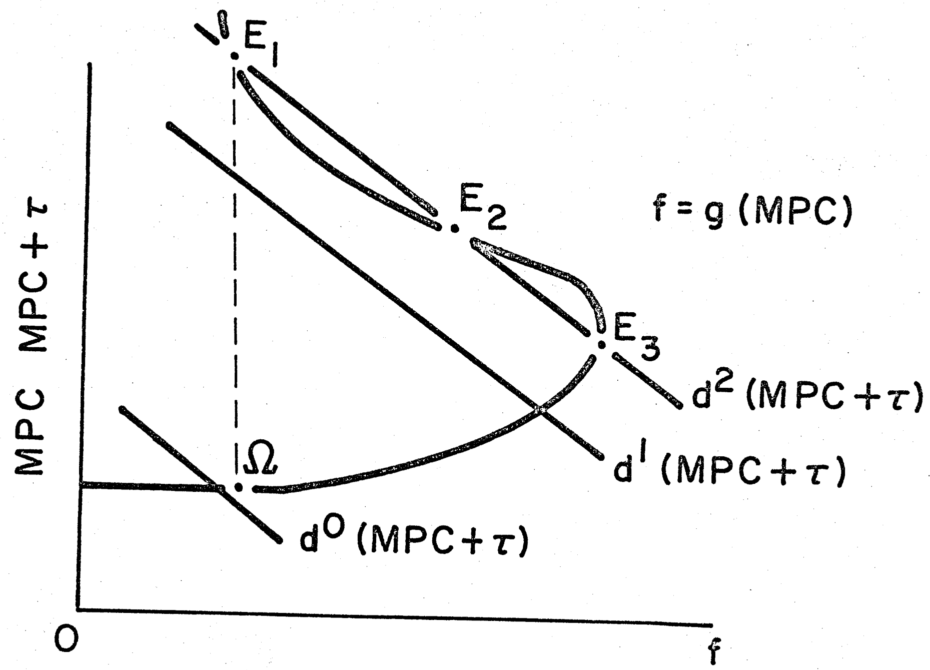

Figure 2: Levels of demand

to the characteristic of travel demand.

We define stationary-state equilibrium to obtain whenever a driver's travel costs inclusive of the toll are such that the flow of the road equals the number of cars wanting to enter the road per unit time. Thus, equilibria are characterized by the points of intersection of $g(MPC)$ and $d(MPC + \tau)$. Exploiting the analogy between market and traffic flow equilibrium, we see that $g(MPC)$ can be interpreted as a traffic supply curve.

II. Low-Peak Demand

We now review the economics of optimal congestion tolls with low-peak demand, using our terminology.

Figure 3 depicts a low-peak demand situation. In the absence of a congestion toll, equilibrium occurs at J, the point of intersection of the

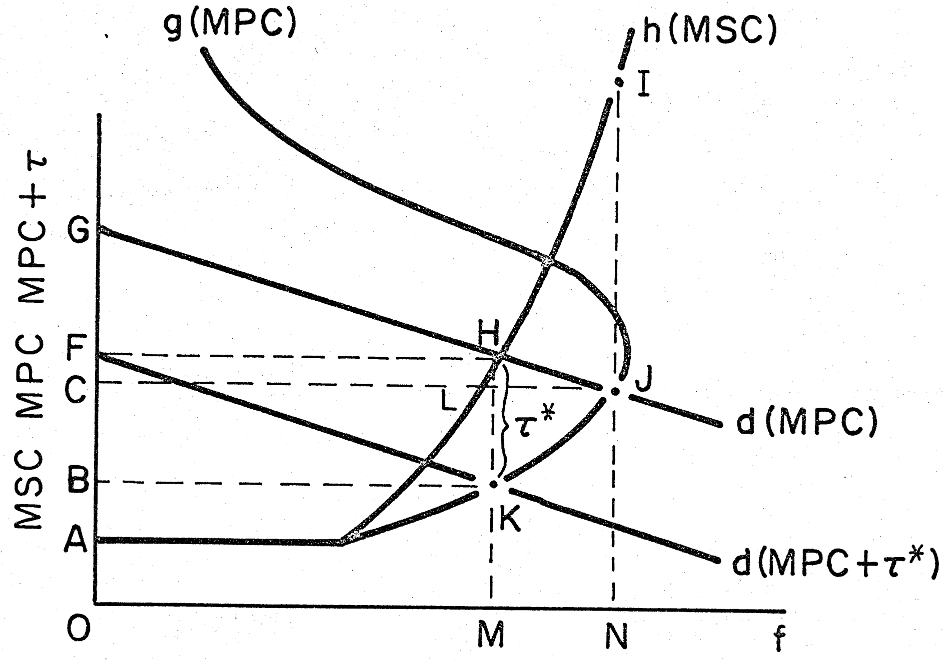

Figure 3: Low-peak demand

supply and demand curves.

It is well-known that, without a toll, equilibrium traffic flow in congested traffic is inefficient because there is an uninternalized congestion externality. In deciding whether or not to travel on the road, a driver considers only his own travel costs and ignores that his travelling on the road slows other drivers down.

There are two necessary conditions for efficient stationary-state utilization of the road. First, hyper-congestion may not occur. Second, the social cost of increasing traffic flow by one unit must equal the social benefit; i.e., the efficient flow level corresponds to the point of intersection of h(MSC) and d(MPC). These two conditions together imply that efficient travel occurs at the point K on g(MPC) in Figure 3.

We have already established that equilibria occur at the points of intersection of $g(MPC)$ and $d(MPC + \tau)$. In Figure 3, for the equilibrium to coincide with the optimum it is necessary to charge a toll of such a magnitude that $d(MPC + \tau)$ intersects $g(MPC)$ at K. The efficient size of the toll is therefore τ^*, the vertical distance between H and K, which equals marginal social cost minus marginal private cost at optimal flow.

In the absence of the congestion toll, social benefit from travel on the road equals OGJN, the area under $d(MPC)$ up to equilibrium flow, since the demand curve is also the marginal social benefit curve. Social cost can be measured as flow multiplied by time costs per driver, OCJN. Thus, social surplus equals CGJ. With the optimal congestion toll, social surplus is BGHK. The excess burden from not imposing a congestion toll therefore equals BGHK - CGJ = BCLK - HLJ.[6] And the toll revenue collected is BFHK. Toll revenue less excess burden therefore equals BFHK - (BCLK - HLJ) = FHJC > 0. Thus, with low-peak demand, if demand is less than infinitely elastic, toll revenue from an optimal toll always exceeds the excess burden from not imposing the toll.

III. High-Peak Demand

The analysis of high-peak demand is similar to that for low-peak, but there are some complications. First, there may be multiple equilibria, as is shown in Figure 2. Which of these equilibria are stable? To answer this question completely satisfactorily requires an explicit treatment of the dynamics of non-stationary-state traffic flow, which is beyond the scope of this paper.[7] We can, however, provide a casual quasi-dynamic analysis using the tools employed in the paper. We may reasonably assume that when demand exceeds (falls short of) supply, travel time increases (decreases). Out of stationary state, one may interpret demand as the flow rate onto the road and supply as the average flow along the road. The assumption is therefore that with increasing (decreasing) flow onto the road, travel time increases (decreases). It follows that in Figure 2, E_1 and E_3 are stable equilibria, while E_2 is unstable.

Why, with equilibrium at E_1, is it not possible to have a sudden switch to Ω, with the same flow rate on the lower part of g(MPC)? The fundamental identity of traffic flow is flow $\equiv$ velocity x density. A sudden switch from E_1 to Ω, holding flow constant, requires a sudden fall in density which is not possible. This line of argument establishes that which of E_1 or E_3 is the stationary-state equilibrium depends on the path of adjustment to the stationary state. If the demand curve were initially above d^2 and intersected g(MPC) only once, above E_1, and then shifted down smoothly towards d^2, E_1 would be the stationary-state equilibrium; otherwise, there would have to be a discontinuous decrease in traffic density which is not possible. Similarly if the demand curve were initially below d^2 and intersected g(MPC) only once, below E_3, and then shifted smoothly upwards, E_3 would be the stationary-state equilibrium.

There are two high-peak demand configurations to treat. The first and the easier to analyze is shown in Figure 4. The second, to be treated later, is shown in Figure 5.

Following the procedure used to derive the optimal toll with low-peak demand, we find the optimal toll to be τ^*, the vertical distance between

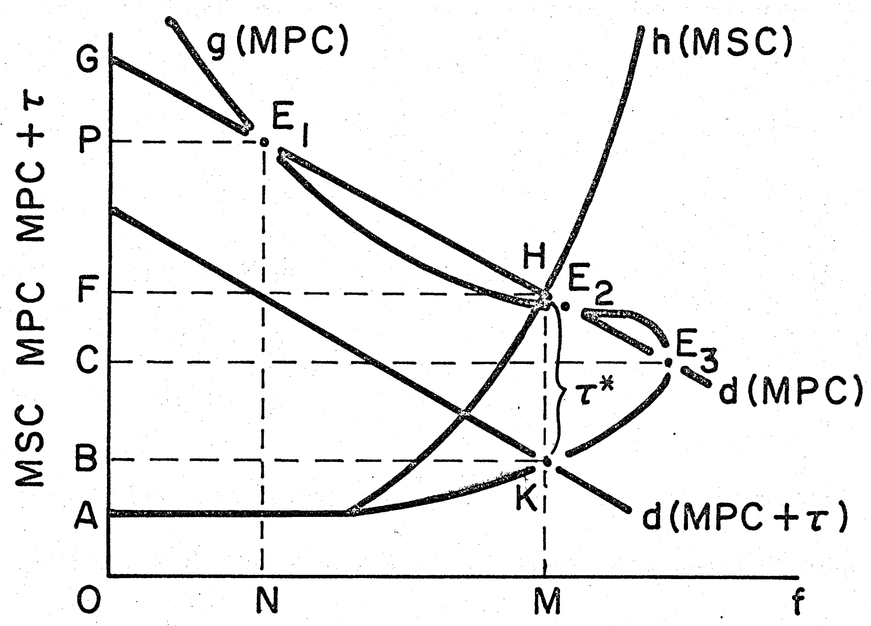

Figure 4: High-peak demand - simple case

the point of intersection of d(MPC) and h(MSC), H, and the point with the same flow rate on the congested portion of g(MPC), K.

If the initial equilibrium were at E_3 in Figure 4, the social surplus from the road would be CGE_3; and if it were at E_1, social surplus would be PGE_1. After the imposition of the optimal toll, social surplus is BGHK. The excess burden from not imposing an optimal congestion toll is therefore BGHK - CGE_3 with initial equilibrium E_3, and BPE_1HK with initial equilibrium E_1. Toll revenues are BFHK. Now, BGHK - CGE_3 -

$BFHK = FGH - CGE_3 = -CFHE_3 < 0$ while, $BPE_1HK - BFHK = FPE_1H > 0$. Thus, with initial equilibrium E_3, toll revenues exceed the efficiency gain that results from imposing the optimal congestion toll. But with initial equilibrium E_1, toll revenues are less than the efficiency gain.

When demand is less than infinitely elastic, the relationship between toll revenues and the efficiency gain from imposing the optimal toll is that: if imposition of the optimal congestion toll causes traffic flow to fall (rise), the excess burden associated with not imposing the toll is less than (is more than) toll revenue collected. When demand is infinitely elastic, toll revenues equal the efficiency gain. These propositions can be demonstrated by simple geometric argument and imply that if the imposition of the optimal congestion toll causes traffic flow to increase, or if demand is infinitely elastic, then if toll revenues more than cover collection costs, the toll should be imposed (according to the conventional cost-benefit criterion).

The above line of reasoning indicates that the excess burden from not imposing a congestion toll when travel is hyper-congested may be very large, and may even exceed the benefits from the use of the road in the absence of the toll.

The high-peak travel demand configuration in Figure 5 illustrates an additional complication that has to be considered. The optimal stationary-state toll, computed as in the previous case, is τ^*. However, if the pre-toll equilibrium is at E_1^b, then application of this toll during the period of adjustment to the new stationary state results in the post-toll equilibrium being E_1^a and not θ.[8] Imposition of a toll

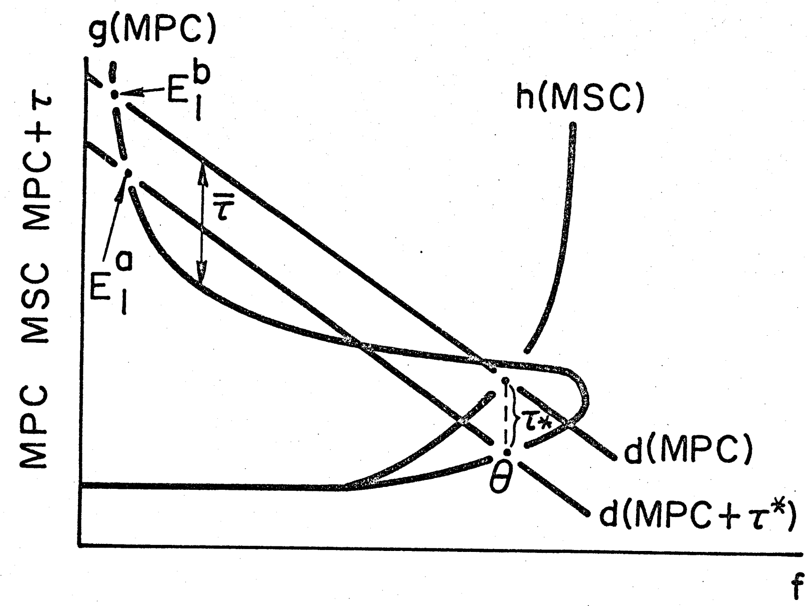

Figure 5: High-peak demand - complication

of magnitude at least $\bar{\tau}$ in the period of adjustment, so that d(MPC + $\bar{\tau}$) does not intersect the hyper-congested portion of g(MPC), circumvents this problem. After some period of time with the toll $\bar{\tau}$ hyper-congestion ends, and then the toll can be reduced to τ^*.

IV. Concluding Comments

The standard analysis in urban economics of traffic congestion is strictly correct only in stationary state and ignores hyper-congestion. We have extended this analysis in examining the economics of congestion tolls, taking hyper-congested travel into account.

Several interesting results have been obtained. First, under high-peak demand conditions, multiple stable hyper-congested stationary-state equilibria are possible. Which of these possible equilibria occurs depends on the pattern of demand prior to establishment of the stationary state.

Second, with high-peak demand, the efficiency gain from imposing the optimal toll may exceed both toll revenues and the benefits from road usage in the absence of the toll. Finally, in moving from an initial hyper-congested equilibrium to efficient utilization of the road, it may be necessary to impose a temporarily larger toll to eliminate hyper-congestion, after which the toll may be lowered to the efficient stationary-state level.

Actual traffic flow is inherently non-stationary-state. Over some sections of the road during some periods of the rush hour, there is hyper-congestion; over other sections of the road and during other periods of the rush hour, there is not. The existing literature makes the assumption that, for most purposes, nothing essential is lost by treating traffic flow as uniform over the rush hour. On the basis of our analysis here we have doubts concerning the validity of this assumption. Whether our doubts are well-founded will have to await an explicitly non-stationary-state analysis of the economics of traffic flow.

FOOTNOTES

* We would like to thank David Pines for helpful comments on an earlier draft.

1. William Vickrey has presented several seminars on this topic. This note both commits his important contribution to paper and extends it.

2. Marginal private cost may actually decrease with flow at very low levels. For instance, a driver will be alerted to stop signs by cars ahead of him slowing down and stopping. Thus, he may safely travel at a slightly higher speed if there are a few rather than no other cars on the road. To simplify the discussion, however, we treat the region of uncongested travel as being horizontal.

3. That is, we work in terms of the inverse of the conventional function. Note that flow is a set-valued function of marginal private cost with uncongested travel. This is a notational inconvenience, but does not concern us since we are interested in the other regions of the curve.

4. What happens if the hourly flow onto the entrance of the road exceeds capacity flow by one car per hour? If capacity flow on the road still occurs, then a queue develops which increases in length by one car per hour. The cost of adding an extra car to the queue is the time taken to move up in the queue by one car, times the number of cars which enter the queue after the car in question,

which is infinite with stationary demand. But this is a non-stationary not a stationary state. Hence, marginal social cost is undefined at capacity flow.

5. We do not use that portion of the function $f = h(MSC)$ corresponding to hyper-congestion.

6. The more usual measure of the excess burden from not imposing a congestion toll is HIJ, the amount by which the social cost associated with the excess flow, MN, exceeds the private cost. We measure excess burden in the way we do to facilitate comparing it with toll revenues and treating hyper-congestion.

7. Arnott, Robin Lindsey, and Kenneth Small are engaged in research on the economics of non-stationary-state traffic flow.

8. Otherwise, there would have to be a physically impossible discontinuous change in density.

Bibliography

Vickrey, W.S., "Pricing in Urban and Suburban Transport", American Economic Review 53, 452-465 (1963).

Vickrey, W.S., "Congestion Theory and Transport Investment", American Economic Review 59, 251-260 (1969).

Walters, A.A., "The Theory and Measurement of Private and Social Cost of Highway Congestion", Econometrica 29, 676-699 (1961).