Congestion Toll Pricing for Public Transport Facilities

MARTIN WOHL, The Urban Institute, Washington, D.C.

Economists, policy-minded administrators, and planners are turning increasingly to congestion toll pricing as a practical and efficient instrument for solving the traffic congestion problem. However, in their advocacy of short-run marginal cost pricing as a replacement for the present method of roadway pricing (which for the public highway system approximates short-run average variable cost pricing), economists and others have generally relied on an over simple theoretical world in reaching their conclusion about the efficacy of marginal cost pricing. Rather than contend that marginal cost pricing is not the best method of pricing, this paper argues that it is not yet clear that marginal cost pricing is better than our present type of pricing policy.

•IT WOULD APPEAR that the case and particularly the public arguments for marginal cost pricing have sometimes, if not often, failed (a) to view short-run marginal cost pricing within a longer-run context and thus to view our pricing and investment policies as an inseparable package; (b) to properly consider some money and non-money costs and effects stemming from the abandonment of our existing pricing system in favor of a marginal cost pricing system; and (c) to consider the incidence of one pricing system versus another (in the sense of "who gains and who loses").

Within these introductory remarks three other aspects of the so-called pricing problem deserve mention. First, although most of the discussion in the transportation literature, and certainly in this paper, deals with roadway pricing, the principles and issues should be regarded as a more general matter having applicability to virtually all types of transport systems. Second, it should be recognized but too seldom is that congestion in urban areas is not confined solely to highways and does not affect just the users of automobiles, trucks, and buses. In fact, it can and does confront even the users of transit systems—not only as they wait in queues when entering or exiting from subways, rapid transit stations, or buses, but also as they are crowded into and thus congested within transit cars or buses. The latter type of congestion, unlike highway traffic congestion, will not add greatly to the passenger's travel time (other than an increased wait to get onto and off transit cars or buses and an increased time for intermediate station stops), but it can and often does markedly affect his level of discomfort by virtue of the additional crowding. Furthermore, in these crowding situations, just as for highway congestion, each additional individual transit user (for a given time period) is causing the total costs to all transit users to increase more than the private cost he faces. In short, if a person jams onto a subway car, not only does he suffer the discomfort of the "crush", but he also causes the other occupants of the car to suffer additionally from his entrance. Thus, arguments about highway users causing congestion for all other highway users are entirely analogous to additional transit users with respect to their crowding and comfort levels.

Third, some supporters of marginal cost pricing (often referred to as congestion pricing) assume that traffic congestion is intolerable and therefore the problem is to

Paper sponsored by Committee on Highway Engineering Economy and presented at the 49th Annual Meeting.

reduce traffic congestion, regardless of all else, and that any measure to reduce traffic congestion—whether it is the imposition of congestion tolls to reduce traffic flow or to shift users from automobiles to transit modes or to shift their travel hours from the peak to off-peak periods—is better, a priori, than existing conditions. To the contrary, there are three choices open to society: (a) continue to endure congestion, whether on jammed highways or in crowded transit vehicles; (b) reduce the traffic flow or passenger ridership, whether through pricing mechanisms, administrative controls, or physical barriers; or (c) expand the highway system capacity or number of transit lines, vehicles, or trains. The wisdom of the second or third alternative relative to the first will depend principally on the extent to which congestion will be reduced, on the value of this reduction to those enjoying faster travel or reduced crowding, on the disbenefit to those "forced off" or affected by those "forced off" the facilities should the second alternative be adopted, on the extra travel benefit accruing to new users, and on the cost or resource commitments necessary to expand the system capacity should the third alternative be adopted. Furthermore, there will be equity and income redistribution considerations that should enter the decision-making process.

The remainder of this paper will examine the consequences of different pricing policies by focusing on (a) identification of the "gainers" and "losers" with respect to marginal cost versus average cost pricing, (b) consideration of the costs associated with implementing different pricing schemes, (c) the efficacy of imposing marginal cost pricing in situations when the roadway system is considered as fixed or unexpandable for all time, and (d) the short- and long-run adjustment problems stemming from expansion and the abandonment of average cost pricing in favor of short-run marginal cost pricing. Importantly, though, this discussion will deal only with the non-backward-bending or non-reducing capacity case for transport facilities (1, 2, 3).

SHORT-RUN CONSIDERATIONS

"Gainers" and "Losers" for Different Pricing Policies

Most economists argue that short-run marginal cost pricing will lead to maximization of net benefits. This consideration is generally founded on the assumption that marginal cost pricing prevails throughout the economy, and that the costs of implementing such a pricing scheme and of countering any adverse effects on employment and income distribution are negligible. In such a simplistic world, even I would agree. But such is not the case and, as often noted, other objectives (4, 5) might be ill served by a sharp reversal of the existing average cost type of pricing policy and a move to marginal cost pricing.

To explore this last point, we can make use of the cost, price, and demand relationships shown in Figure 1. Assume that the variable cost curve includes not only the variable supporting-way and vehicle operating and maintenance costs (and other similar expenditures that vary with the flow rate) but also the costs for travel time, effort, discomfort, and safety

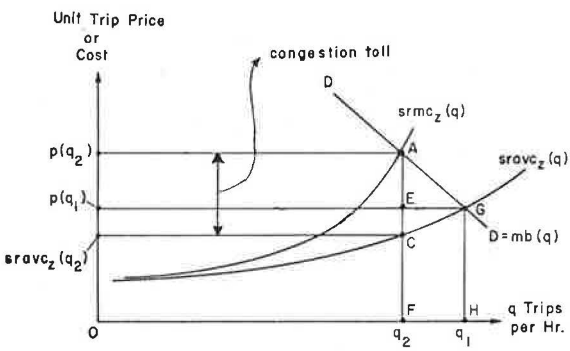

Figure 1. Cost, price, and demand relationships: $srmc_z(q)$ = short-run marginal costs for flow q on facility z; $sravc_z(q)$ = short-run average variable costs for flow q on facility z; DD = demand function for facility z = mb(q); p(q) = price for flow q (for certain pricing policies); $sratc_z(q)$ = short-run average total costs for flow q on facility z; and mb(q) = marginal benefit for flow q on facility z.

NOTE: These curves can be applied with equal validity to both highway and transit travel. In the former case, q can be regarded as the hourly vehicular flow on a particular highway; the increase in variable costs with increases in q stems mainly from increased travel time. In the latter case, q can be regarded as the hourly passenger flow arriving at a particular bus stop or rail transit station; the increase in variable costs with increases in q stem partially from increased travel time but mainly from increased crowding and discomfort as buses or trains become overloaded (relative to the number of seats). In the short run, the highway vehicular capacity should be regarded as fixed.

hazard that are incurred by travelers; assume further that, under the pricing policy now practiced for highway and transit systems, the traveler in deciding whether to travel and how much to travel perceives his trip payments to total and be equal to the short-run average variable cost. More specifically, I have assumed the following:

1. For a given highway trip, users perceive the variable user gasoline, tire, and parts taxes equally and as part of their complete money-time-effort payment. These taxes are assumed to be variable with respect to the quantity and length of trips but uniform per vehicle-mile. Also, the unit variable user gasoline-tire-parts taxes are assumed to be just equal to the unit variable costs for operating and maintaining the highway (which are assumed to be constant for all levels of flow).

2. For transit travel, it is assumed that the transit fare is perceived by transit users and that it is just equal to the unit variable costs for operating and maintaining the transit system. Also, for highway travel, the complete money-time-effort variable costs increase at an increasing rate because of the increase in congestion and travel time with increases in flow q, whereas for transit travel the increasing rate stems from increases in congestion and passenger discomfort (resulting from crowding) with increases in flow q. In the latter case, as a bus or as the cars of a rail transit train become more and more crowded with increases in the passenger load (that is, with increases in q), not only will crowding and discomfort increase but travel time will also because of the stops required.

These sets of assumptions are made mainly for convenience and may do some violence to the actual facts; that is, it may well be that the highway-user tax in some instances more than covers the highway maintenance and operating costs and in some cases less. Similarly, transit fares may sometimes cover more than the transit maintenance and operations expenses and others less. For many dense urban situations, however, I suspect that these assumptions are not far from the truth.

With reference to Figure 1, it is evident that if we switched from the present average variable cost type of pricing policy to marginal cost pricing (while ignoring the costs of implementing a workable marginal cost-pricing system), some of the existing q_1 travelers (or q_1 minus q_2 travelers) would be unable or unwilling to pay the toll AC that would be required to bring about marginal cost pricing. In the short run these q_1 minus q_2 users would suffer a loss as they switched to a less preferable mode, to a less preferable hour of travel, to a less preferable route, or to less travel (e.g., they decided to travel less often or not at all). Furthermore, all of the q_2 users—even though able and willing to pay the toll of AC—would suffer a unit loss of AE with marginal cost pricing (relative to an average variable cost-pricing policy); that is, although their unit congestion costs would be reduced by an amount EC when the volume was reduced from q_1 to q_2, their money payments would be increased by AC, an amount that would exceed the reduction in congestion costs by unit amount of AE. Thus, all of the users in the short run—those continuing to use the facility and those forced off of it—are worse off. (It is presumed throughout that motorists are homogeneous with respect to time and congestion values and costs.)

The obvious question arises: How can net benefits to society be increased when the benefits to all users of the facility are decreased by a switch to marginal cost pricing? The problem is simply that the increases in net benefit have been extracted from the q_2 users in the form of tolls and (presumably) have been transferred to other parties, political groups, or the like. Thus, the users have suffered a loss, individually and in aggregate, but those receiving the tolls and society in aggregate have received a gain. Needless to say, the imposition of marginal cost pricing would hardly induce the users to look on the results as "optimum" or "efficient" and it is likely that they would view the matter as highly inequitable, at least in the short run. Finally, I would point out that there may be other losers as well; that is, some, if not most, of those users unwilling or unable to pay the toll required by marginal cost pricing would shift to other routes, other modes, or other times of day. In the process, they would usually increase the travel time, congestion and/or crowding on those other routes or modes, both for themselves and for the other users.

Practicalities and Costs of Different Pricing Policies

The Highway Case—When considering the wisdom of abandoning average variable cost pricing in favor of marginal cost pricing, it is necessary to develop a pricing system that will inform users, in advance of their making a trip, about the marginal cost prices they will face if they decide to travel. To develop a system in which the prices would be hidden or in which users are billed later in some aggregate fashion (e.g., by receiving monthly bills) would tend to defeat a major purpose of marginal cost pricing, namely, to ensure that users are aware of the costs stemming from additional trip-making. Ideally, of course, the pricing system would be pervasive with respect to all facilities, modes, times of travel, and so forth. Prices for given facilities would change from hour to hour and from day to day as the equilibrium flows and marginal costs would change in response to fluctuations or changes in demand. And certainly the system should reflect the demand relations for each mode of travel during each hour of the day as well as the cross-relations with respect to other modes and times of day for trip-making. However, were such a flexible and pervasive pricing system to be instituted—and one can hardly argue otherwise if the case for marginal cost pricing is to have a solid basis—it seems evident that an extensive and expensive toll-collection system would be required.

Although many types of electronic toll-collection systems have been talked about, whether they can be practically developed and applied remains to be seen. (Again, a necessary requirement would be that prices for trip-making be known in advance.) In the interim, of course, it would be possible to use the usual tollgate-type system for implementing marginal cost pricing. On the one hand, the use of tollgates on roadway systems would ensure that users are confronted with the actual short-run marginal costs rather than permit some people to travel even though their marginal benefit is less than the marginal costs for volume rates of that magnitude. But, on the other hand, to institute de novo a tollgate or other system would require resource expenditures not only for the construction and operation of tollgates, but it would also cause the system users to suffer additional travel-time delays and time costs while waiting to be serviced at the tollgates. [Of course, delays at tollgates could be reduced to virtually nothing if sufficient extra tollbooths were provided and operated. There is an obvious trade-off between short-run travel-time delays and long-run tollgate costs, although it is presumed that the most efficient gate capacity would result in some queuing delays.]

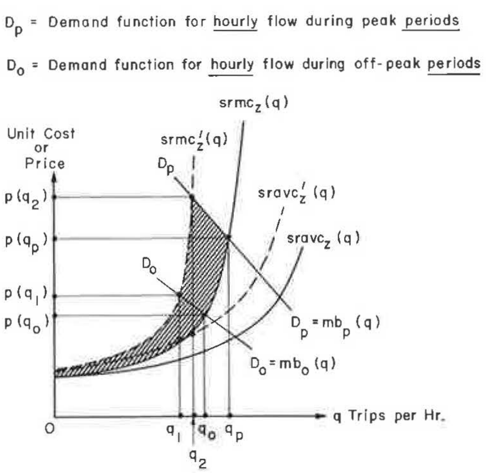

Figure 2. Short-run cost, demand, and pricing relationships for marginal cost pricing: D_p = demand function for hourly flow during peak periods; D_o = demand function for hourly flow during off-peak periods.

The full marginal costs, to include the variable costs of implementing and being delayed by a marginal cost-pricing system, may be represented somewhat as shown by the $srmc'_z(q)$ curve in Figure 2, where $srmc'_z(q)$ equals the short-run marginal cost, including pricing implementation and delay costs, for facility z at volume rate q. These costs together with the fixed costs required for the tollgates can be compared with $srmc'_z(q)$, the short-run marginal costs for "costless" marginal cost pricing, and with the short-run average variable costs both with and without the costs of implementing the pricing system, $sravc'_z(q)$ and $sravc'_z(q)$ respectively. Relative to costless marginal cost pricing, the full marginal costs will cause the peak and off-peak period costs and prices to be increased, thus reducing the hourly flow from q_p to q_2 during peak periods and from q_o to q_1 during off-peak periods. The comparisons between hourly flows and between net benefit totals for different pricing policies will be made by using the costless marginal cost curve and pricing policy as a base. This base is used in order to simplify the graphical illustrations of the changes in total net benefit. (This assumes, of course, that the hourly demand throughout the day can be represented by hourly demand functions for just two time periods. Also, this representation ignores the cross-relations between peak and off-peak time periods, two obvious oversimplifications.)

At this stage of the analysis it is appropriate to ask whether, in light of the additional travel delays and toll collection costs, marginal cost pricing still appears to be the most efficient pricing policy. Obviously, the answer depends on what other alternative pricing schemes and pricing policies are available. But it is of utmost importance to note that, given these broader and more realistic conditions accompanying the advocacy of marginal cost pricing, one can no longer argue a priori that marginal cost pricing—even in a perfect economic world—is or is not better than even average cost pricing or quasi-average cost pricing.

Alternatively, the present pricing policy for most public roads might be looked on as a quasi-average total cost pricing scheme in which the perceived roadway price is equivalent to the uniform user tax plus the short-run average variable cost or $sravc_z(q)$. This view is somewhat different from that adopted earlier for the general comments about the effects of different pricing policies. Here it is implied that the present highway user taxes cover more than the variable facility operating and maintenance expenses. In fact, many would estimate that, for the highway system in aggregate or for that within central cities (again, taken as a whole), user taxes in total cover both the fixed and variable costs for the facilities and their use (6, 7, Tables 4 and 5). For these assumptions, the price to the user would be represented by the $srtvc_Z$ (q) curve shown in Figure 3. Of considerable importance, however, the costs of administering this pricing policy for highways and of collecting the user prices (uniform tax plus the user's expenditures of time and effort) are virtually nil, particularly because travelers are not delayed by the collection scheme. Thus, as a practical matter, the average variable cost plus uniform tax pricing policy is costless for highways. Relative to costless marginal cost pricing, as represented by $srmc_Z(q)$, we would usually find that hourly flows and congestion during peak periods would be too high and during off-peak periods would be too low. [Of course, the uniform tax level could be set high enough to result in a marginal cost price being charged during peak (and off-peak) periods, with an overall price of $p(q_2)$ and flow q_2 as shown in Figure 2. However, flow during off-peak periods would be far too low. On an a priori basis, no conclusion can be reached about the "best" or "better" level for the uniform tax.] Relative to costless marginal cost pricing, this average variable plus uniform tax cost pricing scheme would cause hourly losses in net benefits during the peak and off-peak periods roughly as shown respectively by the right and left shaded areas in Figure 3. [The losses in net benefits are equal to the difference between the total net benefits for the pricing policy in question and those for the costless marginal cost pricing policy. Total net benefits for any given pricing policy are defined as the difference between total benefits and total costs, the latter to include any fixed costs. The total benefits may be calculated by summing the marginal benefits—or integrating under the demand curve—up to the equilibrium flow level. Total costs may be calculated by summing marginal costs up to the equilibrium flow level and then by adding them to any fixed costs, where applicable, for new or existing facilities (2, 3).] These losses can be compared to those stemming from full marginal cost pricing, which are represented by the shaded areas below the

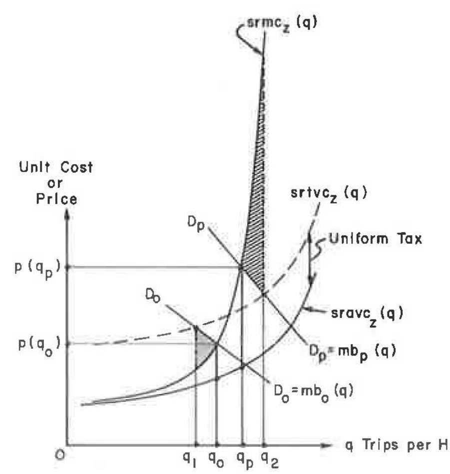

Figure 3. Short-run cost, demand, and pricing relationships for uniform user tax pricing policy.

peak period (D_p) and off-peak period (D_O) demand functions in Figure 2 plus the fixed costs required to install the tollgate or other pricing system.

On balance, it is not clear which pricing policy will result in the smallest net benefit losses, and thus only a full-scale benefit-cost analysis will indicate which of these (or other) pricing policies will be most efficient economically or will be better on other grounds. A second but equally important conclusion can be drawn with respect to pricing highway travel. That is, because the costs of administering uniform user taxes are nil, or virtually so, some user tax is preferable to no user tax or to "free" highway travel. For example, referring to Figure 3, it can be seen that a uniform tax set high enough to bring about a price of $p(q_O)$ and an hourly flow of q_O during off-peak hours will result in lower peak and off-peak hourly losses in net benefits—relative to costless marginal cost pricing—than would a zero uniform tax.

The Transit Case—The situation for transit facilities is somewhat different from that for highways, other than for the case of "free transit". That is, except for free use of transit facilities, there is no way of administering either a uniform or differentiated price that will be costless (or virtually so) and that will not delay passengers during collection. If a uniform-fare (plus short-run average variable costs) pricing policy were to be used for the transit system, the situation would be similar to that shown in Figure 4. For this situation, the $srmc'_z$ (q) curve represents the short-run marginal costs for facility z operating at hourly flow q, to include the variable costs of implementing the collection system and of delaying passengers while collecting fares. Curve $srfvc_z(q)$ represents the price function faced by travelers and is the sum of the uniform fare and short-run average variable costs (to include those caused by crowding and discomfort as vehicles become more and more crowded). During peak periods, the hourly flow would be q_2 and the price would be $p(q_2)$ for this uniform fare policy and, relative to the net benefits resulting from costless marginal cost pricing as represented by the $srmc_z(q)$ curve, the resulting hourly losses would be equal to the entire dashed shaded area plus the fixed costs for the collection facilities. During off-peak periods, the hourly flow would be q_1, the price would be $p(q_1)$, and the hourly losses (again, relative to the net benefits obtainable with costless marginal cost pricing) would be equivalent to the entire dashed and dotted shaded area lying below D_O (the off-peak demand curve) plus the fixed costs for the collection facilities.

One should ask whether the use of differential peak and off-peak period transit fares would help to reduce the above losses in net benefits. The answer, almost certainly, is yes. (This conclusion would need no qualification if the costs of administering and collecting differential fares, including delays to users, were equal to those for a uniform fare system. Although differential fares may entail slightly higher collection costs, the remarks that follow will assume that the increase is negligible. If the peak-period fare were increased so as to bring the total user price up to point A shown in Figure 4, the loss in net benefits lying between D_P (the peak-period demand curve) and $srmc'_z$ (q) could be eliminated. Similarly, if the off-peak transit fare were reduced so as to bring the total user price down to point B, the loss in net benefits lying between D_O (the off-peak demand curve) and $srmc'_z$ (q) could be eliminated. Although differential prices would almost certainly be more efficient than a uniform fare policy for transit facilities (in contrast to the case for highway facilities in which such a result may or may not be true), one may not assert that such a pricing policy is definitely preferable to "free transit" (that is, to a zero transit fare policy). For the free transit case, the hourly loss in net benefits relative to the

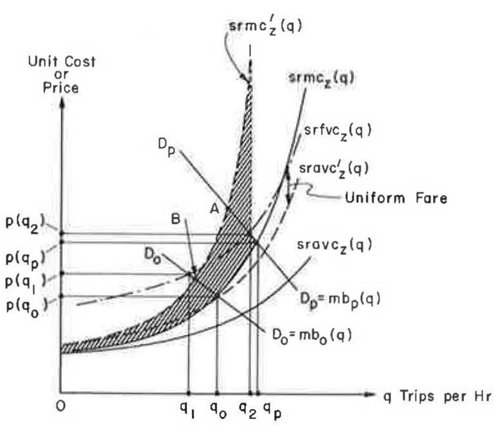

Figure 4. Short-run cost, demand, and pricing relationships for uniform transit fare pricing.

costless marginal cost pricing case would be equal to the dashed shaded area in Figure 5 during peak periods and to the dotted shaded area during off-peak periods. On an a priori basis, it cannot be argued that the sum of these shaded areas is either larger or smaller than the combined sum of fixed costs of the collection facilities and the shaded areas for either the uniform or differentiated transit price policies discussed earlier and shown in Figure 4.

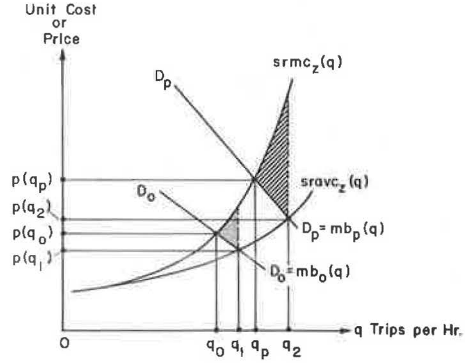

Figure 5. Short-run cost, demand, and pricing relationships for "free transit" pricing.

LONG-RUN CONSIDERATIONS

Practical Aspects of Pricing and Facility Expansion

Considerable attention should be devoted to the efficacy of imposing marginal cost pricing in situations when further expansion of the roadway system is regarded as impossible for whatever reasons and regardless of whether or not its expansion is more efficient economically. Gabriel Roth (8) seems to adopt this attitude, for example, in saying "In most cases access is provided by one road only, and the provisions of further roads is impossible because of the technical layout of built-up areas. In these circumstances competition in any area is effectively impossible. Any firm or individual owning an access road would be in a monopoly position." Perhaps to put the matter in more technical terms, one should consider pricing in broader perspective and as just one of several links in the investment-pricing-operation chain rather than as a matter for short-run consideration at the present. (To simplify the discussion of this point, the costs of implementing pricing systems will be ignored.) Also, innumerable studies and articles on the subject of roadway and congestion pricing suggest that a fundamental purpose of pricing is to reduce congestion, to lessen peaking, or to shift people from automobiles to transit. More pertinently, pricing is simply a mechanism to guide consumers in making decisions among the choices open to them, to help establish the values of products or services, and thus to aid private or public firms or agencies in their decisions about investments, operations, products, and services. (These comments are made in the context of maximizing the public's economic welfare and are aside from matters of equity and income redistribution.) In the context of the public transport problem, pricing is merely one of the instruments available to help determine how many and which kind of modal services should be provided, the extent to which the services should be offered, and the appropriate levels of operation and congestion (or, say, performance) for those modal systems. For highways, the appropriate use of pricing and the concurrent analysis of the additional costs of expansion and extra value of increased trip-making and reduced congestion can serve to guide decisions about the amount of highways and number of lanes needed. Similar trade-offs should be used in making decisions about the wisdom of direct pricing mechanisms, about the number of tollbooths (should they prove to be warranted), and so forth. For transit systems, good pricing and incremental benefit-cost analysis can aid significantly in making decisions not only about the feasibility of various modal systems, together with their route coverage and trackage requirements, but also about schedule frequency and train lengths. For the latter, transit operators can more sensibly make trade-offs between different crowding levels (and the associated loading and unloading times) within buses or trains and different bus or train frequencies, bus sizes, and train lengths rather than rely on arbitrary load factors or operating rules of thumb to make such decisions.

Pricing is a short-run proposition and it is a matter of determining day-to-day prices, given the market (as represented by appropriate demand functions), given a pricing policy, and given the actual day-to-day variable costs stemming from operation and use of a particular facility. And it is a problem of determining which pricing policy will result in the most efficient (or "maximized" net benefits) day-to-day operation and use or volume (and thus congestion) levels. From day to day, one can do no better than maximize the daily efficiency (i. e., net benefits) of the system that is in place and in operation. But of equal clarity and importance, it is all too obvious that the system size, its service, and its operating characteristics can be altered over time, thus changing the cost functions and the resultant use, prices, net benefits, and so forth. Consequently, our concern with pricing should rest not only with the present-day, short-run operating circumstances of a given facility or system but also with the effects of long-run changes to the facilities and their operations.

To view this problem more directly, I would note that many economists, when analyzing the roadway pricing problem, have concluded that present uniform user tax levels and pricing policies have led to pricing highway travel (particularly in core areas of cities) far below short-run marginal costs and thus have led to economic inefficiency. Crucially, though, not all of these economists have appropriately analyzed either the long-run implications of this finding or conclusion or the practical considerations and costs stemming from the implementation of a marginal cost pricing system (the latter in the sense discussed in the first part of this paper). And it is here that the contradiction seems most apparent. Much to the point, economists have continually concerned themselves with economic inefficiencies stemming from poor short-run pricing policies but have sometimes failed to examine and pinpoint the manner and extent to which economic inefficiencies also can and do result from poor investment and expansion policies. Inefficiency with respect to pricing is no more onerous or important than that with respect to expansion.

Furthermore, one is inescapably led to conclude that economists, by arguing for the institution of congestion tolls set equal to the short-run marginal costs without specifically considering the long-run possibilities and implications as well, are perhaps unwittingly lending support to the usual contentions that traffic congestion must be abated; that traffic is strangling and choking the downtown and central city core areas; that the construction of more highways will only lead to more traffic, more congestion, and more strangling; that the urban highway-expansion program (particularly the Interstate highway links) should be slowed if not halted; and that more rapid transit should be constructed in place of highways. The unrest with respect to the contention that highway expansion should be slowed if not halted probably stems more from past failures to recognize and take account of certain externalities and income redistribution problems (for those dislocated or affected by specific highway locations), from failure to compensate nonusers for their social costs, from poor aesthetic designs, and from an inability (or failure) to trade off these social and external costs with higher highway location or design costs than from matters of pricing and long-run investment policy. (It should be obvious, of course, that most economists would hardly endorse such bald and unqualified contentions, and certainly not without having been provided the appropriate benefit-cost analyses to validate the views.)

It may be, for example, that traffic is strangling and choking the CBD or the central city. But such a conclusion must rest on deeper consideration of many aspects other than the mere existence of traffic congestion on streets and expressways. In the first place, the fact that workers or shoppers will actually put up with traffic congestion during certain times of day and will continue to travel by automobile—rather than forego trips altogether, rather than use mass transit (where it is available), or rather than travel at some less congested time of day—is ample proof that automobile trips during those times of day are highly valued, both in an absolute sense and relative to other available opportunities, and it is proof that reduction of automobile travel will lead to a reduction in benefits as well as in congestion costs. A related matter of some considerable importance is that automobile and mass transit can hardly be regarded as highly or even reasonably substitutable services for either work or shopping trips and that significant shifts from automobile to mass transit are not in the offing. [The evidence on these matters is far from complete and is, of course, limited to the presently available set of modal opportunities. However, the few competent studies thus far conducted bear this out. See, for example, the automobile and transit demand elasticities, with respect to automobile and mass transit travel times and costs, tabulated elsewhere (9).] As a consequence, a shift to congestion toll pricing (at short-run marginal cost) for highways not only would, in all likelihood, reduce traffic and its congestion but also would reduce the number of person trips (and the associated benefits). Thus, a mixed blessing may result, and the city may well be crippled more by the loss of workers and shoppers (and their expenditures and contributions) than by the reduced congestion for those willing to pay the tolls. In fact, it is difficult to envision how any significant gains —other than less noise and air pollution, if these are that important—will accrue to the city itself or to its businessmen merely from having traffic congestion reduced, unless, of course, there are excess toll revenues that will be distributed to the city and its businessmen or used to reduce their taxes. In reference to excess revenues, if—as is probably the case in many and perhaps most urban core situations—the demand for highway travel is high enough to produce considerable congestion and to result in equilibrium flows above the level corresponding to that at which the short-run average total costs are at a minimum point (i.e., if equilibrium flow q is above the level q_O shown in Figure 6), abandoning the average variable cost plus user tax pricing in favor of marginal cost pricing will result in excess money revenues being generated. As can be seen in Figure 6, the revenues in excess of total costs to society would be equal to q_A times the difference between $p(q_A)$ and $sratc_x(q_A)$. Also, if the situation shown in Figure 6 is representative of urban core highway conditions at present, then so long as facilities are not improved or expanded, we can expect the excess revenues or profits to continually increase over time. These increases in profits will occur, of course, so long as demand shifts (upward and to the right) in response to population shifts, income increases, and so forth. The excess revenues alone should induce the economist and the engineer (and the "city fathers") to consider the obvious question: Should the facilities be expanded?

Clearly, to talk about the necessity or economic wisdom of instituting marginal cost pricing and to ignore government policies with respect to investment planning is tantamount to being negligent or unobjective. To put the matter more strongly, I suspect that some planners and analysts would (for noneconomic and subjective reasons) be delighted to see short-run marginal cost pricing instituted and to have the urban highway program halted even if long-run considerations indicate that it is more efficient economically to expand the highway system. In a sense, this suggests that they feel it is "impossible" or would do irreparable harm to expand the highway system within urban core areas, and thus they forego consideration of this possibility.

Effects of Expansion Under Nonconstant Returns to Scale

In a rough way, the short- and long-run effects of following different pricing and investment (or, say, expansion) policies can be examined by referring to the cost and demand relationships shown in Figure 7. Of course, even these situations and functional relationships are oversimplified, particularly with respect to the characterization of demand as static, both during the day and over the years. On the other hand, both increasing and decreasing returns to scale cases are represented.

Given cost and demand functions, such as those shown in Figure 7, one can determine which facility size or capacity will maximize net benefits over the long run by noting the specific facility associated with the intersection between the demand and the long-run marginal cost curves. (If the demand function intersects the long-run average total cost curve, one may be sure that net benefits will be positive, fixed costs included.) Or, put in other terms, facilities and, in turn, the output should be continually expanded so long as the incremental benefits are larger than the additional fixed and variable costs stemming from increased

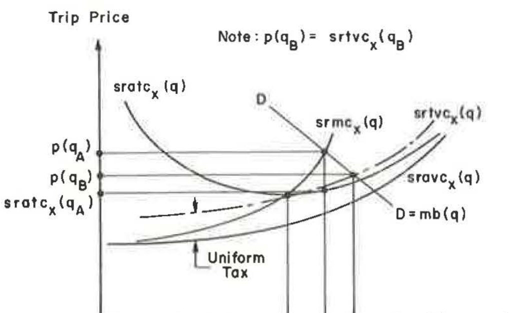

Figure 6. Short-run cost and demand relationships for a fixed facility and high demand.

capacity and output (these costs being represented by the long-run marginal cost curve). Similarly, one should reduce facilities and output so long as the loss in benefits is less than the reduction in fixed and variable costs. At the point when the marginal benefit just equals the long-run marginal cost, or point E for demand curve DD in Figure 7 and point F for demand curve D′D′, the long-run marginal cost will also equal the short-run marginal cost for the proper facility (i.e., the facility of lowest total cost for that output or q level, with facility A having lowest total cost at output level q_A and with facility C having lowest total cost at output level q_C, and will determine the most efficient day-to-day operating price. Admittedly, idealized conditions are embodied within these cost and static demand functions, and the costs and other effects of implementing workable marginal cost pricing systems are not included.

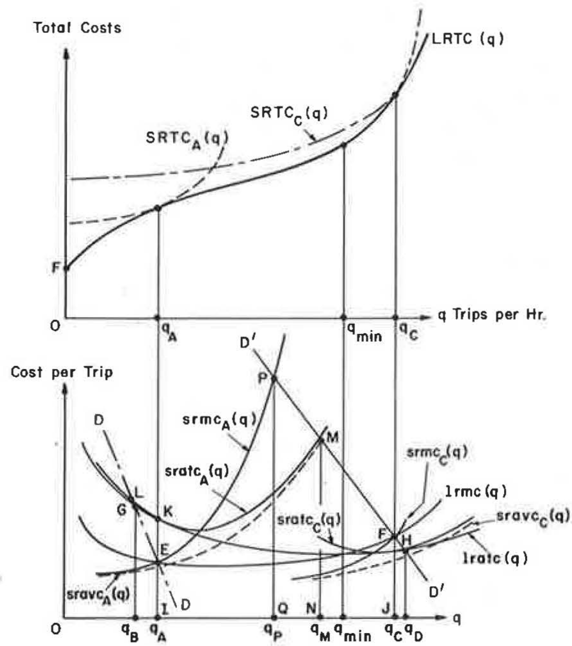

Figure 7. Long-run cost and demand relationships for other than constant returns to scale.

To follow the criteria outlined in the preceding, and to adopt facility A and output level q_A (with a price equal to EI) when the demand is DD or to adopt facility C and output level q_C (with a price equal to FJ) when the demand is D′D′, will result in maximizing net benefits to the public at large, regardless of whether a public or private facility is involved. [Using the same assumptions as before, the price is the combined time, effort, and money expense the user must forego to make a trip; furthermore, it is assumed that, aside from any toll, users perceive their time, effort, and money expense to be equivalent to the short-run average variable cost. Thus, for short-run marginal cost pricing, the toll for facility x and flow q must be set equal to $srmc_x(q)$ minus $sravc_x(q)$.] However, such idyllic planning and operating decisions will not always be forthcoming, either because of the lack of competition (both in the private and public sectors) or because of financial feasibility and pricing constraints (or, of course, because demand is hardly so static and predictable).

Were the demand level to be at DD, for example, no firm or public authority could build facility A, price its use so as to maximize net benefits (i.e., set its toll so that the total user price was equal to short-run marginal cost or EI), and cover its total costs. In fact, and even with some competition, firms or public authorities operating under these demand conditions would tend to build the minimum total cost facility for an output level of q_B or less rather than facility A. But if facility A were built by a firm or public authority, and if total costs had to be recovered (i.e., financial feasibility were a requirement), a high toll—equal to the difference between short-run average total cost and short-run average variable cost—would have to be charged. The total user price would correspond to the level indicated by point L, and the flow would be reduced to an amount slightly below q_B. Clearly, from a public point of view (i.e., from that of attempting to maximize net benefits to the public, regardless of who incurs the costs and who accrues the benefits), it would be in the interest of society to subsidize either public or private firms or authorities faced with similar demand and cost conditions (as shown in Figure 7 and indicated by demand DD) to the extent required in order to encourage proper planning and pricing. More specifically, both private and public firms and agencies should be encouraged to build facility A (for the DD demand case) and to set the toll equal to marginal cost minus average variable cost. In this case, the total user price would be EI, the toll would be equal to EI minus $sravc_A(q_A)$, the subsidy per trip would be KE, and the total (hourly) subsidy to the firm or agency would be equal to the product of KE and q_A. Just to place this discussion on more realistic grounds, it is likely that certain, though not many, low use and high fixed-cost turnpike and bridge authorities (e.g., perhaps the Massachusetts Turnpike Extension) find themselves in this increasing returns to scale situation and are required (because of commitments to bondholders) to price so as to cover total costs, thus causing the public at large to forego the extra net benefits accruing from a lower toll and higher use. In such instances, the local, state, or federal governments would do well to consider a subsidy (assuming, of course, that the authority would adopt marginal cost pricing).

The short-run facility A cost and D′D′ demand relations shown in Figure 7 are probably more typical or representative of present-day conditions for public highways in many dense urban core areas and thus can be used to focus attention more generally on the pricing and investment policy questions. Also, the short-run average variable cost curve for facility A can be regarded as the price function now being faced by travelers. As a result of these assumptions and conditions, use of and congestion on facility A is high, with an equilibrium flow of q_M and a total user price of MN. Clearly, this pricing policy and the resultant flow level cause serious short-run economic inefficiencies because some of the trip-makers (those represented by the demand function between flow q_p and flow q_M) will have marginal benefits that are less than the marginal costs attendant with an increase in hourly flow rate from q_p to q_M. With equal clarity it is evident that considerable long-run economic inefficiencies also result because for these conditions (i.e., facility A and demand D′D′) capacity is in short supply and grossly underexpanded. For this case and the relations shown in Figure 7, expansion to the level of facility C would bring about the following:

1. More trip-making would be permitted and thus total travel benefits would increase.
2. Total (fixed and variable) travel and facility costs would increase with expansion, but to a lesser extent than would the travel benefits. Net benefits would thus increase.
3. The price of travel for an individual trip-maker would decrease, whether the present-day user tax plus short-run variable cost pricing policy were to be continued or marginal cost pricing were to be adopted.
4. Congestion would be markedly reduced, even though the (hourly) volume rate would be increased.

At this point it is worth pointing out an anomaly that can result from considering only the effects of pricing policy changes rather than those of pricing and investment changes. (Again, let us assume that facility A is presently in existence and that the D′D′ demand curve is representative of the market for travel.) On the one hand, if the present pricing policy and its price function can be typified by the short-run average variable cost curve and if one restricts his attention to the efficiencies stemming from a switch to short-run marginal cost pricing and regards facility expansion as impossible for whatever reasons, the result will be to increase the total price of a trip from MN to PQ and to reduce the (hourly) volume rate from q_M to q_p. On the other hand, if both pricing and investment (or expansion) policy changes were to be adopted, thus switching to marginal cost pricing and expanding the capacity level to that of facility C, opposite long-run effects would result; after expansion, the trip price would be reduced from MN to FJ and (hourly) volume rate would be increased from q_M to q_C. Obviously, during the expansion period, prices—if adjusted to marginal cost—would fluctuate considerably. In the early years before new capacity was available, congestion, and thus trip price, would be high and flow would be reduced considerably (from that now witnessed with short-run average variable cost pricing). Later, after new capacity was made available and congestion dropped, the price would fall markedly and flow would increase.

The price fluctuations noted and the resultant shifts in trip-making (to other modes, to other routes, to other times of day, or to less trip-making) could, of course, cause some considerable anxieties for the traveling public as well as the business and employment groups they are dealing with. Among the numerous arguments against price fluctuation, particularly where the differences are large and when facilities have been seriously underexpanded, there are two worth noting. First, it should be recognized that certain individuals have taken jobs and businesses have established locations based on many factors including an expectation of the continuation of existing highway pricing policies. Given a switch in pricing policy, many of these individuals or firms might find themselves in untenable economic circumstances even though they were not (or only partially) responsible for the bad predictions about policy changes. Thus, it seems wise to ask whether it is fair for them to suffer the costs stemming from the switch in pricing policy (10). Second, Boiteux (11), among others, has argued the desirability of maintaining steady rates over periods of changing demand and expansion and of setting a constant price equivalent to one that would result if the facility capacity were always in perfect adjustment. This would seem particularly important when facilities have been seriously underexpanded.

REFERENCES

1. Walters, A. A. The Theory and Measurement of Private and Social Cost of Highway Congestion. Econometrica, Oct. 1961, pp. 678-681.
2. Wohl, M., and Martin, B. V. Traffic System Analysis. McGraw-Hill Book Co., 1967.
3. Wohl, M. Development of a Rationale for Transportation Investment. Univ. of California, Berkeley, unpublished DEng thesis, 1966.
4. Wohl, M. The Short-Run Congestion Cost and Pricing Dilemma. Traffic Quarterly, Vol. 20, No. 1, Jan. 1966, pp. 46-70.
5. Zettel, R. M., and Carll, R. R. The Basic Theory of Efficiency Tolls: The Tolled, the Tolled-Off and the Un-Tolled. Highway Research Record 47, 1964, pp. 46-65.
6. Lyle C. Fitch and Associates. Urban Transportation and Public Policy. Chandler Publishing Co., 1964, pp. 32 and 129.
7. St. Clair, G. P. Congestion Tolls—An Engineer's Viewpoint. Highway Research Record 47, 1964, pp. 66-112.
8. Roth, G. A Self-Financing Road System. The Institute of Economic Affairs, 1966, p. 31.
9. Charles River Associates. An Evaluation of Free Transit Service. Dept. of Transportation, Rept. for Office of Policy Development, Aug. 1968, Tables 3-1 and 3-2.
10. Meyer, J. R., et al. The Economics of Competition in the Transportation Industries. Harvard Univ. Press, 1959, Chap. 1.
11. Boiteux, M. Peak Load Pricing. In Marginal Cost Pricing in Practice (J. R. Nelson, ed.), Prentice-Hall, 1964.