TRANSPORTATION RESEARCH RECORD 791

Transportation Finance, Equity, and Cost Allocation

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1981

Transportation Research Record 791
Price $4.20
Edited for TRB by Sandra Vagins

modes
all

subject area
14 finance

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Transportation finance, equity, and cost allocation.

(Transportation research record; 791)
Reports prepared for the 60th annual meeting of the Transportation Research Board.
1. Transportation–Finance–Addresses, essays, lectures.
2. Transportation–Costs–Addresses, essays, lectures. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 791 [HE196.5] 380.5s [380.5'9] 81-9548
ISBN 0-309-03205-9 ISSN 0361-1981 AACR2

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Leon M. Cole, Library of Congress, chairman

Management and Finance Section
Ira F. Doom, Huntsville Department of Transportation, chairman

Committee on Taxation, Finance, and Pricing
Damian J. Kulash, Congressional Budget Office, chairman
Frank A. Maresca, Washington State Department of Transportation, secretary
Robert P. Aex, Bruce Barkley, Kiran Bhatt, A. R. Castello, Ronald E. W. Crisman, Lawrence D. Dahms, Morton Ehrlich, Robert E. Fagley, Barry M. Goodman, John S. Hassell, Jr., Alice E. Kidder, Thomas D. Larson, Bruce D. McDowell, Walter F. Mead, Lance A. Neumann, Robert G. Prestemon, Kant Rao, Richard M. Shackson, Fred Lee Smith, Jr., Joseph C. Smith

Kenneth E. Cook, Transportation Research Board staff

The organizational units, officers, and members are as of December 31, 1980.

Contents

Authors of the Papers in This Record

Beliveau, Russell, Crosbro, Inc., 1100 Pearl Street, Brockton, MA 02401
Collura, John, Civil Engineering Department, University of Massachusetts, Amherst, MA 01002
Lee, Douglass B., Transportation Systems Center, U.S. Department of Transportation, Kendall Square, Cambridge, MA 02142
Rock, Steven M., Department of Economics and Finance, Illinois Institute of Technology, Chicago, IL 60616
Rogers, Anthony D., Cape Cod Regional Transit Authority, 275 Mill Way, P.O. Box 318, Barnstable, MA 02630
Stahl, Dale O., II, Institute of Transportation Studies and Department of Economics, University of California, Berkeley, CA 94720
Warren, Robert P., Cape Cod Regional Transit Authority, 275 Mill Way, P.O. Box 318, Barnstable, MA 02630

Recent Advances in Highway Cost Allocation Analysis

DOUGLASS B. LEE

A wide range of proposals has been advanced over the last two decades to deal with the general problem of cost allocation, but highway cost-allocation practice has stuck to a relatively narrow framework of equity. The choice at the present time is whether to continue to treat highways as a tax-supported public service or to recognize that the highway system is a major economic enterprise. Recent policy shifts in transportation and in other sectors strongly suggest that highway user charges be designed explicitly to meet efficiency as well as equity objectives.

Highway cost allocation has been confronted in the past as a problem of how to raise revenues from selected groups of taxpayers so as to meet a given budget in a fair and equitable manner (1). Economists have urged that the problem be viewed as one of pricing highway services in order to achieve efficient use of scarce resources (2-6), but this perspective has never explicitly been put into practice. Current efforts seek to integrate the two approaches in a way that will preserve the best of both (7-9).

CRITERIA FOR SELECTING AMONG USER-CHARGE STRUCTURES

Normative standards against which to evaluate user-charge instruments and rates are fundamentally two: efficiency and equity (10,11). A third criterion can be the effectiveness with which stated goals are achieved, but the goals themselves usually relate to selected aspects of efficiency or equity.

Efficiency

Although efficiency is not mentioned in the congressional mandate (Surface Transportation Assistance Act of 1978, Sections 506 and 507) for the federal highway cost-allocation study now under way, efficiency considerations are strongly implied. The notion that vehicles should pay for the costs they occasion is described as equitable, but it is also efficient that they do so because it encourages them to reduce these costs and make sure that the benefits they derive are greater than the costs created.

Short-run efficiency assumes a given set of capital facilities and seeks ways to secure the best possible utilization of those facilities. The theoretical mechanism for this optimization is pricing--interpreting the concept of a price broadly to include such factors as travel time and risk of accident, as well as user charges. Highway-user charges are the most direct means for achieving short-run efficiency with respect to the highway system.

Long-run efficiency deals with finding the best program of investment in fixed facilities while also satisfying the short-run efficiency criterion. Analytically, the path to long-run efficiency is followed by first comparing the incremental costs and benefits of alternative projects and then investing (or disinvesting) in the appropriate links of the highway network. Although user charges inevitably have an influence on the actual pattern of maintenance and investment, the theoretical linkages are indirect.

Equity

The concern addressed by equity is the distribution of costs and benefits among groups within society. In contrast to efficiency, equity is a term that is frequently mentioned, yet one that provides very little positive guidance. It is essential that the redistributional impacts of alternative user-charge schemes be thoroughly illuminated, and the imposition of equity constraints on efficiency solutions will be necessary both analytically and politically; however, there is no hard and fast way to assert that some user charges are equitable and others inequitable. Equity is inherently a matter of political choice, although technical analysis can contribute to the political debate by formulating equity constraints and displaying their consequences. For example, the requirement that users pay the full costs (or some prespecified share) is an equity constraint.

Horizontal equity is most directly related to popular ideas of fairness; it urges that equals be treated equally. Vehicles in equal circumstances--from the standpoint of the highway provider--should be charged equally; however, there may be instances in which price discrimination is useful for achieving other efficiency and equity objectives. Vertical equity describes the distribution of net gains among income classes, a factor of major concern but one on which highway-user charges have only a minor impact. Equity impacts are of prime interest in designing user charges, but equity objectives only make sense in conjunction with efficiency objectives.

THE NEED FOR AN IMPROVED FRAMEWORK

Earlier studies have agreed that some share of highway costs should be borne by users, that using "cost-occasioned" distribution is a fair way to allocate costs among users, that the amount of cost occasioned by a vehicle class can be determined by disaggregating items of expenditure and assigning them to vehicle classes, and that equity lies in the method for allocation rather than in the distribution of the tax burden it produces. An algorithm known as the incremental cost method has been popular in recent studies, but there is very little professional consensus on which are the best methods for the practical determination of highway-user cost responsibilities. The incremental cost method takes as its starting point a basic highway, usually one designed for automobiles alone. Additional classes of vehicles cause additional increments of cost, and these increments are apportioned among the members of each class.

A thorough exposition of these concepts and others that have been considered at some point for use in highway cost-allocation analysis would be an enormous and not very rewarding task. Without exception, the methods are ad hoc and unsupported by theory. Unfortunately, there is no pragmatic test for these methods that would tend to select the workable ideas from the mistaken ones, so the absence of theory is a serious handicap to improvement. Without an attempt to criticize previous efforts, some arguments can be offered for rethinking the overall framework within which highway-user-charge analysis is conducted.

1. Efficiency should be explicitly recognized: It has already been noted that payments in accordance with costs occasioned can be consistent with an efficiency goal. Many other policies that relate to highway transportation, e.g., reduction in

fuel consumption, deregulation of trucking to encourage competition, and the cost-benefit evaluation of highway investment projects, imply an efficiency orientation. Failure to incorporate efficiency concerns on the pricing side seems pointlessly myopic.

2. Equity is too ambiguous: Fairness, or equity, even when applied as rigorously as possible, is a criterion that relies very heavily on value judgments and leaves a great deal of room for discretion. The costs of climbing lanes, for example, can be assigned to light vehicles or to heavy ones, depending on whether the starting point for analysis is taken to be an automobile or truck highway. Fairness can be resolved politically in such a manner, but planners should recognize that a pure equity approach does not lead to technically stable answers. Much of the previous highway cost-allocation work reveals the underlying ambiguity by drifting superficially over numerous alternative concepts or by being inflexibly arbitrary.

3. Intermodal policy should be consistent: Deregulation is proceeding in the airline and railroad industries; minimal user charges have been initiated for inland waterways. In all of these actions, intermodal price competition was an important concern. Electric power, telephone, postal service, and intercity bus enterprises face similar types of problems, and there is no reason for highway policy to stand out as incompatible with the concepts used in analogous industries. Whatever policy direction is taken next, private transportation modes and publicly owned modes should be priced comparably and treated analytically as similar sorts of beasts.

DESIGN OF A HIGHWAY COST-ALLOCATION STUDY

Although precise methods and techniques are still incomplete, the outline of a reasonably robust conceptual framework can be presented at this time. The problem of federal user charges will be taken as a prototype, the same framework being also applicable to state user-charge studies. Much of the knowledge gained from previous cost-allocation studies, as well as parallel work in related fields, will prove to be useful, but there are also many areas in need of further development. In the following, the cost-assignment problem rather than the selection of user-charge instruments will be emphasized, in part because the choice of instruments depends on empirical and pragmatic matters that are too detailed for general treatment

An outline of the problem (Figure 1) indicates the major tasks to be accomplished. Costs should be broken into (a) variable and (b) fixed, variable costs being those on which variable user charges (e.g., fuel tax, weight-distance fees) are based. If these charges do not raise sufficient revenues to cover costs, then a residual will remain that can be met from access charges (e.g., registration fees) or general revenues (e.g., property taxes). Because practical realities will force many compromises, both the prices and the assignment of residual costs will need to be evaluated (in the form of a number of alternatives) against efficiency and equity criteria. Once a workable set of user charges has been constructed, the federal portions of these charges can be broken out and matched against the budget.

Total Costs

In general, expenditures and costs are not the same thing. Even if expenditures represent the social value of the particular resources covered by the expenditure, many costs do not appear as expenditures. Some examples are (a) exemption from paying a tax and (b) no interest charged on capital funds. In addition to capital and maintenance expenditures, the following costs should be tabulated.

1. Hidden costs: Some costs appear in public budgets but not in the budget of the agency responsible for highway expenditures. Vehicle code enforcement and traffic control may be hidden in police budgets, electricity consumption may be buried in a utility budget, and payroll administration may be centralized rather than included in the transportation agency budget.

2. Negative externalities: Negative externalities in the form of air pollution, noise, water pollution, and other unpriced effects on the physical, natural, and human environment constitute real costs to society. Even though we may never be able to place accurate dollar values on these costs, present policies can be (and are being) improved on by a recognition of such costs.

3. Interference costs: Private costs in the form of delay time, vehicle wear, fuel, and accidents are relevant to the correct pricing of highway services. These relationships will be explained below.

4. Tax expenditures: Exemption of fuel from general sales taxes and exemption of highway property from local property taxes result in a favorable treatment of highways in comparison with other activities that are not exempt. To the extent that these taxes pay for general government services (as opposed to income transfers), highway users are being subsidized by those engaged in other activities.

5. Interest forgone: The pay-as-you-go philosophy, in which each year's expenditures are matched with the same year's revenues, hides the fact that invested capital has an opportunity cost represented by the rate of return (i.e., interest) that the money would earn in another activity. Money spent from the Highway Trust Fund does, in fact, lose the interest it would be earning if left in the fund.

Prices

Theory tells us that the price charged for use of the highway should be equal to the (short-run) marginal use cost. If less is charged, the user may not value the use as much as society values the resources used up; if more is charged, some potential users are deterred, even though they would gain more from the travel than it costs society. This principle only applies to charges, such as a fuel tax or a weight-distance tax, that vary directly with usage. Access charges (such as an annual weight fee) and general taxes are subject to different considerations. Partly because of the particular nature of variable highway costs and partly because of the general tendency for variable charges to be more costly to administer than access charges, the design of practical mechanisms for imposing correct prices on highway users presents a major challenge.

1. Pavement damage: Probably the most easily accepted basis for user charges is pavement wear; there is a clear connection between expenditures and the cost imposed by particular vehicles. Additional empirical research is needed to better establish the relationship between axle weight and the cost of damage to a particular road, but a solid information base has been developed (12,13). The preferred user charge would be a weight-distance fee based on equivalent single-axle load repetitions for each vehicle, but some approximation based on averages will undoubtedly be necessary for most vehicle classes.

2. Interference costs: The concept of the con-

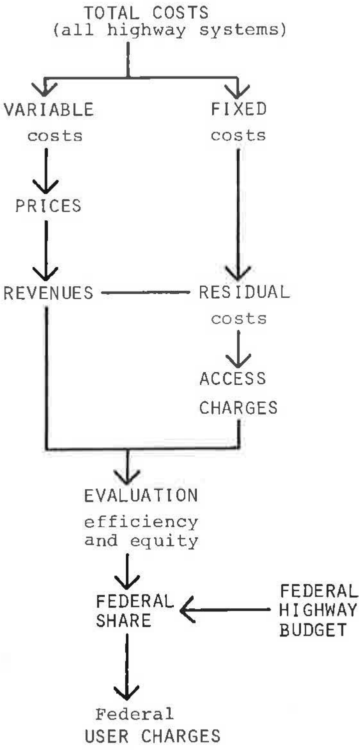


Figure 1. Outline of federal highway user-charge determination.

gestion toll has been known for some time; numerous direct and approximate means for collecting it have been proposed. With the exception of a few bridge tolls that vary with time of day and direction, no highway-user charges in the United States are in any way related to congestion.

Interference costs include such congestion-related costs as accidents between vehicles, excess vehicle and tire wear, and excess fuel consumption, as well as travel delay. In contrast to pavement wear, interference costs rise in the short run (i.e., on a given highway) with increases in vehicle volumes of travel, which results in a deviation between the marginal social cost of travel and the price paid by the user. The price paid by the user is in the form of delay time and other private congestion costs, rather than in the form of a money price. Because marginal cost is above average cost, a price (or toll) needs to be charged (presumably by the facility operator) to bring congestion down to the efficient level. The relative contributions of different vehicle types to a congested traffic stream can be measured in passenger-car equivalents, a measure of road space effectively occupied by a vehicle of a given type under given terrain, vehicle-mix, road-type, and congestion conditions (14).

Congestion pricing is not taken very seriously by noneconomists, apparently on the basis of the twin notions that congestions is (a) something that is confined to users and (b) not a real cost anyway. Yet congestion reduction is regarded as a real benefit when expenditures for additional highway capacity are evaluated, i.e., when money is spent for the purpose of reducing congestion. From a theoretical perspective, congestion pricing is the only efficiency rationale for recovering any of the fixed costs of highway construction and maintenance (5). In the case of telephone calls, for example, a congestion toll is reflected in the price per minute during peak (business) versus off-peak periods. The telephone system is subject to large fixed costs (the capacity and extent of the network cannot be varied by time of day), and the marginal cost of inputs is not significantly greater during peak periods. Thus, the peak premium is a measure of the opportunity cost of not having time available for some potential user. The important difference between telephone calls and highway travel, however, is the relative ease with which the telephone caller can be charged according to time of day and duration of call.

3. Externalities: The imposition of noise, asbestos dust, fumes, litter, and the like on persons who have not been compensated for these injuries by persons who are not required to pay an emissions or damage charge can be regarded as a "taking" by the polluters from the pollutees. In legal terms, the perpetrators of the externalities should be required to purchase pollution rights or easements in order to engage in polluting activities. Efficiency clearly calls for highway users to pay a charge for the externalities they create (15) according to the amount and nature of the damage caused so that the correct charge will be a function of use and emission rate.

Besides the lack of precedent, implementation of externality charges is hindered by the facts that the damage caused varies by location and vehicle type and that it is difficult to place a value on the damage. We should be able to improve, however, on the price of zero that is currently charged. For example, the cost of noise barriers could be collected from motorists along all highways wherever barriers are or should be constructed.

4. Other variable costs: Annual operating costs that bear some rough relationship to volume of use--highway and traffic police, management and administration, and perhaps snow and ice control--can be collected on a variable-charge basis. The relative magnitudes are small and the efficiency incentive negligible, so the precise instrument is not of critical importance.

Residual Costs

Under conditions of constant returns to scale, optimal investment in capacity, correct marginal-cost prices, and a "first-best" world, revenues from highway users would exactly equal long-run costs. Both variable and fixed costs would be fully recovered without the necessity of ever contemplating the allocation of any fixed-cost components. Congestion tolls would generate enough of a surplus over variable costs in the short run to pay the fixed costs in the long run. Under ideal conditions, the above revenues would completely solve the cost-allocation problem.

With reasonable confidence, we can assert that none of the ideal conditions are satisfied with regard to the U.S. highway system at the present time, and the likely direction of devation from the ideal is to lead toward a shortfall between costs and revenues from efficient user charges. If no congestion tolls are imposed, revenues cannot be expected to recover fixed costs, and empirically, user revenues of all kinds are less than total public expenditures on highways. We are left, then, with an awkward problem.

Before considering how to approach the problem of residual costs, we can make sure the problem is minimized by taking advantage of variable cost pricing to the fullest extent in the following ways.

1. Even though congestion tolls are not actually levied, in the ideal sense, congestion or interference costs can be a basis for setting some other

user charge, even an access charge. If it is known which vehicles are responsible for congestion or increase the risks of accident, user fees can be imposed on those vehicles. Untying the charge from the actual amount of congestion removes the efficiency incentive, but price inelasticity or fairness may still warrant such an indirect instrument.

2. Similarly, externality charges can be levied, even though the instrument may be indirect and compensation is not actually paid for the damages caused. A fuel tax is only very roughly related to either congestion or externalities, but it does reflect usage for a given vehicle and it could be combined with an annual surcharge that is based on average emissions and contribution to congestion.

3. An expenditure budget may be covered by pricing variable costs that are not actually included in the expenditures. A correct charge for pavement damage, for example, might raise more revenues than the actual expenditures to repair the damage, both because (a) the pavement was being allowed to depreciate and (b) revenues were collected from trucks for the user costs imposed on light vehicles although compensation was not made to light vehicles. Such a strategy, however, implies investment and equity policies that are unacceptable for anything other than an emergency regime.

Allocation of Residual Costs

The residual-cost assignment problem can be structured in several ways, ultimately reconciling on pragmatic grounds those approaches that seem strongest. A typical procedure will involve allocating a set of costs to a vehicle class and then allocating the costs to individual vehicles. Some of the boundaries and possible starting points for residual-cost assignment are described below.

1. Subsidy from general revenues: All, none, or some portion of residual costs may be covered by revenues from general taxes. A decision to use general revenues for highway purposes should be made on the basis of the justification for a subsidy, such as increasing returns to scale, and not on spurious arguments about nonuser costs and external benefits.
2. Incremental fixed costs: Many professionals and policymakers regard the assignment of certain expenditure items to associated vehicle classes as equitable. For example, weigh stations can be assigned to trucks and guardrails to automobiles. Such assignments have an inherent degree of arbitrariness to them, but methods are available for placing bounds on reasonable solutions. No group of users should have to pay more overall than it would pay for a separate system of its own (16-18). The preferred revenue instruments are access charges (19).
3. Benefits: If emphasis is placed on the resource allocation that results from marginal-cost pricing of the existing system, then residual costs should be covered by taxes that change resource use to the smallest degree. To the extent that taxes are imposed on users, they can be scaled according to ability to pay, consumer surplus, the inverse of the consumer's demand elasticity, or benefits to the user (20,21). These strategies have a great deal in common.
4. Long-run marginal cost: If the capital stock is far from what would be optimal for expected demand conditions, efficient prices may be nowhere near ideal prices in the long run. Instead of reliance on the combination of short-run marginal cost for setting prices and efficient investment programs for adjusting the scale of the highway system, prices can be set directly at long-run marginal cost. The rationale for this strategy is to avoid misleading signals for investment in related activities, as well as highway transportation, when long-run costs are understated by current prices. In practice, a long-run pricing objective can be supported by estimating future capital stocks and basing short-run prices on those rather than on what currently exists. Long-run incremental costs can be recovered from access charges based on criteria in the subsidy, cost, and benefit approaches described above.
5. Distributional equity: Alternative access-charge instruments can be compared as to which groups of the population ultimately bear the costs. Choices would depend on judgments about which of the population groups are most deserving.
6. Effectiveness: Choices among user-charge instruments can also be based on the effectiveness with which each one attains various transportation and nontransportation public goals. Effects on mobility of the elderly, employment, or urban revitalization may be considered, but the relationships are likely to be weak and the cost-effectiveness poor.

User Charges to Recover an Expenditure Budget

To select a single budget for cost recovery from the many government agencies that participate in financing the highway system is highly arbitrary. It is like having the makers of copper wire recover their portion of telephone system costs by imposing their own set of telephone service charges. The price of a telephone call would thus include the part charged by the copper makers, the part levied by the operators, the part imposed by pole erectors, etc. Each portion of the price would be set independently of all the other portions. The arrangement would be universally regarded, quite properly, as lunatic.

The highway enterprise lacks a single authority (for any portion of the system) that can establish a complete user-charge structure. If such authorities existed, they could purchase inputs from suppliers (who currently are government agencies) and charge consumers in accordance with costs and the characteristics of demand. Without this institutional structure, the only possible surrogate is federal government initiative.

How the federal government can arrive at its own share of highway user charges, even by expedient means, is unclear. Some alternative strategies are to (a) collect the same proportion of user charges on each system as the federal government pays in costs, (b) assign costs to vehicles according to the purpose of the travel and let the federal government impose user fees on interstate travel, and (c) establish a floor of federal charges that are uniform across the country and let state and local governments supplement the federal user charges for their own needs. The third of these is the only one that appears remotely workable in the near term, and it offers few appealing features.

CONCLUSIONS

To a degree we have come full circle, in that several of the residual-cost strategies sound something like the incremental-cost method that was dismissed at the beginning. Nonetheless, a great deal of progress has been made: Variable costs have been separated from fixed costs and given suitable treatment, the scope of the analysis has been expanded to include all costs (not just government expenditures) on all systems (not just the federal portion), and a framework has been constructed that allows for a clear statement of the problem and an informed evaluation of alternatives.

Some of the improvements over previous methods are not immediately obvious. A form of partially distributed cost analysis was used in the past, but it was justified on the basis of the nonuser cost-responsibility myth (I should be taxed to help pay for my grocer's store because it serves me). When the assignment of residual costs is handled as a constrained optimization problem, at least we are informed of the criterion used and the consequences. The incremental-cost method was applied indiscriminately to variable as well as fixed costs without assessing its suitability. The emphasis on allocation of budgets rather than pricing of costs has meant that user charges have fallen along with expenditures (in real terms) at the same time that costs have been rising. Finally, attention is directed at the effects of alternative user charges on efficiency and equity, not at the largely pointless exercise of labeling expense items with vehicle-class names.

REFERENCES

1. Final Report of the Highway Cost-Allocation Study. 87th Congress, 1st Session, U.S. Government Printing Office, House Doc. 72, 1961.
2. M.Z. Kafoglis. Highway Policy and External Economies. National Tax Journal, Vol. 16, No. 1, March 1963, pp. 68-80.
3. A.A. Walters. The Economies of Road-User Charges. Johns Hopkins Press, Baltimore, MD, World Bank Staff Occasional Papers, No. 5, 1968.
4. A.E. Kahn. The Economics of Regulation, Volume 1: Principles and Institutions. Wiley, New York, 1970.
5. H. Mohring. Transportation Economics. Ballinger, Cambridge, MA, 1976.
6. J. Slavet, K. Bradbury, and P. Moss. Financing State-Local Services. Heath, Lexington, MA, 1975.
7. Wilbur Smith and Associates. Development of Strategies and Procedures for Highway Cost Allocation. Federal Highway Administration, Interim Rept. 1, Nov. 1975.
8. Congressional Budget Office. Guidelines for a Study of Highway Cost Allocation. Congress of the United States, Washington, DC, Feb. 1979.
9. First Progress Report on the Federal Highway Cost Allocation Study. Federal Highway Administration, U.S. Department of Transportation, March 1979.
10. W. Nicholson. Intermediate Microeconomics and Its Application, 2nd ed. Dryden Press, Hinsdale, IL, 1979.
11. R.A. Musgrave and P.B. Musgrave. Public Finance in Theory and Practice, 2nd ed. McGraw-Hill, New York, 1976.
12. Consequences of Deferred Maintenance. NCHRP, Synthesis of Highway Practice 58, 1979.
13. The AASHO Road Test. HRB, Special Rept. 73, 1962.
14. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
15. W.J. Baumol and W.E. Oates. The Theory of Environmental Policy. Prentice-Hall, Englewood Cliffs, NJ, 1975.
16. J.C. Bonbright. Principles of Public Utility Rates. Columbia Univ. Press, New York, 1961.
17. S.C. Littlechild and G.F. Thompsson. Aircraft Landing Fees: A Game Theory Approach. Bell Journal of Economics, Vol. 8, No. 1, Spring 1977, pp. 186-204.
18. T. Sandler and J.T. Tschirhart. The Economic Theory of Clubs: An Evaluative Survey. Journal of Economic Literature, Vol. 18, No. 4, Dec. 1980, pp. 1481-1521.
19. Z. Haritos. Theory of Road Policy. Transportation Journal, Vol. 13, No. 3, Spring 1974, pp. 53-64.
20. F.P.A. Ramsey. A Contribution to the Theory of Taxation. Economic Journal, Vol. 37, 1927, pp. 47-61.
21. W.J. Baumol and D. Bradford. Optimal Departures from Marginal Cost Pricing. American Economic Review, Vol. 60, June 1970, pp. 265-283.

Publication of this paper sponsored by Committee on Taxation, Finance, and Pricing.

Proposed Fare Policy for Advance-Reservation Bus Service

ROBERT P. WARREN, ANTHONY D. ROGERS, JOHN COLLURA, AND RUSSELL BELIVEAU

This paper reviews the present fare policy of the advance-reservation bus service in Barnstable County, Massachusetts, examines several alternative fare policies, and, finally, proposes a new policy. The present fare policy allows individuals to ride an unlimited number of times at a flat rate for any purpose during a three-month period. Four alternative fare policies are described. The alternative proposed for implementation would charge riders on the basis of the number of trips taken and the length of each trip. Reduced rates would be available for elderly or handicapped persons and for those who made group trips. Riders would be sent bills at the end of each month like telephone bills—the trips and miles traveled would be detailed as long-distance calls are. These invoices would be prepared by the existing computer system, which currently maintains complete client listings and generates detailed drivers' schedules. The cost of this minicomputer system, including hardware and software, was about $50 000. The development of the billing and invoicing system would cost an additional $5500. The paper also recommends further research into alternative fare policies, including their effects on travel behavior, revenue generation, and subsidy requirements. Other recommended topics are alternative mechanisms for implementation of such fare policies, such as sale of tickets, punch passes, manual invoicing, and (as proposed) implementation as a component of a comprehensive computerized management information system.

As a result of increasing fiscal austerity at the federal, state, and local levels, government subsidies for public transportation services are expected to decline, although the need for and the costs of such services are increasing dramatically. As a result of this, consumers will be called on to pay higher proportions of total costs. As the amounts to be paid by consumers increase, the equity of the fare policies used will become of paramount importance. If public transportation is to maintain its feasibility in the 1980s, equitable fare policies, and means for implementing them, must be developed.

It has been proposed that the Cape Cod Regional Transit Authority (CCRTA) implement a fare policy for its advance-reservation demand-responsive service (the b-bus system) that is based on use, as measured by number of trips taken and the length of trips. This policy would require computerization; however, the CCRTA has a computer-based management information system (MIS) on line at its b-bus operations center that, with little modification, could accommodate the new fare policy.

The purpose of this paper is to introduce this proposed fare policy, describe the situation from which it developed, and present recommendations for further research into the equity and practicality of alternative fare policies.

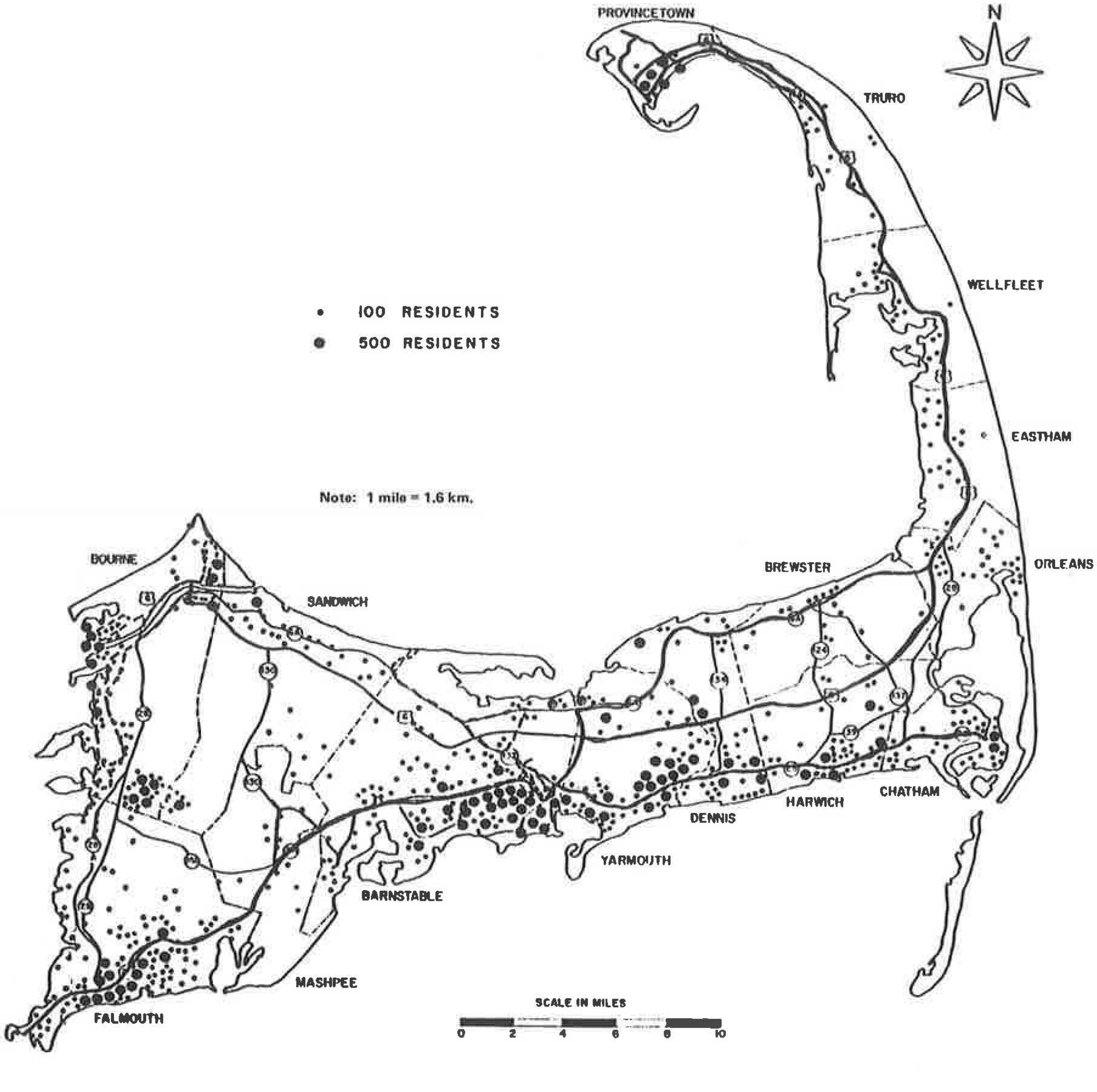
HISTORICAL REVIEW

Barnstable County (Cape Cod), Massachusetts, has a year-round population of 140 000 and a summer population, due to a seasonal influx of tourists, of 450 000. The 15 communities that make up Barnstable County cover 1008 km² (389 miles²), giving a population density of 138 persons/km² (see Figure 1). It is significant that more than 33 percent of the Cape's population can be described as either elderly or handicapped. This high proportion of transit-dependent persons explains the high priority that is given to the provision of paratransit services by local officials.

On March 17, 1976, the first step was taken by the region to develop a system to meet the needs of the Cape's elderly and handicapped residents. A proposal was submitted to the Federal Highway Administration (FHWA) under Section 147 of the Federal-Aid Highway Act of 1973 for a rural public transportation demonstration grant.

Among the innovative techniques that were proposed for development within the grant application was use of a serially coded flash pass (1). The pass numbers were to be used for monitoring and evaluation of socioeconomic and trip (origin-destination) information to produce comprehensive transit-rider profiles. A secondary function was to serve fare collection. Clients were to pay flat quarterly fees for use of the service. These payments were to be made by mail and tracked by flash-pass serial number. The potential for more-sophisticated revenue collection techniques by using the pass was recognized; however, it was impractical to

Figure 1. Barnstable County.



implement them without on-site computer assistance.

On July 20, 1976, FHWA approved the $868 750 grant request. As a result of this initial investment by FHWA, the county was able, for the first time, to provide a comprehensive regionwide public transportation service, the b-bus system. Today some 28 vehicles provide b-bus service, 8:00 a.m. to 4:00 p.m., five days a week. Clients request service by phone three to five days in advance and are given portal-to-portal service.

On February 9, 1979, the county's demonstration project was turned over to the newly formed CCRTA. CCRTA's initial mandate was to bring about coordination of all public transportation services that existed in Barnstable County. This coordination effort culminated in the full consolidation of services on June 1, 1979 (2). The elimination of duplications of effort in such areas as management, dispatching, and marketing resulted in substantial cost savings. Vehicle use was also increased, to yield higher productivities and increase the cost-effectiveness of the b-bus program.

Having brought about substantial cost savings and service improvements through consolidation, CCRTA turned its attention to refinement of its management tools. One of its primary concerns was to develop a procedure to equitably allocate costs among member municipalities (3). Consolidation presented the opportunity for development of a centralized MIS that could perform multiple functions, among them compilation of information on origins and destinations. Such information would allow use of a multivariable formula for assessments to the member municipalities that would be based on actual use by residents. It was the attractiveness of this possibility that led to the overwhelming support of the CCRTA Advisory Board for development of such a system.

On April 10, 1980, CCRTA awarded a contract to Crosbro, Inc., of Brockton, Massachusetts, for development of such a system. The computer went on line in December 1980 and is currently performing all necessary data-collection functions. In addition, the system aids in dispatching the vehicles, essentially eliminating pen and paper from the process, as described in a later section of this paper.

As early as 1975, when the grant proposal for the b-bus program was submitted to FHWA, two items were recognized to be of paramount importance to the development and continuance of local support for a regional public transportation system on Cape Cod. The first, and the most urgent, was the development of an assessment formula for ensuring the equitable allocation of costs among member municipalities. Such a formula had to result in payment only for services received by the residents of each municipality. Of equal importance, but not as immediately essential, was the development of a fare-collection mechanism that would result in the equitable distribution among riders of the portion of the total cost to be paid by consumers.

Such a mechanism has been proposed for use by CCRTA and is, in fact, the subject of this paper. Simply stated, riders would be charged according to a cost formula designed to proportionally approximate the actual cost of the service received. A rate would be developed that would include a cost per trip and a cost per mile of travel. Travel distance would be estimated by using a zone-to-zone distance matrix, and invoices would be generated automatically by the computer and mailed to riders.

CURRENT FARE POLICY

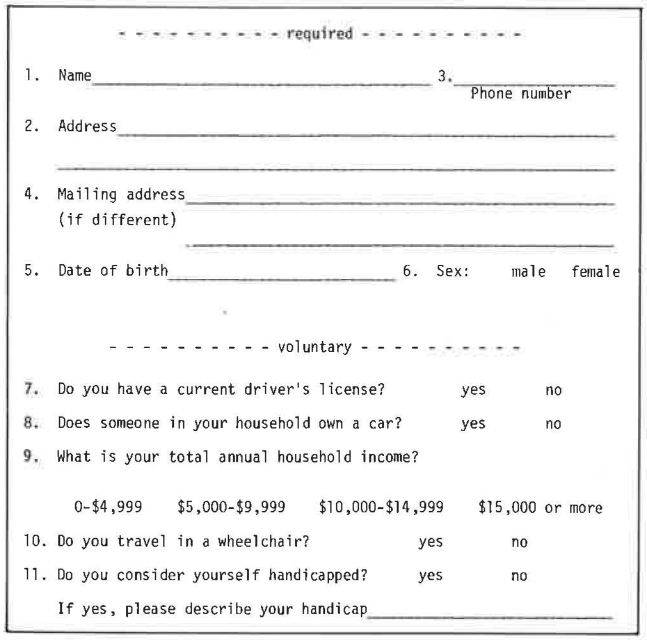
Travelers who use the b-bus system are currently being charged on a quarterly flat-rate basis. An integral part of the current fare policy is the use of a serially numbered rider identification pass. Socioeconomic information is obtained as part of the pass distribution process (Figure 2). The possession of a pass qualifies elderly or handicapped persons to free health-care-related service. If the passholder would like additional service, or is neither elderly nor handicapped, he or she mails CCRTA a check or money order each quarter. When the passholder telephones to schedule a trip, the passholder identification number and trip data (e.g., pickup and drop-off times and origin and destination) are obtained and the information is entered into the computer. The MIS eliminates the need to write trip data onto request sheets and then onto driver logs, because the computer has the capability to print schedules. This information is also used for allocation of costs among towns (which meets the reporting requirements of Section 15 of the Urban Mass Transportation Assistance Act of 1970) and in billing and accounting.

ALTERNATIVE FARE POLICIES

During the last 12 months, CCRTA has considered several alternative fare policies to replace the current policy. Four alternatives reviewed are summarized in Table 1.

Each alternative can be described in terms of type of payment, method of fare collection, and the basis of the charge. Alternative 1, a free-fare

Figure 2. Questionnaire for b-bus client.



- - - - - - - - - - required - - - - - - - - - -

1. Name ______________________ 3. ____________ Phone number

2. Address ______________________

4. Mailing address (if different) ______________________

5. Date of birth ______________ 6. Sex: male female

- - - - - - - - - - voluntary - - - - - - - - - -

7. Do you have a current driver's license? yes no

8. Does someone in your household own a car? yes no

9. What is your total annual household income?

0-$4,999 $5,000-$9,999 $10,000-$14,999 $15,000 or more

10. Do you travel in a wheelchair? yes no

11. Do you consider yourself handicapped? yes no

If yes, please describe your handicap ______________

Table 1. Alternative fare policies.

Alternative	Type of Payment	Method of Collection	Basis of Charges
1. Free fare	None	None	None
2. Fare box	Cash	On board	Flat rate or sliding scale per trip
3. Prepaid tickets	Cash or money order	Mail	Flat rate per trip
4. Mail-in method	Cash or money order	Mail; bill sent after use	Base rate per passenger trip plus the rate per passenger mile

system, was discarded because of financial restrictions.

The second alternative included the collection of fares on the vehicle by means of a fare box, as is typically done on conventional fixed-route services in urban areas. This alternative was viewed unfavorably because of the need to hire special personnel to store and handle lock-type fare boxes and the potential for pilferage and theft.

Alternatives 3 and 4 both employed a mail-in method to collect fares; however, they differed in the type of payment. Alternative 3 required that the fee be paid before use. Tickets would be sold at a fixed rate on a per-trip basis. If a passholder purchased 10 tickets, the passholder would be entitled to make 10 one-way trips of any length for any purpose. Alternative 4, on the other hand, collected payment after the fact and charged passholders on the basis of the number of one-way trips and the length of the trip.

Alternative 4 was considered more favorably than alternative 3, because persons were charged not only for the number of trips but also for the distance traveled. The billing would be carried out by the computer by means of an accounts receivable system developed by Crosbro, Inc. (4). The necessary data files would include (a) the passholder file, which provides the user's identification number, name, and address; (b) the trip file, which has the user's identification number and each trip by origin and destination; and (c) the trip-distance file, which stores the minimum travel distance between each origin and destination pair. Use of the trip-distance file would eliminate the need for the driver to record odometer readings for each trip.

This fare structure takes into account the level of use, the rider's ability to pay (elderly or handicapped persons receive a 20 percent discount), and group-trip riding. Hypothetically, the discount for group riding would encourage this practice and lead to increased vehicle use.

IMPLEMENTATION OF PROPOSED FARE POLICY

The MIS installed at CCRTA provides for on-line scheduling of b-bus vehicles. The system provides various operational, managerial, and statistical reports. In addition, the data that are gathered and maintained by the system would enable CCRTA to install a billing and payment system of the type necessary to support the proposed fare-collection system.

The MIS is operated on a Data General Nova 4/S computer with 64K of metal oxide semiconductor (MOS) memory. The hardware includes 20 megabytes of disk storage, a 180-character/s printer, and three cathode-ray tube (CRT) terminals. This configuration is sufficient to handle up to 7500 clients and 40 000 trips/month (i.e., three times the current load). The hardware is highly expandable and can be altered to support new applications, as well as greater volumes. The programs were written in Data General Business BASIC. The data files are all index-sequential files, some of which require multiple keys. All data files and indexes have been assigned contiguous disk space to provide the fastest possible access.

The MIS itself can be broken down into four major functions or components, sometimes called subsystems:

1. File maintenance and inquiry routines,
2. Scheduling and trip-related data entry,
3. Monthly and annual routines, and
4. Client billing (i.e., the proposed fare system).

File Maintenance and Inquiry Routines

The file maintenance and inquiry routines allow the operator to add, delete, and modify master file records. The routines allow on-line inquiry against particular records, as well as report capabilities that produce a hard-copy listing of all the records in each file. The file maintenance and inquiry routines support the master files that form the data base required by the other major components of the system. The master files include (a) a vehicle file that contains equipment and maintenance data, (b) a town file, (c) a location file that identifies villages within the towns, (d) a trip matrix that contains the distances between villages, (e) a client file that contains residence and socioeconomic data on each client, (f) a purpose-code file that categorizes trip purposes, (g) a standard trip file that contains data on repetitive trips, and (h) a vehicle-schedule file that contains the current driver-vehicle-schedule assignments.

Scheduling and Trip-Related Data Entry

Scheduling and the trip-related data entry routines are the heart of CCRTA's daily operations, which include booking trips, printing the schedule, entering vehicle data, and entering changes to previously booked trips. There are eight routines required to perform these tasks, each of which is described below.

The system provides three methods for booking trips. The first method, the request-for-service routine, is used to book advance reservations taken over the phone. If the request is made by a new client, the operator will create a client record and issue a pass number before attempting to book the request. The booking itself is accomplished by entering the date, time, origin, destination, and trip purpose. The operation will then enter the most geographically appropriate schedule. The trips already booked on the selected schedule will be displayed. The operator will then analyze the schedule to see whether the request can be accommodated. The operator can review several schedules in an attempt to accommodate a single request. Both outgoing and return trips can be booked through the same routine. The routine ends when the operator confirms or denies the request. Regardless of confirmation, a trip record is added to the trip transaction file. The record includes all the data entered in the request for service, as well as a pickup code that indicates whether or not the request was confirmed.

The flag-stop and nutrition-trip routine is the second method of booking a trip. This type of booking allows the operator to create a trip record for a trip that has already been taken. In these situations, the trip data are not available to the operator until the driver returns the schedule listing at the end of the day. Data on any nonscheduled trips are written on the listing by the driver. The operator then enters these data by using the flag-stop and nutrition-trip routine.

The final booking method is the standard trip-scheduling routine. This routine is executed once a month. The routine converts the day of the week, found in the standard trip file records, into dates that occur in the forthcoming month. The routine then creates one trip record for each converted date. This routine accounts for approximately 25 percent of all trips.

The schedule print routine produces the schedule listing. It is printed each night for distribution to the drivers the next morning. A schedule update routine is available to allow the operator to make last-minute adjustments to the schedule. Each time the schedule print routine is executed, the vehicle-schedule file is used to post the appropriate vehicle and driver to the schedule listing and to the individual trip record.

The trip-by-client inquiry allows the operator to print or display on the screen all trips on file for a given client.

The trip-transaction maintenance or inquiry allows the operator to add, delete, and modify trip records. This routine's primary function, however, is to allow changes to the pickup code when a client cancels a trip or fails to show at the scheduled pickup location.

The daily vehicle log allows the operator to enter the vehicle mileage, fuel consumption, and maintenance data that are turned in daily by the driver of each vehicle. Each entry results in the creation of a vehicle log record that is posted in the daily vehicle log file.

Monthly and Annual Routines

The monthly and annual routines (a) produce management and statistical reports, (b) invoice the various social-service agencies that purchase service from CCRTA, (c) purge old trip and vehicle log records, and (d) reset monthly accumulators in the vehicle and town files during year-end processing.

The monthly reports generated by these routines accumulate trips, passenger miles, vehicle miles, and vehicle hours and report the totals by town and by vehicle. In addition, a socioeconomic report is produced that breaks down trips by age, family income, availability of other transportation (i.e., client possesses a driver's license or owns a car), and physical disabilities. This report can be further broken down by trip purpose.

Another monthly routine invoices the Department of Public Welfare for authorized medical trips taken during the month. In addition, this routine produces a report that summarizes medical trips taken by clients of more than 60 years of age during the past month. This report is used to justify Elder Service invoices, which are produced manually.

The month-end purge deletes the past month's records from the trip-transaction file and the daily vehicle log. Before they are purged, daily vehicle data are summarized and posted to the vehicle record. The operator can elect to produce a year-to-date vehicle performance report from these data at any time.

The annual or year-end routine clears vehicle maintenance and revenue data from the vehicle and town files.

Client Billing

The client billing system as called for by the proposed fare-collection system begins in the request-for-service routine. When a client requests a trip, the system will automatically decide whether the trip is billable, based on the type of client and the trip purpose. If the trip is billable (e.g., a shopping trip), the system will determine the trip distance from the trip matrix file. A trip cost will be calculated and displayed on the screen. This allows the client to cancel prohibitively expensive trips. Another feature of the request-for-service routine is the ability to offer group discounts such that a discount percentage can be applied while the trip is being booked. The discounted cost will also be displayed. Finally, the date of the oldest unpaid invoice will be displayed, which gives the operator an opportunity to inquire about payment of overdue invoices.

The next step in client billing is the actual invoice printing. In the proposed system, an invoice will be produced that gives a line-by-line breakdown of all charges incurred by the client during the past month, along with all past-due charges. These charges will be summarized to show a total due charge. The invoice will also cite any nonbillable trips and any group discounts that may have been granted. In addition to the printed invoice, an invoice record that summarizes the past month's charges will be added to the invoice file. All invoices remain on file until they are paid.

A cash-receipts-or-payment routine will allow the operator to enter payments and post them to unpaid invoices. For each entry, a payment record will be created and it will be stored in the invoice file. At the same time, the payment will be added to the appropriate revenue accumulator in the town file. A cash-receipts journal will be printed that shows all payments received during the day's processing.

An adjustment routine will allow the operator to enter credit and debit memos to adjust for overcharging and undercharging. For each such entry, a memo record will be created and posted in the invoice file. The debit memo will be treated as an additional charge to the client, and the credit memo can be used as a payment to be applied to an open invoice.

During month-end processing, payments will be matched up with invoices and those invoices that are paid in full will be purged along with the associated payments. After the invoice file has been purged, an "aging" report will be generated. This report will show the age of all open invoices by client. The report can be used by management to initiate appropriate dunning action on delinquent accounts.

CONCLUSIONS AND RECOMMENDATIONS

1. Fiscal austerity at all levels of goverment, increasing operating costs, and increasing demands for service will lead to the need for greater fare-box revenues.
2. Higher costs to consumers will increase the importance of equity in generation of fare-box revenues.
3. Equitable fare structures should ideally take into account the number of trips taken, miles of service received, ability to pay, and group riding (3).
4. Fare policies that take these four factors into account may be cost effective when they are implemented as part of a comprehensive computer-based MIS.
5. Such fare policies will be more cost effective when (a) trip lengths vary significantly, (b) both the general public and elderly or handicapped clients are served, (c) vehicles are centrally dispatched, and (d) many vehicles are dispatched from a single office.

Recommendations for further research include

1. Evaluation of the equity and practicality of alternative fare policies;
2. Evaluation of possible strategies for implementation of such fare policies, e.g., punch passes, tickets, fare box, manual invoicing, invoicing by means of off-site batch processing, use of single-function dedicated computer systems, and (as recommended) use of a comprehensive MIS;
3. Evaluation of alternative hardware and software options for implementation of alternative fare

policies by using a MIS strategy, e.g., micro, mini, and main-frame computers and alternative programming languages;

4. Evaluation of the suitability of implementation strategies for alternative fare policies under various constraints, e.g., fleet size, labor rates, and decentralized or centralized dispatching;
5. Evaluation of the effects of alternative fare policies on consumer behavior, e.g., ridership, trip lengths, and travel patterns; and
6. Evaluation of the feasibility of alternative fare policies from a public policy standpoint, e.g., rider acceptability, acceptability to policymakers, acceptability to funding agencies, overall effect on subsidy requirements, and efficient use of available subsidy funds.

The research of such topics could be extended through consideration of such concepts as demand elasticity, utility maximization, social benefit, market segmentation, service coordination, and funding coordination.

It should be noted that the existence of the computer-based MIS now serving CCRTA's b-bus program presents a tremendous opportunity for the research efforts recommended above. Grant funds would not be required for purchase of hardware and could be spent entirely on the research recommended.

ACKNOWLEDGMENT

The guidance and assistance of Alice Kidder of Syracuse University and Paul Bushueff of the Transportation Systems Center are gratefully acknowledged.

REFERENCES

1. R.P. Warren and J. Collura. Data Recording and Evaluation: The Barnstable County Experience. TRB, Transportation Research Record 696, 1978, pp. 58-65.
2. R.P. Warren, A.D. Rogers, and J. Collura. Coordination Through Consolidation: The Barnstable County Experience. Proc., 21st Annual Meeting of Transportation Research Forum (Philadelphia), Richard B. Cross Co., Oxford, IN, 1980, Vol. 21, No. 1, pp. 21-29.
3. J. Collura and D.F. Cope. Applications of the Concept of Equity to Cost Allocation. Civil Engineering Department, Univ. of Massachusetts, Amherst, 1979.
4. Cape Cod Regional Transit Authority: Accounts Receivable System Feasibility Study. Crosbro, Inc., Brockton, MA, Aug. 1980.

Publication of this paper sponsored by Task Force on Local Transportation Finance.

Income Equity of Two Transit Funding Sources

STEVEN M. ROCK

Currently, a number of funding sources are used to subsidize public transit. These originate at all levels of government, and their mix differs greatly among regions. Each source or combination has implications for equity that are often overlooked since each has a unique incidence, i.e., pattern of who pays by income group. The purpose of this paper is to examine the incidence of two commonly used sources: a sales tax and a motor fuel tax. Previous studies of the incidence of these taxes are not comparable; what is necessary is a single source of data on which to examine them. Suitable data to calculate incidence are available from the 1972-1973 Consumer Expenditure Survey of the Bureau of Labor Statistics, a comprehensive source of information on consumption expenditures by detailed items and income for 40 000 U.S. families. These data allow the relative percentage of income paid as sales or motor fuel tax to be calculated. The results indicate that both sources are regressive. Use of the S-index of progressivity for comparison suggests little short-run difference in income equity between the two (although exactly what items are subject to the sales tax can affect the results). The study points out that the equity impact of potential funding sources should be understood, available, and part of the decision-making process.

Transit systems throughout the United States have become increasingly dependent on subsidies from various levels of government. Each system tends to have a unique set of funding sources that is usually determined by law and politics in a particular geographic area. As new and expanded sources of transit funding are sought, the equity issue of who is paying from each source (the incidence) is often overlooked.

In addition, great concern is placed by federal agencies to ensure that their funded activities comply with Title VI of the Civil Rights Act of 1964. As an example of this concern, the Urban Mass Transportation Administration (UMTA) issued Circular 1160.1 (December 1977). A number of the objectives of this circular relate to this issue of equity of federally funded activities. Although most of the emphasis of Title VI has been on the distribution of benefits, a less obvious but related potential inequity involves the distribution of burdens. That is, Who pays for transit and what are the equity implications of different funding sources? A complete examination of equity would thus involve analysis of both who pays and who benefits. This paper attempts to shed light on a portion of the former aspect of this issue, recognizing that it is only a piece of the total equity problem.

Recent legislastion has changed the funding mechanism used to provide subsidies for public transit in the Chicago area. The essence of the change was that a 5 percent tax on motor fuel was eliminated; a general sales tax increase was substituted (1 percent in Cook County, 0.25 percent in the adjacent five counties). The main purpose of this change was apparently to generate more funds. In addition, an issue of geographic equity (the relationship between the funds raised and the funds expended in an area) was addressed. However, very little analysis has been undertaken to determine the income equity (who pays versus who benefits by income groups) of the funding switch.

The purpose of this paper is to explore the equity of two common sources of transit subsidies suggested by the Chicago Area Regional Transit Authority's funding switch from a sales tax to a motor fuel tax. A recent survey by the American Public Transit Association (APTA) (1) listed 24 regions that use a sales tax and 5 areas that obtain transit funds through a gasoline tax. It will be assumed that the revenue raised under each source would be structured to be similar. Since the basic groups that benefit from the subsidy funds remain the same, only the groups that pay for the subsidies will be examined. In economic terms, one wishes to compare the differential tax incidence of one source (e.g., motor fuel tax) with the incidence of the other source (general sales tax). Incidence refers to who (ultimately) bears the burden of the taxes, i.e., who pays.

Obviously, different sources will have different incidence. The initial distribution of liabilities (statutory incidence) can differ significantly from the final distribution (economic incidence). This will be true to the extent that a chain of adjustments by consumers or firms ensues. For simplicity as well as data limitations, it will be assumed that the sales and motor fuel taxes fall totally on the consumer. In fact, consumers may make some adjustments, such as the amount or location of gasoline or taxable goods purchased. However, previous studies of incidence have allocated sales and excise taxes to those who purchase the taxable products. Since these taxes are assumed to fall completely on consumers, incidence can be determined by noting the amount of each tax paid by consumers in each income level. It is noted that the incidence of multiple funding sources can be determined by combining and weighting the data from the individual sources used. Detailed discussion of theoretical issues in tax incidence can be found in most texts on public finance, e.g., Musgrave and Musgrave (2).

PREVIOUS STUDIES OF SALES AND FUEL TAX INCIDENCE

A number of previous studies have examined the incidence of the sales tax under different bases (i.e., items that are subject to tax). Although most of the studies took place in the 1960s, their conclusions were similar: The sales tax is regressive; that is, the tax paid by a lower-income family represents a larger percentage of income than that paid by a higher-income family. For example, Musgrave and Musgrave (2) used 1968 data to estimate that families in the lowest annual income bracket (under $4000) paid 3.4 percent of their income for general sales taxes; as incomes rose, this percentge fell continuously (to 0.3 percent in the $92 000 and over bracket). A second study by Pechman and Okner (3) reached the same conclusion. By using 1966 data, they found that families in the lowest annual income bracket (under $3000) paid 9.4 percent of their income for general and specific sales and excise taxes. Families in the highest income bracket ($1 000 000 and over) paid 1.0 percent for these taxes. Similarly, a study by the Advisory Commission on Intergovernmental Relations (4) used allowances by the Internal Revenue Service for sales tax deductions to obtain like results.

A handful of studies have looked at the incidence of a motor fuel tax. Most lump this tax together with other goods that are selectively taxed, such as cigarettes, alcohol, and public utilities [e.g., Musgrave and Musgrave (2)], or combine all sales and excise taxes together [e.g., Pechman and Okner (3)]. Probably the most comprehensive analysis of gasoline tax incidence was reported by Freeman (5). Freeman used 1972 household data provided by the Brookings Institution and an assumed tax of $0.20/gal (although the results would be representative of any tax that would be proportional to usage) and obtained a pattern that is slightly progressive except at either end of the income distribution [(5, p. 189); relative incidence compares the implicit tax rate of all income brackets with that of the highest income bracket]:

1972 Income ($000)	Relative Incidence
<2	4.16
2-4	2.71
4-6	2.68
6-8	3.00
8-10	3.23
10-15	3.45
15-20	3.52
20-26	3.42
26-50	2.94
<50	1.00

Zupnick (6) used a four-step process to examine the 1971 incidence of a tax-induced $0.10/gal price rise. Starting with average fuel use by automobile model year, he combined data on average miles driven by income group with ownership of each automobile model year by income class. The results indicate progression in the lower-middle brackets but regression in the income brackets above this (6, p. 412):

1971 Income ($000)	Relative Incidence
<3	1.53
3-4	1.33
4-5	1.56
5-6	1.76
6-7.5	1.85
7.5-10	1.54
10-15	1.26
>15	1.00

Finally, the Institute of Public Administration (7) examined incidence in a much less detailed manner by using data from the Motor Vehicle Manufacturers Association. For five income-bracket quintiles, the incidence of gasoline expenditures (and therefore taxes that would be proportional to expenditures) was found to be extremely regressive.

Other potential sources of data on gasoline expenditures are deficient or duplicative in some manner. For example, the U.S. Department of Energy (8,9) offers data on the distribution of gasoline consumption for households that own vehicles or that use gasoline (but not for all households). Their model employs a synthetic data base; the distribution of gasoline expenditures implied is quite similar to that reported in the Consumer Expenditure Survey used below. The old Federal Energy Administration had a Household Energy Expenditure Model; however, the basic data input source was the 1970 census, and mean income in each income bracket was not reported (10). The Survey Research Center of the University of Michigan has also analyzed household behavior for a number of years. However, incidence of gasoline expenditures was calculated by assuming the same miles per gallon for all vehicles, which clearly would bias the results (11).

The problem with previous studies is that it is very difficult to compare the incidence results of one tax source with the results of any other source. Each study used a different set of data, different time periods, different definitions of income, etc. In order to effectively compare the incidence of two or more taxes, a single set of data is necessary.

INCIDENCE OF SALES VERSUS MOTOR FUEL TAX

In order to draw an income profile of who pays the motor fuel tax and compare it with the impact of a sales tax, a suitable single source of data must be obtained. One source that will allow this to be undertaken is the 1972-1973 Consumer Expenditure Survey of the Bureau of Labor Statistics (BLS). This survey describes itself as the only comprehensive source of detailed information on expenditures and income related to socioeconomic and demographic characteristics of U.S. families. For a sample of 40 000 families, consumption expenditures by detailed items (i.e., the average dollar amount spent by a family on good or service X) were compiled and classified by income bracket. The BLS data sources are the interview survey (20 000 families) (12) and the diary survey (20 000 families (13). The former provided the primary data for this study; the latter supplemented and expanded the available categorization. Although the BLS data and the empirical analysis below used a national focus, it would be straightforward to adapt the technique to a study of incidence in a particular region. The BLS has recently released data that are drawn from and reported by particular standard metropolitan statistical areas (SMSAs).

To determine incidence of the sales tax, it would be necessary to total the consumption expenditures on items that are subject to a general sales tax (by income group). For the motor fuel tax, the amount spent on motor fuel needs to be noted. Since different regions subject different items to a sales tax, two cases were tested. In variant 1, each consumer expenditure item was considered as to whether it was subject to the Illinois Retailers' Occuption Tax (general sales tax). Illinois (and 7 other states) tax both food purchased for home consumption and prescription drugs. A second variant was compiled for an area where food consumed at home and medicine and drugs were not subject to a sales tax, as is the case in 23 states.

The sample population was ranked by income deciles, from the families with the lowest 10 percent of income (decile 1) to those with the highest 10 percent (decile 10). The total dollar amount of spending on taxable items was estimated for each decile. Since the sales tax is included in this spending and represents a flat percentage of the total, it was not necessary to separate the tax out. That is, it is sufficient to look at spending on taxable items as a percentage of income by deciles to determine the incidence of the sales tax. This information is displayed in Table 1. The average consumption expense (expenditures) by income deciles is broken into consumption exempt from sales tax and spending subject to (and including) sales tax. The relative incidence compares taxable expenditures as a percentage of income for each decile with that of the highest decile. Since the sales tax would be a flat percentage included in taxable expenditures, the relative incidence for both total taxable expenditures and sales tax payments as percentages of income will be the same. The results confirm the conclusion that the general sales tax is regressive. That is, those in the lowest income decile pay 2.5-3 times as much of their income in sales tax as do those in the highest income decile.

To derive the incidence of the motor fuel tax is a somewhat simpler task. The BLS data report on dollar expenditures for gasoline (including tax) by income group. Since this tax would be proportional to use, it also does not have to be separated out in order to examine incidence. Table 2 displays this information: average consumer expenditures on gasoline and fuels for vehicle operations by income decile, this expenditure as a percentage of income, and the relative incidence. It reveals a regressive tax; consumers in the lowest income decile pay 3.5 times as much of their income in gasoline tax as those in the highest decile do.

There appears to be some discrepancy between the previous studies of motor fuel tax incidence (text tables above) and the results presented in Table 2. This may be due in part to the choice of income brackets used in the previous studies, which do not match closely either the income of the population deciles of Table 2 or the definition of income. In addition, some of the previous studies' assumptions (e.g., fuel economy being the same for a model year for all cars) could lead to biased results.

A number of caveats exist in the empirical analysis above. First, since each regional sales tax includes or excludes a unique set of goods and services, sales tax incidence can differ somewhat between regions. Next, it is not possible to be completely accurate in excluding those expenditures that are not subject to a general sales tax (or in including those that are) since some expenditure categories listed by the BLS include both taxable items (e.g., parts) and nontaxable items (e.g., labor). However, these ambiguous categories are relatively small compared with those that are unambiguous (e.g., housing expenses), and the bias is apt to be minimal. Third, the data were collected in 1972-1973. To the extent that consumer expenditure patterns have changed, the incidence could change, e.g., How have different income groups responded to the large price increase in gasoline? The use of a single year's income can be criticized as unrepresentative of a longer-run view of income. Unfortunately, no data are readily available to correct this. Finally, it is assumed that all gasoline purchases are made by households or, alternatively, that the tax is levied only on consumer purchases of gasoline. In fact, approximately 68 percent of motor fuel was consumed by automobiles in 1976; most of the remainder was used by trucks.

It is also assumed that, in response to any change in tax levels, households continue to buy gasoline and consumption goods subject to the sales tax in the same proportion and geographic area that they did before. Any other assumptions would vastly complicate empirical calculations. In effect, these assumptions look at short-run incidence, assuming that the price elasticity of demand (sensitivity of quantity demand to price changes) is zero. The price elasticity of gasoline is fairly low (-0.2 to -0.4). A couple of tax studies have attempted to ascertain the impact of competition from firms in areas not subject to a particular tax in proximity to firms in areas that are subject to a tax. This could alter the burden of the tax by affecting how much of the tax gets shifted to the consumer. Unfortunately, these studies did not attempt to determine who bears the unshifted portion of the burden. They did suggest that proximity to a political border where no (additional) tax is levied does reduce the ability to shift the burden to the consumer.

DISCUSSION OF RESULTS

When the relative incidences of both taxes are compared, the following conclusions become apparent. First, both sources are regressive. The motor fuel tax is roughly as regressive as variant 1 of the sales tax; variant 2 is somewhat less regressive. A glance at the lower-middle income deciles of variant 1 and the fuel tax (second through fifth decile) shows that the relative incidence of the latter is close to proportional; the former is more regressive in this range.

A second way of comparing the incidence of different funding sources is to use the S-index developed by Suits (14). The S-index is a quick, convenient, one-number way of comparing incidence. There is no other generally accepted index of progressivity. This index, similar to the Gini ratio of income distribution equality/inequality [cited in Mendershausen (15)], ranges from +1 (extreme progressivity) to -1 (extreme regressivity). A proportional source would have an S-index of zero. Use of this index requires that families be ranked by income percentiles, from lowest to highest, and have their contribution to each source

Table 1. Expenditures and sales tax incidence.

	Income Decile									
Item	1	2	3	4	5	6	7	8	9	10
Variant 1										
Expenditures ($)	3037	4026	5161	6299	7417	8348	9472	10 578	12 168	16 015
Tax exempt	1630	2054	2560	3037	3491	3846	4333	4704	5503	7339
Taxable	1407	1972	2601	3262	3926	4502	5139	5874	6665	8676
Average income ($)	1559	3268	5081	7063	9112	11 244	13 466	16 116	19 747	31 974
Taxable expenditures divided by income (%)	90.3	60.3	51.2	46.2	43.1	40.0	38.2	36.4	33.8	27.1
Relative incidence	3.33	2.23	1.89	1.70	1.59	1.48	1.41	1.34	1.25	1.00
Variant 2										
Expenditures ($)	3037	4026	5161	6299	7417	8348	9472	10 578	12 168	16 015
Tax exempt	2214	2871	3537	4098	4680	5131	5749	6221	7139	9168
Taxable	823	1155	1624	2201	2737	3217	3723	4357	5029	6847
Average income ($)	1559	3268	5081	7063	9112	11 244	13 466	16 116	19 747	31 974
Taxable expenditures divided by income (%)	52.8	35.3	32.0	31.2	30.0	28.6	27.6	27.0	25.5	21.4
Relative incidence	2.47	1.65	1.50	1.46	1.40	1.34	1.29	1.26	1.19	1.00

Note: Consumption and income figures are from the 1972-1973 Consumer Expenditure Survey (11, 12).

Table 2. Motor fuel expenditures and income.

	Income Decile									
Item	1	2	3	4	5	6	7	8	9	10
Expenditures for gasoline ($)	98	132	208	270	336	394	449	480	525	561
Gasoline expenditure divided by income (%)	6.3	4.0	4.1	3.8	3.7	3.5	3.5	3.0	2.7	1.8
Relative incidence	3.50	2.22	2.28	2.11	2.05	1.94	1.83	1.67	1.50	1.00

Note: Source of gasoline expenditure figures is the 1972-1973 Consumer Expenditure Survey (11, 12).

noted. The S-index for any tax can be estimated (for 10 income deciles) as

$$S \approx 1 - (1/5000)\left\{\sum_{i=1}^{10} (1/2)\,[T_x(Y_i) + T_x(Y_{i-1})]\,(Y_i - Y_{i-1})\right\} \qquad (1)$$

where

x = given funding source x,
Y_i = income decile i, and
T = cumulative percentage of funding source paid by cumulative percentage of total income represented by income decile i.

Computation of the S-index for the two variants of the sales tax and the motor fuel tax reveals that all three are regressive. Surprisingly, the motor vehicle tax registers slightly greater regressivity (S = -0.16) than either variant 1 (S = -0.13) or variant 2 (S = -0.09) of the sales tax. However, due to the caveats mentioned above, the differences among S-indices are probably not significant. Previous studies that used 1966 and 1970 data have computed S-indices for sales taxes of -0.15 to -0.16. To give an idea of the S-index range for typical taxes, the variation is from about -0.40 to +0.40.

CONCLUSIONS

The results suggest little short-run impact on the distribution of income from changing funding sources between sales and motor fuel taxes. In addition, the magnitude of these taxes collected from transit is not particularly large. For example, a 1 percent sales tax would cost a first-decile family around $9-15/year or a tenth-decile family $74-93/year (Table 1). A levy of 5 percent on motor fuel usage would cost a family $5-28 in tax. In addition, the data suggest that to exempt food purchased for home consumption and drugs and medicine (variant 2) would be less regressive than one that taxes these items.

A change from a sales to fuel tax (or vice versa) affects not only income groups but also the sectors of society that pay the subsidy. If a motor fuel tax is imposed, redistribution stays within the transportation sector; automobile users pay and transit users receive primary benefits (others may benefit as well). If a general sales tax is used, however, the redistribution involves other sectors: from all consumers to transit users.

Governments have a wide menu of sources available to obtain funds. Although issues such as the revenue-raising potential of each source or how much would be raised in each geographical area by source are important, the income incidence of different taxes should not be overlooked.

ACKNOWLEDGMENT

This research was funded in part by a grant from UMTA. The views and conclusions are mine alone. Helpful comments and suggestions were provided by W.C. Hall, Joel Gibbons, and others.

REFERENCES

1. A Survey of Local Mechanisms for Financing Transit Operating Costs. American Public Transit Association, Wasington, DC, 1979.
2. R.A. Musgrave and P.B. Musgrave. Public Finance in Theory and Practice, 3rd ed. McGraw-Hill, New York, 1980.

3. J.A. Pechman and B.A. Okner. Who Bears the Tax Burden? Brookings Institution, Washington, DC, 1974.
4. Significant Features of Fiscal Federalism. Advisory Commission on Intergovernmental Relations, Washington, DC, 1974.
5. A.M. Freeman III. The Incidence of the Costs of Controlling Automotive Air Pollution. In The Distribution of Economic Well-Being (F.T. Juster, ed.), National Bureau of Economic Research, Ballinger Publishing Co., Cambridge, MA, 1977.
6. J.W. Zupnick. The Short-Run Incidence of a Tax-Induced Rise in the Price of Gasoline. Journal of Economic Issues, Vol. 9, No. 2, June 1975, pp. 409-414.
7. Institute of Public Administration. Financing Transit: Alternatives for Local Government. U.S. Department of Transportation, DOT-I-7916, July 1979.
8. Distributional Aspects of Alternative Energy Price Scenarios. Energy Information Administration, U.S. Department of Energy, Service Rept. DOE/EIA-0182/3, April 1979.
9. Distributional Aspects of Standby Gasoline Rationing Plan. Energy Information Administration, U.S. Deprtment of Energy, Service Rept. DOE/EIA-0176/1, April 1979.
10. Task Force on Distributive Impact of Budget and Economic Policies. Distributive Impacts of Proposed Changes in National Energy Policies. Committee of the Budget, U.S. House of Representatives, 1977.
11. Survey Research Center. Five Thousand American Families: Patterns of Economic Progress. Institute for Social Research, Univ. of Michigan, Ann Arbor, Vol. 3, 1975; Vol. 4, 1976; Vol. 6, 1978; Vol. 8, 1980.
12. Bureau of Labor Statistics. Consumer Expenditure Survey: Interview Survey, 1972-1973. U.S. Department of Labor, Bull. 1997, 1978.
13. Bureau of Labor Statistics. Consumer Expenditure Survey: Integrated Diary and Interview Survey Data, 1972-1973. U.S. Department of Labor, Bull. 1992, 1978.
14. D.B. Suits. Measurement of Tax Progressivity. American Economic Review, Vol. 67, No. 4, Sept. 1977, pp. 747-752; Vol. 70, No. 1, March 1980, p. 211.
15. H. Mendershausen. Changes in Income Distribution During the Great Depression. National Bureau of Economic Research, New York, 1946.

Publication of this paper sponsored by Committee on Taxation, Finance, and Pricing.

Economic Analysis of Transportation Pricing, Tax and Investment Policies

DALE O. STAHL II

In response to the ad hoc nature of current transportation user charges and cost allocations, a rigorous analytical framework is presented based on economic welfare theory. A multimodal transportation system model that has explicit price and tax, investment maintenance, service quality, and externality variables is formulated; the optimal decision rules of equating marginal social benefits and marginal social costs are derived and given operational interpretations. Optimal and administratively feasible aggregate prices by user class and mode are derived in terms of aggregate marginal social costs that are not impractical to estimate. An optimal cost allocation is defined as marginal social-cost pricing followed by general taxation of consumer goods (excluding transportation) to cover any deficit.

Considerable confusion exists about economic principles as they are applied to transportation policy analysis. Although a correct operational definition of marginal cost is hard to find in the literature, it is widely assumed that the marginal-cost pricing principle is not relevant to transportation facilities for a number of alleged reasons, e.g., there is no feasible way to (a) cover full costs or (b) implement ideal marginal cost pricing. The principles that find their way to practitioners suggest ad hoc rules of thumb rather than deduced results from a unified theory.

The purpose of this paper is to present an integrated economic transportation model that will clear up some of the confusion and serve as a basis for policy analysis. The model is set in the framework of welfare economics, and the results can be interpreted as the well-known principle of equating marginal social benefits with marginal social costs. Moreover, these concepts and principles are brought in touch with reality by the detailed structure of the model. All relevant investment and maintenance variables of a multimodal transportation system are incorporated in the model; service quality attributes and externalities are made explicit.

The results reported here are a summary of several aspects of an extensive working paper (1). Optimal decision rules for investments, maintenance programs, and prices are derived and interpreted. "Second-best" issues are discussed. An original contribution is the derivation of optimal and administratively feasible aggregate prices by user class and mode. Finally, an optimal cost allocation is defined as marginal-social-cost pricing followed by optimal taxation of consumer goods (which excludes transportation) to cover any deficit.

INTEGRATED MODEL

The task of this section is to model the transportation system and its effects in a manner that facilitates the application of economic welfare theory to transportation policy issues. The level of detail is sufficient for addressing the issues of investment and maintenance policy, service quality and externalities, pricing and cost allocation, and intermodal effects.

The welfare optimization problem can be stated in operations research terms as maximizing a social objective function with respect to control variables subject to constraints given by the system. The direct arguments of the objective function are called the impact variables. The role of the system model is to relate the impact variables to the control variables.

Description of the Model Variables and Relationships

Social Objective Function

The objective is to invest, maintain, and price in order to obtain the highest level of net social welfare possible. The problem can be formalized by letting W(U) be a Bergson-Samuelson type of social welfare function, where U is a vector of individual utilities. As is often done, and to keep this exposition simple, utility indices in terms of dollar values will be assumed, and the social welfare function will be the unweighted sum

$$W(U) = \sum_{i=1}^{N} U_i \tag{1}$$

It should be noted that, although the principles developed in this paper are generalizable to other social welfare functions, the functional form of specific results are critically dependent on this distribution-neutral social welfare function.

The benefits to businesses that use the transportation system must also be included. If a competitive private sector and the distribution-neutral social welfare function are assumed, it is sufficient to assess these benefits at the stage of the businesses as users rather than attempt to trace the incidence through to customers and stockholders. To keep the notation simple, businesses will be included in the set of N individuals (or agents) over which utilities are summed, with the understanding that (for a business) U_i denotes initial benefits as profits.

Two major complicating features of an economic analysis of the transportation system are that the transportation "good" has multiple characteristics and that consumption of transportation services generates numerous externalities. Lancaster's formulation of consumer theory (2) is well suited to handle the multiple-characteristic aspects, and it can easily accommodate the incorporation of externalities. Suppose individual utility (and business profits) can be represented as an explicit function of two sets of variables: U(q,E). Let q be a vector of trip and service-quality characteristics such as trip destination benefits, travel time, operating expenses, safety, comfort, and aesthetics. Let E be a matrix of nonuser externalities such as pollution and noise, with one column for each link in the transportation network. Since these public-good-type externalities depend more on the total traffic level than on any one individual's travel, it is reasonable to view such externalities as impinging on the individual in his or her capacity as a nonuser. In contrast, the trip and service-quality characteristics affect the individual only when the individual makes a trip. This representation admits the possibility that total travel level affects the service-quality level; the potential interuser externalities will be made explicit later.

In the manner of Lancaster, suppose there is a simple linear relation between trips taken and the amount of trip and service-quality characteristics derived from travel. In particular, suppose that $x_{j\ell}$ is the amount of trip and service quality (j) derived from one trip mile on link (ℓ) of the transportation network. (Throughout this paper, a "link of the transportation system" will denote a link of the transportation network, a structure, or a terminal.) Then q = Zx', where x is the row vector of trip miles by link and Z is the corresponding matrix $[z_{j\ell}]$. Thus, individual utility can be written as $U_i(Zx_i',E)$, where the subscripts denote individual (i).

From the representation of the social objective function, it is clear that the pertinent impact variables are Z and E. Further, from the point of view of the individual, x_i is a control variable. Let X be the matrix of trips by all individuals--one row for each individual.

The typical welfare optimization problem is to maximize the social welfare function subject to a number of constraints, among which is the technology constraint--the mathematical description of all technologically feasible combinations of all goods. In a market economy, a convenient way to represent this technology constraint is to require that the total cost of production of all goods equal a fixed nominal gross national product (GNP): $TC(X) = \bar{M}$, where $\bar{M}$ is the fixed nominal GNP and TC( ) is the total (private and public) cost of production of all goods in the economy. TC(X) should be understood to be the total cost of producing transportation services plus the total cost (net of transportation) of producing all final (or consumer) goods, where the final goods are understood to be implicit in this notation. For convenience of exposition, the total cost function is divided into public transportation costs, which are denoted by C( ), and all other public and private costs, which are denoted by c( ); hence TC( ) = C( ) + c( ). If public costs are expressed as gross costs before tax and fee receipts, then private costs clearly should not include taxes and fees.

For the social welfare function used in this paper, in which the social value of $1.00 accruing to anybody is $1.00, maximizing this social welfare function subject to the technology constraint is equivalent to maximizing

$$W(U) + [\bar{M} - TC(X)] \tag{2}$$

Time is implicit in the expression of the social objective function. It should be understood that U_i represents the discounted present value of the stream of dollar-value utilities and that the cost functions represent the discounted present value of the stream of costs, both computed at the same social discount rate.

Control Variables

A control variable is some quantity that the transportation agency has direct control over, e.g., resurfacing intervals and thickness, but not bumpiness. Although there may be a deterministic relation between transportation agency activities and the bumpiness of the road, it is best to define bumpiness as an intermediate state variable and to specify the engineering relationship between control variables proper and the intermediate state variables.

To avoid undue complications of notation, one variable label (s) will be used for two groups of control variables. The first group consists of changes that are determined by investment decisions (such as construction, widening, and resurfacing) and are fixed in the short run. For example, on a given road segment of the highway network, these would include changes in length in miles, number of lanes, width, pavement type and base, shoulder type, signs and signals, geometric design features, scenery, and speed limit. Variables for structures and terminals can also be included. The second group consists of variables that can be adjusted on a day-to-day basis--essentially maintenance activities. These two groups of control variables will be represented by the vector s_ℓ, which will be referred to as the transportation program strategy for a specific link (ℓ). Let S be the matrix of transportation program strategies--one column for each link of the transportation system. Then, the entire transportation construction, rehabilitation, and maintenance work program over time can be represented by S(t).

Intermediate State Variables and Relationships

Let y_ℓ be a vector of aggregate load on a link (ℓ)--for highways, specifically, the number of equivalent vehicles per hour and the equivalent 18-kip axle loads per hour. Let Y(t) be the matrix of aggregate load flows on the transportation system as a function of time.

Assume a simple linear relationship between trip demand by individuals and aggregate load:

$$y_\ell = Bx_\ell \tag{3a}$$

$$Y(t) = BX(t) \tag{3b}$$

where B is a matrix whose elements are equal to the contribution to aggregate load of one trip by a specific individual.

Let r_ℓ be a vector of pertinent serviceability attributes on a link (ℓ). There are two groups of pertinent serviceability variables that correspond to the two groups of control variables. The first group consists of the state variables that result from cumulative past investment decisions and are fixed in the short run; for highways, state variables include length of miles, number of lanes, width, pavement type and base, shoulder type, signs and signals, geometric design features, and scenery. The second group includes state variables affected by maintenance activities and traffic; for highways, state variables include bumpiness, skid resistance, hazards, condition of signs and signals, litter, and condition of rest areas. The American Association of State Highway Officials (AASHO) present serviceability index is a composite of some of these serviceability variables. Let R(t) be the matrix extension of r_ℓ for all links as a function of time.

The state of the system depends on the history of transportation activities and aggregate load flows on the system. One way to represent this relationship is by the differential equation:

$$\dot{R}(t) = F\,[Y(t), S(t), R(t), t] \tag{4}$$

where the dot over the variable denotes the derivative with respect to t and where F( ) is a general vector function that specifies the rate of change in serviceability as a function of instantaneous load, program strategy, serviceability, and time; F( ) can accommodate any interaction among the variables, including weather (through time).

To complete the model, a relationship is needed between the serviceability variables and the impact variables that has perhaps some dependency on the aggregate load. Specifically, assume that such relationships can be represented in the following forms:

$$Z = G(R,Y) \tag{5a}$$

$$E = H(R,Y) \tag{5b}$$

where, as usual, the capital letters denote the matrix extensions of the vectors to the entire transportation system and time is implicit. An approach similar to that of the Manual on User Benefit Analysis (3) can be employed to determine these relationships.

In summary, the control variables of the model are travel (X), which determines aggregate load (Y), and the transportation-program strategy (S). The serviceability of the system is related to the control variables by a dynamic equation. Serviceability and aggregate load determine the impact variables, which directly affect individual utilities.

Formal Statement of the Optimization Problem and First-Order Conditions for Optimality

A two-state optimization procedure is chosen because it handles the dynamics in a simple manner and is easier to interpret. The first stage is to solve the following public-cost-minimization problem, given a desired serviceability of the system, R(t), and actual loading, Y(t).

1. Minimize C(S) with respect to S(t) subject to $\dot{R} = F(Y,S,R,t)$.

This stage of the optimization process contains most of the complications, in that it contains the dynamics of the serviceability of the system. One could take the Hamiltonian-Lagrangian approach to solving this problem, but a heuristic approach provides more insight. Given R(t) and Y(t), there may be only a few or only one compatible strategy, S(t), i.e., a solution to the constraint equation. Once a set of compatible strategies has been found, it is a simple matter to pick the least costly strategy. Let S*(R,Y) be the least costly strategy compatible with Y(t) and R(t). Note that for some (R,Y) there may be no compatible strategy, in which case the cost is set equal to + ∞. The result of stage one is the minimum cost function

$$\bar{C}(R,Y) \equiv C[S^*(R,Y)] \tag{6}$$

2. The second stage problem is to maximize $\sum_{i=1}^{N} U_i(Zx'_i, E) - \bar{C}(R,BX) - c(X) + \bar{M}$ with respect to X,R subject to Z = G(R,BX) and E = H(R,BX), recalling that Y = BX.

Assuming that the set of feasible control variables (that satisfy the constraints) is compact and convex and that the net social objective function is quasi-concave, the solution to this stage amounts to equating marginal benefits to marginal costs. The solution (R*,X*) can be substituted into the solution of the first stage to obtain the optimal transportation-program strategy S*(R*,BX*). The choice of R* can be viewed as the choice of a set of serviceability standards by which the transportation agency can evaluate its performance and needs. A subset of the standards will provide guidelines for maintenance activities; the other standards will provide guidelines for the planning of new construction, improvements, and rehabilitation.

For the concepts of total cost and marginal costs of a good to be well defined, the good must be specified in terms of relevant characteristics, and these characteristics must remain fixed. This is precisely what $\bar{C}(R,Y)$ does. Given R and Y, all the service-quality characteristics are determined, so $\bar{C}(R,Y)$ can be interpreted as the minimum public cost of providing the transportation good defined by

(R,Y) or, equivalently, as the economic cost of providing the good.

For the purpose of writing the first-order conditions for optimality, it is convenient to introduce a number of simplifying assumptions and definitions. As an intermediate step, let $x_\ell = \sum_i x_{i\ell}$ be the aggregate travel on link (ℓ) and let $a_{h\ell} = (1/x_\ell)(\sum_i x_{i\ell}\partial U_i/\partial q_h)$ be the weighted social value of quality characteristic h, and assume that it is constant. In other words, assume a constant weighted average value of travel time, etc. Define a serviceability value index to be $V_\ell \equiv \sum_h a_{h\ell}z_{h\ell}$, which has the flavor of the imputed social value of a link per trip. In a similar fashion, let $w_{h\ell} = \sum_i \partial U_i/\partial E_{h\ell}$ be the aggregate social value of nonuser externality attribute h, assume it to be constant, and let $E_\ell \equiv \sum_h w_{h\ell}E_{h\ell}$ be an index of nonuser externalities of a link.

By using these definitions, the first-order conditions can be written as

$$x_\ell(\partial V_\ell/\partial r_{k\ell}) + (\partial E_\ell/\partial r_{k\ell}) = \partial\hat{C}(R,Y)/\partial r_{k\ell} \tag{7}$$

for all (k,ℓ), and

$$\begin{aligned}\sum_h(\partial U_{i'}/\partial q_h)z_{h\ell} = {} & [\partial c(X)/\partial x_{i'\ell}] + \sum_j\{[\partial\hat{C}(R,Y)/\partial y_{j\ell}] \\ & - x_\ell(\partial V_\ell/\partial y_{j\ell}) - (\partial E_\ell/\partial y_{j\ell})\}\,b_{ji'}\end{aligned} \tag{8}$$

for all (i',ℓ).

The left-hand side of Equation 7 is the marginal social benefit (including externalities) from a change in the serviceability of the system, and the right-hand side is the marginal public cost of such a change. The left-hand side of Equation 8 is the marginal private benefit of a trip, and the right-hand side is the marginal social cost (including externalities) of a trip by individual (i') on link (ℓ).

OPTIMAL DECISION RULES

The framework developed in the previous section can be used as the basis for deriving optimal decision rules. First, the conditions for the optimal transportation-program strategy and serviceability standards are interpreted and discussed. Next, optimal ideal and aggregate pricing rules are derived. In addition, second-best issues are briefly discussed.

Optimal Strategies and Serviceability Standards

The optimal serviceability standards are determined by the condition in Equation 7, which is in the familiar form of the Samuelsonian conditions for optimal production of a public good. The marginal social benefits (user and nonuser externalities) are summed over all individuals and equated to the marginal social costs, which is reasonable since serviceability has the character of a public good--all users of a particular link enjoy the same serviceability and all nonusers bear the same externalities. Marginal social benefits are equal to the sum of (a) the marginal effect of a particular serviceability attribute on the serviceability value index times the traffic volume and (b) the marginal effect of the particular serviceability attribute on total nonuser externalities. Marginal social costs are determined by first minimizing total costs (given hypothetical serviceability levels and aggregate loads) and then computing the marginal effect of a change in a particular serviceability attribute on the economic costs, given the same aggregate load. Given the optimal serviceability standards, the optimum program strategy is the minimum cost strategy given by the solution to the stage-one problem: S*(R,Y). The optimality conditions are valid if and only if either the aggregate load is also optimal or the aggregate load is independent of the serviceability. Deviations from these conditions will be discussed under second-best issues.

In general, investment and maintenance strategies are not separable. Moreover, in general, it is not possible to arrive at the optimum system solution by seeking a link-by-link solution. In practice, however, it may be reasonable to assume that the network spillover effects are confined to a manageable subnetwork.

Optimal Pricing Rules

First, the ideal optimal pricing rule is derived from the optimality conditions. Second, an aggregation assumption is made that leads to more feasible, but optimal, aggregate pricing rules. Finally, the issue of cost recovery from optimal pricing revenues is briefly addressed.

Ideal Optimal Pricing Rules

The condition in Equation 8 is the basis for ideal optimal pricing rules. Assume that the individual maximizes a utility function subject to a fixed-price budget constraint, where $p_{i'\ell}$ is the price that individual (i') must pay for a trip on link (ℓ). Then Equation 8 can be rewritten as

$$\begin{aligned}p_{i'\ell} = {} & [\partial c(X)/\partial x_{i'\ell}] + \sum_j\{[\partial\hat{C}(R,Y)/\partial y_{j\ell}] - x_\ell(\partial V_\ell/\partial y_{j\ell}) \\ & - (\partial E_\ell/\partial y_{j\ell})\}\,b_{ji'}\end{aligned} \tag{9}$$

for all (i',ℓ) [see Stahl (1) for a detailed derivation]. At this price, the individual would freely choose the amount of travel required by the social optimum.

The optimal pricing rule can be stated as a marginal-social-cost pricing rule. The first term on the right-hand side of Equation 9 is the private out-of-pocket cost of a trip, e.g., gasoline, oil, and fares. The second term is composed of three components. The first component is the marginal public cost of providing a fixed transportation serviceability level with respect to different traffic loads. The second and third components are the marginal externality costs of travel. The second term captures the effects of travel on the value of serviceability, such as travel time and accident rates; thus, this term includes the familiar congestion and safety externalities that users face. The third term is the marginal externality effect on nonusers; it includes such effects as pollution and noise. These three components are multiplied by the contribution of the individual (i') to aggregate loads.

Moreover, the optimal pricing rule can be stated as a short-run marginal-social-cost pricing rule in the sense that the serviceability characteristics (which it will be recalled include such items as the number of lanes) are held constant in Equation 9. One of the clearest arguments for short-run marginal cost pricing of highways was given by Walters (4). This short-run marginal-social-cost pricing rule is strictly valid if and only if either (a) the total transportation program strategy, all serviceability standards, and all other variables in the economy are optimal or (b) all suboptimal strategies, suboptimal serviceability standards, and other suboptimal variables in the economy are independent of travel demand. Deviations from this rule will be discussed further under second-best issues.

The expression for the optimal price given by Equation 9 is deceptively simple. Notwithstanding the practical difficulties of estimating the terms on the right-hand side, it should not be overlooked that the optimal price for an individual is a function of time and the particular link of the transportation system. For example, the optimal price on an urban road during rush hour may be quite high, whereas the optimal price on a rural road at 3:00 a.m. may be zero. Furthermore, the optimal prices are in units such as vehicle miles of travel and passages over structures, which suggests a different tax base than is currently used. The next topic is concerned with devising feasible prices and taxes.

Feasible Optimal Pricing Rules

Considering the infeasibility of the optimal pricing system derived above, it is imperative that a more feasible system be found that also has some optimality properties. The approach is to use prices explicitly as the control variables in lieu of the travel variables (X). Travel demand can be expressed as a function of prices, $x_{i\ell}(\underline{p})$, with all other variables implicit and perceived to be fixed. These demand functions can be substituted into the formal statement of the optimization problems. Since these results are an original contribution of this paper, the essentials of the derivation will be given.

In optimizing the stage-two problem with respect to transportation prices, the condition in Equation 8 is replaced by

$$\sum_{i'}\sum_{\ell'} p_{i'\ell'}(\partial x_{i'\ell'}/\partial p_{i\ell}) = \sum_{i'}\sum_{\ell'} \mathrm{MSC}_{i'\ell'}(\partial x_{i'\ell'}/\partial p_{i\ell}) \tag{10}$$

for all (i,ℓ), where $\mathrm{MSC}_{i'\ell'}$ is equal to the right-hand side of Equation 9, i.e., the marginal social costs of a trip by individual (i') on link (ℓ'). Let $\underline{x}$ be the vector of travel demands formed by stacking the columns of X, let $\underline{p}$ be the corresponding price vector and $\underline{\mathrm{MSC}}$ the corresponding marginal-social-cost vector, and let H be the Hessian matrix of partial derivatives $\left[\dfrac{\partial x_{i'\ell'}}{\partial p_{i\ell}}\right]$ for all (i',ℓ') and (i,ℓ). Equation 10 can then be written compactly as $H\underline{p} = H\,\underline{\mathrm{MSC}}$. The second-order conditions on individual utility maximization ensure that H is invertible, so the optimum prices can be solved for explicitly as $\underline{p} = \underline{\mathrm{MSC}}$, which is identical to the result of Equation 9, as should have been expected.

Unfortunately, this "ideal" price system is impractical. More-useful results are obtainable by imposing administrative feasibility constraints on the price system, such as a single aggregate price for a user class and mode.

Let v denote a particular user class (or vehicle class), and let m denote a particular mode (or subsystem of links). For each user in class v ($i \in v$) that takes a trip via a link of mode m ($\ell \in m$), we want a common price, p_{vm}, i.e., $p_{i\ell} = p_{vm}$ for all $i \in v$ and $\ell \in m$. In addition, let $x_{vm} = \sum_{i\in v}\sum_{\ell\in m} x_{i\ell}$ be the aggregate travel by user class (v) on mode (m), and let $\overline{\mathrm{MCS}}_{vm} = \sum_{i\in v}\sum_{\ell\in m} \mathrm{MSC}_{i\ell}$ (#vm) be the average marginal social cost of a trip by user class (v) on mode (m), where (#vm) is an abbreviation for the number of individuals in user class (v) times the number of links of mode (m). By using these definitions, Equation 10 can be rewritten as

$$\begin{aligned}\sum_{v'}\sum_{m'} p_{v'm'}(\partial x_{v'm'}/\partial p_{vm}) &= \sum_{i}\sum_{\ell} \mathrm{MSC}_{i\ell}(\partial x_{i\ell}/\partial p_{vm}) \\ &= \sum_{v'}\sum_{m'} \overline{\mathrm{MSC}}_{v'm'}(\partial x_{v'm'}/\partial p_{vm})\end{aligned} \tag{11}$$

where the summations over v' and m' mean all user classes and all modes. The second equality holds under the reasonable assumption that the deviations of $\mathrm{MSC}_{i\ell}$ and $\partial x_{i\ell}/\partial p_{vm}$ from their means are independent across all individuals and all links (ℓ). In the manner used for the ideal price system, aggregate vector notation can be introduced, and the aggregate Hessian matrix can be inverted to derive the optimal aggregate marginal-social-cost pricing rule:

$$p_{vm} = \overline{\mathrm{MSC}}_{vm} \tag{12}$$

for all user classes (v) and modes (m). [Aggregation by peak and off-peak periods can be handled in a similar way (1).]

Cost Recovery

Suppose optimal pricing is implemented. Is there any hope that full costs can be recovered? Under the assumption that the private cost function [c(X)] is homogeneous of degree one (i.e., constant returns to scale), full private costs would be recovered by marginal cost pricing. Thus, the issue is whether full public costs can be recovered. At the global optimum (R*,X*), sufficient conditions for full cost recovery are that $\hat{C}(R,Y)$ be homogeneous of degree one, that V(R,Y) be homogeneous of degree zero, and that E(R,Y) be homogeneous of degree zero. Under these conditions, the short-run marginal-cost pricing rule would generate a full cost allocation without worry about imputing any common capital costs to users.

However, there is evidence that suggests these homogeneity conditions are not likely to be satisfied. For example, there is evidence of substantial increasing returns to scale in highway pavement thickness. [As an example of how to estimate a component of marginal social costs, the appendix in Stahl (1) estimates the marginal pavement cost of highways by axle-weight class. It is concluded that the component of marginal cost pricing due to pavement wear is not likely to recover more than 10 percent of the cost of pavement rehabilitation.] Also, there appear to be increasing returns to scale in air pollution that are not offset by design considerations. In general, if the homogeneity conditions do not hold, then marginal cost pricing will not recover full costs. With the increasing returns to scale suggested, there will be a deficit. (This result applies equally to the ideal optimal prices and to the optimal aggregate prices.)

Second-Best Issues

The implications of second-best considerations on optimal decision rules have been addressed in detail by Stahl (1); space permits only a brief summary here. The deviation of second-best rules from first-best rules depends on the suboptimality of the investment and pricing rules actually used by the transportation agency. Sound policy advice consists of advocating both optimal investment rules and optimal pricing. If the transportation agency makes a sincere effort to design its program optimally, then even if the existing system is suboptimal and even if the agency makes (uncorrelated) mistakes, the optimal rules are still the marginal-social-cost rules given by Equations 7-9 and 12.

On the other hand, if the transportation agency has a tendency to persist, for example, in overbuilding the highway system and underbuilding the mass transit system, then the optimal prices are higher for highways and lower for transit than the short-run marginal social costs. As another example, if urban roads are consistently underpriced and mass transit overpriced with respect to the marginal social costs, then, based on actual travel demand, urban roads should be underbuilt and transit should be overbuilt to compensate for the suboptimal prices.

A more recent second-best issue concerns optimal deviations from marginal cost pricing to cover nonallocable costs. The issue arises from the realization that the marginal-cost pricing rules generally require lump-sum taxes to cover deficits and that in reality there exists no such thing as a lump-sum tax (i.e., a tax that does not affect the relative prices of goods). The problem of devising a system of taxes on commodities that covers the deficit and causes the least loss in social welfare has been recently addressed in the economics literature; for an excellent survey, see Sandmo (5). A widely popularized result is the "inverse elasticity rule"; this rule states that, if all cross-price elasticities are zero, then the optimal deviation from marginal cost pricing is proportional to the inverse of the own-price elasticity for each commodity. This result has been loosely applied to the problem of highway cost allocation (6). However, the application of this rule to transportation is invalid for the following reason.

A fundamental result of the theory of optimal taxation is that production efficiency is always desirable. Production efficiency requires that all producers face the true marginal social costs of all inputs; therefore, intermediate goods should not be taxed. Optimality calls for taxes on final goods or primary factors, not both, and not on intermediate goods. For the most part, transportation is an intermediate good. Certainly, all business uses of transportation qualify as intermediate goods, and all work commutes should also be considered intermediate goods. Whether to count shopping trips and recreation-destination trips as intermediate or final is debatable. Perhaps only the classic Sunday drive is unambiguously a final good. Thus, it appears that all but a small and perhaps insignificant portion of travel would be exempt from optimal taxes. Thus, a feasible optimal tax system to cover the deficit of the transportation agency would call for no user taxes. (It is necessary to emphasize that "tax" here means any additional payment above the marginal social cost, not to be confused with the optimal price charged by the government.)

Notwithstanding these remarks, if the government should decide on a user-only tax scheme for transportation, the best scheme (in terms of least welfare loss) could be determined by methods analogous to those used in optimal taxation theory. Optimal taxation of intermediate goods is an unsolved problem because of the complex ways such a tax works through the economy and affects final-good prices. Research on this problem is needed and would have important policy implications.

CONCLUSIONS

The integrated economic transportation model is well suited for considering investment, maintenance, pricing, and tax policies. (Cost allocation will also be discussed.) Optimal decision rules were derived in terms of the marginal social cost of travel. Marginal social costs include the marginal private costs, the marginal public costs defined to be the "economic" costs of providing a given quality of service, and the marginal externality costs, such as congestion and pollution.

To operationalize the optimal prices, the marginal social cost can be broken down into the marginal private cost (e.g., gasoline and oil) and the marginal public cost (government agencies and externalities). It is reasonable to assume constant returns to scale in the private sector, so full private costs will be recovered by marginal-private-cost pricing. The government prices can be set equal to the appropriately defined aggregate marginal public and externality costs. The units of the prices could be chosen to be, for example, vehicle miles by vehicle weight class. Prices based on these units could be assessed as part of the annual registration process. The practical issue of how to cover any deficit will be discussed next.

The common notion of a cost allocation is probably best described by a private-sector accountant's spread sheet in which several products are listed across the top and the customers are listed down the left-hand side. The costs of production are allocated to each customer so that the sum of each column is equal to the total cost of production for each product. The sum of each row is equal to the total charges for each customer. The spread sheet balances when the sum of all the columns (costs) equals the sum of all the rows (charges); then, one has a full cost allocation. In a competitive economy, a full cost allocation can be obtained by simply charging each customer the market price for each product, because price is equal to marginal cost, which is also equal to average cost.

Cost allocation is considerably more complex for public-sector activities because they usually involve public good aspects, externalities, and increasing returns to scale. In the presence of these complications, allocation by price is not likely to give a full cost allocation, and one must devise ways to apportion the deficit. The lack of a solid theoretical basis for a cost allocation underlies the criticisms of previous highway cost allocation (7,8).

The major intention behind a cost-allocation study is to provide information relevant to the formation of price and tax policy. But this objective can be met by the direct approach of determining the optimal price and tax policy. Then if one wants the information presented in a cost-allocation format, it can easily be done because any given price and tax policy implies a de facto cost allocation.

The optimal cost allocation is defined as the de facto cost allocation that corresponds to the optimal price and tax policy. The optimal cost allocation can be determined in two steps. First, allocate by pricing at marginal social cost and calculate the revenues and any deficit. Second, apportion the deficit by the principles of optimal taxation.

With a bit of imagination and study, it should be possible to devise an administratively feasible approximation to the optimal price and tax policy and cost allocation that would be superior to the ad hoc methods currently employed. Since this optimal tax system would exempt transportation as an intermediate good, the deficit would in essence be allocated among agents in their role as nonusers. This is really the only valid argument for nonuser taxes to help finance transportation. It arises not from arguing unsoundly that nonusers should help pay because they benefit from transportation but, rather, from arguing that the deficit should be allocated in such a way as not to distort the efficient use of productive resources. A critical review of the incremental-cost method, the benefit principle, and the newest congressionally mandated highway-cost-allocation study is given in Stahl (1). Before the principles advocated in this paper could be implemented, considerably more research is required to estimate properly defined marginal social costs.

ACKNOWLEDGMENT

I am grateful to the Institute of Urban and Regional Development and the Institute of Transportation Studies for financial support during this research and to Robert G. Harris for his encouragement and helpful criticisms. Naturally, all opinions, results, conclusions, and errors are my sole responsibility and do not necessarily reflect the views of the aforementioned.

REFERENCES

1. D.O. Stahl II. An Integrated Economic Transportation Model for Policy Analysis. Center for Transportation Policy Research, Univ. of California, Berkeley, Publ. Series 80-8, June 1980.
2. K. Lancaster. A New Approach to Consumer Theory. Journal of Political Economy, Vol. 74, April 1966, pp. 132-157.
3. A Manual on User Benefit Analysis of Highway and Bus Improvements. American Association of State Highway and Transportation Officials, Washington, DC, 1977.
4. A.A. Walters. The Economics of Road User Charges. International Development Association, International Bank for Reconstruction and Development, Johns Hopkins Press, Baltimore, Rept. EC-158, 1968.
5. A. Sandmo. Optimal Taxation: An Introduction to the Literature. Journal of Public Economics, Vol. 6, July-Aug. 1976, pp. 37-54.
6. K. Bhatt, M. Beesley, and K. Neels. An Analysis of Road Expenditures and Payments by Vehicle Class (1956-1975). Urban Institute, Washington, DC, March 1977.
7. Reforming Highway Cost Allocation Analysis. Association of American Railroads. Washington, DC, Staff Studies Group Memo 78-3, Feb. 9, 1978.
8. A Critique of Bureau of Public Roads-Federal Highway Administration Incremental Cost Assignment. Association of American Railroads. Washington, DC, Staff Studies Group Memo 78-6, March 15, 1978.

Publication of this paper sponsored by Committee on Taxation, Finance, and Pricing.