Establishing Optimum Levels of Congestion on Toll Roads*

DANE J. COX†

The toll road or turnpike is a controlled-access highway. For the most part the rates charged for turnpike service do not result in the most efficient use of capacity. A more efficient use of turnpike capacity could be achieved if the rate structures were designed according to the following criteria.1

1. No user is allowed to purchase service at less than the full additional cost of serving him, an absolute minimum.

2. The output or quantity of service offered at any point in time is expected to clear the market at a price equal to the short-run marginal cost incurred in the production of that output, subject to the overriding constraint that total costs, including a fair return on capital, are to be covered by the user revenue generated over the anticipated useful life of the turnpike.

A unique characteristic of turnpike service is that an individual user imposes a cost on all other users if his purchases add to traffic congestion.2 Increased congestion is an added cost to the extent that it decreases the quality of service by raising vehicle operating costs, increasing accident costs, and adding to driver discomfort. Thus, the short-run marginal cost minimum paid by any consumer should include the assignable short-run variable cost and the marginal congestion cost for which he is causally responsible.3 Short-run marginal cost, SRMC, is the sum of the short-run marginal variable cost and marginal congestion cost.

The short-run is that period of time over which the product price and investment decision which alters the existing physical capital.

Short-run marginal cost, SRMC, is the marginal or added cost of the last unit of output produced under conditions of fixed capacity. SRMC pricing is the practice of charging a uniform price in the market equal to SRMC.

AVC (average variable cost) is the per-unit variable cost (excluding congestion cost) expressed as a function of output, and SRMVC (short-run marginal variable cost) is the additional to TVC (total variable cost) resulting from production of the last unit.

Congestion cost is an additional cost of output. It can be viewed in terms of total congestion cost, TCC, average congestion cost, ACC, or marginal congestion cost, MCC. In any case, the total operating costs (TCC) are equal to TVC + TCC, average operating costs (AOC) are equal to AVC + ACC, and SRMC are equal to SRMVC + MCC.

In Figure 1 we see a single vehicle class in a hypothetical market with traffic congestion occurring once congestion consumption exceeds Qℓ.4 AOC includes the per-unit congestion cost beyond point C. SRMC price is fixed at the congestion price, which is the marginal social cost of increasing the flow of traffic.

Note: The congestion level is fixed at the point where the additional cost of congestion is equal to the marginal social benefit of additional traffic.

This particular concept of marginal congestion cost is acceptable if it reflects the opportunity costs or foregone benefits of additions to congestion and if it can be demonstrated that turnpike capacity will be used most effi
ciently at the SRMC price, where the marginal congestion cost included in SRMC has been calculated according to the proposed concept.

To the extent that marginal consumption adds to congestion it decreases the quality of service. The real cost of increasing congestion would be the loss of benefits (decreased travel time, lower accident costs, etc.) to pre-marginal users. The costs of congestion, once it is a reality, are the opportunity costs of its occurrence; i.e., the benefits that its reduction would bring to pre-marginal users.

The amount which pre-marginal users, taken as a group, would pay to prevent marginal consumption would be equal to that which they would spend in the market to purchase the benefits they stand to lose if the quality of service declines, or the amount they would be willing to pay to displace marginal consumers and reduce congestion depends upon the value of the benefits they stand to gain if congestion is reduced. The only way in which pre-marginal consumers can convert a loss of benefits and minimize it or realize a gain in benefits in the short-run is to pay a price in the market high enough to discourage marginal consumers from purchasing service. The premium that pre-marginal consumers would be willing to pay just for the purpose of discouraging the consumption of others reflects the value of the loss of benefits that incremental consumption would render.

If the value of the benefits to be lost by incremental consumption, or the value of the benefits to be gained by its reduction, are equal to the aggregate sum of pre-marginal users taken as a group who would pay to prevent the loss or realize the gain, then the marginal congestion cost of incremental consumption can be said to reflect the real and opportunity cost of incremental congestion.

One might view the turbulence as resources which have been committed to a particular use and once these resources become sunk capital they are quite immobile. However, if the demand at current prices (short-run marginal variable costs) increases enough to cause traffic congestion, alternative degrees of congestion or levels of service are available and a particular level of service can be achieved by selecting the appropriate price-quantity combination. There are competing alternatives for the use of fixed capacity. Consequently, in order to have an efficient utilization of a limited resource, short-run fixed capacity, the toll rate must equate the benefits of increased consumption and the costs of increased congestion at the margin.

When we speak of using a limited resource most efficiently we mean that the opportunity cost of the marginal unit in production is equal among the various alternatives, i.e., the value of the marginal unit to production is equal between all competing uses. Turbulence capacity, even though fixed, is divisible, and is therefore transferable in the same that per-unit space or capacity decreases as additional or marginal consumers purchase service. In effect the capacity or space used by the last consumer has been transferred from all the consumers who preceeded him and are still using the facility.

Congestion is really that point where pre-marginal consumers (those having the strongest demand preferences for less congested travel) begin to experience a loss of benefits from the division and transfer of capacity and become willing to pay a premium to prevent the continued decline in the per-unit space. They, the pre-marginal users, would like to discourage the marginal transfer of capacity to marginal users. Consumers would bid against each other for turbulence space. For optimum efficiency the opportunity costs of the marginal or last transfer of capacity should be equal among the various alternatives; i.e., the value of the marginal transfer of capacity should be equal between all competing uses. Actually the only alternatives are to increase or decrease congestion by transferring capacity between marginal and pre-marginal users.

The turbulence is being used most efficiently, the level of congestion is optimum, when the value of the marginal transfer of capacity is equal between the marginal and pre-marginal users. At this point the value of the benefits accruing to the marginal user from the marginal transfer of capacity would be equal to the value of the benefits lost (increased congestion) by the pre-marginal users.

Since no consumer is permitted to pay less than marginal variable cost we can argue that the ideal price should exceed marginal variable cost by an amount equal to the aggregate sum pre-marginal users taken as a group would be willing to pay to displace the marginal consumer, or what they would pay to reclaim the last unit of capacity transferred to the marginal consumer.

At this uniform average price the marginal consumer is paying a premium above marginal variable cost equal to the value of his displacement (the opportunity costs of his consumption) to pre-marginal users, which is marginal congestion cost; and the marginal transfer of capacity is equal among competing alternatives. The ideal price then is at SRMC because marginal variable cost plus marginal congestion cost equal SRMC.

If it were possible to calculate the value that pre-marginal users place on quality changes resulting from marginal changes in congestion and congestion we would have calculated the added costs of marginal congestion, because both marginal congestion cost and the value of quality changes to pre-marginal users are equal to what pre-marginal users would pay to prevent marginal consumption and/or displace marginal consumers.

A framework capable of measuring the value of quality improvements to pre-marginal consumers would enable us to determine congestion costs and thereby establish optimum levels of service.

**The Value of Quality Change**

An improvement in quality makes it possible to sell the same quantity at a higher price, to sell greater quantities at the same price, or to increase both price and quantity somewhere between the two extremes. In situations of congestion an increase in toll rates reduces congestion and at the same time improves quality, but an improvement in quality increases demand for the service. A given price increase will displace fewer consumers when congestion is prevalent than when congestion is not a problem. A price increase will decrease congestion, but less consumption at congested periods also improves the quality of turnpike service by reducing congestion, which in itself tends to stimulate demand, thereby offsetting the decline in consumption resulting from the price increase. Consumers know that higher rates charged during congested periods will improve the quality of service. The "quality effect"—the tendency for demand to increase or decrease as the level of service is improved (or decreased) by a reduction (or an increase) in consumption and congestion—competes with, i.e. works against, the price effect. For example, if a given improvement in quality is considered desirable one must make sure the increase in price, which is designed to reduce consumption, includes a charge or premium that reflects the value of the improvement in service so that any increase in demand based on the improvement will not occur and partially offset or cancel the planned improvement.

Actually the quality effect is very much like the mob effect (the extent to which demand is increased owing to the fact that others are decreasing their consumption) in that both compete with the price effect to reduce the elasticity of demand. Movement along a demand curve will reflect the price effect and the shift in the schedule is the quality effect. The quality effect follows the price effect.

*Harvey Leibenstein, "Handwars, Free, and Vol-
assuming that turnpike capacity is sufficiently large to insure that no change in quality results from a change in the level of consumption. To measure the price effect one simply hypothesizes or calculates the price change required to alter aggregate consumption by a specific amount under the assumption that the quality of service remains unaffected, but the specific change in aggregate consumption related to the price effect is precisely that amount which would be necessary to bring about a given quality change from the level prevailing under existing congested conditions. An offsetting change in consumption due to a shift in the demand schedule following a change in the price and level of service reflects the quality effect.

Suppose Figure 2 depicts a market for turnpike service such that at price $P_2$, consumption is $0 X_2$, but let us assume consumption at $0 X_2$ creates traffic congestion.

As prices are increased above $P_2$ (marginal variable cost) consumption should decrease along the demand curve $dp$. Suppose, however, we find that as the price is increased from $P_2$ to $P_3$, consumption declines to $0 X_3$ on $d_1$ instead of $0 X_3$ on demand schedule $dp$, and at price $P_3$, consumption is at $0 X_3$ on $d_2$ rather than $0 X_3$ on $dp$. The relevant demand schedule when the quality effect is included is $dp$ and not $dp$. Consequently, if it were determined that traffic congestion would occur once consumption exceeded $0 X_3$ the relevant demand schedule between $P_4$ and $P_5$ is the appropriate portion of demand along $dp$. $P_2 - P_3$ is the per-unit premium above $P_2$ that $0 X_3$ pre-marginal consumers would pay to insure that market price does not decrease from $P_2$ to $P_3$ and consumption does not increase from $0 X_3$ to $0 X_3$. At $P_3$ congestion would be $0 X_3$ with some congestion. ($P_2 - P_3$) $0 X_3$ is the market value of the decline in quality stemming from the increase in congestion that $0 X_3$ - $0 X_3$ incremental consumption would impose on $0 X_3$ consumers if average price were set at $P_3$ instead of $P_2$. ($P_2 - P_3$) $0 X_3$ is also the premium that $0 X_3$ pre-marginal consumers as a group would pay to prevent the incremental consumption of $0 X_3$ - $0 X_3$ from occurring, and is therefore the marginal congestion cost associated with $0 X_3$ - $0 X_3$ marginal users.

Homogeneous Traffic

Highway engineers demand the expression for turnpike service in terms of diversion functions.6 Highway users traveling between a given set of origin and destination points will probably face alternative routes. All traffic moving between two areas, origin and destination, on all routes including the turnpike is considered to be the corridor volume. The demand for turnpike service is expressed as a percentage of the corridor traffic diverted to the turnpike. Empirical evidence indicates that the percentage diversion is directly related to the corridor volume and indirectly related to the toll charge.6 Traffic congestion on the turnpike would be more likely to occur when corridor volumes reach or exceed certain levels, unless, of course, the toll rates are raised to prevent it.

6 Hanner E. Davis, Ralph A. Noyer, Norman Kennedy, Howard S. Lathan, Toll Road Development and Their Significance in the Promotion of Expressways (Institute of Transportation and Engineering, Report Number 111, Berkeley: University of California, 1953), pp. 38-44.


To develop the technique through which optimum levels of congestion might be determined, let us begin with the following illustration.
fact to leave the turnpike. The market value of a 5 percent reduction in congestion to those consumers ($fe$) who would purchase the improved service ($P_2 - P_1$) is the aggregate premium that $fe$ pre-marginal users taken as a group would pay to deny $fe$ access to the turnpike and, in addition, the marginal congestion cost of increasing congestion from zero to 5 percent. To begin the case, $fe$ users would not be the merely marginal variable cost at a price of $P_f$. Price should be set at SRMC, which would be equal to $P_f (P_f - P_1) (1/2)(1/f)$. If the stimulation of price shall be no higher than necessary to reduce congestion to the zero level or $P_f$.

Raising price to $P_f$ along $d_f$ would reduce congestion back to $fe$ if the reduced consumption did not at the same time reduce congestion. But the fact that quality has been improved means two things: (1) pre-consumers on $d_f$ would pay a positive premium above $P_f$, $P_2 - P_1$, to consume uninhibited service instead of service at 5 percent congestion and/or to prevent $fe$ marginal consumption from occurring. (3) If the price were not increased from $P_f$ to $P_1$, some of the $fe$ consumers would not purchase service because of the belief that other consumers in the $fe$ group would not purchase service at $P_f$. To the extent that some consumers in the $fe$ group refrain from consumption at $P_f$ congestion will be less than 5 percent; however, the decrease in congestion below the 5 percent level would motivate others in the $fe$ group to purchase service. The result, if price were set at $P_f$, would be to stabilize the congestion level somewhere between 0 and 5 percent. Demand schedules $d_f$ and $d_f$ in Figure 4 are the original and increased schedules respectively, and demand curve $d_f$ shifts to $d_f$ to reflect the increased demand based on the quality effect that would result if quality were improved to zero congestion.

Consumption will not increase above $P_f$, however, if price is increased to $P_f$. $fe$ are marginal consumers in the sense that of the entire group of $fe$ consumers in the market they will begin to leave the market as price increases above $P_f$ and all will have left the market if the price is increased to $P_f$. ($P_f - P_1$) is the aggregate premium that $fe$ pre-marginal users taken as a group would pay to deny $fe$ access to the turnpike and, in addition, the marginal congestion cost of increasing congestion from zero to 5 percent. To begin the case, $fe$ users would be the merely marginal variable cost at a price of $P_f$. Price should be set at SRMC, which would be equal to $P_f (P_f - P_1) (1/2)(1/f)$. If the stimulation of price shall be no higher than necessary to reduce congestion to the zero level or $P_f$.

The probability that the SRMC price will be at $P_f$ may be low because at all prices between $P_f$ and $P_1$, the marginal consumers are paying a premium above marginal variable cost to purchase service with some congestion. Unless ($P_2 - P_1$) ($P_f - P_1$) is equal or greater than $P_2 - P_1$, some congestion will prevail at the optimum level of capacity utilization. In effect the marginal consumers and the pre-marginal consumers are bidding for use of the fixed capacity in the price range between $P_f$ and $P_1$. Most likely the optimum level of congestion will be less than 5 percent but more than zero. The demand schedule upon which we find the optimum marginal-cost price and quantity combination is $d_1$.

**Mixed Traffic**

The method of determining congestion costs under conditions of mixed traffic, i.e., vehicles of differing configuration, is approximately the same as it is when vehicles are of a single type. The costs of congestion are the amounts that pre-marginal users are willing to pay to prevent incremental consumption. Total congestion costs of a heterogeneous group of marginal users are allocated among the various vehicle types according to accepted engineering guidelines that indicate the contribution to congestion of any type of vehicle in terms of a standard vehicle, the passenger car.

**Conclusion**

Toll road prices should be set to reflect SRMC which means that the more conventional flat rate or constant toll schedules would have to be abandoned in favor of variable toll pricing. Assuming that SRMC prices include congestion costs, turnpike rates would increase during peak periods and decrease in off-peak periods.

An argument often advanced against SRMC pricing is that an insufficient amount of revenue generated in the long run to cover all costs. The validity of this argument rests on the fact that the SRMC price may be less than average total and even average variable costs. If turnpike rates were set to include congestion costs in the appropriate short-run periods over the life of the facility, it is conceivable that SRMC pricing during periods of congestion would exceed average total costs, thereby making a contribution to the fixed costs which might not be fully covered from the revenue generated at SRMC prices charged during off-peak periods. Certainly, if by pricing at SRMC in each market the total revenue generated over the life of the facility (the long run) is at least sufficient to cover all costs, there is no need to adopt any other pricing program.

Finally, it is significant to point out that the optimum level of service or capacity utilization will not necessarily be at the zero congestion level when rates are set at SRMC. The pre-marginal consumers at zero congestion may not be willing to pay a premium sufficiently large to prevent any additional consumption that would cause congestion. In short, the marginal consumers as a group may be willing to pay more than marginal variable cost to consume service with some congestion than pre-marginal users as a group will pay to keep consumption at the level required to prevent any congestion. It seems possible that during periods of heavy demand or high corridor volumes, especially as the system becomes more congested, the operator of the turnpike or toll road may be better than the quality of the best alternative route, while the decline in the level of turnpike service resulting from incremental consumption might be relatively slight.