HIGHWAY RESEARCH RECORD

Number 476 | Price-Subsidy Issues in Urban Transportation

6 reports
prepared for the
52nd Annual Meeting

Subject Areas

15 Transportation Economics
81 Urban Transportation Administration

HIGHWAY RESEARCH BOARD

DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D.C. 1973

NOTICE

These papers report research work of the authors that was done at institutions named by the authors. The papers were offered to the Highway Research Board of the National Research Council for publication and are published here in the interest of the dissemination of information from research, one of the major functions of the Highway Research Board.

Before publication, each paper was reviewed by members of the HRB committee named as its sponsor and accepted as objective, useful, and suitable for publication by the National Research Council. The members of the review committee were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the subject concerned.

Responsibility for the publication of these reports rests with the sponsoring committee. However, the opinions and conclusions expressed in the reports are those of the individual authors and not necessarily those of the sponsoring committee, the Highway Research Board, or the National Research Council.

Each report is reviewed and processed according to the procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved by the President of the Academy upon satisfactory completion of the review process.

ISBN 0-309-02261-4
Library of Congress Catalog Card No. 74-184
Price: $2.00

Highway Research Board publications are available by ordering directly from the Board. They are also obtainable on a regular basis through organizational or individual supporting membership in the Board; members or library subscribers are eligible for substantial discounts. For further information write to the Highway Research Board, National Academy of Sciences, 2101 Constitution Avenue N.W., Washington, D.C. 20418.

CONTENTS

FOREWORD

Papers in this RECORD examine the use of pricing and subsidy in urban transportation. Because of the inability of public transportation to cover capital and operating costs from its operating revenues, there is growing interest in looking to government to subsidize public transit operation and at the same time to use a pricing or taxing system to reduce urban traffic congestion at peak hours.

Cudahy examines the funding and management of the Massachusetts Bay Transportation Authority. One of the principal problems in financing the operations and improvements to the system is the way that local property taxes have been used to provide the necessary operating subsidy. The constituent communities are assessed according to a formula in which costs are based on general district benefits and specific community benefits. Inadequacies of proposed remedies for fiscal problems are discussed.

Sherman looks at mode choice and asserts that the present allocation of automobiles, buses, and rail transit systems does not react to the ideal economic marketplace because costs and pricing of services do not follow the rationale of the ideal market and present transit choice opportunities do not reflect marginal social costs well enough to allow a coordinated, balanced transportation system. In choosing travel by personal automobile, the highway user pays only the average rather than the marginal cost of his trip.

Roth examines the types of regulations applied to bus services in cities. He concludes that, although safety, noise, air pollution, and timetables may be under public control, route operation, fares, and the introduction of new bus services should not be. Where subsidies are necessary, such grants should be related to passenger mileage. Also, bus subsidies do not justify the public operation or regulation of services.

Tye also addresses the transit subsidy issue and recommends that, if the arguments for transit subsidy are valid, such funds should be allotted as a generalized subsidy and not restricted to capital expenses. He disputes the 4 primary arguments used for restricting federal aid to capital grants: Local governments are capital poor, capital grants prevent aid from being dissipated through labor wage gains, capital grants avoid the open-ended commitments that are entailed in operating subsidies, and capital grants are a highly visible means for demonstrating federal concern.

Vickrey examines the usefulness of pricing and metering techniques in obtaining more efficient use of existing facilities. He states that roadway pricing should not be used for redressing the imbalance between the private automobile and public transportation. Rather, road pricing could lead to the improvement in the efficiency in the use of existing transport facilities. He contends that congestion pricing for roadways is not mainly an issue in the battle between automobile and transit interests.

Kurnow, Brief, and Silberman present an "enterprise" approach to the analysis of problems relating to the financing of urban transportation services in the New York metropolitan region. This approach to accounting for the regional transportation systems includes revenues from direct user charges and indirect user charges reflected by general fund appropriations for transportation services in excess of local government user charges and special assessments. Expenditures are measured in terms of operating expenses, debt service, and capital outlay for highways, toll facilities, transit, commuter railroads, and private bus companies. This approach may be used to derive a transportation surplus for the region.

Hoch introduces the papers with an analysis of the issues discussed and the proposed strategies for dealing with them.

INTRODUCTION

Irving Hoch, Resources for the Future, Inc.

•PAPERS in this RECORD are of high quality and contribute a good deal of useful information. With that overall judgment rendered, I turn to specifics.

Tye reviews and demolishes arguments for capital grants and establishes some good empirical evidence on the rational economic behavior of transit companies. The argument for capital grants to transit might be associated with the general rule that an enterprise ought to cover operating costs or else it would be rational to go out of business. Of course, that argument makes sense only after capital has been sunk. Before capital funds have been expended, they constitute a variable input, and I see this as the essence of Tye's message: What is fixed ex post is variable ex ante. If we view both capital and operating costs as variable inputs and if we assume that an expansion of transit is a given policy, then the theory of the firm implies that cost minimization will occur on the firm expansion path. In the special case of constant returns to scale, or of any Cobb-Douglas function, the expansion path is a straight line from the origin. If transit firms are currently at an optimum, as indicated by Tye's evidence, then in these special cases both capital and operating expenditures should be expanded in the same proportion. Of course, the real-world expansion path may not be a straight line, but I offer the observation as a first approximation to best policy.

Roth has a good review of regulation and subsidies for urban buses and gives a worldwide review of cases. He makes an interesting point on the scheduling of buses. Rivals competing for a share of the market will tend to leave at the same time. This seems analogous to spatial competition where 2 rivals tend to locate next door to one another in an attempt to capture a maximum share of the market: Witness gasoline stations or supermarkets on adjoining corners. The situation might call for regulation to minimize costs to consumers, and this is the position Roth takes on scheduling. Roth's evidence on public versus private ownership in transit furnishes ammunition to advocates on each side of the question, but the preponderance of evidence favors the private ownership side. In Calcutta and Caracas, the private performs better than does the public; however, in Manila, the private engages in "cutthroat" competition (which could be in the public interest) and also imposes a great many negative externalities, e.g., via the deplorable driving habits of the drivers.

Kurnow, Brief, and Silberman pull together data from many sources to develop an integrated set of transportation-dollar magnitudes for the New York region. They demonstrate that the New York region pays more for transportation than it receives; it is then argued that this "surplus" could and should be applied to the financing of transit. They state the following basic propositions that support their position: "Adequate funds for financing all public transportation services in the region would be ensured if (a) all funds earned in transportation at each level of government were pooled; (b) transportation services had priority in the use of funds earned in transportation; (c) the allocation of funds earned in transportation to a region were in rough proportion to the money earned in it; and (d) all transportation planning and administration were coordinated in one agency for the region or in a group of cooperating agencies."

Point c argues that spending in the region should equal earnings in the region. But it is plausible that regional residents use more road mileage (or, more precisely, more value of road mileage) outside the region than nonresidents use within the region. There may be more New Yorkers driving in Montana than Montanans driving in New York; the excess may occur even in value terms, i.e., after we account for higher road costs per mile in the New York region. Hence, there may be something of an ethical case for a surplus.

Point b, that transport services should have priority on transport funds, is worth contrasting with point a, that all transport funds should be pooled. But if we accept pooling, why can we not pool transport with education, welfare, and so on? If we believe in applying funds to the activities that yield those funds, why not spend the funds on the modes that are their source? The case for transit subsidy is not made explicit by Kurnow, Brief, and Silberman, but Roth and Sherman do present a number of arguments for such subsidy.

One argument is that of redistribution to the poor. Roth points out that there may be more effective means of redistribution than transit subsidy and suggests travel coupons, including their use for taxi service. If society insists on the use of transit by the poor (taxis being too "rich"), it might be cheaper to subsidize the relocation of poor people to locales near profitable transit lines rather than to subsidize losing lines. Finally, there is the usual argument that we would do better to give the poor money to spend as they please rather than coupons limited to a specific commodity.

A second major argument for transit subsidies involves negative externalities imposed by automobile drivers on others (air pollution, noise, and accidents) and on themselves (the previous items plus congestion time losses). If we admit the case of externalities on others, it is not obvious that transit subsidy is the best, or even an effective, response. A case can be made for emission regulation (and EPA regulation will have major impact) and for effluent charges. But the latter policy might yield higher transit fares, for many forms of transit impose a considerable amount of negative externalities.

Turning to congestion externalities imposed on themselves by automobile drivers, Sherman gives a clear and detailed statement of the case for congestion tolls in terms of a marginal social cost argument. Transit subsidy is seen as a second best approximation to congestion tolls and has greater political feasibility. He argues that raising automobile travel prices or lowering transit prices or both will move us toward setting marginal social cost equal to marginal social benefit.

I have some qualms. We may well ask why it is that people are so resistant to congestion tolls. The analysts insist that people will be better off, but their prospective beneficiaries resist the argument (see St. Clair's statement, for example, in Sherman's footnote 3). It is somewhat worrisome that the analysts are going to make people better off in spite of themselves. (Of course, people may not act in their best interests or assess risks correctly; very few automobile occupants fasten their seat belts, for example.)

It is my hunch that the recommended congestion toll for automobile travel may be too high and may indeed correspond to a monopolistic price. (This could be the source of some of the resistance.) On the product market, competitive equilibrium involves the intersection of marginal cost or supply and average revenue, not marginal revenue. On the factor market, competitive equilibrium involves the intersection of value of marginal product or demand and average factor costs, not marginal factor costs. Hence, the intersection of 2 marginal curves seems suspect.

The proposed equation of marginal social costs and marginal social benefits seems to involve costing on the consumer side and neglect of cost of production or cost on the producer side.

Consider the application of the marginal social cost idea to housing. If a new family comes into an area, they will bid up the price of housing. The new price obtained is the marginal cost of housing to producers, but it is an average cost to consumers. Because all previous consumers now pay a higher price, there is a marginal social cost that is above market price. It follows by the traditional social cost argument that we ought to set a higher price than the market price to the new entrant: In effect, this will help keep the rascals out. I think the argument applies in the setting of transportation as well. In both cases, the higher price appears to involve a monopolistic approach to policy.

This is not to deny that congestion tolls can improve matters, for they will allow people to trade money for time. Put another way, tolls will move us out of a situation where the charge for road use is 0 dollars and all time cost. Vickrey illustrates some of the gains to be had by such user tolls. However, though I believe that people in general will be better off through the institution of tolls, I doubt Vickrey's assertion that everyone will be better off. Initially, with both a time charge and a money charge, everyone will be faced with a higher total price. After redistribution of trips, some people will change their time pattern of travel. I expect this will make some people worse off. If one has to leave work earlier, this should be a cause of disutility. I would go further and question Vickrey's assumption of a constant value of leisure time; value per minute is quite likely to be a function of hour of the day. Finally, Vickrey does not devote much concern to the distribution of the funds collected through tolls. We move back to the ethical questions of pooling, noted in response to Kurnow, Brief, and Silberman. Here, I think Sherman is on stronger grounds when he advocates a lump-sum return of tolls to the road user. Further, why not consider using the funds to build more roads, improve existing roads, and compensate outsiders harmed by externalities imposed by road users?

FINANCING TRANSIT: THE BOSTON EXPERIENCE

Brian J. Cudahy, Massachusetts Bay Transportation Authority

Operational and economic administrational experiences of a transit system are described. The history of the present Massachusetts Bay Transportation Authority is presented, and its operating expenses, which are the source of the major problems, are discussed in detail. A brief discussion of statutory issues and proposed legislation conclude the paper.

•THE IMPORTANCE of transit to urban America requires no elaboration. So let me catalog and describe some of the experiences we have had in Boston in attempting to both run our trains and pay our bills. My remarks and comments are based not on rigorous and reflective analysis of carefully gathered data but on real and genuine experiences with a bona fide transit system in a major American city.

MASSACHUSETTS BAY TRANSPORTATION AUTHORITY

The Massachusetts Bay Transportation Authority (MBTA) came into existence in 1964 in a metropolitan area that contains the City of Boston and 78 other cities and towns. The present authority succeeded an earlier public authority that was created in 1947 and was roughly similar to the MBTA, except that its service area included the City of Boston and only 13 surrounding communities. This predecessor—the famous MTA of song and story—itself succeeded an earlier private corporation, The Boston Elevated Railway. But even the Boston El, as it was called, was not a simon-pure example of liassez-faire capitalism.

Since 1918, by one or another statutory technique, public money was spent to meet the company's annual operating deficits. Indeed the very use of the term "deficit"—which we insist is an inappropriate and indeed misleading and inaccurate description of the ledger books of a public authority—was branded into Boston consciousness during the later days of the Boston Elevated Railway.

Mention of these historical facts raises another matter that is crucial to an understanding of today's fiscal plight of the MBTA: Both the physical plant and the practices that private carriers within the district have developed in the area of labor are quite old. The principal bargaining agent has been in existence since 1913, and the subway system opened its initial line in 1897, the first subway in North America. Other lines and routes were quickly added, so that the basic plant we operate today is essentially a product of the pre-World War I era.

Another result of Boston's early entry into the subway business was that new lines and routes were constructed to what were then the latest state-of-the-art advances. The consequence of what was honestly thought to be progress in 1901, 1904, 1911, and 1924 is that today we operate 4 separate and different rapid transit lines, all built to totally incompatible standards. I will not dwell on the implications this raises for operational efficiency. I simply mention it here to give some picture of the property we will be discussing.

Although the basic rail system is old, one line was recently extended and is as modern and up to date as any in the nation, although it must interface with the older network and also be operated by a labor force whose work rules predate the automated technology that the line features. Our rapid transit network uses some 700 vehicles and carries slightly more than 525,000 passengers daily. Despite new transit extensions, we, like other transit systems, are losing passengers year by year.

One additional facet to our operations is the subsidizing of private carriers within the district. Today, we operate about 1,200 surface buses on 198 different lines. Bus subsidies are rather insignificant, but we have 2 commuter railroads: the Penn Central and the Boston and Maine. Both are in bankruptcy, making dealings with them is very difficult, and the total number of daily passengers this network carries is very small by comparison with rail commuter operations elsewhere in the nation. Penn Central carries 6,000 one-way riders daily on 5 separate routes, and the B&M carries 11,000 on 8 routes.

Control of the MBTA is vested in a 5-member board of directors, who are appointed by the governor and who meet weekly. In addition, there is a larger body composed of representatives of all 79 cities and towns that constitute the district. This board is mandated to review and approve our budgets: its vote is weighted by population, so that in essence Boston and a few of the larger communities can, if they so choose, effectively control the entire body.

FISCAL ISSUES

And now we get to fiscal problems specifically. Let me say two things initially. First, the MBTA is generally perceived in Boston and environs as a "problem"; its equipment is old and not so reliable as it soon will be. But the chief focus of discontent centers around the question of cost. I cannot underestimate the critical problem that cost and its control represent for us today. Second, the present status of things may well be on the verge of change. Total reorganization of state government is in the wings. And even shy of this, as I will shortly outline, there is considerable pressure to effect major alteration in our basic system of funding.

CAPITAL EXPENSES

We have a physical plant whose unreliability and age make for extraordinary maintenance costs—not to mention substandard service. Although better maintenance practices have already significantly improved performance, the final solution here must await completion of new repair facilities, currently abuilding, and delivery of new rolling stock. Although delays are frustrating and inevitable, we can say that the solution is on the way. But in many ways, Boston is nothing if unexceptional in this area. Our own legislature has provided us with necessary bonding during the years to qualify for an absolute maximum of UMTA support, and the general tone of our capital program these days calls for modest expansion of the system, coupled with major renovation of the existing plant.

OPERATING EXPENSES

The area of operational expenses is where the problems lie and where our experiences may prove to be illuminating.

Source of Operating Subsidy

One feature of transit financing in Boston is that there was an early recognition of the need for operational subsidy. As I mentioned earlier, in 1918 provision was made to meet fare-box deficits out of the public treasury. When the MTA was created in 1947, the modality adopted presumed that the fare box was not to be the sole support of the system. The tragic flaw in the whole process, in retrospect, may well have been the selection of the tax source to meet expenses. What was selected was the property tax. And although the formula was changed at the time of the MBTA take-over in 1964, the simple fact today is that the employment of the local property tax as an operational subsidy for transit is creating what may well be insurmountable problems. The property tax is fast being recognized as a regressive levy, and the fact that local municipalities within our district have little control over the cost they are annually assessed for MBTA purposes helps to focus the problem.

There has never been any tendency in Boston to prune operational costs by sleight-of-hand bookkeeping. We receive no relief from the reduced fare we charge the elderly or students; the cost of maintaining our police force appears on the books as a straight operational expense; and, although the state pays most of the costs of our capital bonds, even some of this goes on the ledger as a cost of running the system. Parallel with this general tendency is a feeling now in the community that transit fares should not be further raised and, if anything, should be reduced on a selective basis to act as an incentive to increased riding, principally in off-peak periods. We intend to experiment along these lines soon. Consequently, by default you might say, we can boast of having a rate of fare that is quite low by comparison with other major cities—20 cents on surface lines and 25 cents on rapid transit (with, however, no transfers at all).

Allocation of Operating Losses

The formula by which our costs in excess of income are assessed on the constituent communities is a marvel of complexity. It does have this simple and univocal feature about it, though: Everyone is completely dissatisfied with it and feels—not without justification—that it specifically discriminates against each one's particular interests.

The formula distinguishes between rapid transit service and bus operations and attempts to assess costs on a twofold basis: general benefit to the district and specific benefit to the particular community. And although I earnestly hope to avoid delving into specifics of the formula, a good example of a disincentive subsidy is the feature whereby a given town is partially assessed on the basis of the number of passengers who board transit stations within its town limits. The problems we now face in selecting sites for new stations need not be discussed; they can simply be imagined.

Bus costs are assessed roughly this way: Fifty percent is charged to the entire district on the basis of proportional population, and the remaining half is assessed on the community where the loss is incurred. However, this factor is developed solely on the basis of mileage and an average cost per mile. Consequently, because a given community is only paying specifically for half the loss of a given service, there is a pronounced tendency to regard poorly patronized services as something "the other guy" is paying for, and not to act cooperatively with MBTA management in curtailing or eliminating services that fulfill no genuine transportation need (I hasten to point out that we do not regard service reductions as our most important mandate). We continually receive the message from throughout the district that we run too many buses on poorly patronized runs. But the suggestion in general is always coupled with the warning in particular that the offending vehicles are not those from the town that the critic represents.

The converse of this situation is equally perplexing. Because particular towns are only assessed on the basis of an average cost per mile, there is no incentive by way of assessment relief for a community to assist in the promotion and marketing of a poorly performing service. Any new riders that might be attracted are not credited against the assessment for a given town; the new dollars are merely poured into the larger "kitty." Although they do ultimately affect the average cost-per-mile figure, there is virtually no perceived relief from the so-called "defici." And, to repeat, there is no built-in incentive for the town to put any of its own money into advertising, for instance, or take stern measures against illegally parked automobiles that impede performance of bus lines or to institute preferential traffic flow for our vehicles.

STATUTORY ISSUES

I mentioned earlier that the present statutory problems may be on the verge of solution. Last spring, the authority submitted a supplementary budget to the Advisory Board for approval—a routine and annual ritual that frequently produces angry rhetoric but eventually passage of the money requested. Last spring, however, the Advisory Board flatly refused to approve the money—not, as it was careful to point out, to show any specific dissatisfaction with our management of the system but to underscore its contention to the legislature that the state should be paying a larger share of MBTA operation costs and thereby effecting some measure of relief from the problems that the property tax has been causing. The MBTA made an especially handy target. For one thing, there are the perceived problems that the public associates with the system, and any anti-MBTA campaign has a built-in populist tone to it. Second, the MBTA assessment is one of the few—and certainly the most visible—cost items that are not of the local government's own making, but which it must ante up regardless.

We were quickly thrown into a crisis. We would run out of money at a certain date. Notices were posted saying all service would be discontinued. There was a last minute flurry in the legislature, the advisory board was temporarily stripped of its power to disapprove our budgets, and the trains continued to roll. This was followed by the inevitable study committee that met all summer long, reviewed our problems, and has now made its recommendations. The governor has pledged his willingness to have the state underwrite a heavier proportion of our annual costs, and a raft of bills have been filed for consideration by this year's legislature.

CONCLUSION

The various legislative remedies now pending to rectify the many problems that exist in our fiscal structure are, sadly, inadequate and incomplete. I say this not in criticism of the work that went into their drafting, but in the light of the peculiar condition that dictates the nature and style of corrective measures that we can expect in the present situation. The overriding issue is the likelihood of passage. Because the entire MBTA district comprises just less than 50 percent of the state's legislators, something better than unanimity is required from all the representatives and senators in the 79 cities and towns. To obtain this type of support requires that any correction in our funding formula provide rigidly equivalent benefits for the entire district. Yet the unfortunate fact is that the reforms we so badly need, for example, the elimination of boarding counts and a totally new assessment for bus operations, remain hopelessly utopian because they will clearly treat one community differently from another and, thus, not receive the kind of support needed for passage.

I do not wish to seem overly negative. There are distinct benefits and advantages to our system of transit financing. Fares are kept low, and the entire metropolitan area participates in cost allocation, although on a basis other than actual use of service. I have painted a bleak picture for 2 reasons: First, the picture is bleak and, second, once established, mistakes are very difficult to eradicate.

INCENTIVES FOR THE COORDINATION OF DECENTRALIZED TRANSIT CHOICES

Roger Sherman, University of Virginia

In a well-functioning competitive market system, decentralized choices by individuals can serve their interests well; but in a typical contemporary setting, such choices, at least about transit, do not. Flaws in the allocative consequence of decentralized transit choices are illustrated here by reference to an ideal set of price signals and description of some of the ways in which real-world prices differ from the ideal ones. One consequence of this distortion between real and ideal price signals is that bus companies may not be able to survive a profit test in cities even though bus service actually may be desired and could survive under ideal prices. Tax and subsidy incentives that can approximate the ideal price signals are described, and incentives for efficient operation of bus services also are briefly noted.

•IN A well-functioning market economy, each person makes choices about what to consume based on prices that reflect fully the marginal cost to society of providing one more unit of each good or service. Inefficient producers will lose money and automatically will be forced to withdraw from their markets, while profit opportunities will attract more resources to those activities that we consumers wish to have expanded. These properties of a competitive market system are so well known and we are so accustomed to their benefits that we may think decentralized choices always are efficacious. If for any reason, however, prices do not reflect the marginal social costs of goods and services, the signals that individuals act on will be wrong, and resources will not be attracted automatically where they can serve consumers best. We must then appropriately consider alternative incentives that might correct the wrong prices and thereby enable us to use our resources efficiently.

Every day, we make transit choices on a decentralized basis, but are guided in our choices by prices that are faulty. We squander certain of our resources as a result because the faulty prices that we rely on lead us to use too much of some things and not enough of others. Our problem is a very difficult one, however. We fail to solve it efficiently not because we are lazy or dumb but because it is a very difficult problem that still is only partly understood. Our purpose here is, first, to illustrate the kind of problem we face in trying to choose together through decentralized private decisions an efficient solution to our transit problem and, second, to indicate ways that the problem might be solved. Means of controlling the transit agency in case subsidies are given are also considered.

PROBLEMS IN COORDINATING DECENTRALIZED TRANSIT CHOICES

Road-use patterns are commonly accepted and are so similar in U.S. cities that we can easily fail to appreciate their faults. And alternative patterns are hard to imagine. To bring out problems in our present system, let us begin with a scheme for optimal control of highway transit and then move gradually to the system we actually find today in U.S. cities. We consider automobile and bus modes, both using the same road space, and we concern ourselves primarily with the peak hours, when most travel occurs.

One Road, Automobile, and Bus Company

Let us consider an illustrative city. Highways offer the only means of travel, and they accommodate both automobiles and buses. All automobiles, buses, and highways are owned, however, by the city's Grand Road, Automobile, and Bus Company (GRAB). An individual can either hire a car or ride a bus and in each case must pay a specified fee per mile traveled. We shall first assume that GRAB is a truly benevolent welfare-maximizing organization. For such a situation, we ask: What prices should GRAB charge per automobile-mile and per bus-mile? What quantities of road, automobile, and bus facilities should it provide?

We obviously must make some heroic assumptions in posing this problem, pretending for instance that automobile operating and maintenance costs will not be very much affected whether automobiles are privately owned or leased from GRAB and that all the other administrative problems of operating such an unusual organization in a welfare-maximizing way can be handled. But let us first skip such problems to focus on important characteristics of urban road transit, namely, the excessive use of roads by private parties, the consequently high level of congestion, and the fact that automobile passengers contribute more to such congestion than bus passengers do.

To maximize social welfare, GRAB should choose prices that equal the marginal costs of each of the 2 transit modes it can supply. Following this rule may lead to surpluses of profit if average cost is increasing or to losses if average cost is decreasing, and the long-run consequences of these profit or loss results are important. The distribution of profits will affect income distribution and possibly efficiency, and raising funds to cover losses will affect efficiency under any feasible taxation scheme.

Let us set these problems aside for the moment, though, to focus solely on efficiency and the ideal price-equals-marginal-cost rule that will ensure it. The logic of this well-known welfare-maximizing rule rests on the fact that, if one person pays less than the marginal cost to society for a service he consumes, the remaining cost must be borne by others, a result that is not only unfair but that also leads the person to consume a nonoptimal quantity of service. The total cost of serving one person will be higher when the person does not face the marginal cost of his actions, for he will tend to consume more of the service when he pays less for it. His excessive consumption, beyond what he would choose if he paid marginal social cost, is inefficient. Any resources that are used by a person who values them at less than their value in alternative uses ought really to be shifted where they are valued more. In principle, the person now using them could be paid an amount equal to the value he places on them, so that he would remain as well off, while others who value them more could pay that amount and enjoy a net benefit from use of the resources. Using resources where they are valued more will add to welfare.

It is possible in a simple illustrative situation to sort out congestion effects in a reasonably straightforward way and provide GRAB with ideal welfare-maximizing price choices. Let us consider only daytime traffic, which we shall assume to be of uniform density at all times and places so that we can concern ourselves with just one problem and one solution. The ideal price per automobile passenger-mile will reflect the marginal social cost of an additional automobile passenger-mile. That cost not only includes the effects on other automobile travelers but also must take into account the effect that an automobile passenger-mile will have on the marginal and average cost of a bus passenger-mile. This second element captures the increase in congestion due to automobiles that actually affects buses. Similarly, the ideal price per bus passenger-mile will reflect the marginal social cost of another bus passenger-mile, plus an additional element to take into account the increase that a bus passenger-mile will cause in the congestion that affects automobiles.

These ideal prices can be summarized algebraically. If automobile passenger-miles are M_A, the average cost of an automobile passenger-mile is AC_A, and the total cost of automobile passenger travel is TC_A, then the marginal cost to automobile passengers of an automobile passenger-mile, MC_A, is

$$MC_A = \frac{\partial TC_A}{\partial M_A} = \frac{\partial(AC_A \cdot M_A)}{\partial M_A} = AC_A + M_A\frac{\partial AC_A}{\partial M_A}$$

The same definition of marginal cost to bus travelers alone can hold for bus travel if we replace A subscripts with B subscripts to represent buses:

$$MC_B = AC_B + M_B \frac{\partial AC_B}{\partial M_B}$$

Rather than focus on the cost of automobile or bus travel alone, though, we must consider them together because an additional automobile or bus affects costs in both modes. The effect of a marginal automobile passenger-mile on total cost, $TC_A + TC_B$, is

$$\frac{\partial(TC_A + TC_B)}{\partial M_A} = AC_A + M_A \frac{\partial AC_A}{\partial M_A} + M_B \frac{\partial AC_B}{\partial M_A} = MC_A + M_B \frac{\partial AC_B}{\partial M_A}$$

Because the ideal price per automobile passenger-mile, P_A, will equal marginal social cost, we have

$$P_A = AC_A + M_A \frac{\partial AC_A}{\partial M_A} + M_B \frac{\partial AC_B}{\partial M_A} = MC_A + M_B \frac{\partial AC_B}{\partial M_A} \quad (1)$$

The term, $M_B \partial AC_B/\partial M_A$, represents the addition to marginal cost felt by bus passengers but not experienced by automobile passengers and therefore omitted from the definition of MC_A. Similarly, the marginal social cost and ideal price, P_B, of a bus passenger-mile will be

$$P_B = AC_B + M_B \frac{\partial AC_B}{\partial M_B} + M_A \frac{\partial AC_A}{\partial M_B} = MC_B + M_A \frac{\partial AC_A}{\partial M_B} \quad (2)$$

These expressions reflect the fact that each transit mode ideally will recover its marginal cost and, in addition, the contribution it makes to the marginal cost of the other mode. Let us note that, by taking up more road space, an automobile passenger contributes more to congestion than a bus passenger. More precisely, it is almost certainly true that $\partial AC_A/\partial M_A > \partial AC_A/\partial M_B$ and $\partial AC_B/\partial M_A > \partial AC_B/\partial M_B$. [Support for these conditions is given in another paper (1).] There is an average of only 1.5 passengers per automobile in cities, even during congested hours; and, although a bus may require as much road space as 3 or more passenger cars, it carries many times more passengers.

There is one consequence of such welfare-maximizing pricing policies that might seem embarrassing to the managers of GRAB: Although it pursues wholeheartedly a socially efficient solution, the transit agency will be very profitable. If it merely charged a price equal to the marginal cost of each of the transit modes it controls, the agency would do well because marginal cost exceeds average cost in each case. But the ideal prices are higher still to take account of the contribution each mode makes to marginal cost in the other, and so GRAB will be even more profitable. GRAB achieves economic efficiency with its ideal prices, but in achieving efficiency no attention has been given to income distribution. And yet how GRAB's abundant profit is distributed is an important question.

The presence of congestion means that road space is scarce and that charging a price to ration its use can yield a profit. Let us tackle the question of how to distribute this profit equitably by posing this question: Among all members of society, what sort of bargain might be struck to determine who would use a scarce facility that no one person or group owned? Quite obviously, some would pay others not to use the facility, and those most keen to pay to use it thereby would be accommodated. Now in practice such a bargain may be impossible to arrange. But if we charge a price for the use of roads and distribute any resulting profits to all citizens, road-users and non-road-users alike, we accomplish—crudely—the same sort of bargain. For by charging each traveler the marginal social cost of his travel decision, we discourage from traveling those who do not value travel very highly, and we ration use of a scarce facility to those who value it most. Because the profits of GRAB amount really to the proceeds from such a rationing price, they reasonably could be distributed to all citizens.

Early criticism of road-pricing arguments overlooked this possibility of redistributing the proceeds of a congestion toll or tax. It is true that road users usually will "lose" if taxes are imposed to control congestion and the road user never benefits in any way from the proceeds of the tax. [In his criticism of congestion tolls, for instance, St. Clair (2) remarked, "It seems adding insult to injury first to recognize that that time delay is a cost [to the road user], and then to say he must pay, in order to maximize his own benefits, a tax that will leap from the average to the marginal point." Actually, if the congestion toll proceeds never benefit the road user in any way, a more apt description of the toll is that it "adds injury to efficiency." The toll can achieve efficient road usage. But if it also takes income from road users and transfers it to others, the road user can still lose. In unusual cases it is true that the efficiency gain can make him better off despite the loss he suffers in making toll payments. Moore (3) gives descriptions of such cases. Vickrey (4) reviews evidence on how significant the efficiency losses can be.] But the tax should not work that way if it is to be equitable. The road user pays the tax per unit of travel so that usage of the road can be controlled marginally and thereby can be made efficient, but the proceeds of the tax should be returned to him in a lump-sum form, perhaps as a reduction of his other tax obligations, so the main effect of the tax is to accomplish the marginal adjustment without vastly redistributing income away from automobile travelers at the same time. If we can assume that automobile travelers will share in the proceeds of the congestion toll, or tax, it becomes an equitable as well as an efficient way to deal with traffic congestion.

The long-run decision about the quantity of road, automobile, and bus facilities can be a very difficult one. If the short-run marginal costs of automobile and bus travel can be identified for any existing level of road supply and trip demand, then optimal prices can be determined that will result in using existing facilities efficiently. Such short-run marginal cost prices also can indicate the value consumers place on existing facilities. By their size, those prices therefore can tell whether the cost of expanding road space and other facilities ought to be incurred. Here there are especially serious difficulties, however. First there is a discreteness in road expansion that requires large rather than small changes. Expansion calls for adding new lanes and possibly new roads rather than making small marginal adjustments within the range of existing usage, where evidence of marginal values can be more reliably estimated. Second, these changes can affect the demand for trips as well as their supply. Knocking out stores to make room for an additional road lane may remove (or relocate) the causes of some trips (to go to those stores) and thus can cause a reduction in trip demand at the same time it makes an increase in supply possible.

The main point is that reasonably efficient short-run prices (or tolls or taxes) are needed first, for they will give some indication of the value road users place on increased system capacity. Congestion itself can never justify road expansion; the congestion may be due to faulty pricing of existing roads rather than inadequate provision of roads. And if road capacity expansion is characterized by increasing cost, as is probably the case in real cities, efficient short-run tolls or taxes should allow net proceeds in excess of existing road costs. A break-even rule is not enough. Road operations ideally will be "profitable" instead, and the net proceeds should be returned to citizens through some sort of payment that will not affect their marginal travel choices. Then transit choices can be efficiently and also equitably coordinated.

Private Ownership of Automobiles

Now let us weaken GRAB's control of all metropolitan transit and permit private ownership of automobiles. Doing so can interfere seriously with the efficiency of urban transit. First, private automobile ownership allows, in effect, a 2-part price for automobile travel, while GRAB presumably will continue to collect a single fee for each bus trip. Automobile owners pay fixed outlays to become owners and then can make smaller outlays per mile traveled because they pay some costs (e.g., insurance, license fees, part of depreciation) independent of the amount they travel. The consequent 2-part price for automobile travel enables those who travel more to achieve a lower average cost because they can spread the fixed costs of automobile travel over more miles. Those who travel more are therefore more apt to commit themselves to automobile ownership to obtain the benefits of a 2-part price (5, 6). In what follows we shall ignore this slight bias toward automobile ownership brought about by lower average cost through a 2-part price, although it actually would tend to bias our private choices more in the direction of the automobile.

Apart from this ownership bias, in the absence of special efforts to coordinate the perceived automobile and bus travel costs, automobile travel will no longer be decided based on marginal social cost. An automobile owner can join a traffic stream and share in the average delays that are experienced, so his or her decision will be based on average social cost. There will be an excessive, inefficient use of roads by automobile passengers as a result because, when it is equal only to average cost, the marginal benefit for automobile passengers will lie below marginal cost. Indeed, the well-known traffic congestion problem that we see daily in real-world city streets is caused by our failure to charge a proper price to those automobile travelers who cause highway congestion. Each of us reckons with average costs and delays when we decide to join in road traffic, but we do not consider the delays that we cause others on the road; those delays to others nonetheless are part of marginal social cost. In terms of Eq. 1, we pay only AC_A per automobile passenger-mile and ignore $M_A \partial AC_A/\partial M_A + M_B \partial AC_B/M_A$, so our effective automobile travel price is below the ideal P_A. Because average cost rises with an increase in traffic, marginal cost must be higher than average cost; and the equilibrium we reach is not optimal because it occurs where the benefit from a marginal trip equals its average social cost and thus lies below marginal social cost. That is why the marginal social cost of trips on congested roads exceeds their social benefit.

Now what should GRAB do? If automobile travel decisions are based on average rather than marginal social cost, what is the optimal price for bus trips? If GRAB tries to maximize welfare by dealing as efficiently as possible with congestion, it now may set a bus price below the average cost of providing bus service. Such a choice can be appropriate because without it too many resources will go to automobile travel, and automobile passengers pay only average rather than marginal social cost and yet contribute more to congestion. Drawing passengers away from automobiles with a lower price for bus transit can offset this misallocation. If a marginal bus passenger would eliminate a marginal automobile traveler, for instance, the effect of their attendant contributions to congestion could be to reduce overall congestion because the automobile traveler's contribution is greater. In that case of induced substitution from automobile to bus travel, the net social cost of a marginal bus passenger would be low because by choosing to travel by bus he actually would reduce congestion. His addition to congestion as a bus passenger would be lower than the reduction in congestion caused by his withdrawal as an automobile traveler. Such an effect depends again, of course, on some substitutability of bus travel for automobile travel.

Thus, in the presence of nonoptimal pricing of road use for automobiles, the price that GRAB should set for its bus passenger-miles will depend on the tendency of travelers to respond to price changes by substituting one mode of travel for another. If they do not respond to prices at all, there is little we can do through bus pricing alone about excessive automobile use and its consequences for the level of congestion. But evidence that individuals respond to relative prices in their other decisions is so abundant that we cannot fail to expect it here. So let us describe the conditions that determine whether the optimal price per bus passenger-mile is above or below average cost.

In a simple model it can be shown that, when automobile users make their private travel decisions at average rather than marginal social cost, we can use our transit resources more efficiently by pricing bus travel below its average cost if the following condition is met:

$$\frac{\partial AC_B/\partial M_B}{\partial AC_B/\partial M_A} < \frac{\partial M_A/\partial P_B}{-\partial M_B/\partial P_B}$$

[This condition is a slightly modified version of the requirement derived in other papers (1, 7). The derivation actually assumed that $(\partial AC_A/\partial M_B)/(\partial AC_A/\partial M_A) = (\partial AC_B/\partial M_B)/$

($\partial AC_B/\partial M_A$). Also the $\partial M_A/\partial P_B$ and $\partial M_B/\partial P_B$ responses are "income-compensated" responses that reflect effects of substitution but not income change. Because income effects resulting from transit outlays probably are small, this proviso should not be crucial.] We have assumed that a marginal automobile traveler causes more congestion for both automobile and bus modes, and when that is true $\partial AC_B/\partial M_B < \partial AC_B/\partial M_A$, so the left side of this inequality will be less than 1. Because its own price usually has more effect than prices of other services on the quantity demanded of any one service, we can expect the denominator of the right side to be larger than its numerator, too. But, although each side is less than 1, we cannot be sure whether this inequality condition will be satisfied. It certainly can be satisfied though (7), which means it can be more efficient to subsidize a bus service in order that it can set its price below its average cost.

We might refer here to the much-heralded experiment in Rome, where free public transit failed to solve a problem of traffic congestion that was due primarily to automobiles. Observers now concede that, perhaps because of the design of existing transit service in Rome, reducing its price brought former pedestrians to use it rather than automobile travelers. In terms of the inequality above, this would mean that $-\partial M_B/\partial P_B$ was large but $\partial M_A/\partial P_B$ was not large, so the right side probably was too small to satisfy the condition that would warrant subsidizing the transit service. A different kind of service, however, that would attract automobile travelers instead of pedestrians still might have warranted a subsidy.

There actually is a way to approach much more closely the original optimal solution chosen by GRAB when it controlled prices of both automobile and bus passenger-miles. Suppose that, in addition to setting P_B, GRAB is given authority to impose a tax, t, on all the inputs that account for operating costs AC_A and AC_B (such as gas, oil, tires, vehicles, and repairs). We assume that any tax must affect both modes because it is not feasible to tax gasoline at different rates for buses and automobiles without incurring prohibitively high enforcement costs. Thus, the input tax necessarily would raise costs in both modes to $AC_A(1 + t)$ and $AC_B(1 + t)$. By adjusting the input tax rate and the level of P_B, the transit firm can achieve prices exactly equivalent to marginal social cost for each mode, just as if it chose P_A and P_B. The only remaining difference is that individuals (rather than GRAB) now decide the quantity of cars, and because of the 2-part price advantage of automobile ownership more automobiles may be chosen by individuals.

It is worth emphasizing, however, that with the input tax the optimal level of P_B is even more apt to be below the average cost of bus travel, which now includes the cost of the input tax. So a subsidy to GRAB will be needed. The ideal P_B will lie below $AC_B(1 + t)$ as long as the ratio of bus to automobile contribution to congestion is lower than the ratio of bus to automobile average cost (7). That is, the ideal level of P_B will be below this average cost, $AC_B(1 + t)$, when

$$\frac{\partial AC_B/\partial M_B}{\partial AC_B/\partial M_A} < \frac{AC_B}{AC_A}$$

This requirement appears very likely to be fulfilled in a city; so on its bus operations alone, GRAB would appear to lose money. It is possible to show, however, that proceeds from the input tax will always exceed these apparent losses on bus operations. Some of the input tax proceeds thus can go to provide the subsidy without causing new tax burdens elsewhere. Because bus passengers contribute proportionately less to congestion but the input tax cannot be adjusted for that difference, a rebate for them on the input tax is appropriate on both equity and efficiency grounds.

Private Ownership of Automobiles With GRAB as a Private Monopolist

Suppose now that we permit private ownership of automobiles, turn operation of the bus service over to GRAB, suspend its welfare-maximizing purpose to let it maximize profit, have local government decide on input taxes and the quantity of roads, and leave the road to the use of both automobiles and buses. The result then will approximate more the sort of situation we have now in many cities. As long as the automobile can serve as a good substitute for bus travel, the monopoly position of GRAB is not really very lucrative, for its attempts to raise price will be met by substitution of the automobile mode for bus travel. (As a recent case in point, the Louisville Transit Company recently gave notice that it will give up its bus franchise in 1974, although it has a right to keep it until 1981.) Thus, GRAB has a monopoly of bus service, but not a monopoly in the classical sense of producing a service for which no good substitutes are available. And as the automobile, its main substitute, becomes more popular, the service a bus can offer is apt to become more costly and less attractive because of greater congestion delays.

Any effect of the 2-part price opportunity available through automobile ownership now will be harmful to GRAB, for a lower average cost for automobile travel via the 2-part price will tend to increase M_A and reduce M_B. Moreover, because the upper limit that the private automobile ownership option places on the bus travel price may be below the average cost of bus service, it is very possible that GRAB will lose money. Imposing an input tax to reduce congestion probably will be slightly prejudicial against buses, too; automobile and bus costs will tend to go up proportionately because of the tax, but the contribution of automobile passenger-miles to congestion exceeds that of bus passenger-miles. As the input tax is set higher, the ideal price for bus travel will fall relative to the automobile travel price that is perceived by travelers because the input tax alone tends to affect bus passengers more, relative to their marginal contribution to congestion. A lower fare for bus travel can offset that input tax bias. In principle, when combined with input taxes, a subsidized bus service can permit final effective automobile and bus travel prices that equal marginal social costs, equivalent to the ideal P_A and P_B values considered earlier and given in Eqs. 1 and 2. Thus, the money-losing bus service we see in the real world may not be a sign of faulty pricing because the ideal price per bus-mile probably should be below its average cost. A subsidy to make up this loss properly should be paid, however, out of proceeds of the input tax.

To the extent that a profit-maximizing bus service could survive profitably in a city, probably those services that did not cover their average costs would be suspended. Yet such services conceivably could add to overall transit efficiency by easing road congestion and might be the most valuable to a community. The profit test simply is inappropriate in the urban transit setting and can yield perverse results. The final outcome also would handicap a city government in choosing what quantity of roads to provide because the prices implicit in road use no longer would reflect faithfully the value of roads to consumers. The wrong price for automobile travel—at average rather than marginal social cost—has induced too many of us to choose the automobile as a mode of travel. We have driven profit-seeking, private-enterprise bus services out of business at the same time we have brought ourselves a large inefficiency burden in the form of excessive traffic congestion.

INCENTIVES TO COORDINATE DECENTRALIZED TRANSIT CHOICES

Urban transit choices are made by individuals on a decentralized basis, just as choices in our market economy generally are made on a decentralized basis. Complicated interdependencies occur in the urban transit setting, however, and they affect both the demand and the supply of transit in complicated ways. These interdependencies alone would make price-setting for decentralized coordination decisions (about what transit modes to use and, thus, about how to allocate resources) very difficult because they can prevent some persons from facing the full cost consequences of their actions. But, in addition, prices cannot serve their familiar purpose in the transit setting because the relatively high cost of actually carrying out the necessary transactions prohibits the use of prices to ration road space. As a result much road use is not properly priced and road use is inefficient; we usually have more congestion than is optimal. Moreover, because of the interdependence of automobiles with other transit modes, this failure to coordinate one aspect of road usage—if uncorrected—can have serious perverse effects on other transit modes.

Pricing of Road Use

Various forms of user taxation—primarily fuel taxes, excise taxes, license fees, parking fees, and tolls (road, bridge, and tunnel)—combine along with other costs of vehicle operation and the time cost of traffic congestion delays to form a crude price for road usage. Where there is little or no congestion most of the time on rural roads and late at night or early in the morning on urban roads, taxes and tolls are not appropriate and they actually will reduce welfare. But they can reduce congestion and invite more efficient road use where there is congestion during peak usage periods in urban areas. Of course such taxes cannot be effective at one time of day without being effective at another. The failure of such taxes to distinguish peak from off-peak time periods may not be very serious, however (8). A greater problem seems to be due to the currently low level of user taxes where congestion is most serious. Fuel taxes, for example, may be at less than half the level needed to ensure efficient use of roads in large urban centers in the United States (9). As a result, excessively high congestion can prevent the most efficient use of the road. A crude way to improve coordination of transit choices is to raise substantially the tax on inputs necessary for road use.

As input taxes go up to ration the use of existing roads and thereby improve the efficiency of road use, a problem arises with respect to the relation between revenues from bus operations and the total cost of those operations. We have already noted that, if we apply a percentage tax on inputs, the average cost of bus operations will tend to go up at the same rate as average cost on automobile operations. The contributions these 2 modes make to traffic congestion are not necessarily in those proportions, however, for automobiles contribute more per passenger-mile to congestion than buses do. Relative to the effect of automobile usage on marginal social cost, the increase in marginal social cost due to greater bus usage will be less than the initial ratio of average bus cost per passenger-mile to average automobile cost per passenger-mile. That means that relative to the ideal price for automobile travel, the ideal price for bus travel will go up by less than the ratio of bus average cost to automobile average cost, and yet the tax will raise both costs by the same percentage. So if we rely on input taxes for achieving an effective automobile travel price, we shall tend to have a higher effective price for bus transit than would be ideal.

It is possible to correct this probable bias of an input tax against buses in several different ways. One way is to subsidize bus operations because, if there is any profit from bus operations, the price for bus transit probably is too high relative to an ideal effective price for automobile travel. For, if the bus travel price is near its optimum level, losses are likely, and a subsidy will be needed to cover the total cost of providing the bus service. An alternative arrangement would reduce the cost interdependency brought on by automobiles and buses contributing to each other's congestion, perhaps by providing exclusive lanes for bus travel. Then a marginal automobile passenger-mile would be less apt to add to the congestion delays felt by bus passengers and vice versa. But even without congestion interdependence—automobile passengers slowing bus passengers and vice versa—there may be reason to price bus transit below its average cost when input taxes are in effect to control congestion. For the automobile still can contribute more to its own congestion and, thus, for efficiency can warrant a higher congestion toll than buses. Yet because the tax on inputs cannot easily be adjusted based on how the inputs are finally used, a tax level that will control automobile usage ideally can make the effective tax on bus travel too high. A subsidy can lower the levy per bus-mile then, however, and bring the effective price for bus travel more in line with its marginal social cost.

New Jersey, Pennsylvania, and on a less complete scale several other states now provide significant operating subsidies to urban transit. There also is systematic relief for publicly owned transit operators from the federal tax on gasoline and from the federal excise taxes on buses and parts. And privately owned transit operators have been granted relief from fuel tax increases that were approved in 1956 under the federal highway program (10). All of these efforts to ease the impact of fuel taxes on urban transit operators probably are consistent with efforts to maximize welfare. Their effectiveness is limited, of course, where fuel taxes tend to be too low to control urban traffic congestion adequately.

Several cities are now turning to transit subsidies (11). Atlanta's reduction in fare from 40 to 15 cents already has increased bus patronage by more than 20 percent, and the adjustment to the new fare may not yet be complete. The service also is being improved by a rapid transit system that is being financed in part by a 1 percent city sales tax. Denver voters passed a $4 million bond issue in order to reduce all bus fares, including a lower off-peak bus fare. And Akron, Dayton, and Toledo have recently approved property tax increases in order to fund transit improvements.

By lowering the price of bus usage more in line with that of automobile usage, these subsidy efforts will be able to reduce the bias that otherwise would favor automobile travel. And because automobile travelers cause more congestion, that will be somewhat relieved. All road travelers typically still will pay less than the marginal social cost of their travel, however, so neither their payments nor the level of congestion they experience will serve as a reliable indication of the value they actually place on road capacity. Only an increase in input taxes, or some other means of collecting a fee for road use, can offer such an indication reliably.

Quantity of Roads

Proceeds from user fees now seem adequate on average to meet the costs of highway construction, and that fact may suggest to some that road-user fees already are high enough. But equality in these magnitudes is not necessarily in order. A user fee, which could be an input tax, serves first the purpose of rationing existing roads so they may be used efficiently. The level of fees or taxes that accomplishes that result also can indicate whether the existing stock of roads should be increased. For instance, in paying their fees, users may reveal a willingness to pay more than the cost involved in expending roads, and then roads should be expanded. This expansion question is a difficult one, though, as we have noted. In urban areas especially, expansion of the highway system is apt to encounter diminishing returns, and an optimal price for road usage then should provide a surplus of revenue over past road construction cost to go as a rent to the advantageously located but scarce land already devoted to road use. Because this land is publicly owned, all citizens reasonably can claim a share in such rents. And so some of the payments by users may go appropriately as nonmarginal transfers to other users and nonusers alike, rather than as investments in new highways. Fees and taxes that make the effective price paid for highway usage equal to the marginal social cost of highway usage will not necessarily also make the total proceeds from fees and taxes equal to construction costs because the long-run average and marginal road construction costs will not necessarily be equal. A surplus of user fees over highway construction expenditures can be perfectly appropriate, especially in cities where land is valuable.

Suppose, as is true in U.S. metropolitan areas, that user fees are not great enough to make the usage of existing roads efficient during daytime hours. As a consequence there will be excessive congestion on roads. And the presence of serious congestion could lead to political pressure to build roads, even though the main reason for congestion is a faulty low effective price for road use (12). In connection with the Interstate highway program, we have an arrangement called the Highway Trust Fund, which urges that fees from users be used for construction of roads. Indeed, it offers to local communities the return of user tax proceeds only if the communities use them for roads. And each community need pay as little as 10 percent of its roads' costs; the balance comes from the Highway Trust Fund. All users already have contributed to this fund through federal gasoline taxes, but can benefit from it only by building highways in their communities. The trust fund arrangement thus systematically allocates enormous resources to highways by distorting extraordinarily the relative costs faced by local communities as they choose between roads and other modes of transit. Although perhaps justified temporarily during a one-time expansion of the national network of highways, the extreme distortion introduced in our long-run transit choices by such a scheme certainly should not be allowed to persist. As a result of such a distortion in relative prices, communities will be persuaded to build more roads when other uses of the resources actually would benefit the society more.

If input taxes are raised to a level that will reduce congestion in urban areas to an efficient level, there will be a surplus of current tax revenue over amortized road construction cost. All tax proceeds thus should not be spent on roads. On equity grounds the surplus of tax proceeds over road construction cost ought to go to highway users in a way that will not influence their marginal road use (as a reduction in input tax would do), and some of it reasonably could go to nonusers as well. After all, a rent for scarce public land can be shared by all citizens. In particular, those who have been discouraged from using the road—the tolled-off—can claim a share in these excess payments. And there is a perfectly sound equity argument for supporting hospitals with the surplus, for instance, because motor vehicle air pollution has been shown quite convincingly to aggravate bronchitis, emphysema, and lung cancer (13). Because input taxes affect bus travelers more than automobile travelers, relative to their respective contributions to congestion, some additional return also should go to bus travelers in some form of subsidy. Because many citizens benefit from local government expenditures, it seems perfectly appropriate to approximate nonmarginal transfers by using proceeds of these travel input taxes as local revenues. Doing so would even alleviate the inefficiencies associated with other taxes. If they do not trust their political institutions, automobile travelers may treat suspiciously any claim that the use of motor vehicle input taxes will permit a reduction in other local taxes. But apart from such distrust it is a perfectly reasonable solution.

INCENTIVES FOR TRANSIT AGENCY OPERATING EFFICIENCY

Setting the price for a transit service deliberately below its average cost by providing a subsidy might destroy any incentive for operating efficiency. So to set prices below average cost in order to coordinate better our decentralized transit choices, we may have to sacrifice the efficiency incentives that exist when an agency must seek a profit or even break even. On the other hand, the profit test may never have provided good incentives. To make a profit a firm might have cut costs not by reducing waste and operating inefficiencies but by terminating services that were socially the most desirable. The profit test serves in a well-functioning competitive market to distinguish efficient firms that satisfy consumers from inefficient ones that do not. The inefficient ones will be driven from the field, and this harsh penalty serves as an incentive to encourage efficient operation. When the full competitive market profit test is inappropriate, however, as in the urban transit setting, its spur to operating efficiency can be lost.

Government action to remedy an excess profit or a loss position in a transit agency cannot then accept as an entirely reliable indication of what is possible and efficient the cost that the transit agency reports. For example, if any loss, no matter how large, is to be made up automatically out of government funds, no incentive will remain for the transit agency to control its costs; and the agency's reported loss may be larger as a consequence. Other bases for determining a subsidy payment, such as an allowable return to private owners based on capital employed, suffer also from biases they may introduce in the choice of inputs, which in turn will increase costs. But different bases for subsidy payment are worth examining for their effect on efficiency incentives because alternative incentives for efficiency are needed when the profit test is not available.

Let us consider 4 main ways that a government may provide subsidy payments to a public or private transit agency: (a) make up any deficit, (b) share fractionally in any deficit, (c) allow a subsidy payment to be "earned" based on an input measure, and (d) allow a subsidy payment to be "earned" based on an output measure. Any one of these methods may provide subsidy funds to a transit agency that is deemed to warrant them, but the methods can have very different effects because of the opportunities and constraints they place on the managers of the transit agency.

Although it may offer some sort of relief in unusual circumstances, the first type, the make-up-any-deficit scheme, removes all incentive for operating efficiently. It can be complicated in practice by the organization of public agencies, for the deficit will depend on the scope of the operating agency and the costs that are included in its operations. Some agencies do not consider capital costs, for instance, and an additional deficit would no doubt be incurred, at least implicitly, to cover the cost of such capital. Where capital is privately owned and a payment must be targeted for its use, we would have the third type of subsidy payment scheme, which we shall discuss in a moment.

Under the second basis for subsidy payments, deficits are shared fractionally by the transit agency, and the government subsidy source can preserve an incentive for efficiency; the transit agency will then enjoy a larger residual sum for it is more efficient (14). Such incentives are common in government contracting and seem useful, although they cannot be claimed as unqualified successes. In typical applications, a price and cost plan is established in advance for a particular project, and departures from the profit implicit in such a plan are shared by customer and vendor according to a preset formula. In a continuing service operation like transit, determining such profit targets for a privately owned transit agency would be difficult without moving to some extent into the third or fourth methods of subsidizing, where some reference to an input or an output is made. But the idea of sharing alone is a useful one and can salvage some efficiency incentive by making the agency suffer when its costs are higher.

The third type of subsidy is keyed to the usage of an input. Public utilities are regulated in the United States by rate-of-return constraint, which relies on the amount of one input, capital, to determine a profit figure that the firm is allowed to earn. In principle, this allows a subsidy to be earned per unit of the capital input, as the third subsidy method requires. In practice as it is applied to public utilities, this method is much like the first method, however, because, although the targeted profit depends on the capital input, that profit is then almost assured the utility, and cost increases lead merely to higher prices; very little incentive for efficiency remains. Indeed, the public utility firm can enhance its allowed profit by distorting its mixture of inputs away from the most efficient blend in favor of more capital on which its allowed profit depends (15). Because of long delays that occur in the administration of this regulatory scheme, its more perverse aspects may actually be muted, for a firm cannot win price increases promptly enough to make relief from financial penalties for inefficient operations immediately available. If no relief were granted and the firm were merely credited with payments based on the amount of capital (or other input) it employed, an incentive for keeping costs low would remain. Earning a subsidy payment based on usage of a particular input is almost certain to cause distortion in the mixture of inputs, however, and so it is undesirable for that reason. The urban transit capital grants have been shown convincingly to introduce a bias toward the use of capital, and thus to distort input mixtures away from the most efficient ones (16).

The earning of a subsidy payment per unit of output avoids any distortion away from efficient input mixtures and also can invite the expanded provision that typically is desired of public services (17, 18) A transit subsidy per passenger, or per passenger-mile, is therefore to be preferred over a subsidy based on any one input or on the size of the agency's loss. In the case of a privately owned transit agency, a subsidy per unit of passenger service can be effective not only because, with fixed price, it leaves cost reduction incentives in the firm but also because it urges expansion of service by the firm in order to win greater subsidy payments.

If ideal input taxes are in force, an ideal subsidy per bus passenger-mile can be estimated. The subsidy can be derived from an admittedly simple model of urban transit, and it takes the following form (7, p. 29):

$$S = M_A\left(\frac{AC_B}{AC_A} \cdot \frac{\partial AC_A}{\partial M_A} - \frac{\partial AC_A}{\partial M_B}\right) + M_B\left(\frac{AC_B}{AC_A} \cdot \frac{\partial AC_B}{\partial M_A} - \frac{\partial AC_B}{\partial M_B}\right)$$

To illustrate application of this formula, consider \$0.0002 a rough estimate of $\partial AC_A/\partial M_A$ or $\partial AC_B/\partial M_A$, where M_A and M_B represent trips that take roughly 1 hour [this crude estimate is based on relations given in the Highway Capacity Manual (19) and a time value of \$1/hour], and let $\partial AC_A/\partial M_B$ and $\partial AC_B/\partial M_B$ be about ⅓ as great (7). Then, if $AC_B = AC_A$ and $M_A + M_B = 800$ trips/hour, S = 10 cents/passenger trip. The appropriate value of S could vary for each bus route depending on traffic volume, the levels of average costs, and the effects of each mode on congestion costs.

Thus, when a profit test is foregone, as we have argued it should be with respect to urban transit agencies, a subsidy must be given to the agency providing transit service, and efficiency incentives can be lost or distorted as a consequence. The most direct way to determine an appropriate amount of subsidy is to estimate effects per passenger or per passenger-mile based on differences between average and marginal costs due to bus and automobile passengers. The ideal price, including effects due to any input tax, can be determined that way, and implicit in an ideal price is a subsidy per unit of service. Because a subsidy per unit of service also can offer an effective incentive for efficiency, it is to be preferred over alternative methods of providing a subsidy.

CONCLUSION

It is doubtful that present transit choice opportunities in cities reflect marginal social costs well enough to enable us to coordinate our actions efficiently. In choosing to travel by automobile, we pay only the average rather than the marginal social cost of our trip, with the result that together we create more than an efficient amount of traffic congestion. Ideally we would employ devices, which only now are technically feasible, to record the presence of vehicles and charge optimal tolls by time and place in order to ration road space. But we have been unable politically to implement such congestion controls. Because bus passengers contribute less marginal road congestion than automobile passengers, we can relieve this congestion problem somewhat if we can lower the price of bus travel and at the same time offer a service that will persuade some travelers to use buses rather than automobiles.

But higher taxes should be imposed on the inputs needed for road use, too, in order to make the perceived private cost of all road use approach more closely its marginal social cost. And the proceeds from such road use taxes reasonably can subsidize a substitute bus service and still should yield net proceeds beyond the cost of constructing roads. There is no compelling logic to support break-even rules in the urban transit setting. Of course, an attempt to achieve efficient use of town roads is apt to require a subsidy to the provider of a bus or transit service, and yet it is difficult to provide a subsidy and still encourage efficient operation of the agency. An award per unit of service probably will preserve an internal efficiency incentive better than alternative bases for the subsidy payment, and it can provide a better incentive for welfare-maximizing actions than the profit goal can offer.

ACKNOWLEDGMENT

Helpful comments by Charles Hedges and Irving Hoch and financial support from the National Science Foundation to the Thomas Jefferson Center at the University of Virginia are gratefully acknowledged.

REFERENCES

1. Sherman, R. Congestion Interdependence and Urban Transit Fares. Econometrica, May 1971.
2. St. Clair, G. P. Congestion Tolls—An Engineer's Viewpoint. Highway Research Record 47, 1964, p. 75.
3. Moore, J. H. Congestion and Welfare—Comment. Economic Jour., March 1968.
4. Vickrey, W. Pricing as a Tool in Coordination of Local Transit. In Transportation Economics, Columbia Univ., New York, 1965.
5. Sherman, R. A Private Ownership Bias in Transit Choice. American Economic Review, Dec. 1967.
6. Sherman, R. Club Subscriptions for Public Transport. Jour. of Transport Economics and Policy, Sept. 1967.
7. Sherman, R. Subsidies to Relieve Urban Transit Congestion. Jour. of Transport Economics and Policy, Jan. 1972.
8. Mohring, H. Relation Between Optimum Congestion Tolls and Present Highway User Charges. Highway Research Record 47, 1964, pp. 1-14.
9. Walters, A. A. The Theory and Measurement of Private and Social Cost of Highway Congestion. Econometrica, Oct. 1961.

10. Feasibility of Federal Assistance for Urban Mass Transit Operating Costs. U.S. Department of Transportation, Nov. 1971.
11. The Christian Science Monitor. Dec. 28, 1972.
12. Meiburg, C. O. An Economic Analysis of Highway Services. Quarterly Jour. of Economics, Nov. 1963.
13. Lave, L. B., and Seskin, E. P. Air Pollution and Human Health. Science, Aug. 21, 1970.
14. Cross, J. G. Incentive Pricing and Utility Regulation. Quarterly Jour. of Economics, May 1970.
15. Averch, H., and Johnson, L. L. Behavior of the Firm Under Regulatory Constraint. American Economic Review, Dec. 1962.
16. Tye, W. B. The Economics of Urban Public Transit Capital Grants. Published in this Record.
17. Roth, G. The Regulation of Buses in Cities. Published in this Record.
18. Miller, D. R The Urban Transit Industry: Deficits and Subsidy Mechanisms. Paper presented at Engineering Foundation Conference on Public Transportation, Henniker, New Hampshire, July 14, 1972, and to be published by the Highway Research Board.
19. Highway Capacity Manual. U.S. Government Printing Office, Washington, D.C., 1950.

REGULATION OF BUSES IN CITIES

Gabriel J. Roth, International Bank for Reconstruction and Redevelopment

This paper examines the main types of regulation—control of routes, rates, timetables, and standards—applied to the provision of bus services in cities and considers their effects in the light of the requirements of travelers. Conclusions are that controlling standards of safety, noise, and fumes to avoid the infliction of excessive costs on the public is generally desirable; regulating timetables, possibly in the form of subsidies to operators who keep to them, may have merit; controlling route operation and fares may not serve a useful purpose; and restrictions on the introduction of new bus services are not logical. The paper also discusses the case for bus subsidies and concludes that the need for subsidies does not justify the public operation or economic regulation of bus services. Grants related to passenger mileage on all or selected routes appear to be the most desirable form of subsidy, for they directly encourage the provision of services desired by travelers.

•THE 1972 National Transportation Report states (1):

> The present regulatory environment in urban public transportation, including obsolete franchise limitations and market-entry barriers for taxicabs and jitneys, restricts the efficient operation of the urban transportation system. The removal of such regulatory constraints is likely to lead to more efficient use of the transportation system and increase the options available to its users.

What are the main regulatory constraints affecting buses in cities? To what extent are they in the public interest? Should any be relaxed?

REGULATIONS

Bus regulations may be classified as follows:

1. Control of routes—specification of a route or routes that may be served by an operator, including the power to deny entry to and terminate service on any route;
2. Control of fares—specification of maximum and sometimes minimum fares;
3. Control of timetables—specification of first and last buses and of service frequencies along different routes; and
4. Control of standards—specification of the type, safety, and appearance of vehicles and the competence of their drivers.

Classes 1 and 2 may be regarded as economic regulations in that they determine who may provide bus services and at what price. Classes 3 and 4 are not selective as between suppliers, but determine the quality of the service that may be offered to the public. They may be regarded as physical regulations, but they also have economic elements in that they affect the costs of the items controlled.

Control of Routes

Most city authorities take for granted that bus routes in their areas should be controlled in the sense that a municipal authority should have the power to decide which operators may serve any particular route. There are different kinds of control. In the United States, control is typically exercised by granting a monopoly franchise to one operator and requiring him to work certain routes as a condition of his franchise. In some cities the control is exercised directly by virtue of municipal ownership. But route control as defined above requires neither municipal ownership nor a monopoly franchise to a private operator. A city can license any number of operators to serve any number of routes, and this is the typical pattern in developing countries. For example, in Bogotá in 1966 some 1,500 proprietors, organized in 19 companies, were licensed to serve 114 routes (2).

To what extent is route control necessary? To what extent are fixed routes necessary? Could the public be efficiently served if buses were to operate like taxicabs, with no fixed routes?

To take the last question first, a small group of people hiring a taxi have no difficulty in agreeing on its destination. On the other hand, 40 people hailing a passing bus would not likely agree on its routing. Let us therefore accept that bus routes are necessary, at least until the advent of demand-responsive systems with computer-controlled routing.

Travelers undoubtedly derive important advantages from the availability of fixed bus routes, but that in itself does not justify public route control. Some may argue that, if the operators understood that serving known routes at fixed schedules also serves their interests, public regulation is not required. But suppose the operator is stupid or fickle or greedy for quick profits or short-term advantages.

Route control is used not only to force existing operators to keep to their routes but also to protect them from competition from other operators and in particular to restrict the entry of other bus operators into the industry. Even the most rigid bus operating franchise, however, does not protect the operator from competition by other modes. Depending on their circumstances, travelers may have the option of walking or using taxis, jitneys, or their private automobiles. The latter alternative has particularly serious implications for public policy, and urban transport planners give high priority to measures that discourage the use of private automobiles, particularly for the journey to work.

The need to discourage the use of the private automobile might explain the acceptance in many cities of competition to bus services by minibuses and shared taxis. For example, in Hong Kong, minibus operation in competition with the established bus operators was legalized in September 1969, after some years of illegal but popular operation. A year later 3,800 minibuses were carrying 1.2 million passengers daily. Fares were not fixed but remained fairly stable at about twice the level of the buses. The minibus drivers could switch from one route to another in response to demand. The cost of the original license to operate a minibus was equivalent to $500 in the United States; but the number of licenses was limited, and their market value soon rose to about $8,500 (3). Other cities in which shared small vehicles serve a major transport role include Singapore, Caracas, Lagos, Istanbul, and Mexico City. None of these cities allows similar competitive freedom to the operators of full-sized buses. An operator of minibuses in at least one of them is interested in running full-sized buses at the minibus fare, which is double the normal bus fare, but has been forbidden to do so.

Much can be said both for and against route control, and it is not the purpose of this paper to take a stand on one side or the other. It is suggested, however, that the following propositions might be accepted by all protagonists:

1. A case may be made for giving a public authority powers to prevent bus operators from arbitrarily abandoning routes on which they provide service;
2. This objective might be achieved in a number of ways and does not necessarily require the establishment of a monopoly operator, either municipal or private (for example, operators might be required to give a 6-month notice before abandoning or reducing a service); and
3. The logic of allowing competition from minibuses while prohibiting competition from full-sized ones is not clear.

Control of Fares

Bus fares are politically sensitive, and most city authorities regard it as their duty to control them. If a bus company has a monopoly, control of some kind is assumed to be in the public interest.

Unfortunately, because of the political sensitivity of this issue, fares are often insufficiently responsive to costs. Thus, many cities insist on a level of fares that is the same for short and long journeys so that the former are discouraged and the latter encouraged. Vickrey argued cogently that uneconomic pricing of this kind can be very damaging to the viability of bus and subway services (4). In a number of cases, fares are fixed at levels that are too low to allow the operators to replace their equipment, and only rarely can one find bus companies being allowed to charge an excess fare for the provision of high-quality services. A high-quality and high-price service is, however, being provided on the outskirts of Manchester and is proving to be successful.

An economist has difficulty resisting the argument that, given free entry into the business, control of fares is unnecessary and can be harmful. To the extent that there is no free entry into the industry, there is a case for controlling fares; but such control does not require the industry to be owned publicly or even by one operator (taxi fares are controlled in most cities, despite the competitive organization of the taxi business). Even if the level of fares were uncontrolled, fares could be publicly displayed and arbitrary and sudden changes prohibited.

Control of Timetables

The timing of bus services can be of crucial importance to passengers and operators alike: to passengers because regularity in service can bring about economy in waiting time and to operators because variations in timing can result in variations in loadings. Suppose, for example, that an operator provides a scheduled service that departs from a certain station always on the hour. If free competition were allowed, a competitor might be tempted to run his bus 5 minutes before the hour and collect most of the first operator's traffic. The first operator might react by advancing his schedule to 10 minutes before the hour, and the resulting confusion would deprive the public of dependence on a regular service. It is the need to protect the scheduled operator that has persuaded many people that competition in bus services is not in the public interest. However, there are methods of control that would allow an element of competition and yet give the public scheduled services. For example, operations along a certain route might be open to all comers provided that they fit into a timetable laid down by a scheduling authority. In that case, the second operator would be told that he had to provide the service at the half hour, or not at all. In that situation the public would enjoy a service every 30 minutes instead of every hour. In Calcutta a few years ago some bus routes were shared by 20 or more independent operators, most of whom owned just one bus. The group formed an association that allotted timetables to all its members, and operators who deviated from their timetables had to pay a financial penalty to the other members of the group.

An alternative to fining operators who do not keep to timetables is rewarding operators who do keep to them with subsidies. The subsidy, which has been compared to a club subscription (5), would reflect the public's preference for an operator to offer service on a scheduled timetable rather than when conditions seem to offer sufficient profit. In both cases, but especially if subsidies are paid, problems must arise in deciding whether an operator does or does not keep to his timetable, particularly when traffic is slowed by congestion. Where bus frequencies are high, for example, 5 minutes or less between buses, precise scheduling may not be important and regulation of timetables may be unnecessary.

Control of Standards

Only public authorities can establish and police vehicle safety standards and safeguard the public from air, noise, and visual pollution. Thus, there are good reasons for cities to require operators to maintain their vehicles to acceptable standards of safety, cleanliness, quietness, and pollution-free operation and to employ competent drivers. To the extent that these regulations are designed to protect the interests of third parties, they should be applied to all vehicles using city streets. To the extent that they are designed to protect vehicle users, standards should probably be higher for vehicle operators serving the public than for vehicle owners who serve only themselves.

Summary

This brief review suggests that for municipal authorities to have powers to prevent bus operators from arbitrarily abandoning routes at short notice and suddenly raising fares and to control timetables may be in the public interest and to control standards of safety, appearance, noise, and pollution is certainly in the public interest. Cities can obtain these powers without having to operate bus services themselves or without having to appoint monopoly operators. But what are the financial implications?

UNREMUNERATIVE SERVICES

Many argue that bus services in cities have to be organized on a monopoly basis because only in that way can "essential" but unremunerative services be provided. The argument is as follows:

1. Some "essential" services can never cover their costs;
2. Therefore, they should be paid for by a process of "cross subsidization" whereby profits from remunerative routes are used to meet losses on unremunerative ones;
3. Therefore, some operators have the obligation to provide some loss-making services; and,
4. Therefore, those operators deserve protection that can only be provided by a monopoly franchise.

There are respectable arguments for subsidizing public transport in cities. In the first place, in the absence of congestion pricing, users of private automobiles are implicitly subsidized when traffic is congested in the sense that they are not required to pay the congestion and pollution costs arising from their trips. Sherman has pointed out that private automobile passengers are not required to meet all their costs and it may not be efficient to require bus passengers to meet theirs (6). Then there are economics of scale in urban public transportation. Mohring argues that, even in the absence of congestion, the time savings to passengers resulting from increased service frequency bring about substantial economics of scale that society should encourage by subsidies (7).

However, it does not follow from this that the subsidies should be provided by cross subsidization from the profits of remunerative services. Ponsonby (8), Hibbs (9), and others have shown that cross subsidization is a particularly inefficient way of supporting public transport, for it adds to the difficulties of the operators of the profitable routes. Furthermore, there is no reason why subsidies that may be justified for "weak" routes bear any relation to the profits earned on the "strong" ones. If there is a case for giving financial support to some routes, the subsidy should surely come out of the general revenues of those who demand it and not from the operators or users of other bus services.

Ponsonby illustrates the results of cross subsidization by considering the relation between the London buses and underground railways (10). Since the end of World War I and until fairly recently, the bus side of the underground group of companies (later the London Passenger Transport Board) made profits beyond what would have been required to keep them in business and used them to aid the underground railways. This meant that, for almost half a century, the road service side of the business had something to "give away" in the sense that bus fares were higher or the quality of bus service was lower (or both) than they could or would have been had the road services been developed apart and financially separate from the underground railways. No wonder the independent bus companies, unrestrained by the obligation to earn a surplus over and above the profit required to keep them going, found it easy immediately after World War I to expand at the expense of the old London General Omnibus Company (financially linked with the underground group) until their further success and expansion, in the central parts of London at any rate, were virtually brought to an end by the London Traffic Act of 1924. In this case, cross subsidization worked against the coordination of transport in the sense that it prevented the fullest possible development of all forms of transport.

How then should such subsidies be paid? A common method in the United States and in Britain is for a public authority to pay the difference between revenues and expenditures of the entity being subsidized. This policy does nothing to encourage efficiency. Lenthall (11) reports that the Middlesex and Boston Bus Company had an operating deficit of $34,000 in 1963. One of the first acts of the newly formed Massachusetts Bay Transportation Authority was to agree to subsidize it, and within 5 years the loss was running at around $500,000, although the service had been cut by 15 percent.

It may, however, be desired to pay the subsidy not to bus travelers in general but to certain classes such as the unemployed, schoolchildren, old-age pensioners, or other groups deemed needful of special assistance. In that case the food-stamp program may provide a suitable analogy. When the U.S. government wishes to subsidize the nutrition of some sections of the population, it does not nationalize the food industry nor require that profits earned on some sales be used to offset losses incurred on others. The authorities distribute food stamps, which do not deprive the recipients of the opportunity to use their bargaining power as consumers to shop around and buy the items most suitable for them in the cheapest markets. This analogy suggests that one way of subsidizing the transport needs of particular classes of users would be to give them coupons of a certain value redeemable by bus companies, taxicab owners, and others providing transport to the public.

This method will not confine the subsidies to unremunerative routes. This is as it should be and leads to another and more fundamental criticism of subsidies that are designed to close the gap between costs and revenues: If the reasons for subsidy are those put forward by Sherman (congestion) or Mohring (scale economies), the criterion of profit or loss becomes irrelevant. The services to be subsidized should be those that show the greatest excess of economic benefits over costs (strictly speaking, those services where marginal social benefits exceed marginal costs).

Because the excess of benefits over costs is likely to be greatest in situations of urban congestion and because some bus services in congested urban areas may be more profitable than some services in lightly trafficked rural areas, subsidies to profitable services may be more justified than subsidies to unprofitable ones (12).

If the object of the subsidy is to reduce the loss of passengers from public to personal transport, it should be based on passenger-miles carried by services that provide a substantial excess of social benefits over costs. If the timing of bus services is considered important, the subsidies could be confined to operators who provide regular services. Where fares are proportional to distance, a subsidy based on revenues would not differ from one based on passenger-miles. A subsidy based on operating costs would also be similar to one based on passenger-miles, but subsidies based on costs are likely to encourage wasteful expenditure. The implications of subsidies based on costs and revenues have recently been described by Nelson (13).

This brief review does scant justice to the problem of bus subsidies and does not deal at all with the question of the level of subsidies or their effects. Its main purpose is to suggest that the need to subsidize bus services is irrelevant to the problem of regulation, that subsidies can be paid as easily in competitive as in monopoly situations, and that they might be used to encourage operators who keep to publicly supervised timetables, although in conditions of congestion the control of timetables may be unnecessary and unenforceable.

SOME EXAMPLES

Before general conclusions are reached, it may be instructive to look at examples of bus operation and regulation in cities of several countries.

Istanbul

The dolmus (shared taxi) is generally indistinguishable from a taxicab, and drivers provide a dolmus service or taxi service according to fluctuations in demand. They may also switch from one of the recognized dolmus routes to another at will. The intending passenger has to shout his destination to the cruising driver who will stop if he is going that way, which is rather inconvenient for the passenger (and other motorists). In certain areas private buses operate in competition with the municipal system. A feature of these is that they are generally older vehicles and their condition varies from the impeccable to the doubtful. At major stops the conductor descends and extols the virtues of his bus service to the bus queue.

Ankara

The dolmus follows settled routes, which are marked on the vehicles, and stops only at official stops. The service is therefore less flexible but easier for a stranger to understand. The municipal bus service does not attempt to compete on frequency, only on price. Much more than in Istanbul the bus is the "working class" convenience. Dolmus fares are graduated, starting at 50 percent higher than the flat fare of the buses, but they offer higher standards of speed and comfort. In Ankara any restriction on the dolmus would probably increase travel by private car rather than by bus. In both Ankara and Istanbul driving standards vary, but accidents are few. The vehicles have to be inspected for safety every 4 months.

Nicosia

A few years ago, urban passenger service vehicles, of which there were some 126 licensed, were owned by 32 different people. By far the greatest proportion of the owners failed to make an adequate return on their investment. There was an obvious need for new vehicles, there was pressure on the government for a substantial increase in fares, there were demands for subsidy, and there were inadequate services. The buses were licensed by route, by number, and by time of departure. The system had many of the restrictions without the advantages. By the formation of a company with monopoly privileges to provide the somewhat limited services required, a better service is now offered to the public, the company makes money, fares have not been increased for more than 3 years, and as financial provisions are made in the new company's accounts new buses are introduced. Cooperation exists where competition had failed, perhaps because of the restrictions placed upon it.

Buenos Aires

The bus system provides saturation service to virtually every part of the area and accounts for 54 percent of daily trips. It is interesting to note that, after 30 years of public ownership, the bus system was returned to private ownership. Today the 14,200 buses in the metropolitan area are typically in the hands of owner-drivers, some buses having 3 owners. These owner-drivers form associations (empresas) to operate a given line. The empresas make schedules and provide some administrative services in return for a percentage of fares. Official regulation of rates and fares is provided at 3 levels: the Ministry of Public Works (for federal capital and interprovincial services), the Province of Buenos Aires (for intermunicipality services within the province), and municipal governments (for services within a municipality or partido). The proportions of the 14,200 buses controlled at these 3 levels are 73, 13, and 14 percent respectively. Fares are too low in relation to the rapidly inflating costs. Although a study by the Ministry of Public Works indicated that a 25 percent increase would be justified, the operators did not at the time raise their fares.

Calcutta

Bus services were nationalized in 1948 and vested in the Calcutta State Transport Corporation. For various reasons the corporation was unable to meet public demand, and 300 privately owned buses (mostly one-bus firms) were allowed to operate in the city in 1966 at the same fares as those of the state corporation. Although the state corporation had the best routes, it ran at a heavy deficit; the private buses made profits, and their owners clamored for more licenses. The comparative success of the private buses has been attributed to their superiority in repairing their vehicles and keeping them on the road and in collecting fares (bus crews are paid on a commission basis).

Manila

Grava (14) gives the following account of the Manila transportation system.

> The mass transportation system of Manila consists basically of public utility bus and jeepney line networks, with the two services operating in almost complete overlap. Both run on all major streets, and there are only a few sections of the city where one or the other predominates. It is interesting to note that in Manila, unlike some other cities, the acceptance and use of both systems is completely equal, i.e., one does not have a higher social status than the other, and the fares are the same. Specific choices by passengers are quite personal, with the only difference being that jeepneys are recognized as faster, while buses are more comfortable on longer journeys. During peak hours, when all vehicles are overcrowded, such fine differentiation is not made. However, a consideration is that when a jeepney has an accident or breaks down, which is not a rare occurrence, the passengers are on their own; while under a similar situation with a bus there may be a following unit of the same company that will pick up the riders without the payment of an additional fare.
>
> Since there is direct competition, the jeepney industry and its associations have a completely negative attitude toward the corresponding bus organizations. This feeling is fully reciprocated, and there is no cooperation between these two major components of the mass transit system, both of which consist of privately operated business ventures. Each side regularly makes proposals that the other should be eliminated.
>
> Negative reactions do not come from the riding public but from owners of private cars who are inconvenienced by the general traffic situation and who also, of course, are in leading positions with access to official agencies and the mass media. It can also be surmised that many well-meaning government officials and local transporation administrators feel faintly embarrassed by the whole system because it does not resemble any of the standard modes found in industrialized countries, and because there is an air of improvisation and limited resources about it.
>
> All investigators have noted that there are many problems associated with jeepney operations in Manila. These include deplorable driving habits and disregard for traffic regulations, lack of loading and unloading areas or their proper utilization, safety and insurance problems, abuses of labor practices, and many others. Yet, it is also apparent that these are not shortcomings of the system per se but can always be traced back to management and operational control. We have here a rampant free-enterprise endeavor that engenders cutthroat competition for fares on the part of drivers and operators alike. It would seem that proper policing, both on the street and of administrative factors, together with minimal physical improvements of channels, could go far in expediting the performance of jeepneys for the benefit of the riding public, the community at large, and the operators and drivers.

Singapore

In 1968 public transport in Singapore was provided by some 1,250 buses (operated by 11 private companies), 3,800 legally registered taxis, 5,000 illegal "private" taxis, 900 school taxis, and 222 school buses. In 1970 the government approved a public transport reorganization plan aimed at increasing the number of buses and reducing the number of operating companies, increasing the number of legal taxis and school buses, and eradicating the illegal taxis. These objectives appear to have been achieved, and in April 1972 Singapore was served by 2,000 buses operated by 3 private companies, 4,800 legal taxis, 186 school taxis, and 1,309 school buses (15). The illegal taxis have virtually disappeared from Singapore, but there is evidence that they operate intercity services between Singapore and neighboring cities (16). The school buses, all privately owned, generally provide door-to-door service. On the basis of private monthly contracts (the casual pickup of passengers is not permitted), 250 of the school buses are also allowed to transport commuters to work. (Working hours in Singapore do not coincide with school hours so that the same vehicles can serve workers and scholars on separate journeys.)

Caracas

Caracas is distinguished by a relatively high level of private automobile ownership (1 automobile for every 10 persons) and a correspondingly heavy reliance on private automobiles and taxis for urban transport. It is estimated that 46 percent of daily trips are made by private automobile, 16 percent by shared taxi, 30 percent by bus, and 2 percent by regular taxi. (During the 7 to 8 a.m. peak hour, 36 percent of trips are by private automobile and 22 percent by shared taxis.) Bus services are provided by 23 companies operating some 1,300 buses over 73 routes. Two of the bus companies are publicly owned and operate at substantial deficits, mainly as a result of poor fleet utilization caused by poor maintenance and old equipment. The 21 private companies, which transport 57 percent of bus passengers in 54 percent of the city's buses, earn enough to remain in business, but possibly not enough to make it worthwhile to replace their fleets. Of special interest in Caracas are the 11,000 taxis and 5,000 shared taxis. (During peak hours, many of the regular taxis operate illegally as shared taxis.) The shared taxis are organized into about 50 associations serving 85 different routes. The number of licensed taxis and shared taxis has been fixed, and the profitability of the business is reflected in the market value of licenses, which is equivalent to about $1,500 in the United States. Most of the shared taxis are U.S. 6-seat sedans, but some of these have been replaced by 9-seat microbuses, which are much cheaper to operate and carry more passengers.

Lagos

Lagos City Transport Service, owned by the Lagos City Council, operates 380 buses and is the largest public transport operator in the area. Private buses are also allowed to operate in Lagos on the same routes as the public buses, except that only 200 private buses are licensed to enter Lagos Island. Minibuses of the Volkswagen type are used to provide service to passengers outside Lagos and are called kia-kia or quick-quick because they make their way with speed and dexterity between the larger and heavier buses. There are some 2,300 of these minibuses operating in Lagos state. Their drivers are all members of a powerful association that operates its own form of route licensing by requiring members to specify their route when they join. New applicants are charged a fee equivalent to $70 for lifetime registration plus about 25 cents per vehicle per day for the use of vehicle parks provided by the union. Fares are fixed by convention and are sometimes lower than the bus fares. About 75 percent of the kia-kia buses are owned by their drivers. In addition, Lagos is also served by some 600 taxis and a large number of "molue" buses that are built locally on truck frames and have all-weather protection for approximately 25 seated passengers.

CONCLUSIONS

The first conclusion is that conditions vary so much from one city to another that no single solution to the problem of bus regulation is likely to suit all circumstances. For example, in the short term the establishment of a monopoly organization may bring substantial improvements to urban transport, particularly where the development of private services is hampered by fare ceilings or restrictions on entry. The success of a monopoly service is likely to depend to a large extent on the efficiency of the municipal administration in general and on the qualities of the bus administrators in particular. Hamburg and Stockholm, for example, are reported to enjoy excellent municipal services. But in the long term the disbenefits of such a monopoly often become more apparent as changes in demand are not reflected in services organized by a management that lacks the incentives of competition.

The second conclusion is that the use of shared small vehicles—taxis, minibuses, jitneys, or jeepneys—can make a very large contribution to urban public transport. By providing fast, flexible service and assured seats, these shared vehicles are not unlike the private car, and the fare is well below that of a taxi. This service is likely to appeal particularly to the dissatisfied bus passenger who is tempted to desert the bus for his automobile. It is, thus, potentially of great value in communities with a large and growing car-owning population.

Third, public transport controls that are likely to be in the public interest are the physical and not the economic ones. They do not usually require the granting of monopoly franchises, still less the operation of bus services directly by municipal authorities. The experience of many cities outside North America confirms the statement quoted at the outset from the report of the U.S. Department of Transportation: "The present regulatory environment... restricts the efficient operation of the urban transportation system. The removal of such regulatory constraints is likely to lead to more efficient use of the transportation system and increase the options available to its users."

ACKNOWLEDGMENT

In preparing this paper, the author benefited from the assistance of Jeffrey Balkind, V. W. Hogg, Alfredo Soto, Rena Zafiriou, M. S. P. Kerridge, Gilbert J. Ponsonby, and W. J. Tyson. The author is solely responsible for the views expressed and the conclusions reached.

REFERENCES

1. National Transportation Report. U.S. Department of Transportation, Executive Summary, 1972, p. 16.
2. Bogotá Transport and Urban Development Study: Phase I. Freeman, Fox; Wilbur Smith and Associates; and Restrepo y Uribe Ltda., Vol. 6, April 1970.
3. Buses. March and Dec. 1971.
4. Vickrey, W. Revision of the Rapid Transit Fare Structure of the City of New York. Jour. of Operations Research Society of America, Vol. 3, 1955, p. 38.
5. Sherman, R. Club Subscriptions for Public Transport. Jour. of Transport Economics and Policy, Vol. 1, No. 3, Sept. 1967.
6. Sherman, R. Subsidies to Relieve Urban Traffic Congestion. Jour. of Transport Economics and Policy, Vol. 6, No. 1, Jan. 1972.
7. Mohring, H. Optimization and Scale Economies in Urban Bus Transportation. American Economic Review, Vol. 63, No. 4, Sept. 1972.
8. Ponsonby, G. J. What Is an Unremunerative Service? Institute of Transport Jour., London, March 1963.
9. Hibbs, J. Transport for Passengers. Institute of Economic Affairs, London, Hobart Paper 23.
10. Ponsonby, G. J. Transport Policy: Coordination Through Competition. Institute of Economic Affairs, London, Hobart Paper 49, 1970.
11. Lenthal, T. Bad Model for British Transport. The Times, London, March 1, 1968.
12. Tyson, W. J. A Critique of Road Passenger Transport Subsidy Policies. Manchester School, Dec. 1972.
13. Nelson, G. R. An Econometric Model of Urban Bus Transit Operations. Institute for Defense Analyses, Paper P-863, Sept. 1972.
14. Grava, S. The Jeepneys of Manila. Traffic Quarterly, Oct. 1972.
15. Lim, L. G. Development of Public Transportation in Singapore 1967-1972. Unpublished.
16. New Nation. Singapore, Feb. 27, 1973.

ECONOMICS OF URBAN TRANSIT CAPITAL GRANTS

William B. Tye, Charles River Associates, Inc.

Four arguments support the restriction of federal grants to capital expenses of public transit: The transit industry is "capital poor"; a capital grant restricts the power of transit unions to dissipate most of the grant through wage gains; a capital grant limits the federal government's liability by avoiding an open-ended commitment such as an operating subsidy support for labor costs; and a capital grant is a highly visible means of showing federal concern for transit. Each of these arguments is shown to be without merit. The uneconomic incentives inherent in a capital subsidy suggest that, if the arguments for a federal subsidy to transit operations are accepted, the funds should be allocated as a generalized subsidy to transit service rather than restricted to capital expenses.

•THIS PAPER examines each of the following 4 major arguments supporting the restriction of federal grants to capital expenses of public transit and provides evidence to show that they are not compelling. [Details of the economic arguments and empirical evidence used to reach these conclusions are given in other papers (1, 2).]

1. To support this restriction to capital expenses, proponents have argued that the transit industry is "capital poor." An examination of actual and optimum investment in 2 major transit systems does not support the hypothesis that the transit industry was undercapitalized prior to the inception of the Capital Grant Program.

2. Supporters have also argued that a capital grant restricts the power of transit unions to dissipate most of the grant through wage gains. On the contrary, a capital grant does nothing to immunize the recipient from wage demands. If the recipient prefers to pass on most of the subsidy benefits through wage gains and featherbedding rather than through fare decreases and service improvements, the capital grant mechanism is shown to be no impediment.

3. A capital grant limits the federal government's liability according to program adherents because funds are disbursed on a project basis rather than through an open-ended commitment such as operating subsidy support for labor costs. As a matter of fact, an analysis of the incentives of the subsidy to capital shows that it encourages a wasteful use of capital through inadequate maintenance, overcapitalized technology, and premature replacement. Far from limiting the federal government's liability, the capital grant technique created such an inefficient incentive to waste capital and such an open-ended demand for capital funds that increases in scope of the program seem inevitable.

4. A capital grant is a highly visible means of showing federal concern for transit. This very powerful consideration is pernicious in its effect, for it encourages the wrong measures of a program's effectiveness: the installation of facilities rather than the ultimate but less tangible program objectives. Although a capital grant does provide a highly visible medium for the expression of federal generosity, this attribute is paid for by a considerable increase in the cost of providing transit service because of the incentives to waste capital.

The uneconomic incentives inherent in a capital subsidy would suggest that, if the arguments for a federal subsidy to transit operations are accepted, the funds should be allocated as a generalized subsidy to transit service rather than restricted to capital expenses.

THE URBAN MASS TRANSIT CAPITAL GRANT PROGRAM

The Urban Mass Transportation Act of 1964 (as amended through October 15, 1970) authorizes the U.S. Department of Transportation to make grants to state and local governments and their instrumentalities to finance as much as two-thirds of the cost of equipment, buildings, rights-of-way, and the like to improve transit. According to the act, the recipient cannot finance the remaining share from fare-box revenues, nor may any part of the funds be used to defray operating expenses. From 1964 through 1971, $800 million was committed by the federal government under the Capital Grant Program (3).

The Urban Mass Transportation Assistance Act of 1970 granted the Secretary of Transportation obligational authority of $3.1 billion during the next 5 years and expressed an intention to commit $10 billion during a 12-year period. This vast commitment to a long-term program of aid to transit makes an examination of the original program decision to exclude operating expenses particularly compelling.

ARGUMENTS SUPPORTING A RESTRICTION TO CAPITAL EXPENSES

Proponents of the Capital Grant Program have enlisted an extensive battery of objectives and justifications: Increase the mobility of those who are disadvantaged (because of income, age, physical handicaps) with respect to the use of the automobile; encourage low fares to ease the plight of low-income persons who patronize transit; impede urban sprawl and shape urban growth in a more pleasing and orderly way; discourage the use of automobiles and thus reduce the undesirable side effects of their use; reduce the total costs of meeting urban travel demands; prevent local communities from choosing highways over transit because only the former is federally funded; ease fiscal imbalance of resources and needs in the federal system by sharing the local government's financial burden of deficit-ridden transit systems; and so on. The wisdom of these objectives and the effectiveness of federal subsidies in achieving them have been extensively dealt with in the literature. It is not the purpose of this paper to address these objectives because they are relevant only to the idea of federal aid; they cannot be used to defend the use of a capital grant mechanism to deliver this aid. Proponents have separated the issue of whether transit should be subsidized from the issue of the best delivery mechanism.

Needed by Capital-Poor Transit Industry

Supporters of the capital grant mechanism have relied on another set of arguments to justify this particular method of subsidizing transit. The first of these is the notion that the transit industry is capital poor. According to this argument, the most promising method to revive public transportation was massive expenditures on high-speed rail transit systems, but local governments were not interested because of the very heavy installation expenses. This faith in a highly capital-intensive technology did not, of course, require that the subsidy be limited by federal law to durable capital expenses; the choice of technology could have been left up to program administrators and local decision-makers.

However, proponents were convinced that local governments generally "lived from hand-to-mouth" and failed to provide for their long-term capital needs and that this was especially true for urban transit systems. [According to a 1964 publication (4), "The greatest needs at the present time are for system improvement, modernization of suburban railroads, modernization and extension of present rail rapid-transit systems, and replacement of obsolete buses. Most of these needs require capital outlays, a fact which somewhat diminishes the strength of the argument for service supports, as opposed to support for capital improvement."]

The inefficiency due to a scarcity of durable plant and equipment would persist if discretion were given to local government; unrestricted grants for service or subsidies for operating expenses would be dissipated by wasteful management practices (4). On the other hand, a capital grant would (one infers) improve the efficiency of the transit industry through the "law of equal cheating": The capital grant's incentive to make lavish use of capital would offset the transit industry's inclination to undercapitalize.

Not Usable for Wages

As a second consideration, supporters of the capital grant mechanism were confident that a restriction to capital expenses would immunize grant recipients from the power of transit unions to dissipate most of the grant through outrageous wage gains (4) and "featherbedding." If the grant were restricted to capital expenditures, clearly it would be channeled to benefits for transit patrons rather than to giveaways to employees. The beauty of this feature was that, because the program's design ensured arm's-length negotiations between the recipient and transit employees, the program could be administered without any federal involvement in local labor negotiations. This would be a sharp contrast to the prior federal experience with operating subsidies.

Advocates of federal aid to transit argued that a vicious cycle of fare increases, service cuts, and patronage losses could be reversed by a program of massive federal aid to transit capital improvements, after which local governments could make transit viable (5, 6). On the other hand, supporters wished to placate the opposition's fear that the subsidy program would become uncontrollable. A capital grant seemed to offer something to everyone—a bold thrust to reverse the decline in transit and prove its worth, yet a program that did not promise to be a commitment forever to subsidize a declining industry.

Limited Federal Liability

A capital grant limits the federal government's liability because the aid is delivered on a discontinuous, project basis. Once the goal of the program has been accomplished (or abandoned), the funding requirements can be reduced. On the other hand, a subsidy for service or operating expenses is ongoing and open-ended. Recipients begin to believe that they are entitled to these periodic payments as a matter of right. Operating costs are primarily labor costs, and vested labor interests create great pressure to expand the program beyond any reasonable bounds. As a result, Congress loses any discretion in determining funding levels. The Secretary of Transportation expressed this viewpoint (3): "There is just no bottom to the barrel."

Highly Visible

An unspoken (but nevertheless powerful) consideration was the high visibility of a capital grant. Both the federal government and the transit constituency could see tangible evidence of federal concern, congressmen could issue press releases and cut ribbons, and the results of the program could never be questioned because every appropriated dollar produced an actual "improvement." Senator Harrison Williams acknowledged that his operating subsidy bill was deficient in this respect (3, pp. 11-12): "The one drawback of the operating subsidy is that it will only pay for existing inefficiency and poor service. It certainly does not build new facilities nor does it buy desperately needed new equipment."

ARGUMENTS TESTED

Needed by Capital-Poor Transit Industry

To test the hypothesis that the transit industry was undercapitalized prior to the Capital Grant Program, a model of the optimum motor bus replacement decision was developed, and optimum and actual replacement decisions were compared for 2 large transit systems. The conclusion is that for both Cleveland and Chicago bus-replacement decisions were very close to optimal. [The empirical evidence is sketched only briefly in this paper. Details of this and other tests of arguments are given elsewhere (1).] If the undercapitalization hypothesis must be rejected for bus-investment decisions, it seems likely that it would be rejected for other investment behavior as well. There is little reason to believe that the restriction of the grant funds to capital expenses will produce a more efficient industry by offsetting the alleged propensity to undercapitalized transit operations. On the contrary, the evidence from Cleveland and Chicago indicates that the capital grant will encourage a costly overcapitalization of the industry.

The replacement model indicated that Cleveland should replace buses operating 50,000 miles annually at the end of 15 years. Prior to the capital grant, 15-year-old buses were rendering less than 6,000 miles of output on the average. Cleveland was, therefore, replacing equipment at an age earlier than that dictated by cost considerations.

A comparison of actual and optimal decisions for Chicago showed similar results. The indicated optimal replacement dates were earlier for Chicago than for Cleveland, but the actual replacement decisions were very close to optimal.

Not Usable for Wages

This advantage of a capital grant fails to consider what economists call "fungibility." An examination of the possible effects of a capital grant on fares, output, wages, and the demand for labor will show that, if the recipient prefers to hold fares and service at the same level as would prevail in the absence of the grant and to pass along the benefits of the grant to the workers through wage gains and featherbedding, the capital grant mechanism is no impediment. In fact, the only conceivable way a capital grant could possibly discourage giveaways to transit workers is that its extreme inefficiency could so reduce the benefits to the recipient that little would be available for diversion.

The two-thirds federal subsidy to capital expenses provides a benefit to the recipient by reducing the cost of providing transit service. The grant recipient could respond to this price decrease in a variety of ways. To pass the benefits on to the transit patron, the recipient would increase service or decrease fares (compared with the levels that would prevail without the subsidy) or do both. On the other hand, the recipient could maintain the same fares and service as would have prevailed without the subsidy and channel the savings in capital costs to the workers. (As an illustration of this point, one big city transit general manager noted that depreciation reserves were no longer immune to wage demands after the Capital Grant Program was initiated.) The point is not that the Capital Grant Program has not resulted in gains in service and lower fares but that the fungibility of the cost savings makes a mockery of the assurances that a capital grant guarantees that no funds can be diverted to giveaways to the workers. Furthermore, the ease of diverting these benefits is facilitated by a prohibition in the Urban Mass Transportation Act of 1964 against federal involvement in local decisions in matters such as fares, service, and wage negotiations. As a result the capital grant mechanism provides no protection against such diversions to employees.

One feature of the program actually encourages such diversions. The 1964 act provides that any part of capital expenses that is financed from operating revenues cannot be available for federal subsidy. Suppose a transit system that generated revenues greater than operating expenses was considering an application for a federal capital subsidy. Such a situation would not last for long: The local system loses 66⅔ cents in federal subsidies for each $1.00 that revenues exceed operating expenses. One solution would be to raise wages until the excess was eliminated: For each $1.00 increase in wages, the local government receives an additional 66⅔ cents in capital grant funds. The federal government has financed two-thirds of the wage gain, a result directly contrary to the intent of the program. The fact that practically every grant under the program has forecast no fare-box funding of capital expenses should give little cause for wonder.

Hence, it must be concluded that the capital grant mechanism provides no constraint on a recipient's ability to allocate the benefits as the local political process sees fit. However, the capital grant does discourage giveaways to workers in one significant way. Because it encourages inefficiency through higher costs, the total benefits to be divided between transit employees and patrons are reduced accordingly. The greater the waste is, the fewer are the potential benefits available for diversion to transit employees.

Limited Federal Liability

Evidently, it never occurred to anyone to question whether a capital subsidy would provide the same incentives to waste capital as an operating subsidy would provide incentives to waste labor. As demonstrated later, a capital subsidy encourages a significant overcapitalization of the transit industry. As a result, the demand for capital grant funds has greatly risen, and in response the scope of the program has been tremendously increased by recent legislation. Far from limiting the federal government's liability, the capital grant technique created such an incentive to waste capital that increases in the scope of the program seem inevitable.

Certainly there is something to the argument that aid delivered on a project basis does not create vested interests as entrenched as does continuous aid delivery that becomes treated as a right. But such arguments carry the greatest weight at a program's inception when lip service is paid to the myth of temporary aid. The commitment to a long-term aid program in the 1970 legislation makes this argument irrelevant.

Highly Visible

The visibility of a capital grant is an extremely powerful argument and is difficult to analyze. However, an operating subsidy is probably equally visible. The continuing dependence of the recipient under an operating subsidy could easily be used to turn the argument around: The threat that aid would be discontinued and the continuing control over the recipient that this threat implies could well give an operating subsidy more "visibility." Furthermore, the capital grant's encouragement of the wrong measures of the program's effectiveness is to be deplored. Whatever advantages may accrue from the visibility of the project form of aid delivery, we shall shortly see that these advantages are purchased at a considerable cost.

INEFFICIENCY OF A CAPITAL GRANT

According to the law of equal cheating, the capital grant was supposed to offset the alleged undercapitalization of the industry. Once the capital-poor hypothesis has been rejected, a subsidy to capital alone is seen to be a tremendous incentive to achieve a costly overcapitalization of the transit industry. This is best illustrated by reference to the bus-replacement model developed to test the undercapitalization hypothesis. The two-thirds subsidy to capital costs fully halves the least cost replacement age as seen by the grant recipient. Unfortunately, if the recipient follows his best interests, he will create a tremendous inefficiency because the cost of the subsidy to the federal government will be much greater than the benefit to the recipient through reduced costs. The source of the inefficiency is the recipient's incentive to incur subsidized capital costs rather than unsubsidized operating expenses. One way to achieve this is by premature replacement, which increases the capital costs of purchasing new buses but decreases the operating expenses of maintaining an older fleet. [A 1970 study noted (7): "Since capital grant-in-aid funds are available from the federal Department of Transportation, a strategy of capitalizing operating and maintenance costs wherever possible permits maximizing the use of available public funds . . . it will in effect be using federal dollars to provide operating assistance." This study was done independently of the present author's research, but the findings are virtually identical. The recommended replacement age under the Capital Grant Program was 8 years.] Whenever a $1.00 increase in capital costs will save more than 33⅓ cents in operating expenses (the local share of the capital grant), the trade-off will be made. Any difference between the increase in capital costs (both local and federal shares) and savings to the recipient is waste.

The cost figures from Cleveland and Chicago show that a recipient who attempts to minimize his costs after implementation of a two-thirds subsidy to capital expenses will choose a bus-replacement date such that more than 22 percent of federal appropriations is wasted through increased costs. Further waste will accrue from neglected maintenance and the choice of extremely capital-intensive technology. Such incentives to waste federal appropriations are a significant threat to the program's ability to accomplish its objectives.

IMPLICATIONS FOR PUBLIC POLICY

The recent vast expansion of the Capital Grant Program makes a reexamination of the original program decision to exclude operating expenses particularly compelling. The findings of this paper are that the arguments for restricting aid to capital expenses are not compelling. Furthermore, the capital grant's incentives to overcapitalize the industry create a source of tremendous inefficiency. Therefore, it is recommended that, as an alternative to a vast expansion of the Capital Grant Program, federal grants to transit operations, whether publicly or privately owned, be allocated among states and municipalities as a generalized subsidy to transit service without a restriction to capital expenses.

REFERENCES

1. Tye, W. B. The Economic Costs of the Urban Mass Transportation Capital Grant Program. Department of Economics, Harvard Univ., PhD dissertation, 1969.
2. The Capital Grant as a Subsidy Device: The Case of Urban Mass Transportation. In The Economics of Federal Subsidy Programs: Transportation. To be published by the Joint Economic Committee of the U.S. Congress.
3. Feasibility of Federal Assistance for Urban Mass Transportation Operating Costs. Urban Mass Transportation Administration, U.S. Department of Transportation, Nov. 1971.
4. Lyle C. Fitch and Associates. Urban Transportation and Public Policy. Chandler, San Francisco, 1964, p. 211.
5. Hodges, L. Urban Mass Transportation. Hearings before a Subcommittee of the Committee on Banking and Currency, 87th Congress, 2nd Session, U.S. Government Printing Office, Washington, D.C., 1962, pp. 71-72.
6. Weaver, R. C. Urban Mass Transportation. Hearings before a Subcommittee of the Committee on Banking and Currency, 87th Congress, 2nd Session, U.S. Government Printing Office, Washington, D.C., 1962, pp. 71-72.
7. Optimum Bus Age Study: Report to Cleveland Transit System. W. C. Gilman and Company, division of Barton-Aschman Associates, Inc., Sept. 29, 1970, pp. 4, 32.

PRICING, METERING, AND EFFICIENTLY USING URBAN TRANSPORTATION FACILITIES

William Vickrey, Columbia University

This analysis is intended to demonstrate that roadway pricing need not, and indeed should not, be viewed primarily as a matter of redressing a balance between the private automobile and other forms of transportation, especially transit. Road pricing, properly applied, can in the most severely congested situations lead to great improvement in the efficiency with which the roadways are used and even, in some cases, to a reduced reliance on transit. Where heavy investment in separate-right-of-way transit facilities is under consideration, roadway pricing may well tip the balance in favor of less, not more, investment in transit services. Roadway pricing is not, therefore, mainly an issue in the battle between automobile interests and transit interests, in which the institution of roadway pricing would be regarded as a victory for transit and a defeat for the automobile. Rather, roadway pricing has a great deal to offer all of the parties involved and should be considered not a defeat for anyone but a victory for the cause of rational and efficient urban transportation.

•IN MUCH of the discussion of the pricing of urban transportation, emphasis is placed on the role of pricing policies on the division of traffic among the various modes to the relative neglect of the equally if not more important role that proper pricing can play in the promotion of efficient patterns of utilization within each mode. Thus, for example, roadway pricing is often thought of primarily as a means of diverting traffic from private automobiles to transit or possibly of suppressing traffic entirely. Actually, as will be shown more specifically below, control of traffic through pricing will in many cases have the effect of increasing rather than decreasing the number of person trips made by private automobile through increasing the efficiency with which the roadway system is used by eliminating or at least greatly reducing the capacity-reducing queues and "hypercongestion" that characterize many areas during peak hours. Again, as long as there is no adequate differentiation within modes between peak and off-peak charges and costs are averaged over peak and off-peak conditions, the relative costs of private automobile transportation, which tends to carry a relatively much smaller share of the peak load, will be understated relative to transit and especially commuter rail transit that tends to carry a much greater share of the peak loads. Proper solution of modal-split problems can thus not be achieved without first ensuring that the intramodal utilization patterns and costs are correctly handled.

SINGLE VALUE OF TIME

Another major difficulty with many discussions of urban transportation is the prevailing use of a single uniform or average value of time for purposes of analysis or of evaluation. Although there is often at least nominal acknowledgment that the value of time may vary from one person to another, or even from one occasion to another, there has been thus far little recognition of the role that variations in the value of time play in determining daily patterns of travel and especially in determining the way these patterns may be expected to vary in response to changes in patterns of pricing and in patterns of congestion. An extra 5 minutes of waiting while one is standing on an exposed platform is not the same thing as an extra 5 minutes of waiting while one is sitting in a stalled commuter train or sitting in a car inching along in bumper-to-bumper traffic or straphanging on a crowded subway train. Even more important, 5 minutes at the office before the start of work or downtown before the stores open is not the same thing as 5 minutes during the working day or at the breakfast table or in bed. And it is hardly possible to develop a motivated model of urban transportation, as distinct from a mere empirical description, without bringing these variations in the value of time explicitly into account.

MODEL OF TRIP-TIME CHOICE

The importance of considering variations in the value of time can be seen in the following trip-time choice model. Let

$$V_b(t) = \int_o^t v_b(x)dx$$

and

$$V_d(t) = \int_o^t v_d(x)dx \qquad (1)$$

be respectively the total value of time spent at the base (home) and the destination (office) from time x = o to time x = t. In each case, these values of time can conveniently be measured relative to a 0 base taken as the value of time spent in travel from one location to the other. (If time value is defined in some absolute sense—though it is not clear what operational meaning would be given to this—then the value of time spent in travel could be subtracted uniformly from all of the time-value functions without affecting the results.) Further, let

t_g, t_r = times of the going and return trips respectively, in terms of the time of arrival at the destination;
n_g, n_r = normal running times for the going and return trips in the absence of delay;
q_g, q_r = amount of delay or queuing involved in the 2 trips, dependent on t;
c_g, c_r = money cost of the trips, in the absence of tolls; and
P_g, P_r = tolls charged, dependent on t.

The total value to be maximized for the day is then

$$U = V_b(t_g - n_g - q_g) - V_d(t_g) + V_d(t_r - n_r - q_r) - V_b(t_r) + V_b(T) - c_g - c_r - p_g(t_g) - P_r(t_r) \qquad (2)$$

where T is an arbitrary terminal time. Because $V_b(T)$, c_g, and c_r do not depend on t_g or t_r, they can be dropped from the maximand. Moreover, the terms involving t_g are distinct from those involving t_r so that the problem can be decomposed into 2 independent problems, one of determining t_g and a separate one for t_r. Maximize

$$\left.\begin{aligned} U_g &= V_b(t_g - n_g - q_g) - V_d(t_g) - p_g(t_g) \\ U_r &= V_d(t_r - n_r - q_r) - V_b(t_r) - p_r(t_r) \end{aligned}\right\} \qquad (3)$$

The problem is shown in Figure 1. The function V_b rises more or less steadily, reflecting the relative absence of critical times at the base, and the function V_d rises steeply during office hours and is relatively flat outside office hours. The individual leaving home at time $t_g - n_g - q_g$ as indicated at a will arrive at his destination at the time indicated at c and leaving the destination at h will arrive back at home base at l. The gain from the trip is the time value realized, ba + (jh - ec) + (rs - ml) less the cost of the 2 trips fg + nq and less the time value that would have been realized had the entire period been spent at home base, rs. This in turn is equal to (ba - ec) - fg + (jh - ml) + rs - nq - rs = (cd - fg) + (lk - nq). The outward trip can thus be timed to maximize (cd - fg) and the inward one to maximize (lk - nq). Either of these may be negative, depending in part on the arbitrary vertical displacement of the curves V_b and V_d. Their sum, however, is independent of vertical shifts in either or both of these curves; if this sum is not positive, the round trip would have no gain and would not, in principle, be undertaken.

Figure 2 shows the outgoing trip decision in somewhat more detail. For purposes of further analysis in cases where a toll is levied at a salient bottleneck, it is convenient to conduct the analysis in terms of the time of passing this bottleneck and to divide the time n_g into time from the base to the bottleneck and the time from the bottleneck to the destination: $n_g = n_b + n_d$. We can then shift curve V_b to the right by n_b to give the curve V_b' and shift curve V_d to the left by n_d to give V_d'. The toll rate in terms of time at the bottleneck $p_g(t_b)$ can now be added to the curve V_d' vertically to get the curve B. Now we can add horizontally to the left an amount representing the time delay in the queue prior to reaching the bottleneck at each of the times represented by curve B to get the curve Q. The time at which the curve V_b' is farthest above curve Q is now the optimum time for the trip in question to take place, starting from home base at a, arriving at the bottleneck at b, being delayed in the queue for a time qg, paying a toll qe, and then requiring a time ec to go from the bottleneck to the destination.

This analysis makes it possible to trace the impact of various toll, capacity, and queuing conditions on the choice of travel times, and vice versa, in a manner that has very important implications for transportation policy.

SIMPLE BOTTLENECK SITUATION

One of the simpler situations to which this analysis can be applied is that of the simple bottleneck, where traffic demand during peak periods exceeds the capacity of the facility and leads to the accumulation of a queue that persists fairly regularly throughout the rush hour until it is gradually worked off during a subsequent period of relatively slack demand. Such situations are increasingly frequent and furnish a dramatic and clear-cut example of how correct pricing in terms of efficiency criteria can increase the efficiency with which facilities are used and actually benefit the users without requiring any substantial modal shift. It is, indeed, in situations such as these that proper pricing can yield extremely high returns in very short lead times with an extremely small investment.

Because we can concern ourselves with the going and return trips separately, we will simplify the notation by dropping the subscripts g and r and also deal with the shifted curves $V_i(t - q) = V_b'(t_g - n_g - q_g)$ and $W_i(t) = V_d'(t_g)$, where t is now the time of leaving the bottleneck and V_i and W_i respectively are the values of time for individual i at origin and destination in terms of times of arrival at and departure from the bottleneck. In this representation and elsewhere, q represents not the actual time spent in the physical queue but rather the time spent in the queue in excess of the time that would have been required to cover the distance occupied by the queue in the absence of congestion. In other words q corresponds to the net delay in the queue.

Equation 3 then becomes

$$U_i = V_i(t - q) - W_i(t) - p(t) \tag{4}$$

and differentiating with respect to t, we have

$$U_i'(t) = (1 - q')\,V_i'(t - q) - W_i'(t) - p' \tag{5}$$

For any given value of t, those for whom $U_i'(t)$ is negative will travel before time t, and those for whom it is positive will travel after time t. Equilibrium will be established when functions p(t) and q(t) are determined in such a way that, whenever q(t) is positive, the number of those choosing to travel so as to pass the bottleneck between t and t + dt is equal to the capacity of the bottleneck at that time, $C'(t)dt = c(t)dt$. Ordinarily $c(t)$ may be taken to be a constant, though in full generality it may be considered to be a function of time, affected by scheduled interruptions or even by variations in the queue length. Obviously $q(t)$ cannot be negative, and, whenever q is 0, travel through the bottleneck will be less than or equal to the capacity c. [It is physically impossible for $q(t)$ to have a substantial upward discontinuity because, if car A following car B through the bottleneck had in fact been waiting substantially longer than car B, it would have been ahead of car B in the queue. On the other hand, it is possible to have a gap in the traffic approaching the bottleneck so that car A that arrived just after the gap would have a substantially shorter wait than car B that just preceded the gap. In this latter case, however, car B would have an incentive to delay his departure so as to approach the bottleneck just ahead of A.]

To proceed with the solution of the problem, we need to know the distribution of the functions V_i and W_i. One fairly simple class of cases that have been partially discussed elsewhere (1) is that in which travelers differ only in the time at which they would wish to pass the bottleneck in the absence of queue or price constraints. They have 1 constant value of time at their home base, so that $V_i = bt$ for all i, and 2 constant values of time at their destination, before and after the desired arrival time respectively, $W_i = et$ for $t \le s_i$, the desired time of arrival, and $W_i = es_i + f(t - s_i)$ for $t \ge s_i$. There are 2 polar cases to consider: the case of no time variation in the toll, $p' = 0$, so that the entire adjustment takes place in q, and the case of optimum pricing and no queue, with the price adjusted so as to accomplish this. In each case it is fairly easy to see that there will be in effect 2 regimes: one in which p' or q' are positive and all travelers pass the bottleneck at or before the desired time and one in which p and q are declining and all travelers pass the bottleneck at or after the desired time. In the constant price case, $p' = 0$, Eq. 5 becomes

$$\left.\begin{aligned} U_i'(t) &= (1 - q')b - e \\ U_i'(t) &= (1 - q')b - f \end{aligned}\right\} \qquad (6)$$

for the 2 regimes respectively. Because the right sides of these expressions are independent of i, they cannot diverge from 0 over any finite range without causing a gap in the traffic to correspond; hence, we must put them equal to 0. This gives

$$\left.\begin{aligned} q' &= 1 - \frac{e}{b} \\ q' &= 1 - \frac{f}{b} \end{aligned}\right\} \qquad (7)$$

The result is shown in Figure 3; the queue rises at the rate $(b - e)/b$ from the onset of the queue at time $t = m$ to the maximum queue at time $t = z$ and then falls at the rate $(f - b)/b$ to the extinction of the queue at $t = n$. Because there must be no discontinuity in the queue function at $t = z$, we must have

$$\left.\begin{aligned} \frac{b - e}{b}(z - m) &= c_m = \frac{f - b}{b}(n - z) \\ \frac{n - z}{z - m} &= \frac{b - e}{f - b} \end{aligned}\right\} \qquad (8)$$

Let $N(t)$ be the number of trips having desired times $s_i \le t$ and $C(t)$ be the total capacity of the facility up to time t. It is then fairly obvious that the total capacity from m to z and also from z to n must equal the number of travelers in this period.

$$\left.\begin{aligned} C(z) - C(m) &= N(z) - N(m) \\ C(n) - C(z) &= N(n) - N(z) \end{aligned}\right\} \qquad (9)$$

The 3 conditions in Eqs. 8 and 9 are sufficient to determine m, z, and n, and the solution is complete.

For the no-queue case, Eq. 5 becomes

$$\left.\begin{aligned} U_1'(t) &= b - e - p' \\ U_1'(t) &= b - f - p' \end{aligned}\right\} \tag{10}$$

Putting these expressions equal to 0 gives $p' = b - e$ and $p' = b - f$ respectively, or in both cases $p' = bq'$. The solution is thus exactly parallel to the solution in the queuing case; the difference is that, instead of leaving home at time $t - q$ and waiting in the queue for a time q, the individual leaves home at time t, encounters no waiting, but instead pays a toll $p = bq$ (in addition to whatever constant toll p_0 may be in effect in off-peak hours), which is the equivalent of the time gained at his home at the value b. Elimination of the queue by means of the variable toll thus leaves the user exactly as well off as before, the revenue from the toll being clear gain obtained for public purposes at no cost to the users.

In this case, the resulting queue and toll rate pattern over the period of capacity flow are entirely independent of the way in which the desired travel times are distributed between m and n, provided only that

$$\left.\begin{aligned} N(z) - N(t) &\geq C(z) - C(t) \\ N(t) - N(z) &\geq C(t) - C(t) \end{aligned}\right\} \tag{11}$$

for all t such that $m \leq t \leq z$ and for all t such that $z \leq t \leq n$ respectively. This ensures that every user is pushed away from z, rather than toward z, as is required by the slopes of the q and p functions. Thus, there is no necessary relation between the time of peak demand and the time of peak queue or peak price, except that the peak demand must perforce lie somewhere in the interval (m, n).

The corresponding case of the evening peak is somewhat more complex and not quite symmetrical. If we let the value of time at the office after working hours be e, the same as time before working hours, then instead of Eq. 6 we have

$$\left.\begin{aligned} U_1'(t) &= (1 - q')f - b \\ U_1'(t) &= (1 - q')e - b \end{aligned}\right\} \tag{12}$$

If now we put n^*, m^*, and z^* for the times of onset, end, and peak of the queue as before, we now have

$$\left.\begin{aligned} q' &= (f - b)/f \\ q' &= (e - b)/e \end{aligned}\right\} \tag{13}$$

as the growth and decline rates of the queue, the relative duration of the growth and decline periods being

$$\frac{n^* - z^*}{z^* - m^*} = \frac{(f - b)e}{(b - e)f} \tag{14}$$

On the other hand, if we consider the alternative of pricing to eliminate the queue, then instead of Eq. 10 we have

$$\left.\begin{aligned} U_1'(t) &= f - b - p' \\ U_1'(t) &= c - b - p' \end{aligned}\right\} \tag{15}$$

which gives $p' = f - b$ and $p' = c - b$ as the growth and decline rates of the variable toll. The relative duration of the growth and decline rates,

$$\frac{n - z}{z - m} = \frac{f - b}{b - c} \tag{16}$$

is symmetrical with Eq. 8 but differs significantly from Eq. 14.

In this case, comparing the results from queuing with those from optimal pricing requires more specificity about the parameters. Suppose, for example, a bottleneck has a capacity of 3,000 cars per hour and the desired times of passage are distributed at a steady level of 2,000 cars per hour between 4:00 and 6:30, except that demand is at 6,000 cars per hour between 5:00 and 5:30. Let the values of time be f = 4 cents/min, b = 2 cents/min, and e = 1 cent/min so that $q' = \frac{1}{2}$ and -1 and $p' = 2$ and -1 cent/min on the up and down sides respectively. Then under queuing, the queuing begins at $m^* = 4{:}00$, increases during the next 80 min to a maximum of a 40-min wait for cars leaving the bottleneck at $z^* = 5{:}20$, and thereafter diminishes to 0 at $n^* = 6{:}00$. Under pricing, the toll begins to rise at m = 4:30, reaches a maximum of 80 cents above the off-peak level at z = 5:10, and falls off again to the off-peak level at n = 6:30.

In this instance those passing the bottleneck before 5:17 will gain and the others will lose. Those with desired travel times s_i between 4:00 and 4:30 will travel at their desired times at no toll under pricing, but will be faced with conditions under queuing that for them are equivalent to traveling at 4:00 with neither queue nor price. They will therefore gain an average of the equivalent of 15 min of time at the office at the value f = 4 rather than at home at b = 2, or a total of 15(4 - 2) 1,000 = $300.

Those making 2,000 trips with desired travel times between 4:30 and 5:10 will be faced with conditions under queuing equivalent to making the trip freely at 4:00 as compared with conditions under pricing equivalent to making the trip freely at 4:30, producing a gain for these trips of (30)(4 cents - 2 cents)(2,000) = $1,200. Similarly, the loss is (15)(2 - 1)(1,000) = $150 for trips with desired times between 6:00 and 6:30 and is 30(2 - 1)(2,000) = $600 for trips with desired travel times between 5:20 and 6:00.

For those with desired trip times between 5:10 and 5:20, the difference ranges from a gain of 60 cents per trip at 5:10 through no change at 5:17 to a loss of 30 cents per trip at 5:20. The 667 trips with desired times between 5:10 and 5:17 gain 30 cents on the average for a total gain of $100, and the 333 trips between 5:17 and 5:20 lose 15 cents on the average for a total loss of $50.

Thus, in the aggregate, 3,667 trips gain a total of $1,600 and 3,333 trips lose a total of $800 for a net gain of $900. This is over and above the revenues from the tolls during the 2-hour period from 4:30 to 6:30, involving 6,000 trips at an average toll of 40 cents or $2,400.

If the same parameters were applied to the morning case, the results under queuing would be the same, and the pricing pattern would follow the queuing pattern with the same toll revenues of $2,400, but no gains or losses aside from these tolls. This is a reflection of one of the many asymmetries between morning and evening rush hours that often escape attention (Fig. 3).

VARIATIONS IN TIME VALUE AMONG INDIVIDUALS

One simple way in which the above model can be extended to cover variations in the value of time among individuals is to postulate that in addition to variations in s_i, the desired time of travel, there may be variations in the coefficients b, e, and f. One not too unreasonable simplifying assumption at this point is that these coefficients vary proportionally from one individual to another so that we can write $b_i = k_i b$, $e_i = k_i e$, and $f_i = k_i f$. If this is done, the factors k_i cancel out in the expressions for q', and the solution for the queuing case remains unchanged.

More specifically, let us take the morning case of the above model but consider the users divided into 5 time-value classes as follows: k = 0.5, 20 percent; k = 1.0, 50 percent; k = 2, 20 percent; and k = 5, 10 percent. If the peak desired travel times are at a level of 6,000/hour between 8:00 and 8:30 and 2,000/hour at other times between

7:00 and 9:00, with b = 2, e = 1, and f = 4, the queuing results are for the queue to begin to build up at 7:00, reach a peak of 40 min at 8:20, and decline thereafter to 0 at 9:00. Those with desired travel times before 8:20 will be on the margin of indifference as between traveling so as to pass the bottleneck at any time between 7:00 and their desired travel times; the rest will be on the margin of indifference as between traveling at any time between their desired travel time and 9:00.

Under pricing, the equilibrium results will be that those with $k > 1$ will travel at their desired times, the price differential being insufficient to shift them and the capacity of the facility being sufficient to accommodate them. Those with $k = 1$ and desired travel times between 7:34 and 8:43 will fill up the capacity in this interval not used by those with $k > 1$; their relative values of time will determine the values of p' over this range that are required to make this pattern acceptable to them in preference to any other. From Eq. 10 we have $p' = 1$ cent/min and -2 cents/min for the periods before and after 8:20. Those with $k = 0.5$ will travel between 7:00 and 7:34 and between 8:43 and 9:00, and the corresponding values of p' for this range are $p' = 0.5$ cent/min and $p' = 1$ cent/min. The toll thus rises from 0 (relative to the constant off-peak toll, if any) to 17 cents at 7:34, to 63 cents at 8:20, to 17 cents at 8:43, and to 0 at 9:00.

Those with the minimum value of k, in this case $k = 0.5$, will still be as well off as before because they will still be on the margin of indifference as between traveling at the time they actually do and traveling at either 7:00 or 9:00 with neither toll nor queue. The gains for the others are given in Table 1. This entire gain results from the elimination of the queuing time and is independent of the amount of shifting in travel time that may occur in terms of time of passing the bottleneck. Indeed, given any pattern of travel that is in equilibrium under the optimal pricing structure, in this case this same pattern of travel would be in equilibrium under queuing.

VARIABLE TIME VALUES

The rather remarkable results of the above models are traceable in part to the fact that the various values of time used, b, e, and f, were stipulated to be constant, and the time value function $W_i(t)$ has a sharp corner at s_i, being otherwise composed of straight-line segments. This in general implies that, if an individual is willing for a payment of p to shift his time of travel in a direction away from s_i by an amount h, he would then also be willing to shift by 2h for a payment of 2p and so on. This is not inconsistent with the way time value is commonly treated in transportation models, but it is a bit extreme. In particular, given the residual uncertainty in the time required to travel from the bottleneck point to the ultimate destination, it would be more realistic at least to provide for the function $W_i(t)$ to have a smooth curve at s_i rather than a corner so as in effect to allow for an increasing marginal disutility of shifting travel time 1 minute as one goes farther and farther from the preferred time s_i.

One simple way to do this is to write

$$W_i(t) = bt - a - h(t - s_i)^2 \tag{17}$$

Then Eq. 5 becomes

$$U_i'(t) = (1 - q')\,b - b + 2h(t - s_i) - p' \tag{18}$$

Individual users will use the facility at capacity level but in sequence according to their respective values of s_i, with $t = s$ at the beginning, at the end, and at some point in the middle of the period of capacity flow.

If one applies this value of time formulation to the morning pattern of traffic demand, which can be slightly generalized to a density of desired travel times of $x_p c$, $x_p > 1$ during the peak period from s_1 to s_2, $x_a c$, $x_a < 1$ prior to s_1, and $x_b c$, $x_b < 1$ after s_2, then for s in the range $s_1 \le s \le s_2$ we will have

$$t - t_p = x_p(s - s_p) \tag{19}$$

where $t_p = s_p$ is the time dividing those who travel before their desired time from those

Figure 1.

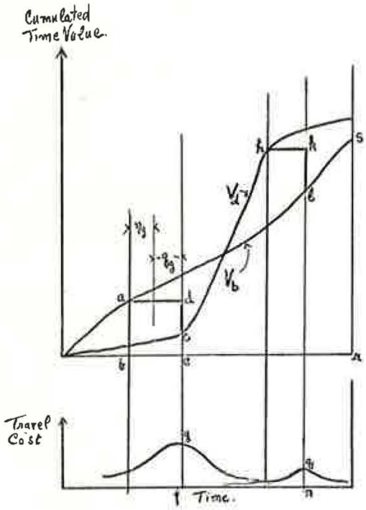

Figure 2.

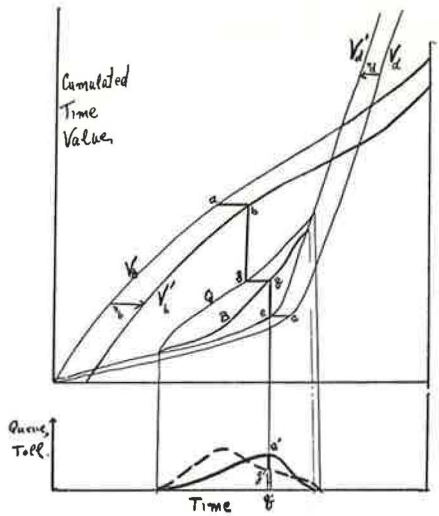

Figure 3.

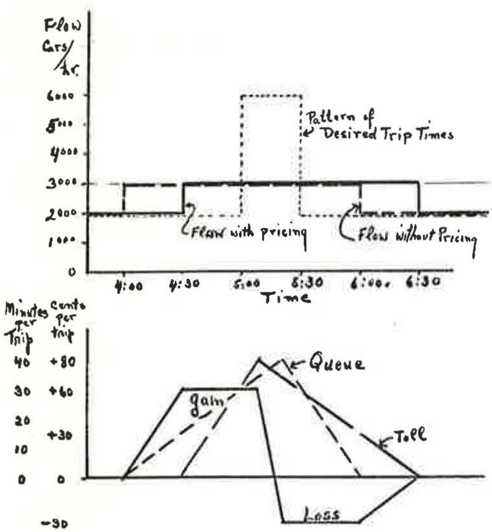

Table 1.

Time Intervals	Range of Price (cents)	k	Value of Queuing Time (cents) Range	Average	Average Net Gain per Trip (cents)	Number of Trips	Total Gain (dollars)
7:00 to 7:34 and 8:43 to 9:00	0 to 17	0.5	0 to 17	8.5	0.0	1,200	0.00
		1.0	0 to 34	17.0	8.5	840	71.40
		2.0	0 to 68	34.0	25.5	340	86.70
		5.0	0 to 170	85.0	76.5	170	130.05
7:34 to 8:00 and 8:30 to 8:43	17 to 43	1.0	34 to 60	47.0	17.0	1,560	265.20
		2.0	68 to 120	94.0	64.0	260	166.40
		5.0	170 to 300	235.0	205.0	130	266.50
8:00 to 8:30	43 to 63	1.0	60 to 80	70.0	17.0	600	102.00
		2.0	120 to 160	140.0	87.0	600	522.00
		5.0	300 to 400	350.0	297.0	300	891.00
Total						6,000	3,301.25
Toll revenues							1,596.75
Total efficiency gain							4,898.00

traveling after. Similarly, we have

$$\left.\begin{aligned} t - t_1 &= x_a(s - s_1) \\ t - t_2 &= x_b(s - s_2) \end{aligned}\right\} \quad (20)$$

for $s < s_1$ and for $s > s_2$ respectively, where t_1 and t_2 are the times at which those with desired travel times of s_1 and s_2 actually travel (in terms of time of emerging from the bottleneck, as always). If now we put Eq. 18 equal to 0 with $p' = 0$, we get

$$q' = -\frac{2h}{b}(t - s) \quad (21)$$

Putting Eq. 19 into Eq. 21, we get, for $t_1 < t < t_2$,

$$q' = -\frac{2h}{b}\left(1 - \frac{1}{x_p}\right)(t - t_p) \quad (22)$$

and integrating, we get

$$q = q_p - \frac{h}{b}\left(1 - \frac{1}{x_p}\right)(t - t_p)^2 \quad (23)$$

where q_p is the constant of integration determined by the length of the queue at t_p.

To continue for the values of s outside the range $(s_1 s_2)$, we obtain from putting $s = s_1$ and $s = s_2$ in Eq. 8

$$\left.\begin{aligned} t_1 &= t_p + x_p(s_1 - s_p) \\ t_2 &= t_p + x_p(s_2 - s_p) \end{aligned}\right\} \quad (24)$$

Using Eq. 24 in Eq. 20 and putting $t_o = s_o$ and $t_e = s_e$, we have

$$\left.\begin{aligned} t_o &= s_1 - (t_p - s_1)\frac{(x_p - 1)}{(1 - x_a)} \\ t_e &= s_2 + (s_2 - t_p)\frac{(x_p - 1)}{(1 - x_b)} \end{aligned}\right\} \quad (25)$$

for the times of the beginning and end of the queue respectively. Equation 20 can now be rewritten in terms of deviations from t_o, s_o and from t_e, s_e, as follows:

$$\left.\begin{aligned} t - t_o &= x_a(s - s_o) \\ t - t_e &= x_b(s - s_e) \end{aligned}\right\} \quad (26)$$

Putting Eq. 26 into Eq. 22 we get

$$\left.\begin{aligned} q' &= \frac{2h}{b}\frac{(1 - x_a)}{x_a}(t - t_o) \\ q' &= \frac{2h}{b}\frac{(1 - x_b)}{x_b}(t_e - t) \end{aligned}\right\} \quad (27)$$

and

$$\left.\begin{aligned} q &= \frac{h}{b}[(1/x_a) - 1](t - t_o)^2 \\ q &= (h/b)[(1/x_a) - 1](t_e - t)^2 \end{aligned}\right\} \quad (28)$$

The constant of integration is 0 because $q = 0$ at $t = t_e$ and $t = t_o$.

The remaining unknowns q_p and t_p can now be determined by equating the values of $q(t_1)$ and $q(t_2)$ given by Eqs. 23 and 27 because $q(t)$ must be continuous, giving

$$t_p = (1 - A)s_1 + As_2 \tag{29}$$

where $A = \dfrac{1}{1 + \left[\dfrac{(x_p - x_a)(1 - x_b)}{(x_p - x_b)(1 - x_a)}\right]^{1/2}}$

In the particular example at hand, we have $x_a = x_b = 2/3$ so that $A = 1/2$ and $t_p = 8{:}15$. Also, $x_p = 2$ so that $t_1 = 8{:}45$, $t_2 = 7{:}45$, $t_o = 7{:}15$, and $t_e = 9{:}15$. Then the length of the queue at each point of time will be given by putting these values in Eqs. 23 and 28: $q = (h/2b)(t - 7{:}15)^2$ for $7{:}15 \le t \le 7{:}45$ and $q(7{:}45) = 450(h/b)$; $q = (h/2b)[1{,}800 - (t - 8{:}15)^2]$ for $7{:}45 \le t \le 8{:}45$ and $q_p = 900(h/b)$; and $q = (h/2b)(9{:}15 - t)^2$ for $8{:}45 \le t \le 9{:}15$. The pattern of the toll required to eliminate the queue is in this case exactly similar, again, with $p = bq$; the net overall gain is equal to the toll revenues, and users are left exactly as well off as before.

INTERPERSONAL VARIATION IN TIME VALUES WITH FINITE ELASTICITY

The problem becomes considerably more interesting and complex if in addition to an increasing marginal value of time for individuals we consider differences among individuals in the overall value of time. This can be done fairly simply by putting $b_j = k_j b$ and $h_j = k_j h$. As before, we can allow k_j to take on the values 0.5, 1.0, 2.0, and 5.0 for 20, 50, 20, and 10 percent of the users having each of the desired travel times s_i.

In the queuing alternative, the results are unchanged from the preceding case because the ratio h/b remains unchanged. For the pricing case, the results are different, however. Instead of Eq. 18, we now have, for $q' = 0$ and $U_i'(t) = 0$,

$$p'(t) = 2k_j h[t - s_j(t)] \tag{30}$$

for $j = 1, 2, 3, 4$. This implies

$$0.5[t - s_1(t)] = t - s_2(t) = 2.0[t - s_3(t)] = 5.0[t - s_4(t)] \tag{31}$$

We also have

$$\frac{dC(t)}{dt} = c(t) = \sum_j y_j(s_j)\frac{d}{dt}s_j(t) \tag{32}$$

as the condition that capacity is just fully utilized between t and t + dt, where $y_j(s_j)$ is the density of desired travel times at time s_j for the j th class of users.

If we apply this to the piecewise constant density case and put t_{24} for the time at which users in class 4 with desired travel times at $s_2 = 8{:}30$ will be traveling, then each y_j will be constant for those actually traveling between $t_p = 8{:}15$, and t_{24}, in fact, will take on the values 20, 50, 20, and 10 for $s_j < 8{:}30$ and $^{20}/_3$, $^{50}/_3$, $^{20}/_3$, and $^{10}/_3$ for $s_j \ge 8{:}30$. Also by differentiating Eq. 31 and using this in Eq. 30, we have

$$c = \sum_j y_j(s_j)\left\{1 - [1 - \frac{d}{dt}s_2(t)]/k_j\right\} \tag{33}$$

By letting successively $s_4 = 8{:}30$, $s_3 = 8{:}30$, $s_2 = 8{:}30$, and $s_1 = 8{:}30$, we obtain the values given in Table 2.

On line 1, the values of t are given in the even columns at which users in the various time-value classes will travel when their desired travel times are at the critical point at which the intensity of demand changes, $s_i = 15$, where all times in the table are for simplicity referred to $t_p = 8{:}15$ as the origin.

Table 2.

Line	Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12
1	t	0.0	18.32	16.63	18.19	19.75	23.24	26.73	32.02	37.30	48.65	60.00
2	s_1	0.0	0.16	0.33		0.75		3.27		15.00		60.00
3	s_2	0.0		8.47		10.25		15.00		26.15		60.00
4	s_3	0.0		12.55		15.00		20.86		31.72		60.00
5	s_4	0.0		15.00		17.85		24.38		35.07		60.00
6	z_1	0.0	8.15	16.30	17.65	19.00	21.23	23.46	22.88	22.30	11.15	0.00
7	z_2	0.0		8.15		9.50		11.73		11.15		0.00
8	z_3	0.0		4.08		4.75		5.87		5.56		0.00
9	z_4	0.0		1.63		1.90		2.35		2.23		0.00
10	p'	0.0	0.5435	1.087	1.777	1.267	1.415	1.564	1.525	1.487	0.7435	0.00
11	p'		0.2717		1.032		1.341		1.545		0.3717	
12	Δt		16.63		3.12		6.98		10.57		22.70	
13	$\Delta t/2$		8.32		1.56		3.49		5.28		11.35	
14	Δp		9.04		3.67		9.88		16.12		16.87	
15	Δp		2.26		1.61		4.57		8.15		4.22	
16	p	55.59	53.33	46.55	44.99	42.88	38.31	33.00	24.84	16.88	4.22	0.00
17	z_1^2	0.00	66.42	265.69	311.52	361.00	450.71	544.98	523.49	497.29	124.32	0.00
18	Z_1	0.00	2.21	8.86	10.38	12.03	15.02	16.17	17.45	16.58	4.14	0.00
19	Z_2	0.00	1.11	4.43	5.19	6.02	7.51	9.08	8.72	8.29	2.07	0.00
20	Z_3	0.00	0.55	2.21	2.60	3.01	3.76	4.54	4.36	4.14	1.04	0.00
21	Z_4	0.00	0.22	0.89	1.04	1.20	1.50	1.02	1.74	1.66	0.42	0.00
22	C_{1p}	55.59	55.54	55.41	55.37	54.91	53.33	51.17	42.30	33.46	8.36	0.00
23	C_{2p}	55.59	54.44	50.98	50.18	48.90	45.02	42.08	33.57	25.17	6.29	0.00
24	C_{3p}	55.59	53.88	48.76	47.59	45.89	42.07	37.54	29.21	21.02	5.26	0.00
25	C_{4p}	55.59	53.55	47.44	46.03	44.08	39.81	34.82	26.59	18.54	4.64	0.00
26	N_1		6.6		8.4		50.4		234.6		300.0	600.0
27	N_2		423.5		89.0		237.5		185.8		564.2	1,500.0
28	N_3		251.0		49.0		39.1		72.4		188.5	600.0
29	N_4		150.0		9.5		21.8		35.6		83.1	300.0
30	ΣN		831.1		155.9		348.8		528.4		1,135.8	3,000.0
31	t_1	0.00	0.33	0.66	1.80	1.50	4.02	6.54	18.27	30.00	45.00	60.00
32	t_2	0.00	8.47	16.94	18.72	20.50	25.25	30.00	33.72	37.44	48.72	60.00
33	t_3	0.00	12.55	25.10	27.55	30.00	31.95	33.90	37.52	41.15	50.58	60.00
34	t_4	0.00	15.00	30.00	30.90	31.80	34.03	36.26	39.80	43.35	51.68	60.00
35	q_1	30.00	30.00	29.99	29.98	29.96	29.73	29.29	24.45	15.00	3.75	0.00
36	q_2	30.00	28.80	25.23	24.15	23.00	19.36	15.00	11.51	8.47	2.12	0.00
37	q_3	30.00	27.38	19.51	17.32	15.00	13.11	11.33	8.43	5.92	1.58	0.00
38	q_4	30.00	26.25	15.00	14.17	13.24	11.25	9.40	6.82	4.62	1.16	0.00
39	Q_1	30.00	30.00	29.99	29.98	29.96	29.73	29.29	24.45	15.00	3.75	0.00
40	Q_2	60.00	57.60	50.46	48.30	46.00	38.72	30.00	23.02	16.94	4.24	0.00
41	Q_3	120.00	109.52	78.04	69.28	60.00	52.44	45.32	33.72	23.68	5.92	0.00
42	Q_4	300.00	262.50	150.00	141.70	132.40	112.50	94.00	68.20	46.20	11.55	0.00
43	Z'_1	0.00	0.00	0.00	0.01	0.02	0.14	0.36	2.78	7.50	1.88	0.00
44	Z'_2	0.00	1.20	4.77	5.85	7.00	10.64	15.00	11.51	8.47	2.12	0.00
45	Z'_3	0.00	5.24	21.02	25.36	30.00	26.22	22.66	16.86	11.84	2.96	0.00
46	Z'_4	0.00	18.75	75.00	70.85	66.20	56.25	47.00	34.10	23.10	5.78	0.00
47	C_{1q}	30.00	30.00	29.99	29.99	29.98	29.87	29.65	27.23	22.50	5.63	0.00
48	C_{2q}	60.00	58.80	55.23	54.65	53.00	49.36	45.00	34.53	25.41	6.36	0.00
49	C_{3q}	120.00	114.76	99.06	94.64	90.00	78.66	67.98	50.58	35.52	8.88	0.00
50	C_{4q}	300.00	281.25	225.00	212.55	198.60	168.75	141.00	102.30	69.30	17.33	0.00
51	g_1	-25.59	-25.54	-25.42	-25.38	-24.93	-23.46	-21.52	-15.07	-10.96	-2.73	0.00
52	g_2	+4.41	4.36	4.25	4.47	4.10	3.54	2.92	0.96	0.24	0.07	0.00
53	g_3	64.41	60.88	50.30	47.05	44.11	36.59	30.44	21.37	14.50	3.62	0.00
54	g_4	244.41	227.70	177.56	166.52	154.52	128.94	106.18	75.71	50.76	12.69	0.00
55	$\bar{g}_1$		-25.53		-25.31		122.38		-15.44		-3.65	
56	$\bar{g}_2$		+4.35		4.37		3.53		1.17		0.08	
57	$\bar{g}_3$		59.70		47.10		36.82		21.57		4.83	
58	$\bar{g}_4$		222.13		166.36		129.41		76.63		16.92	
59	G_1		-1.68		-2.13		-11.28		-36.34		-10.95	-62.38
60	G_2		+18.42		3.89		8.38		2.17		0.45	+33.31
61	G_3		149.85		23.08		14.40		15.62		9.10	212.05
62	G_4		333.20		15.80		28.21		27.28		14.06	418.55
63	Total user gains											+663.91
64	Total user losses											-62.38
65	User net gains											601.53
66	R		436.99		70.00		133.20		131.46		63.95	947.04
67	Total community gain											1,548.57
68	$\bar{p}$		52.58		44.90		38.19		24.88		5.63	

Lines 2 through 5 give as s_i the values of the desired travel times for the members of the various classes who will actually travel at the time in line 1.

Lines 6 through 9 give as z_i the number of minutes by which the actual travel time deviates from the desired travel time as a result of the price variation, p', given in line 10. z_i, s_i, and p' are piecewise linear functions of t, with corner points at the values of t in the even numbered columns. The odd numbered columns give the values for the midpoints of these linear segments. The z_i are inversely proportional to k_i in each column, as is required by the relative effectiveness of the p' at a given point in pulling them away from their most desired travel time. Values of p' are also given in line 11 for a quartile point of each line segment, to facilitate the calculation of p at the midpoints.

Lines 12 and 13 give the length of each of the intervals and half intervals. Multiplying by the corresponding midpoint values of p' gives the increments of p in lines 14 and 15, which are cumulated to give the price p at each instant of time as given in line 16.

Lines 18 through 21 give $Z_i = k_i h z_i^2$, the value of the loss per trip due to travel at a time other than the desired time, and adding this to p gives the total loss of satisfaction per trip as compared to making the trip without delay at the most desired time without toll.

Lines 26 through 30 give the total number of trips in each time-value class made in the intervals between the times given in the even columns.

For the situation without tolls but with a queue, lines 31 through 34 give the times the trips that would have been made at the times in line 1 with pricing and that have the desired travel times given in lines 2 through 5 will be made under queuing conditions, lines 35 through 38 give the length of the corresponding queues, lines 39 through 42 give the value of the loss of time in the queue, and lines 43 through 46 give the loss due to traveling at times other than the desired time by reason of the queue. Added together, these comprise the total cost of traveling under queuing conditions, again relative to the no-toll and no-queue situation. Comparison of this with the cost under pricing gives the gain from pricing given in lines 51 through 54.

This gain from pricing is a quadratic function of time over the intervals between even-numbered columns, and the mean gain $\bar{g}_i$ for each such interval can be obtained by Simpson's rule: $\bar{g} = [g(t_1) + 4g(t_{1.5}) = g(t_2)]/6$. The results are given in lines 55 through 58. Multiplying by the number of trips in each category gives the aggregate gains and losses in lines 59 through 62. An average price for each period can also be obtained by Simpson's rule, as given in line 68. Applying this to the total traffic in each period gives the total revenues in line 66. The total net community gain of $1,548.57 given in line 67 is thus the gain for each peak period. At 650 peak periods a year, the total gain could well be evaluated at $1,000,000 per year.

METERING VERSUS PRICING

The above examples involved a facility with an exogenously fixed capacity. In many cases, of course, traffic flows at near capacity levels are highly unstable, for example, at a "chokepoint" where traffic approaching a bottleneck is backed up (and frequently traffic flow through the bottleneck is 5 to 10 percent below the maximum flow level) and a "triggerneck" where a queue backed up from one bottleneck gets in the way of traffic headed in another direction and thus creates a bottleneck where none existed before.

One approach to such sensitive congestion spots is to limit by metering signals or other means the rate of access of vehicles to that which can be adequately dealt with. The objective is to control the accumulation of traffic in such areas to a level compatible with maximum flow, or perhaps with a flow somewhat less than maximum at a more satisfactory rate of speed. Thus, in the off-ramp example specified above, if the main on-ramps upstream from the critical off-ramp are provided with metering signals arranged to limit the total flow on the expressway at the off-ramp point to 4,000 vehicles per hour, then the setting up of the triggerneck situation can be avoided, and the flow can be maintained at 4,000 vehicles per hour.

The pricing approach, on the other hand, would be to set a price for the use of the off-ramp sufficiently high to ensure that those desiring to use it would not exceed 600 per hour for any appreciable length of time and thus ensure that the capacity of 5,000 is maintained. Because pricing, at least of the electronic variety, does not require significant physical space and can be established at any point where traffic conditions warrant, it can be much more flexible and can bring traffic flows much closer to the full capacity of the system than is possible with metering.

More fundamentally, metering, even under simple conditions where it can in principle achieve as full utilization of capacity as can pricing, imposes at the metering point delays that represent a net waste of resources. The deterrent or rationing provided by pricing, however, is a transfer: The deterrent to the motorist is in terms of giving up purchasing power, and the resources not consumed can be used for public purposes. Alternatively, the revenues derived from the roadway pricing can be used to reduce other taxes. Metering does not eliminate or even reduce queuing, except to the relatively minor extent that capacity flows are increased, but merely transfers the queuing and delay from locations where the queue may have triggerneck effects to locations where the accumulation of the queues has less serious effects on the traffic flows. Indeed, where the metering queue is remote from the bottleneck point, metering may result in some traffic being delayed that would not otherwise have been held up in that it is destined for points other than those involving the bottleneck.

METERING, PRICING, AND TRANSIT

In many situations, if transit vehicles can operate without delay over free-flowing expressways and connecting local streets, the service that can be provided may be as convenient as or even more convenient and far less costly than rapid transit service provided on a separate right-of-way for substantial portions of the trip. One method that has been proposed for achieving this has been to preserve free-flowing conditions on expressways by means of metering at a sufficient number of on-ramps while providing for transit vehicles separate ramps that will afford them access to the expressway without delay (2).

The pricing alternative is to set up pricing at a level that will keep not only the expressways but also the local streets reasonably free flowing and operate transit vehicles in the general traffic stream, except possibly at bus stops, transfer points, and terminals. Although this does require taking on the fairly substantial overhead costs associated with the establishment of a sophisticated pricing system, it eliminates several fairly serious difficulties and inefficiencies that are associated with the metering system and are beyond the scope of this paper. Moreover, there is danger that, in setting up a metering system, considerable investment may be committed to a form of traffic control destined to become obsolete in a fairly brief period.

REFERENCES

1. Vickrey, W. Congestion Theory and Transport Investment. American Economic Review, May 1969, pp. 251-260.
2. Kain, John F. How to Improve Urban Transportation at Practically No Cost. Harvard Univ., Cambridge, Mass., May 1970.

EFFECTING CHANGE IN PUBLIC POLICY: FINANCING URBAN TRANSPORTATION IN THE NEW YORK, NEW JERSEY, AND CONNECTICUT REGION

Ernest Kurnow and Richard P. Brief, New York University; and
Irwin H. Silberman, Irwin H. Silberman and Associates, Inc.

In this study an "enterprise" approach was adopted that would be applicable to the analysis of problems relating to financing urban transportation systems in any metropolitan area. This approach to accounting for regional transportation systems reflects the view that an overall, coordinated effort is required to obtain a balanced transit network in metropolitan regions. Both the methodology and the statistical findings should be of interest to transportation specialists and to those involved in public administration. In addition, this paper sheds some light on the more general process of influencing public policy. It provides an 8-year history of both successful and unsuccessful efforts to achieve change in transportation financing in the New York metropolitan region.

•THIS STUDY (1, 2) was undertaken to provide basic financial data and to make recommendations to the Tri-State Transportation Commission (now the Tri-State Regional Planning Commission) for its development of immediate-action and long-range programs for financing transportation services in the New York metropolitan region. Intraurban facilities and related activities to be financed in the region include streets, roads and highways, bridges, tunnels, commuter railroads, buses, public transit, street cleaning, parking, lighting, sidewalks, storm sewers, and traffic and highway police. The study does not include facilities devoted solely to interregional travel or to the shipment of goods.

We adopted an "enterprise" approach that is applicable to the study of financing urban transportation systems in any metropolitan area. This approach to accounting for regional transportation systems reflects the view that an overall, coordinated effort is required to obtain a balanced transit network in metropolitan regions. Thus, both the methodology and the statistical findings should be of interest to transportation specialists and to those involved in public administration. [There is, in fact, a paucity of comprehensive data relating to the sources and uses of funds for transportation in metropolitan areas. Some of these sources are cited (3, 4, 5).] In addition, the study sheds some light on the more general process of influencing public policy. It provides an 8-year history of both successful and unsuccessful efforts to achieve change in transportation financing in the New York metropolitan region.

We present in this paper an analysis of the sources of transportation funds, the expenditures on transportation, and the difference between the sources of funds and expenditures. We elaborate on our general approach to the study of financing transportation and the implications of the specific findings for public policy. We then summarize the study recommendations, the institutional constraints that influenced their implementation, and subsequent events.

SOURCES OF TRANSPORTATION FUNDS

The Tri-State region generated $2.6 billion in 1962 to finance transportation (Table 1).

The 3 major sources of funds are classified as direct users, indirect users, and borrowing.[1]

Direct users are those who contribute to the transportation system in direct relation to the use they make of transportation facilities. Their payments for transportation services are direct and their benefits are also direct. All funds provided by direct users are recorded at the level of government that initially receives the funds. The following funds are provided by direct users.

1. User taxes include federal and state motor fuel taxes, motor vehicle license fees, and other motor vehicle-related excise taxes. These taxes are viewed as direct payments for the use of transportation services and, insofar as possible, the allocation procedures attempted to estimate the user charges actually earned in the region. For example, motor vehicle taxes were allocated to the region on the basis of gasoline sales, an index that would be expected to reflect the amount of mileage driven and, therefore, the user taxes paid by motorists for traveling on the region's roads and highways.

2. User charges consist of fares, tolls, and parking fees paid to local governments for the use of transportation facilities.

3. Related income consists predominantly of income from concessions directly associated with the operation of some transport facilities. Interest on investments is also included in this category of funds.

Indirect users include persons, firms, and entire communities—all deriving an indirect benefit from the transport facilities. The contributions of indirect users usually do not bear a direct relation to the use they make of transportation facilities. These contributions are considered to reflect the payments for the social benefits received by localities from the transportation system. The present level of economic and social activity in the region would, of course, be impossible without the vast transportation system entwined throughout the region, and those who may not actually use the facilities or those who make less than average use of them nevertheless benefit from the system. In the absence of a more precise measure, the statistical analysis assumes that the following sources of funds represent the contributions of indirect users.

1. General fund appropriations for transportation services in excess of local government user charges and related income are included in indirect user funds. Local user charges and transportation-related income of local governments are considered as payments by direct users. However, most local governments in the region channel such payments by direct users into their general funds. Therefore, these amounts are excluded from general fund appropriations in order to avoid double counting. The amount recorded under indirect users as general fund appropriations of local governments is the net amount of these direct revenue sources.

2. Special assessments, theoretically, are generally imposed in rough proportion to benefits directly received by the property owner. However, actual methods of assessment vary widely among local governments in the region, and many governments report only one total for special assessments and real property tax yields. Therefore, special assessments are included in measuring the contributions by indirect users. However, this practice tends to understate the contributions of direct users.

A resident of the region, then, contributes to the transportation system in a dual capacity. For example, each time he registers his automobile or purchases gasoline, he contributes as a direct user. Each time he pays his real property tax or a sales tax, he contributes as an indirect user to the extent that these payments may be allocated to transportation.

This system of classifying funds and the methods for allocating them to the region have the advantage of giving a clear picture of the flow of funds generated by the transportation system. It offers a more meaningful basis for evaluating financial policy

[1] The original manuscript of this paper included an appendix that explains the methods of allocation and sources of data for Tables 1 and 2. The appendix is available in Xerox form at the cost of reproduction and handling from the Highway Research Board. When ordering refer to XS-48, Highway Research Record 476.

than a system that restricts sources of funds to those that are actually made available for transportation purposes. Even though certain user payments flow directly into general funds and are then appropriated "back" to transportation, either wholly or in part, it is essential to provide a measure of the total contribution of the users to the region's transportation system.

This approach is at variance with the view that all user taxes returned to a region by the federal and state governments constitute subsidies to its transportation system. It is reasoned that, if user taxes and charges are earned in a region, calling them a subsidy when they are returned is inappropriate. Transportation in a region is considered subsidized only to the extent that federal and state expenditures exceed user taxes earned there. Any particular region may, therefore, have a surplus or deficit of user taxes over federal and state expenditures on transportation.

Although transfer payments among individuals and governments are used to promote equity for all regions, it is important to identify the sources of all such transfer payments. If it is found that the donor is ailing, the transfer payment may not be justified on equity grounds; in fact, it may make no economic sense at all.

Out of the total of $2,579.4 million raised to finance transportation, direct users provided 65.6 percent, indirect users provided 16.9 percent, and borrowing accounted for 17.5 percent. As the source of funds, government roads and highways accounted for 46.9 percent, transit for 24.5 percent, authority toll roads, bridges, and tunnels for 17.0 percent, private bus companies for 6.0 percent, and commuter railroads for 5.6 percent.

EXPENDITURES FOR TRANSPORTATION

More than $2 billion was expended in 1962 to operate, maintain, and expand transportation facilities in the region (Table 2). Operation and maintenance of facilities accounted for 51.3 percent, debt service for 14.7 percent, and capital outlay for 34.0 percent.

In addition to expenditures made by local governments in the region, maintenance includes money spent by the state in the region on state highways including maintenance of condition, snow removal, traffic services, general administration expense for highway-related items of appropriate state departments, and budget appropriations made to the region for highway purposes. Similarly, capital outlay expenditures include money from local, state, and federal governments for right-of-way, engineering, and construction expenses. Thus, the accounting system used here considers all capital outlays in the region financed by the states and federal governments as user taxes that are returned to the region. Because such capital outlays vary substantially from year to year, an attempt was made to ascertain whether the 1962 figures are representative. It was found that capital outlays financed with federal and state user taxes in 1962 overstated average annual expenditures for such purposes. To this extent, the estimate of a typical "transportation surplus" is understated.

In New York State, the average amount ($166.4 million) of contracts let in the region for the 3-year period from 1961 to 1963 was used as the estimate of construction expenditures, the major component of capital outlay. The remaining expenditures for capital outlay in New York include $33.8 million for right-of-way and $8.1 million for engineering.

This figure is apparently high because it reflects the atypical $230 million of contracts let in 1961, primarily for highways relating to the World's Fair. For example, the 6-year average for the period from 1958 to 1963 for contracts let was only $130.9 million. As a further test, both the Tri-State Transportation Commission and the New York State Department of Public Works made estimates of capital expenditures on the basis of work completed during the year. The commission's estimate was approximately $170 million, and the department's estimate was $140 million. Thus, the estimate of $166.4 million for construction costs is toward the upper end of the range.

In New Jersey, the state highway department's estimate of $88.2 million was included as the capital outlay in the region. This figure compares favorably with the average of $90 million of capital outlay for the 3 fiscal years ending in 1962. In

Connecticut, the state highway department's estimate of $39.8 million was included as the capital outlay in the region. This figure is approximately $8 million above the average outlay in Connecticut for the 3 years ending in 1962.

In terms of transportation facilities, roads and highways and allied services accounted for 55.7 percent of all disbursements, public transit for 29.7 percent, commuter railroads for 7.4 percent, and private bus companies for 7.2 percent.

TRANSPORTATION SURPLUS

The data presented above indicate that, although $2,577 million was provided to finance the region's transportation "enterprise," only $2,110.4 million was applied to secure facilities and finance operations. Therefore, from the regional point of view, a transportation surplus of $467.0 million existed in 1962. The composition of this surplus is given in Table 3.

The major components of the regional surplus are $203.1 million of state and $154.9 million of federal user taxes earned on government roads and highways but not expended on transportation in the region. This basic finding concerning the disposition of user taxes is in agreement with research conducted in other metropolitan areas in the United States. User taxes earned in high-density areas are often used to finance transportation and other government services outside of the area (6, 7, 8, 9).

The seemingly profitable position of the private bus industry ($14.4 million surplus) is in part due to numerous tax concessions granted by various government agencies (10). The deficit for public transit ($10.3 million) would have been substantially larger had it been adjusted for $16.1 million provided by New York City to finance the operating deficit.

The last item given in Table 3 ($71.9 million) is a residual that results mainly from temporary differences between sources and uses of funds provided from bond issues and applied to capital construction. An illustration is the temporary increase in funds of the New York City Transit Authority as a result of bond sales that were $19 million in excess of the capital projects undertaken during the year under study. This residual is, therefore, not meaningful for this analysis and is not considered in the study. (A similar increase of $60.5 million is reflected in the operations of authority roads, tunnels, and bridges and one of $3.5 million in operations of local governments. The balance reflects a decrease in accumulated surplus of private bus companies to finance capital outlays during the year.)

An analysis of the $395.1 million current transportation operating surplus by type of facility is given in Table 4. The current operating surplus is computed on a cash-flow basis indicating the excess of current sources of funds over current expenditures. Because bond issues are viewed as noncurrent transactions, neither the funds provided by borrowing nor the capital outlay financed with borrowing is considered as a current transaction. However, debt service on borrowing is viewed as a current expenditure. Where capital outlay is financed out of user taxes, it is regarded as part of current expenditures.

POLICY IMPLICATIONS

This overall approach to the financial analysis of a region's transportation system differs from those that seek to measure the economic profit or loss earned on particular transportation facilities. Such calculations involve a great deal of conjecture with respect to the magnitude of implicit costs and benefits such as depreciation and the value of time. Furthermore, the calculation of economic profits is not of primary importance when the problem of financing transportation is studied. As a practical matter, what is required is an estimate of the system's surplus funds that might be made available to finance the region's transportation needs. This estimate is provided by the procedures outlined in this study.

If the $1,690.9 million of direct user taxes and charges earned in the region had been available to finance all of the $1,731.5 million of current disbursements, i.e., current operating expenses, debt service, and capital outlay financed with federal and state user taxes, the region's transportation system would have operated at a deficit of $40.6 million. This deficit is not distributed evenly among the various facilities. Government

Table 1. Sources of funds by type of facility in 1962.

Source of Funds	Government Roads and Highways	Authority Toll Roads, Bridges, and Tunnels	Public Transit	Commuter Railroads	Private Bus Companies	Total	Percent
Direct users							
Federal user taxes earned	375.9					375.9	
State user taxes earned	430.2					430.2	
User charges earned	22.1	202.2	331.0	122.7	153.3	831.3	32.4
Related income	7.8	23.3	9.2	10.3	2.9	53.5	2.1
Subtotal	836.0	225.5	340.2	133.0	156.2	1,690.9	65.6
Indirect users[a]	268.9	–	161.1	5.7	–	435.7	16.9
Borrowing	106.0	209.6	129.7	5.5	–	450.8	17.5
Total	1,210.9	435.1	631.0	144.2	156.2	2,577.4	100.0
Percent	46.9	17.0	24.5	5.6	6.0	100.0	

Note: Amounts are in millions of dollars. Interfacility transfers and tax relief are not included.

[a]General fund appropriations, special assessments, and other (local).

Table 2. Expenditures by type of facility in 1962.

Expenditure	Government Roads and Highways[a]	Authority Toll Roads, Bridges, and Tunnels	Public Transit	Commuter Railroads	Private Bus Companies	Total	Percent
Operating expenses	332.1[b]	67.2	392.8	149.0	141.8	1,082.9	51.3
Debt service	72.3	114.4	122.6	–[c]	–[d]	309.3	14.7
Capital outlay							
Federal government	220.8	3.0	–	–	–	223.8	10.6
State governments	115.5	—	–	–	–	115.5	5.5
Borrowed funds and cash reserves	102.5	149.1	110.7	5.5	11.1	378.9	17.9
Subtotal	438.8	152.1	110.7	5.5	11.1	718.2	34.0
Total	843.2	333.7	626.1	154.5	152.9	2,110.4	100.0
Percent	39.2	15.8	29.7	7.4	7.2	100.0	

Note: Amounts are in millions of dollars.

[a]Includes expenditures on streets, toll and nontoll highways, ferries, parking, street cleaning, lighting, sidewalks, storm sewers, and police. State expenditures for these purposes as well as costs of tax collection have been allocated to the region.

[b]Financed by state funds, $111.6 million; federal funds, $0.2 million; and local funds, $220.3 million.

[c]Not available.

[d]Interest on debt included in operating expenses.

Table 3. Transportation surplus in 1962.

Item	Amount
Government roads and highways	
Excess of federal user taxes over federal expenditures	154.9
Excess of state user taxes over state expenditures	203.1
Increase in current reserves of local governments	6.2
Subtotal	364.2
Authority toll roads, bridges, and tunnels net operating surplus	40.9
Public transit net operating deficit	-14.1
Commuter railroads net operating deficit	-10.3
Private bus companies net operating surplus	14.4
Total current operating surplus	395.1
Change in working capital due to capital transactions of public authorities, local governments, and private bus companies	71.9[a]
Total surplus	467.0

Note: Amounts are in millions of dollars.

[a]Includes $3.5-million surplus of borrowings over capital expenditures of local governments.

roads and highways, authority facilities, and private bus companies earn surpluses. The deficits of transit and commuter railroads more than offset these surpluses. Thus, transportation facilities may be classified as "winners" and "losers" and, because of the nature of the present underlying demand and the cost structure of each of these facilities, the same pattern will probably continue. However, if all the facilities are viewed as parts of a single enterprise, users alone would appear to contribute almost enough funds to finance all disbursements.

In addition, indirect users contributed $435.7 million to maintain and improve the transportation system. Of this amount, $268.9 million was provided for the maintenance of local streets and roads and the operation of allied services. The remainder, $166.8 million, was provided primarily for transit.

It is, of course, difficult to determine whether the funds provided by indirect users would have been provided to the transportation system if all direct user taxes earned in the region were returned to it. Nevertheless, it is also clear that communities should help to finance the transportation system in return for the social and economic benefits provided to the community at large.

Based on direct user charges and taxes and the assumption that continued financial support for those "externalities" or benefits accrue to the community at large, it can be stated that the transportation system more than pays for itself. In fact it appears to have "earnings" that justify expansion or improvement or both.

The major policy implications of the statistical findings are that adequate funds for financing all public transportation services in the region would be ensured if

1. All funds earned in transportation at each level of government were pooled;
2. Transportation services had priority in the use of funds earned in transportation;
3. The allocation of funds earned in transportation to a region were in rough proportion to the money earned in it; and
4. All transportation planning and administration were coordinated in one agency for the region or in a group of cooperating agencies.

IMPLEMENTATION OF RECOMMENDATIONS

A series of recommendations incorporating these policy implications were made in 1965 and are given in Table 5. These recommendations reflect the legal and institutional constraints that impinge on the decision-making process. Thus, no overall regional agency was proposed because of the legal difficulties in creating an interstate compact among 3 states, and the resistance that the then-existing state and local organizations had to such an arrangment. To date, all the organizational changes have been implemented.

1. A federal department of transportation has been established.
2. Each of the 3 states—New York, New Jersey, and Connecticut—has coordinated the administration of all modes of transportation in one department.
3. In each state a commuter transportation agency has been organized with overall responsibility for transit-facility operations. In fact, the Connecticut Transportation Authority and the New York Metropolitan Transportation Authority have joint responsibility for the operation of the New Haven Railroad, which services both states. The 3 commuter agencies eventually could be combined into a single regional organization responsible for all transit operations in the region.

The recommendations relating to the flow of funds into the transportation system have as yet not been implemented. However, there has been some movement at the federal level in this direction.

1. President Nixon called for the creation of a transportation fund in connection with his special revenue-sharing proposal for transportation, and the requisite legislation was introduced in Congress.
2. The Secretary of Transportation and many senators urged the use of money from the Highway Trust Fund for transit purposes until a transportation trust fund is created.
3. On the regional level, surplus automobile tolls in New York and New Jersey are being used to finance transit.

Table 4. Current operating surplus by type of facility in 1962.

Item	Government Roads and Highways	Authority Toll Roads, Bridges, and Tunnels	Public Transit	Commuter Railroads	Private Bus Companies	Total
Funds provided by direct users	836.0	225.5	340.2	133.0	156.2[a]	1,690.9
Operating expenses, debt service, and capital outlay financed with federal and state user taxes	740.7	184.6	515.4	149.0[a]	141.8	1,731.5
Surplus (deficit) funds provided by direct users over expenditures	95.3	40.9	-175.2	-16.0	14.4	-40.6
Funds provided by indirect users	268.9	–	161.1	5.7[a]	–[a]	435.7
Surplus (deficit) funds provided by direct and indirect users over expenditures	364.2	40.9	-14.1	-10.3	14.4	395.1

Note: All amounts are in millions of dollars.

[a]The following adjustments were made in consolidating the statements of the region's facilities: (a) Tax relief granted to commuter railroads in New York State of $12.7 million and in Connecticut of $1 million was not included in operating expenses nor in funds provided by indirect users; (b) tax relief granted to bus companies in New York State of $3.0 million was not included in operating expenses nor in funds provided by indirect users; and (c) user taxes and charges of $7.8 million paid by private bus companies for the use of toll roads, bridges, and tunnels were eliminated from funds provided by direct users and from the operating expenses of these companies.

Table 5. Proposed organizational framework and flow of funds for financing transportation.

Governmental Level	Agency	Function	Source of Funds	Use of Funds
Federal	Department of Transportation	Coordinate transportation system for nation Administer federal transportation trust fund to be set up by federal government	All revenue from transportation-related taxes	Pay states for transportation projects approved by regional planning agencies (surpluses to general fund of federal government)
State	Department of Transportation	Coordinate transportation system in state Administer state-operated transportation facilities Construct state highways Administer state transportation trust fund to be set up in each state	All revenue from related taxes User charges from state-operated transportation facilities Federal transportation trust fund in proportion to funds raised in the region	Defray costs of state-operated transportation activities and construction of state highways Transfer to a region of funds from federal transportation fund Pay region from state transportation funds for transportation projects in proportion to funds raised by states in the region Pay local governments (surpluses to general fund of state)
Portion of each state in region	Transit Commuter Authority	Operate or contract for the operation of transit and commuter service Plan for improvement and expansion of transit and commuter facilities	User charges from operated facilities State transportation trust fund (including federal funds) Receipts from local governments (station maintenance)	Carry out planning, operating, and investment functions relating to transit and commuter operations
Local government		Operate and maintain local streets and roads and provide for allied services (including capital outlay)	User charges from operated facilities State transportation trust fund General funds	Carry out indicated functions

The slowness in implementing financial recommendations is attributable to the greater impact of institutional constraints on such changes rather than those that are purely organizational. Thus, the automobile lobby on the federal and state level is still effective in its opposition to the creation of a single transportation fund, but it has yielded in its blanket opposition to the use of any automobile funds for other transportation purposes. In addition, the majority in the U.S. Senate and the House of Representatives represent other than metropolitan areas; hence, these legislators have resisted the creation of a single transportation fund and oppose any shift in the present balance in the use of funds for highways and transit facilities.

On balance, the major organizational changes that we proposed have been adopted. Although there has been some progress in the direction of the recommended methods of financing, concerted action is still required in this area. Crisis methods of financing, including the use of questionable bookkeeping procedures, are the dominant practice. A unified approach to financing does not yet exist. Thus, the financial data remain essentially the same, and a study conducted in 1971 reveals the same basic needs (11).

The deficit of the region's transit system increased to $420 million in 1969-70 and can be expected to increase substantially in the future. Although no data are available on the transportation surplus for this period, we would conjecture that an overall surplus existed. Furthermore, even in the absence of a surplus, our financial recommendations would have made it possible to take a more balanced approach to financing transportation and to simplify the problem of allocating funds among individual forms of transportation in order to maximize the effectiveness of the system as a whole. Recommendations for financing public transportation in the 1972 report (11) also reflected this philosophy. Thus, the organizational changes recommended and adopted were only a first step for attaining a balanced and financially sound system.

Studies, however, are only the first step in the slow process of implementing change. One is reminded of the old saying, "You can lead a horse to water but you cannot make him drink." Perhaps if you lead him there often enough he is sure to drink if he wants to survive. The process in government decision-making is similar. Studies of the type presented here may help to show the way. Sooner or later governments must act if metropolitan transportation systems are to remain viable.

REFERENCES

1. Kurnow, E., Brief, R. P., and Silberman, I. H. Transportation Finances in the Tri-State Region—A New Method of Reporting. Tri-State Transportation Commission, March 1968.
2. Kurnow, E., Brief, R. P., and Silberman, I. H. The Financing of Transportation in the Tri-State Region. Tri-State Transportation Commission.
3. Bielak, S. F., and McCarthy, J. F. Highway Income, Expenditure, and User Tax Earnings in Standard Metropolitan Statistical Areas. U.S. Bureau of Public Roads, 1965.
4. Developing Metropolitan Transportation Policies. Committee for Economic Development, April 1965.
5. Netzer, D. The Financial Problem of New York Region's Suburban Rail Service. Regional Plan Association, Inc., Jan. 31, 1961.
6. St. Clair, G. P. Congestion Tolls—An Engineer's Viewpoint. Highway Research Record 47, 1964, pp. 93, 95.
7. Burch, P. H., Jr. Highway Revenue and Expenditure Policy in the U.S. Rutgers Univ. Press, New Bunswick, N. J., 1962, p. 123.
8. Harbeson, R. W. Some Allocation Problems in Highway Finance. Transport Economics, 1965, pp. 138-165.
9. Meyer, J. R., Kain, J. F., and Wohl, M. The Urban Transportation Problem. Harvard Univ. Press, Cambridge, Mass., 1965.
10. Special Report of the Law Committee. American Transit Association, Feb. 1964.
11. Kurnow, E., Brief, R. P., and Reibstein, R. Financing Mass Transportation, A Positive Approach. April 1972.

SPONSORSHIP OF THIS RECORD

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Charles V. Wootan, Texas A&M University, chairman

Committee on Taxation, Finance and Pricing
Charles A. Hedges, U.S. Department of Transportation, chairman
Richard P. Brief, Henry W. Bruck, Lawrence D. Dahms, Irving Hoch, Norman Kennedy, Gerald Kraft, Julius Margolis, John R. Meyer, Leon N. Moses, Gordon M. Murray, James R. Nelson, J. E. Peddy, Joseph H. Stafford, William S. Vickrey, Richard M. Zettel

Kenneth E. Cook, Highway Research Board staff

The organizational units and the chairmen and members are as of December 31, 1972.