NOTES AND COMMENTS

A REFORMULATION OF THE THEORY OF OPTIMAL CONGESTION TAXES

A Comment

By C. A. Nash*

Else (1981) seeks to show that the conventional approach to the theory of optimal congestion taxes is incorrect in three respects. Firstly, by wrongly defining the marginal social cost curve, the optimal tax for a flow at less than its maximum value is overestimated. Secondly, for the same reason, it is wrongly said to be impossible for an optimum to exist on the backward-sloping section of the average cost curve. Thirdly, by concentrating on situations in which the flow of traffic is constant over time, the theory fails to deal with the typical case in which an increase in peak flow on to the road leads to an increase in the duration of the congestion, and thus delays some off-peak traffic.

This comment argues that the first two points are incorrect. Consequently, it can never be optimal to charge a toll at which speed and flow are both less than at the maximum flow. Nevertheless, a rise in peak traffic densities can impose delays on off-peak users for the reasons suggested by Else; and his argument that an ideal toll would diminish gradually at the end of the peak period, rather than ceasing abruptly, appears valid in theory even if of doubtful practical relevance.

A RESTATEMENT OF CONVENTIONAL THEORY

Much of the confusion about the backward-sloping section of the cost curve seems to derive from a belief that evaluation of positions on this section is rendered impossible by the absence of a meaningful marginal social cost curve. Though this argument is *not* put forward by Else, it will be helpful to clarify the theory before dealing with Else's points.

Figure 1 is based on Else's Figure 2b, but with the conventional marginal cost curve (M_F) and with the addition of the "irrelevant" downward-sloping marginal cost curve (M_{F2}), which relates to the portion of T_F above T_G.

For a flow of less than $F_{\max}$, an increase in demand imposes costs solely by raising journey time for all vehicles using the road. The new total journey time may be

* Institute for Transport Studies, University of Leeds.

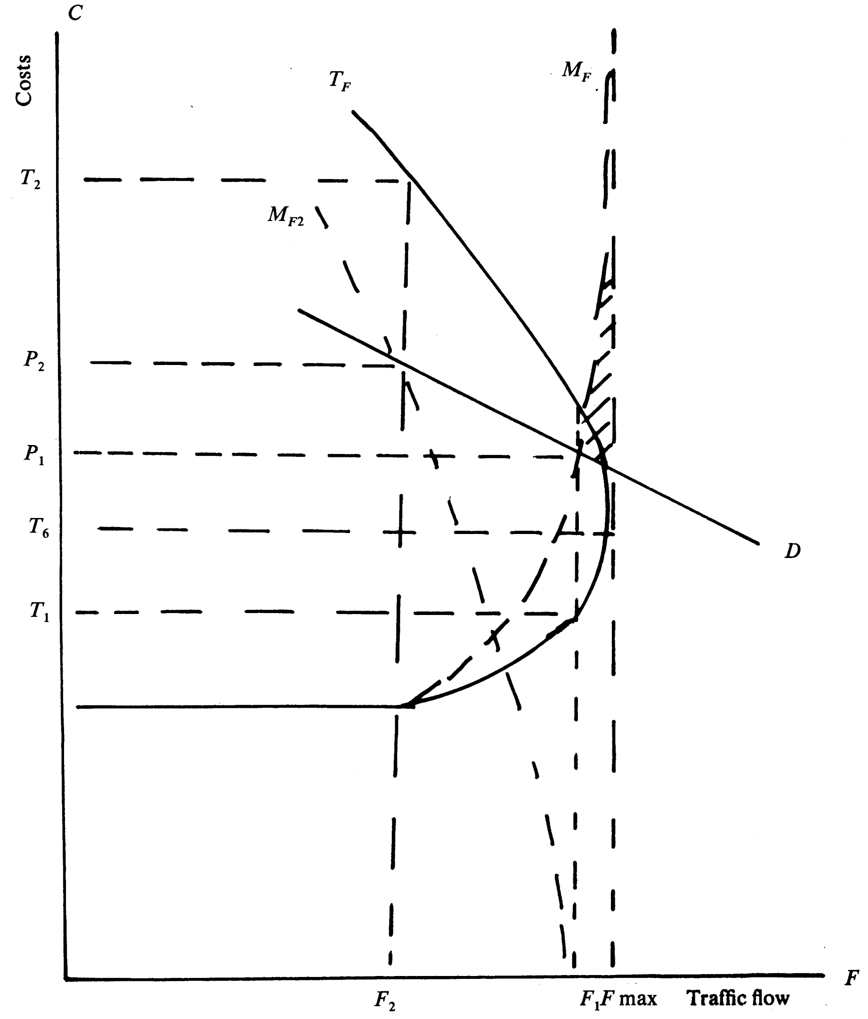

FIGURE 1

derived by multiplying T_F by the initial value of $F(F_o)$ or by finding the area under M_F between zero and F_o. This procedure remains valid up to F_{max}, at which point $T_F \cdot F_{max}$ remains finite and so (must therefore) the integral

$$\int_{o}^{F_{max}} M_F \, dF$$

(and also the deadweight loss, which is the shaded area in Figure 1).

If demand continues to increase, we shall reach a point such as $F_2 T_2$. Here, clearly, the total cost for vehicles still travelling remains finite; thus the costs of vehicle delay are easily calculated. What has also happened, however, is that some vehicles have been priced off the road altogether. Mathematically this may be thought of as an infinite delay, but its social cost is not infinite; rather it may be measured as the loss of consumers' surplus. For instance, a move from a position such as T_1F_1 to T_2F_2 will involve social losses measured as the rise in journey times $(T_2-T_1)F_2$ plus the loss of consumers' surplus $\frac{1}{2}(F_1-F_2)$ (P_2-P_1) (the position T_2F_2 will also require a negative toll of T_2—P_2 to be sustained in time units!). It is now clear why an intersection of a demand curve with the marginal cost curve M_{F2} can never lead to an optimal position. There will always be another intersection with M_F, and the position given by that intersection will involve lower journey times and higher consumers' surplus.[1] Points such as T_2F_2 may satisfy the first-order conditions for an optimum, but they are not optimal.

THE MEASUREMENT OF MARGINAL SOCIAL COST

Else argues that the marginal social cost curve should be defined in terms of the costs imposed by one extra vehicle on the road, rather than by an increase in flow. This appears to be a straightforward confusion of units. Demand curves in economics are conventionally measured in terms of a desired flow (that is, quantity per unit time).

If we follow Else's approach of saying that the demand is really a demand for a number of completed journeys, this must still be expressed per unit time, and the number of completed journeys per unit time is simply the flow off the road. Thus there is no reason to change the argument put forward in the previous section.

In case the reader remains unconvinced, let us look again at where Else's analysis leads us. He obtains in Figure 2b (which is the basis of my Figure 2) a backward-sloping marginal cost curve similar to the average cost curve (C_F), and claims that an intersection of this curve with a demand curve such as d_2 could lead to an optimum on the backward-sloping section of C_F. But again, looking at second-order conditions shows that this could never happen. For *either* M_F is more steeply sloping than d_2 at the intersection (as shown), in which case the second-order conditions for an optimum are not satisfied, *or* the demand curve is more steeply sloping than M_F (as for d_3), in which case there is another (superior) intersection with M_F on its upward-sloping portion, and this gives both lower costs and higher flow (C_2F_2 as opposed to C_1F_1). In the former case, Else's method leaves us with no way of finding an optimum; the only intersection between M_F and d_2 is not one. In the latter case, a supposed optimum is found at C_2F_2, though, as I have argued above, the mis-specification of the marginal cost curve means that this optimum will be incorrectly located.

[1]This argument might not hold if the demand curve cut M_{F2} at a point to the right of M_{F2} (i.e. below the intersection of M_{F2} and M_F). But in that case, the demand would clearly be less steeply downward sloping than M_{F2}, so that the second-order conditions for the intersection with M_{F2} to give an optimum would not hold.

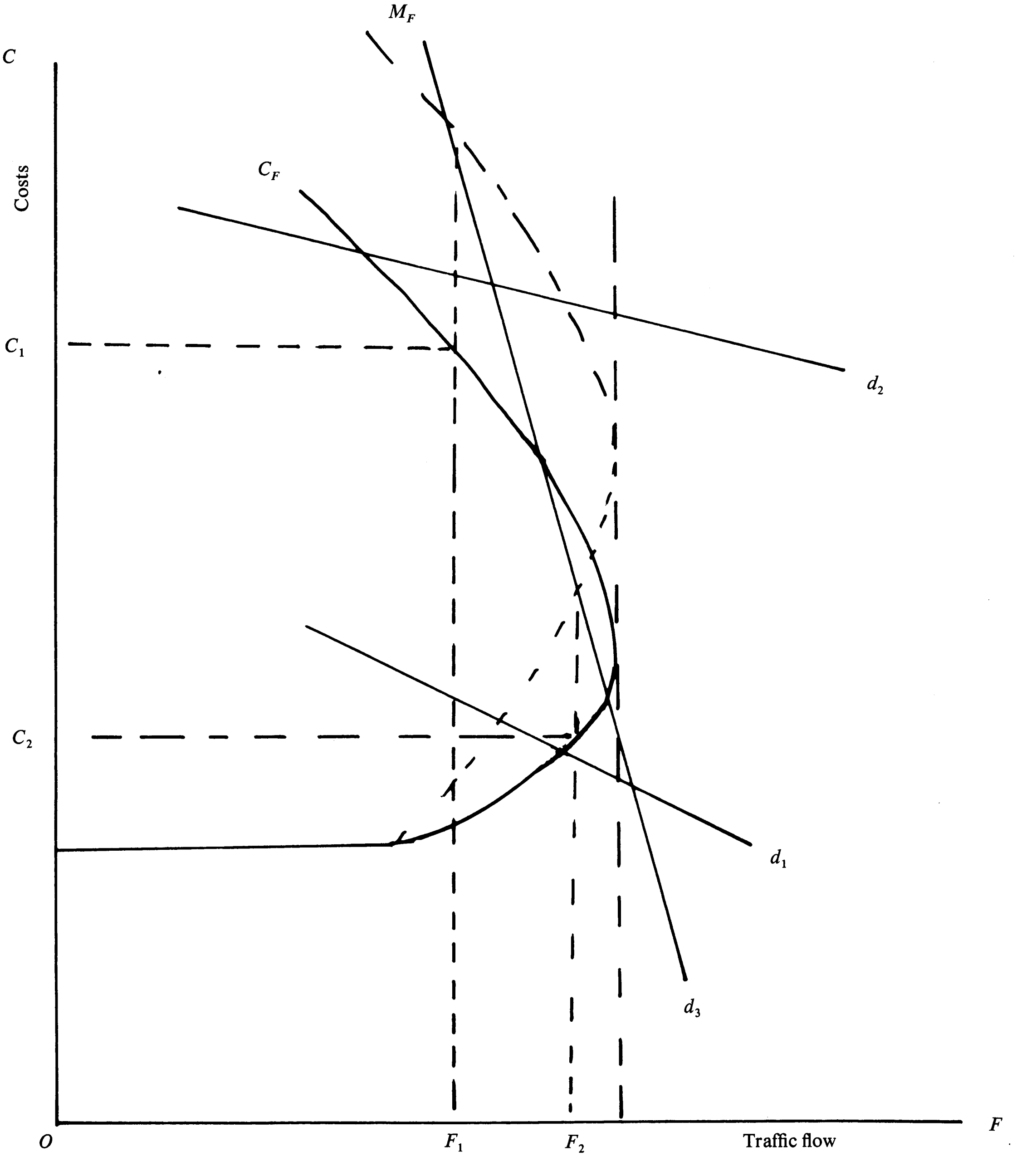

FIGURE 2

VARIABLE TRAFFIC FLOW OVER TIME

Else's criticism of the lack of realism of a model in which flow is assumed constant is valid, and his analysis of what actually happens when an additional vehicle enters the flow is very interesting. However, the point must be made again that whenever the flow may reasonably be regarded as being in a steady state (that is, in the heart of the peak or off-peak) the conventional analysis remains correct. Furthermore, even when the flow is not in a steady state, it is still true that an optimal structure of taxation will never lead to situations such as that envisaged at the top of page 222, where an optimum exists on the backward-bending part of the cost curve. For by limiting the flow on to the road so that the density never rises above that at F_{max}, both the speed of travel and the flow off the road may be raised. Thus it is possible so to arrange the outcome that all journeys—even those at the start delayed by the restraint procedure, assuming that they simply relocate in time rather than ceasing to exist—are completed sooner than if density is allowed to rise above this level. This is the principle of some traffic management schemes which, by causing traffic to queue to enter a main road rather than allowing congestion to build up beyond F_{max}, claim to achieve faster journey times even for vehicles forced to queue to get on to the road.

Subject to this proviso about the measurement of the marginal social cost, Else's analysis of the transition from peak to off-peak is a good one, showing that congestion does not suddenly cease, but disperses over time. Thus the congestion toll should ideally do likewise. Given a sophisticated metering scheme, this result might be of practical relevance. With the types of supplementary licensing which seem to be the most one could hope for in practice, the real question is when is the best time to lift the restriction, and fine analysis is likely to be rendered redundant by the need for a convenient round number.

Our conclusion is, then, that the conventional analysis is correct whenever the flow on a road may be assumed to be at a steady state. Else does add to our understanding on the transition between steady states, but it is doubtful whether his analysis has much practical relevance, since the only forms of pricing with any likelihood of implementation would be far too crude to take this into account.

A Rejoinder

By P. K. Else†

There seem to be two main questions raised in Nash's comment which need reply. The first is the appropriate definition of marginal social cost when there is traffic congestion, and the second is whether an optimum can exist on the backward-bending part of the cost-flow curve.

---

† Division of Economic Studies, University of Sheffield.

THE MEASUREMENT OF MARGINAL SOCIAL COST

The first point can perhaps most easily be dealt with by restating the basis of my original analysis in a slightly different way. It is of course true, as Nash points out, that demand functions in economics are conventionally expressed in terms of a desired flow or quantity per unit of time, and I was careful not to depart from that convention in my original paper. Nevertheless, underlying every demand curve are decisions to buy or not to buy individual units of a good, whether that good be a loaf of bread or a journey along a road. Cost functions are also usually expressed in flow terms, but it has long been recognised that they may also have other dimensions. Alchian (1959), in discussing the relationship between costs and output, observed that "The rate of output is typically regarded as the crucial feature, and concentration on it alone has led to serious error". It seems to me that this has certainly been true in the analysis of traffic congestion.

Putting the argument in fairly basic terms, when an individual makes a decision to buy an extra unit of any commodity, the demand per unit of time in one (unit) period is increased. If suppliers can meet this extra demand by a corresponding increase in output in one period, the resulting additional cost is simply the cost of increasing the flow of output by one unit in that period, or, in other words, marginal cost as conventionally measured. However, as Alchian's analysis suggests, the additional demand may alternatively be met by increasing the flow of output by less than one unit over more than one period of time, or, in the most extreme case, by not increasing the flow of output at all but simply expanding the production run. In such cases the cost of meeting the extra demand may be somewhat different from that indicated by the normal marginal cost curve. Firms may typically be able to choose how to meet any increase in demand. But on a road, the way any extra demand for its use is met is determined by the technical relationship between traffic density and traffic flow. This relationship implies that under congested conditions an increase in demand (e.g. as represented by the entry of one additional vehicle to the road) leads to a less than proportionate increase in the flow of output[2] and a consequent spreading out of its effect over time. Hence the additional costs arising from the extra demand cannot properly be measured by marginal cost as conventionally defined, as that relates to the cost of increasing the traffic flow by one unit. This is the basis of my reformulation. Marginal social costs are defined in terms of the actual costs imposed by someone making an additional journey, as that seems to be the most appropriate basis for determining the optimum price for individual journeys. There is no confusion of units as Nash suggests, but simply a recognition of the wider dimensions of the cost function.

AN OPTIMUM ON A BACKWARD-BENDING CURVE

The reply to the second question follows from this, since, if my analysis of the marginal social cost of extra traffic is correct, the possibility of an optimum position on the backward-bending section of the cost-flow relationship is a logical consequence.

[2] Or, in the terms used in my original paper, $N/F\ (dF/dN) < 1$ (see Else, 1981, p. 222).

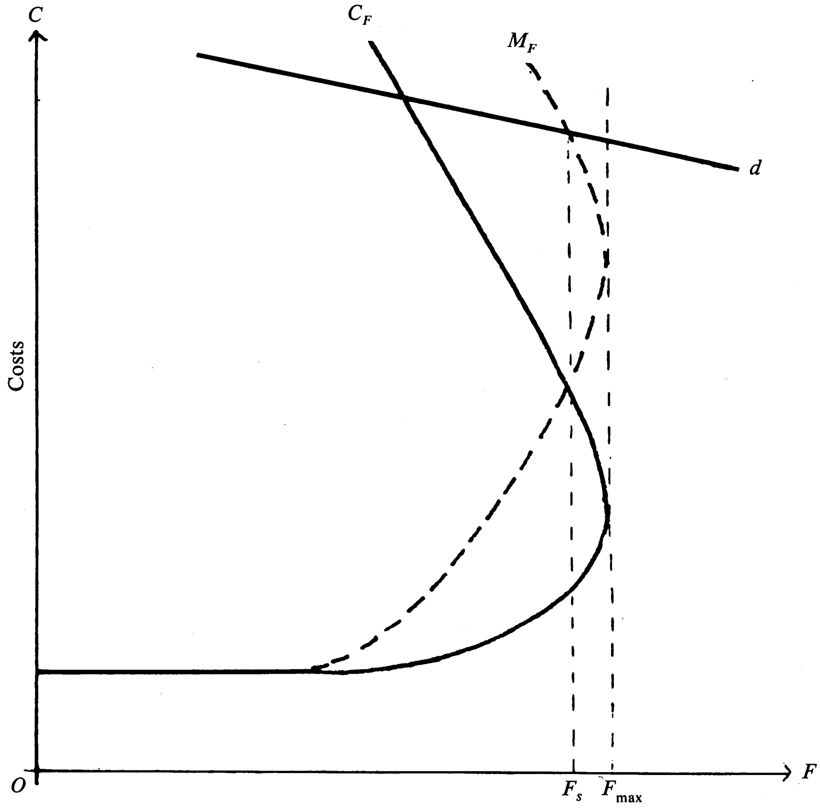

FIGURE 3

Moreover, there would appear to be no problems over second-order conditions. In fact, contrary to Nash's assertions, the second-order conditions are satisfied when the backward-bending marginal cost curve is steeper than the demand curve, as in the case illustrated in Figure 3. This becomes clear as soon as it is remembered that movements up the backward-sloping section of the marginal cost curve, M_F, are associated with increases in the number of vehicles on the road, even though the traffic flow is declining. Hence, in the vicinity of the point of intersection of M_F and demand curve *d*, as the number of vehicles increases marginal costs are increasing faster than marginal benefits, and so the normal second-order conditions for a maximum are satisfied. They would not be satisfied if the demand curve was steeper than the marginal cost curve; but in that case there would always be another point of intersection at which the second-order conditions would be met.[3]

[3]This second point of intersection could be either on the backward-bending part of M_F or on its upward-sloping part, depending on the precise shapes of the two curves.

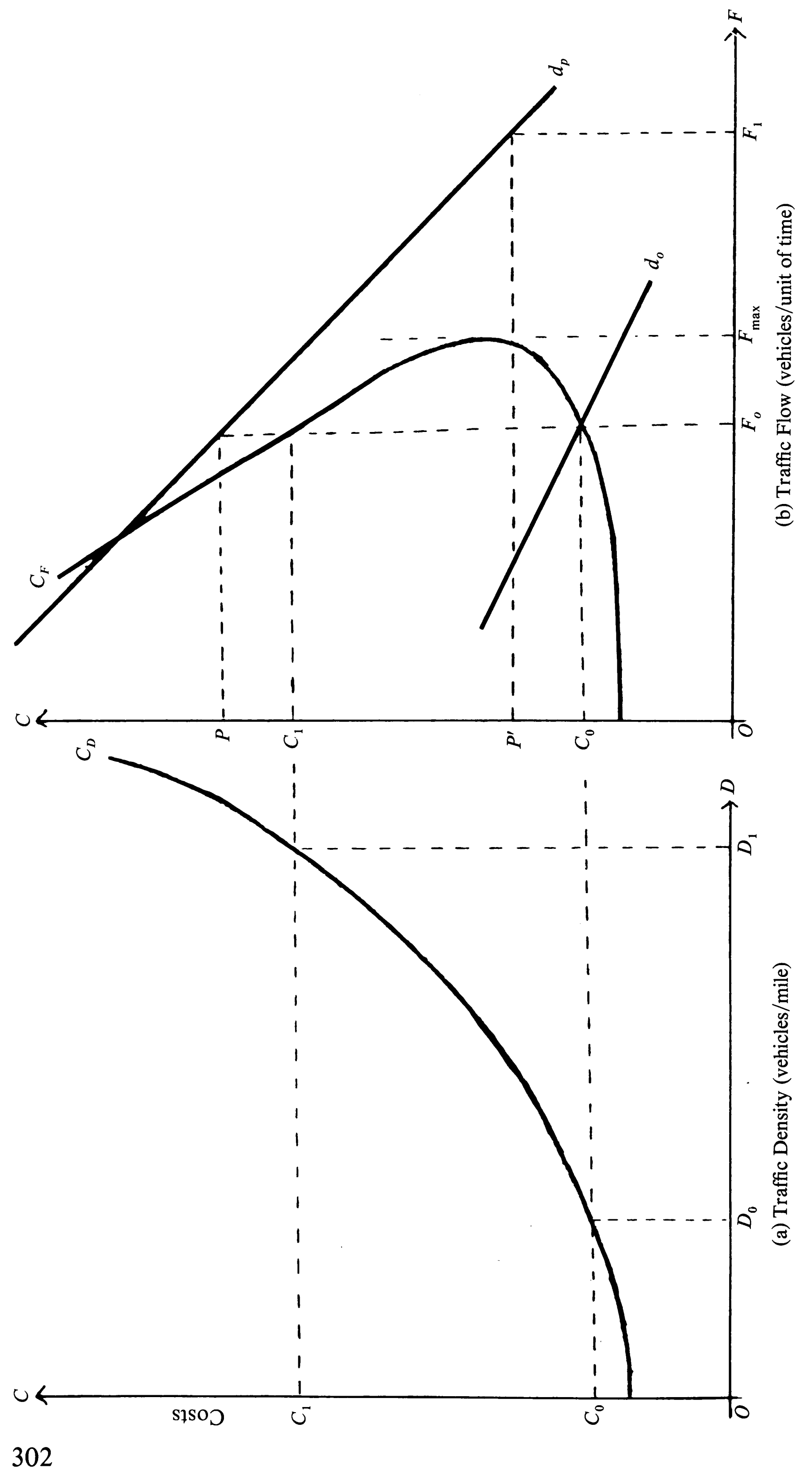
Costs
(a) Traffic Density (vehicles/mile)
(b) Traffic Flow (vehicles/unit of time)

FIGURE 4

Provided therefore that the marginal cost curve is steeper than the demand curve, an optimum position on the backward-sloping part of M_F, and hence C, remains a possibility. Nevertheless, since it is a somewhat counter-intuitive result, some further explanation might be helpful.

In Figure 4, where C_D is the relationship between traffic density (D) and journey costs (C) along a given road and C_F is the derived relationship between the traffic flow (F) and journey costs, suppose that d_o is the off-peak demand curve for the use of the road and d_p the peak-period demand curve. Suppose also that the road authority has decided on a purely arbitrary basis to try to maintain the peak-period equilibrium traffic flow at the off-peak level of OF_o. It could do this most straightforwardly by imposing a charge for peak period use of C_oP. However, it could also achieve the same objective by fixing the peak-period charge at the lower level of C_1P ($= C_oP'$ in the diagram). In this case, faced with an initial overall journey cost of OP', the flow of traffic on to the road would be OF_1, exceeding OF_o, the current flow off the road. Traffic density would therefore increase, and so would journey costs. This would reduce demand until eventually, with total journey costs at OP_1, an equilibrium with the desired traffic flow of OF_o would again have been reached, but with a traffic density of OD_1 rather than OD_o. The question to consider therefore is whether this situation could ever yield any net welfare gains over that with the lower traffic density.

Clearly in this case there would be additional costs of up to C_oC_1 per journey,[4] and the total additional costs would increase with the duration of the peak-period traffic flow. There would also be some benefits, however, because the lower charge for the use of the road would permit some journeys (initially F_oF_1, per unit of time, but declining to zero as overall journey costs approached OP) to be started which would have been priced off with a charge of C_oP. The value of the net benefits arising from these journeys would range from C_oP to zero, and the total net benefits involved would be higher the greater the peak period demand (or more particularly the higher the price road users were prepared to pay with the traffic flow OF_o). Hence it would appear that, if the peak level of demand was sufficiently high and of sufficiently short duration, the benefits from allowing the traffic density to rise to OD_1 could exceed the costs—and this of course is precisely the kind of situation in which the optimum position might be located along the backward-sloping part of C_F.

However, while such a position is possible, its significance should not be exaggerated. In the simplest possible case, with no following traffic and where the only costs are time costs valued at a constant rate per unit of time, it can be shown that at the maximum traffic flow marginal social costs would be double the costs per journey.[5] An optimum on the backward-bending part of the marginal cost curve could then occur only if, at the maximum traffic flow, road users were willing to pay more than twice the cost of each journey to use the road. With following traffic the divergence between marginal social cost and journey costs at the maximum traffic

---

[4] Initially additional journey costs would be zero, but they would rise to C_oC_1 as the equilibrium position was approached.

[5] Equation (4) in Else (1981, p. 221) give the following expression for marginal social costs when there is no following traffic:

$$M = C'(T)T'(D)D + C.$$

When the traffic flow is at its maximum level, $T'(D) = T(D)/D$, and if $C = \theta T$ where θ is the value of time per unit, by substitution $M = \theta T + C = 2C$.

flow would be even greater, requiring an even higher level of demand to produce an optimum on the backward-sloping sections of the cost curves. It thus appears that the situation can arise only where roads are heavily overloaded for relatively short periods of time. The longer the peak level of demand continues, the more likely it is that the optimum will be located on the upward-sloping sections of the cost curves.

This does not, of course, mean that the "conventional analysis" will then apply, since, as argued above, that is based on an inappropriate definition of marginal cost. The conventional analysis is correct only for *permanent* changes in the flow of traffic, because in that case, and in that case alone, it is not possible for the less heavily utilised capacity in later periods of slacker demand to make any contribution to accommodating the extra traffic. But, as I argued in my original paper, bearing in mind the way traffic flows vary in practice from hour to hour on any normal road subject to congestion, that must be rather a special case.

REFERENCES

ALCHIAN, A. A. (1959): "Costs and Outputs". In Abromovitz, M., *et al., The Allocation of Economic Resources*. Stanford.

ELSE, P. K. (1981): "A Reformulation of the Theory of Optimum Congestion Taxes". *Journal of Transport Economics and Policy,* Vol. XV, pp. 217–232.