TRANSPORTATION RESEARCH RECORD 731

Evaluating Transportation Proposals

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1979

Transportation Research Record 731
Price $3:60
Edited for TRB by Susan Singer-Bart

modes
1 highway transportation
2 public transit
3 rail transportation
5 other

subject areas
12 planning
15 socioeconomics
16 user needs

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Academy of Sciences, 2101 Constitution Avenue, N.W., Washington, DC 20418.

Notice
The papers in this Record have been reviewed by and accepted for publication by knowledgeable persons other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The views expressed in these papers are those of the authors and do not necessarily reflect those of the sponsoring committee, the Transportation Research Board, the National Academy of Sciences, or the sponsors of TRB activities.

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Evaluating transportation proposals.

(Transportation research record; 731)
1. Transportation–Planning–Addresses, essays, lectures.
2. Transportation–Economic aspects–Addresses, essays, lectures.
I. Title. II. Series.
TE7.H5 no. 731 [TA1155] 380.5s [380.5'068]
ISBN 0-309-02986-4 ISSN 0361-1981 80-12434

Sponsorship of the Papers in This Transportation Research Recor

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

Social, Economic, and Environmental Factors Section
Clarkson H. Oglesby, Stanford, California, chairman

Committee on Social, Economic, and Environmental Factors
Hays B. Gamble, Pennsylvania State University, chairman
Walter H. Bottiny, Federal Highway Administration, secretary
M. E. Bond, Malcolm F. Brenan, Jesse L. Buffington, Jon E. Burk H. G. Dutz, John W. Fuller, John C. Goodknight, Harold B. Hand James M. Holec, Jr., Evan A. Iverson, Ata M. Khan, Snehamay Khasnabis, Walter H. Kondo, C. John Langley, Jr., Stephen C. Lo Philippos J. Loukissas, Paul R. Lowry, James T. McQueen, Kenne Nelson, V. Setty Pendakur, Owen H. Sauerlender, John H. Suhrbi David L. Weiss

Committee on Application of Economic Analysis to Transportati Problems
Charles W. Dale, Federal Highway Administration, chairman
Gary R. Allen, Martin J. Beckmann, Dietrich R. Bergmann, James Boos, James B. Chiles, Paul J. Claffey, Donald O. Covault, James Fay, Gerald A. Fleischer, Terry L. Friesz, Dan G. Haney, Loyd H Daniel L. Jones, Jr., Edward N. Kashuba, Ramesh Kher, Richard Leavenworth, Whitefield W. Mayes, Douglas S. McLeod, Deloy Ka Peterson, Merrill J. Roberts, Gabriel J. Roth, Martin M. Stein, Wil D. Weiss, Robley Winfrey

Committee on Citizen Participation in Transportation Planning
Kathleen Stein Hudson, Central Transportation Planning Staff, Bo chairwoman
Malcolm F. Brenan, West Virginia Department of Highways, secre
G. Robert Adams, Joseph M. Clark, John L. Coil, Sid Davis, Geor Duffy, Joel Ettinger, John W. Fuller, Beverly A. Harper, Stuart L. Robert B. Hixson, Julie Hetrick Hoover, Leroy E. Johnson, Danie McGeehan, Michael A. Perfater, Jacquelyn R. Smith, John H. Suh John F. Willey

Task Force on Economic Analysis
Dwight M. Blood, Colorado State University, chairman
Gary R. Allen, Virginia Highway and Transportation Research Co secretary
W. Bruce Allen, Daniel Brand, Ken Casavant, John J. Coyle, Curti Cramer, Charles W. Dale, John F. Due, Martin T. Farris, Hays B. G Charles A. Hedges, Lowell D. Hill, Damian J. Kulash, Edward Mar William S. Vickrey, Martin Wohl, John C. Spychalski

Floyd I. Thiel, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. organizational units and officers and members are as of December 1978.

Contents

Common Misunderstandings About the Internal-Rate-of-Return and Net Present Value Economic Analysis Methods

Martin Wohl, Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh

Engineering economy and benefit/cost analysis manuals usually include the net present value and internal-rate-of-return methods for the analysis of mutually exclusive alternatives and, more times than not, contend that both methods, if properly applied, will invariably lead to the same economic decisions. However, it can be demonstrated that such a view is incorrect, as a general rule, and that use of the internal-rate-of-return method can lead to incorrect or ambiguous economic decisions. Accordingly, the purpose of this paper is to define the specific cases and situations in which application of the rate-of-return method will lead to incorrect or ambiguous economic decisions as well as to demonstrate why the net present value method is preferable and to explain the underlying reasons for the differences. Numerous examples will be employed to illustrate the various cases and underlying principles.

Among the more common methods of economic analysis used by engineers to judge the economic worth of mutually exclusive alternatives are net present value (NPV), benefit/cost ratio, and internal rate of return. Economists, however, have long warned about the dangers of using the internal-rate-of-return method for analyzing mutually exclusive alternatives. Specifically, use of the internal-rate-of-return method can lead to incorrect economic decisions when the alternatives are ranked in improper order or when multiple solutions (i.e., multiple internal rates of return) are encountered. Unfortunately, most engineering economy textbooks deemphasize these drawbacks and, as a consequence, practitioners have been misled about the desirability of employing the internal-rate-of-return method in the analysis of mutually exclusive alternatives. Therefore, a clarification of these and other related aspects is desirable.

THE INTERNAL-RATE-OF-RETURN METHOD AND RANKING CRITERIA

Once a set of mutually exclusive alternatives has been specified (to include, implicitly or explicitly, the null alternative), the stream of costs and benefits for each must be estimated year by year over a common analysis period or planning horizon. In turn, the alternatives must be ranked from lowest to highest. The usual (though not necessarily best) criterion is to rank them in ascending order with respect to the costs for the initial year; also, if the costs for the initial year of all (or some) alternatives are equal, then order those that have equal costs for the initial year in descending order with respect to the benefits in the following year.

A cutoff rate or minimum attractive rate of return (MARR) must be specified. This interest rate indicates the effective annual yield of the opportunities that will be foregone if the resources are used for one of the alternatives being analyzed. In essence, the analyst is merely trying to ensure that at least one of the alternatives being analyzed will provide a yield at least that high. Otherwise, other opportunities should not be foregone. Thus, the MARR can be regarded as the opportunity cost of capital for both borrowing and lending situations.

The analysis proceeds in stepwise fashion. We must first determine the lowest-ranked alternative that has an internal rate of return at least as high as the MARR. Thus, we determine the internal rate of return for alternative 1 (i.e., the lowest-ranked one) such that:

$$\sum_{t=0}^{n} B_{1,t}/(1 + r_1)^t = \sum_{t=0}^{n} C_{1,t}/(1 + r_1)^t \qquad (1)$$

where

$B_{1,t}$ = the benefits for alternative 1 during year t,
$C_{1,t}$ = the costs for alternative 1 during year t,
n = the number of years in the analysis period, and
r_1 = the internal rate of return.

It can be seen that the internal rate of return is simply the interest rate at which the NPV of alternative 1 is zero. (The above formulation implies that the null alternative has zero benefits and costs over the n-year analysis period.) If the internal rate of return (r_1) is equal to or greater than the MARR, then the alternative is regarded as acceptable. If not, it is rejected and the next-higher-ranked alternative is examined for its acceptability, and so forth, until the lowest-ranked acceptable alternative is identified.

After the lowest-ranked acceptable alternative is identified, then an incremental analysis is used to determine the acceptability of higher-ranked alternatives. Assuming (for simplicity) that alternative 1 is found to be acceptable, we then determine the incremental rate of return on the increments in benefit and cost between alternatives 1 and 2 ($r_{1/2}$ or Δ rate of return) such that

$$\sum_{t=0}^{n} (B_{2,t} - B_{1,t})/(1 + r_{1/2})^t = \sum_{t=0}^{n} (C_{2,t} - C_{1,t})/(1 + r_{1/2})^t \qquad (2)$$

where $B_{x,t}$ and $C_{x,t}$ are the benefits and costs for alternative x during year t. Rearrangement of the terms in Equation 2 shows that the incremental rate of return is simply the interest rate for which the NPV of alternative 1 is just equal to the NPV of alternative 2. If the incremental rate of return ($r_{1/2}$) is equal to or greater than the MARR, then the higher-ranked alternative is deemed to be better than the lower-ranked alternative (i.e., alternative 1 is rejected in favor of alternative 2). In turn, the incremental rate of return for the next-higher-ranked alternative as compared to alternative 2 ($r_{2/3}$) would be computed to determine which is preferable. However, if $r_{1/2}$ is less than the MARR, then alternative 2 would be rejected and the next paired analysis would be conducted between alternatives 1 and 3. That is, we would determine whether $r_{1/3}$ was equal to or greater than the MARR.

Analysis is continued in pairs for all higher-ranked alternatives until the highest-ranked alternative that has an incremental rate of return at least as high as the MARR is identified. That highest-ranked alternative will then be the best alternative, economically speaking.

MULTIPLE RATE PROBLEM FOR THE INTERNAL-RATE-OF-RETURN METHOD

Multiple solutions for the internal-rate-of-return method can arise in one of two ways. The first can occur when especially heavy costs are expected in the future (for example, rolling stock replacement or guideway resurfacing, rehabilitation, or restoration). The guideway resurfacing, rehabilitation, and restoration situation is especially pertinent for many existing roadways and bridges and provides a typical example in which multiple rates probably would occur; this is particularly true if roadway or bridge repairs cause some or all of the lanes to be closed during resurfacing, rehabilitation, or restoration. The example in Table 1 illustrates this first case.

The maximum number of internal rates of return can be determined from an inspection of the variation in the net cash-flow stream. The right-hand column in Table 1 shows the estimated year-by-year net benefits ($B_{1,t}$-$C_{1,t}$). Applying Descartes' rule of signs, the number of sign changes that occurs over the 30-year horizon indicates the maximum number of positive rates of return that can result. In this case, the net benefits changed signs three times, thus indicating that as many as three positive solutions or internal rates of return could occur.

The second and probably more frequent case in which multiple internal rates of return can occur is with incremental rate-of-return analysis for pairs of alternatives. This possibility is more common than we might be led to believe. It could apply, for instance, when higher initial outlays lead to different benefit-accrual patterns or when the future cost-outlay patterns for two alternatives are different. The example given in Table 2 illustrates the former situation and might be applicable if, say, a firm is deciding between two different oil pumps for the extraction of oil from a well. The more expensive pump would permit the oil to be extracted quicker and slightly increase the total amount of oil extracted. In this instance, there is a single internal rate of return for each alternative (analyzed separately), but there are two solutions, or internal rates of return, associated with the incremental costs and benefits between alternatives 1 and 2.

Table 1. Costs and benefits for two-stage improvement of an existing bridge.

End of Year t	$B_{1,t}$[a] ($000s)	$C_{1,t}$[b] ($000s)	$B_{1,t}$ - $C_{1,t}$ ($000s)
0	-	50	-50
1	61	55	+6
2	63	0	+63
3	65	0	+65
.	.	.	.
.	.	.	.
.	.	.	.
9	77	0	+77
10	79	705	-626
11	81	610	-529
12	83	495	-412
13	85	0	+85
.	.	.	.
.	.	.	.
.	.	.	.
29	117	0	+117
30	119	0	+119

[a]Benefits in year t, net of annual operating and maintenance costs.
[b]Nonrecurring capital outlays in year t.

Table 2. Rate-of-return analysis for two oil pump alternatives.

	Alternative 1 ($000s)		Alternative 2 ($000s)	
Year t	$B_{1,t}$	$C_{1,t}$	$B_{2,t}$	$C_{2,t}$
0	0	100	0	110
1	70	0	115	0
2	70	0	30	0

Note: The internal rate of return for alternative x (r_x) would be the interest rate at which the discounted benefits just equal the discounted costs; the formulation would be the same as that shown in Table 1. In alternative 1, r_1 = 25.69 percent; in alternative 2, r_2 = 26.16 percent. The internal rate of return for the incremental benefits and costs between alternatives 1 and 2 ($r_{1/2}$) would be the interest rate at which the NPV of alternative 1 is just equal to the NPV of alternative 2.

As a general proposition, both of these cases can and do arise. Yet, Grant and others argue (1, p. 560):

> It cannot be emphasized too strongly that [multiple solution] cases such as those illustrated in [our examples] are the exception rather than the rule. They occur chiefly in the mineral industries and the petroleum industry; even there they arise only in rather specialized circumstances.

Similarly, Winfrey assumes away the multiple-solution problem for the internal-rate-of-return method by saying (2, p. 161), "Since the situation of two or more rates of return is so infrequent, there is no need to outlaw the rate-of-return method, a highly useful and understandable method of analysis." Newnan echoes this view, saying (3, p. 138): "In certain rare situations we find that solution of a cash-flow equation results in more than one positive rate of return."

Such instances are not necessarily exceptional, infrequent, or rare. Rather, for highways or bridges, which may require heavy outlays for reconstruction or replacement in future years, as well as for transit systems, which may require costly rolling stock replacement or rehabilitation every 10-30 years, the possibility of multiple rates of return is high, if not the typical expectation.

Analysts tend to regard especially high or low internal-rate-of-return values as being unrealistic or inappropriate. For example, most of the advanced pocket calculators that are preprogrammed to calculate the internal rate of return for a cash-flow stream identify only the lowest positive internal rate of return and thus ignore all others and imply their irrelevance. To the contrary, all multiple rates are valid and should be considered.

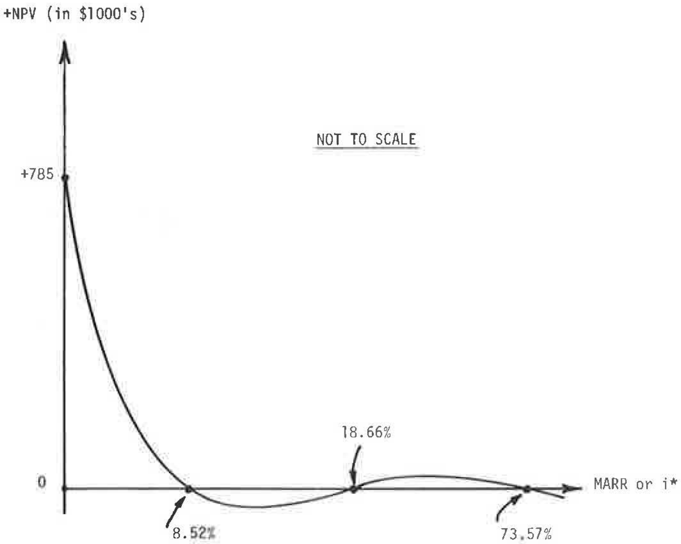
THE FALLACY IN MANY ENGINEERING ECONOMIC TEXTBOOKS

Many engineering economic textbooks incorrectly claim that all analysis methods (such as the NPV, benefit/cost ratio, and internal rate of return), when properly applied, lead to identical ranking of alternatives. For instance, Grant and others say (1, p. 117)

> Once a particular [MARR] is selected for the comparison of alternatives, a correct analysis of relevant rates of return will invariably lead to the same conclusion that will be obtained from a correct annual cost comparison or a correct present worth comparison.

Winfrey echoes the above position, saying (2, p. 123): "When properly applied in accordance with their limitations, each method will give a reliable result for economic evaluation and for project formulation." In a more recent article, he reiterates (4, p. 37): "All methods will give the identical selection of the alternative of greatest economy when the procedures of analysis are correctly chosen and properly used."

Figure 1. NPV at different MARR for the bridge alternative described in Table 1.



To explore this claim of engineering economists, let us consider two examples, the first given in Table 1 and the second in Table 2.

The data in Table 1 represent the expected year-by-year costs and benefits associated with undertaking a specified course of action over a 30-year analysis period or planning horizon. The numbers appear somewhat typical for previously built highways or bridges that now are in need of repair, restoration, or replacement, or they could apply to a transit system that plans to extend its lines in the future. In this case, assume that a community has an old bridge that is in imminent danger of collapse. In turn, the public works department was ordered by the city council to analyze the various repair strategies that would ensure safe operation of the bridge for the next 30 years and to evaluate the economic worth of each, relative to the null or abandonment alternative. Among the possibilities are (a) make minor repairs to the bridge now and a major overhaul 10 years later or (b) completely overhaul the bridge now. The appropriate benefit and cost data for the first of these two alternatives are given in Table 1. Presumably, the second alternative would have higher initial outlays and thus would be analyzed in terms of the incremental benefits and costs after the first alternative has been analyzed in terms of its acceptability.

Accordingly, the data in Table 1 represent the incremental costs and benefits for the first alternative relative to bridge abandonment. In turn, we can calculate the internal rate of return for this lowest-cost alternative. The discounted internal-rate-of-return method, properly applied, would yield three rates of return in this instance: 8.52, 18.66, and 73.57 percent. First, all of these solutions or rates are correct. Internal rates of return (r_1) = 8.52, 18.66, and 73.57 percent, where r_1 is the interest rate (or rates) that satisfies the following identity:

$$\sum_{t=0}^{30} B_{1,t}/(1 + r_1)^t = \sum_{t=0}^{30} C_{1,t}/(1 + r_1)^t \quad (3a)$$

or

$$\sum_{t=0}^{30} (B_{1,t} - C_{1,t})/(1 + r_1)^t = 0 \quad (3b)$$

That is, they represent the interest rates for which the NPV of this alternative is zero. Second, in the absence of any other information, how do we interpret these rates? Suppose, for instance, that the appropriate MARR is judged to be approximately 10 percent. Then, by using just the internal-rate-of-return figures, we will obtain either an ambiguous answer or an incorrect one. That is, we presumably would incorrectly regard the alternative as acceptable (since both 18.66 and 73.57 percent are higher than the MARR) or would incorrectly regard the decision as ambiguous.

By contrast, if we had simply computed the NPV (or discounted benefits minus discounted costs) for the stated MARR of 10 percent, we would have learned that the NPV was negative and thus that the minor bridge repair alternative was economically infeasible and should be rejected. Specifically, for a MARR of 10 percent, the NPV would be equal to -$14 140. In addition, the benefit/cost ratio for this alternative would be 0.981 (or less than 1.0, indicating rejection) at an interest rate of 10 percent. In sum, we see that all methods of analysis do not invariably provide either identical or sound conclusions. The NPV and benefit/cost ratio methods are conclusive and unambiguous; the internal-rate-of-return method is ambiguous and inconclusive.

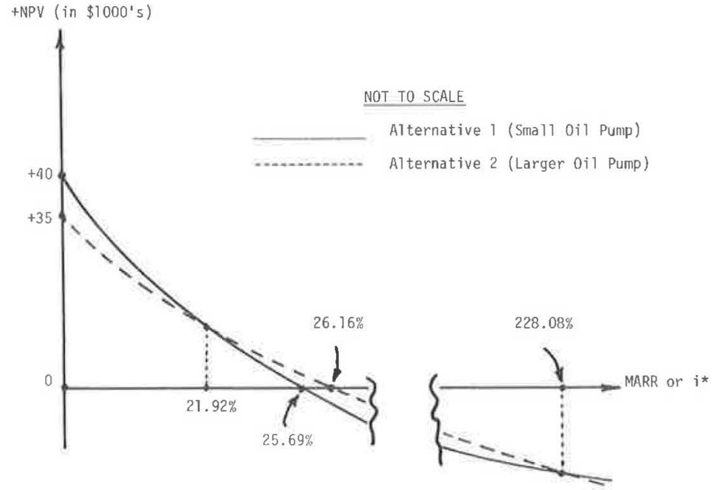
Moreover, in this situation the internal-rate-of-return method could have led us astray for a wide range of circumstances. In Figure 1, a plot of the appropriate NPVs versus the interest rate is shown for the full range of interest possibilities. Clearly, if the appropriate MARR was deemed to be between 8.52 and 18.66 percent, then an analysis that did not reject this alternative would be incorrect—a result that would not necessarily result from strict application of the internal-rate-of-return method.

A Simplified Example

Now, let us review the circumstances for another example situation, one involving overall analysis as well as incremental analysis between a pair of alternatives. The particular example was chosen for clarity and to minimize calculations and is given in Table 2. It deals with two investment alternatives, each having two-year cost and benefit streams, as shown. The additional initial investment (of alternative 2 over alternative 1) will permit earlier recovery of the overall gains, as well as lead to slightly higher two-year gains (measured in current or undiscounted dollars).

To apply the internal-rate-of-return method, we first

Figure 2. NPV at different MARR for the two (oil pump) investment alternatives described in Table 2.



need to specify the MARR. Let us assume it is 15 percent. Next, we calculate the internal rate of return for the lowest-cost alternative and then ask whether the rate is at least as high as the MARR. (This is the first step in answering the question, "Is any alternative worth undertaking?") Since the internal rate of return for alternative 1 (25.69 percent) is higher than the MARR, alternative 1 is judged to be acceptable, economically. In turn, we must calculate the incremental internal rate of return ($r_{1/2}$) associated with the incremental costs and benefits between alternatives 1 and 2, the latter being the higher initial-cost alternative. That is, $r_{1/2}$ is the discount rate that satisfies the following identities:

$$\sum_{t=0}^{2} (B_{2,t} - B_{1,t})/(1 + r_{1/2})^t = \sum_{t=0}^{2} (C_{2,t} - C_{1,t})/(1 + r_{1/2})^t \qquad (4)$$

or

$$\sum_{t=0}^{2} (B_{1,t} - C_{1,t})/(1 + r_{1/2})^t = \sum_{t=0}^{2} (B_{2,t} - C_{2,t})/(1 + r_{1/2})^t \qquad (5)$$

The internal rate of return for the incremental costs and benefits is not a single rate but two of them—21.92 and 228.08 percent. Since both of these rates (examined without any other information) are greater than the MARR, the analyst would presumably regard alternative 2 as being more attractive than alternative 1, economically speaking. Or, the decision would be regarded as ambiguous.

However, neither of the above conclusions would be correct. For instance, the NPV method will show that for a MARR of 15 percent the NPV of alternative 1 is $13 800 as compared to only $12 684 for alternative 2, thus unambiguously indicating the preferability of alternative 1. Similarly, if the benefit/cost ratio method had been used, the ratio for alternative 1 would have been 1.138 for an interest rate of 15 percent; thus, alternative 1 is acceptable. In turn, the incremental benefit/cost ratio for the increments in benefits and costs between alternatives 1 and 2 can be shown to be 0.884, or less than 1.0, thus indicating that alternative 2 should be rejected. Accordingly, it is obvious that the various analysis methods do not invariably lead to the same sound conclusions about which alternative is best. Moreover, the situation would be even more perplexing if the MARR were, for example, about 25 percent. In this instance, use of the internal-rate-of-return method would lead to the acceptance of alternative 1 (since r_1 is greater than the MARR) but would provide little guidance about the acceptability of the higher-cost alternative, since one of the incremental rates of return is less than the MARR and one is higher. By contrast, either the NPV or benefit/cost ratio methods would have shown that alternative 2 is unambiguously the best choice for a MARR of 25 percent (see Figure 2).

Reasons for Different Conclusions from Different Methods

We saw that the internal-rate-of-return and NPV methods sometimes lead to conflicting decisions about which alternative is best. In turn, we should ask, Which method gives correct results? Why can and do the results sometimes differ?

To begin, it seems appropriate to emphasize the objective of the analyst (or investor)—maximize the net gains or profits to be accrued over the analysis period or planning horizon. That is, we wish to identify which project will maximize the surplus that a firm or community will accrue over the analysis period. Accordingly, the economist has argued that the NPV (computed at the MARR) is a simple and unmistakable indicator of a project's profitability and that the project that has the highest positive NPV will be the best, economically speaking. Moreover, the internal rate of return will sometimes prove to be a misleading indicator of profitability.

Consider again the example in Table 2. Given these two alternatives (relative to investing in neither), which would accrue the highest profit or surplus by the end of year 2? The answer involves two aspects: (a) the MARR or opportunity cost of capital, which informs us about the yield possibilities that we must forego if we invest in alternative 1 or 2, and (b) the possible uses of any net revenues or benefits that are accrued prior to the end of the two-year analysis period. For the first aspect, and again assuming that the MARR is 15 percent, to invest $100 000 in alternative 1 would mean that we would forego the opportunity to accumulate $132 250 by the end of year 2. But, by foregoing this opportunity and investing in alternative 1, we would accrue annual net earnings of $70 000 at the end of years 1 and 2. Obviously, though, if we had invested in alternative 1, the first-year earnings of $70 000 would be reinvested during the second year rather than sit idle. A reasonable assumption is (as is implicit with the NPV method) that these early-year earnings would be reinvested at the MARR (which, after all, represents a best estimate of the potential yield of any outside opportunities). Accordingly, if the first-year earnings of $70 000 were reinvested at 15 percent, one year later we would have accumulated $80 500 (or $70 000 plus $10 500 in yield) plus of course the $70 000 that was generated in the second year by the initial investment. Altogether then, an investment in alternative 1 would require us to forego $132 250 and instead to accumulate $150 500 during the same two-year period. The profit or net gains to be accumulated by the end of two years would be $18 250 (or $150 500 less $132 250). A similar analysis can be carried out for alternative 2, again for a MARR of 15 percent. For an investment in alternative 2, we would accumulate total earnings of $162 250 by the end of year 2 and would forego the opportunity to earn $145 475. Thus, the profit accrued by the end of the two-year period would be $16 775.

Both investment alternatives are profitable (relative to investing in neither); however, alternative 1 is more profitable than alternative 2. Moreover, this is the same result (i.e., decision) that was obtained from NPV analysis, a result that is hardly surprising since (in the parlance of engineering economy) the NPV is exactly equivalent to the net future worth when the latter is multiplied by the single-payment present-worth factor. That is,

NPV for alternative 1 (at 15 percent) = (net future worth) $\times$ (P | F, 15 percent, 2) = ($18 250) $\times (1.15)^{-2}$ = $13 800.
NPV for alternative 2 (at 15 percent) = (net future worth) $\times$ (P | F, 15 percent, 2) = ($16 775) $\times (1.15)^{-2}$ = $12 684.

where P | F = the present worth given the future value. That is, if a project has the highest positive net future worth, then it also will have the highest positive NPV. By contrast, for a MARR of 15 percent, the internal-rate-of-return method would lead to the conclusion that alternative 2 was the most profitable or that the choice was ambiguous. Moreover, when using the internal-rate-of-return method, such a confusing result would surely be forthcoming for any MARR value below 21.92 percent (and probably for any value up to 25.69 percent). All in all, the NPV method leads to correct results, whereas the internal-rate-of-return method sometimes provides incorrect or ambiguous ones.

The internal-rate-of-return method sometimes gives misleading results or ones that differ from those obtained by using either NPV or benefit/cost ratio calculations because of different assumptions about reinvestment of early-year benefits or revenues. [For extensive coverage of this point, see articles by Hirshleifer, Lorie and Savage, Renshaw, and Solomon (5).] To use the internal-rate-of-return method is to assume implicitly that earnings accrued prior to the end of the analysis period are reinvested at the internal rate of return for the remaining years. To use the NPV (or benefit/cost ratio) method is to assume implicitly that prior-year earnings are reinvested at the MARR (or opportunity cost of capital) for the remaining years. In the Table 2 example, for instance, use of the NPV method implies that the $115 000 first-year earnings of alternative 2 were reinvested at the MARR (of 15 percent) during the second year; however, the internal-rate-of-return method implied that these same earnings were reinvested at a rate of 26.16 percent during the second year—no wonder the results were different in this case. Also, for this example, the internal-rate-of-return method would imply that the first-year earnings for alternative 1 would be reinvested at 25.69 percent, but those for alternative 2 would be reinvested at 26.16 percent. Such an assumption would be nonsensical on two grounds. For one, Why should the reinvestment rate differ from one alternative to another? Should they not be equal? For another, we should recognize that the MARR is the indicator of our other investment or reinvestment opportunities and that the MARR has no necessary relationship to the internal rate of return.

Some engineering economists argue the inappropriateness of considering reinvestment possibilities for any revenues or benefits that are accrued prior to the end of the planning horizon (such as those accrued at the end of year 1 for alternatives 1 and 2 in Table 2). Winfrey, for example, says (2, pp. 162-163)

> It is most difficult to convince the layman that his rate of return on a given investment is dependent upon how he reinvests return from that investment; neither does it seem logical when comparing possible investment alternatives that the choice of investment could depend upon how the return from each alternative would be reinvested.

Accumulated profit or net gain over the entire investment period is clearly related to and thus dependent on reinvestment of revenues or benefits gained along the way; thus, no wise investor will choose to ignore them. Let us demonstrate the point by a somewhat contrived (yet appropriate) example. [Another interesting example and discussion of this same aspect appears in Mishan (6, p. 225).] Suppose, for instance, that we want to borrow $70-90 now and that, in turn, we go to the ABC Loan Company to request one of the two following loans, as shown in Table 3:

1. A $70 loan now to be paid back in two installments, the first one of $75 one year from now and the second of $70 ten years from now; or
2. A $90 loan now to be paid back in two installments, the first one of $115 one year from now and the second of $5 ten years from now.

Assume that the ABC Loan Company estimates its MARR to be 8 percent and that the company wishes to know which loan plan (if any) would be most profitable. In turn, let us assume that the ABC Loan Company prefers to use the internal-rate-of-return method and calculates the various rates of return, as shown in Table 3. Accordingly, the company notes that loan plan 1 has an internal rate of return (r_1) of 22.84 percent and thus is acceptable (since it is larger than the MARR of 8 percent). Next, the company notes that the incremental rates of return ($r_{½}$) of 16.26 percent and 99.35 percent are both higher than its MARR, thus suggesting that loan plan 2 is better than loan plan 1 or that the choice is ambiguous. However, if the company had used either the NPV method or the benefit/cost ratio method, it would have discovered that loan plan 1 and not plan 2 is clearly best, as shown in the calculations below. In fact, the data in Figure 3 will show that loan plan 1 is better for the ABC Loan Company whenever its MARR value is below 16.26 percent.

NPV Method

Loan plan 1

$[NPV_{1,10}]_{8\%}$ = $31.87.

Loan plan 2

$[NPV_{2,10}]_{8\%}$ = $18.80.

Benefit/Cost Ratio Method

Loan plan 1

$[BCR_{1,10}]_{8\%}$ = 1.455.

Loan plan 2

$[BCR_{2,10}]_{8\%} = 1.209$.

Comparison of loan plans 1 and 2

$[BCR_{½,10}]_{8\%} = 0.347$.

Again, the differences between the NPV and internal-rate-of-return methods (in indicating the best alternative for a MARR of 8 percent) stem from the different assumptions with respect to reinvestment. First, the reinvestment possibilities for the payback amounts received in year 1 should not be ignored. The $75 received in year 1 for plan 1 or the $115 received for plan 2 would not be ignored or placed in a drawer for the remaining nine years. Rather, they would be reinvested in other investment opportunities or in early (rather than later) year enjoyment. The most reasonable assumption (in the absence of other information) is that these yearly earnings will be reinvested at the MARR.

Second, if the year-1 payback amount of $75 for plan 1 were to be reinvested at our assumed MARR of 8 percent for the remaining 9 years, then by the end of year 10 the ABC Company would accumulate $149.93, which can then be added to the 10th year payback amount of $70. Thus, the accumulated earnings will be $219.93 by the end of year 10 if the early-year earnings are reinvested at 8 percent. These accumulated earnings ($219.93) when discounted to their present value at 8 percent and balanced against the $70 initial investment will be exactly equal to the NPV of $31.87 shown in Table 3. This proves that the NPV method implicitly assumes that any early-year earnings are reinvested at the MARR. That is, if r is the reinvestment rate for the year-1 earnings, then at 8 percent the NPV of the accumulated 10-year earnings less initial costs would be as follows:

$$\{[\$75\,(1+r)^9 + \$70]/(1.08)^{10} - \$70\} = \$31.87.$$

This identity would hold only for a reinvestment rate (r) equal to 8 percent.

Third, it can be shown that, for the internal-rate-of-return method, early-year earnings are assumed to be reinvested at the internal rate of return. If this is true, then for plan 1 the first-year payment of $75 is reinvested for the remaining 9 years at 22.84 percent, thus accumulating $477.66 by the end of 10 years, to be added to the 10th year payment of $70. The accumulated 10-year earnings will be $477.66 + $70 = $547.66 [e.g., $\$547.66 = \$75(1.2284)^9 + \$70$]. These accumulated earnings when discounted to their present value will be identical with the initial investment of $70 only for an interest rate of 22.84 percent, the internal rate of return. In short, this proves that the early-year earnings were assumed to be reinvested at the internal rate of return, an assumption that is clearly different from that used for the NPV method. Moreover, if the early-year earnings were reinvested at any reinvestment rate other than the internal rate of return, the 10-year accumulated earnings when discounted at the internal rate of return would not be equal to the discounted costs (which in this case were equal to the initial loan amount).

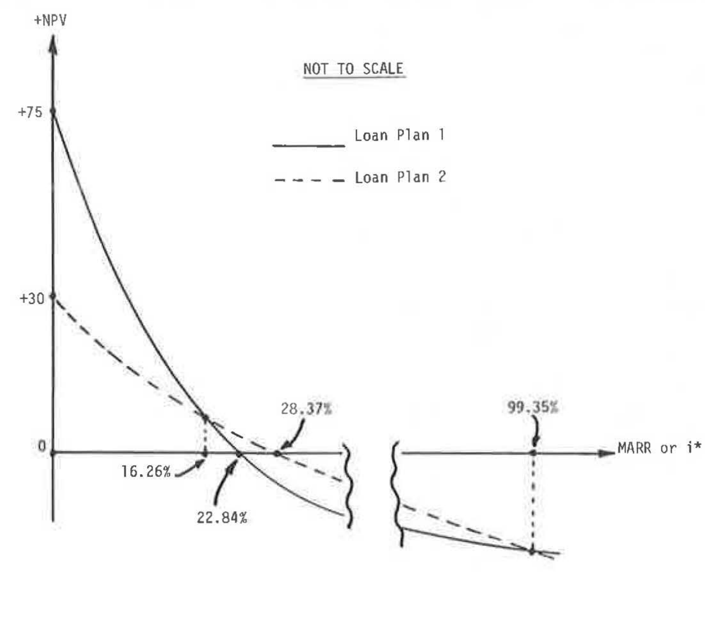
Given that we have proved that the internal-rate-of-return method uses the internal rate of return as the reinvestment rate and that the NPV method used the MARR as the reinvestment rate, we can be more explicit about the confusion in applying the internal-rate-of-return method to the selection of the best loan plan in Table 3. Note first that plan 1 has an internal rate of return of 22.84 percent, and plan 2 has a rate of 28.37 percent. As a consequence, the method assumes that the year-1 payment for plan 2 can be reinvested at a higher rate than can the year-1 payment for plan 1. What rationale is there for assuming different reinvestment rates for

Table 3. Two loan and payback possibilities from the ABC Loan Company's viewpoint.

	Loan Plan 1		Loan Plan 2	
Year t	Loan Amount ($)	Payback Amount ($)	Loan Amount ($)	Payback Amount ($)
0	-70		-90	
1	0	+75	0	+115
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	+70	0	+5

Note: For loan plan 1, r_1 = 22.84 percent; for loan plan 2, r_2 = 28.37 percent; $r_{1/2}$ = 16.26 and 99.35 percent for loan plans 1 and 2.

Figure 3. NPV at different MARRs for the two ABC Loan Company plans described in Table 3.



earnings that are accrued at the same point in time? None. By the same token, in the computation of the incremental rates of return, we assumed that the difference in first-year payments (or \$115 - \$75) was reinvested at either 16.26 percent or 99.35 percent–again, an assumption without rationale and one that is very different from that used to analyze separate projects.

Rate of Return for a Reinvestment Rate Equal to MARR

If identical assumptions are made about reinvestment for all methods, then invariably the same conclusions will result. As noted before, both the NPV and benefit/cost ratio methods assumed that the MARR would be the proper reinvestment rate. Thus, let us use the same reinvestment assumption for the rate-of-return method and then compare the results. That is, we will assume that early-year earnings are reinvested at the MARR and then determine the interest rate at which the discounted accumulated earnings just equal the discounted costs. However, the resultant interest rate, strictly speaking, will reflect more than internal earnings and thus will be designated by an R_x instead of an r_x. This R_x value will represent the effective yield to be obtained over the analysis period and will be equivalent to the internal rate of return only in exceptional cases. [This adjusted rate-of-return value has been termed the "equivalent rate of return" by Solomon (5, p. 74) and the "reinvestment-corrected internal rate of return" by Mishan (6, p. 228).]

Let us apply this procedure to the loan example in Table 3. Again let us assume that the MARR, and thus the reinvestment rate, for early-year earnings is 8 percent. Accordingly, the calculations would be as follows:

Effective Rate of Return (R_1) for
Loan Plan 1

Accumulated 10-year earnings = $\$75(1.08)^9 + \$70 =$ \$219.93.

In turn, find R_1 such that the discounted earnings are just equal to the discounted costs, or

$\$219.93/(1 + R_1)^{10} = \$70;$

thus, R_1 is equal to 0.1213 or 12.13 percent.

Effective Rate of Return (R_2) for
Loan Plan 2

Accumulated 10-year earnings = $\$115(1.08)^9 + \$5 =$ \$234.89.

In turn, find R_2 such that the discounted earnings are just equal to the discounted costs, or

$\$234.89/(1 + R_2)^{10} = \$90;$

thus, R_2 is equal to 0.1007 or 10.07 percent.

In this case, one in which the modified- or effective-rate-of-return method uses a reinvestment-rate assumption that is identical to that used in the NPV and benefit/cost ratio methods, the outcome and conclusions will be identical for all methods. That is, loan plan 1 provides an acceptable rate of return (i.e., one that is higher than 8 percent) and has a yield that is higher than that for plan 2. We also could have computed the modified or effective rate of return on the increments in costs and benefits between plans 1 and 2, although the step is unnecessary. The resultant incremental return figure would be -2.86 percent; this result is obvious when we note that the extra initial-year cost of \$20 led to extra accumulated 10-year earnings of only \$14.96.

OTHER CONFUSING ASPECTS ABOUT REINVESTMENT OF EARLY-YEAR BENEFITS

First, analysts have been troubled about benefits or gains that are accrued in nonmonetary rather than monetary terms. How does the concept of reinvestment apply, for example, to time savings accrued in a year prior to the end of the analysis period? The answer is simple and straightforward. Reinvestment principles (broadly construed) apply with equal validity to monetary and nonmonetary benefits that are accrued prior to the end of the analysis period because time savings accrued in earlier years are more valuable to people than the same amount of time savings accrued in a later year. Or, put somewhat differently, enjoyment (or consumption of earnings accrued earlier) is more highly valued than that accrued later. Also, the MARR, rather than the internal rate of return, reflects (in part) the strengths of people's tastes and preferences with respect to the importance of enjoyment now versus enjoyment later. Specifically, the MARR reflects the trade-off between people's time preferences and the rate of productivity of investments and thus the marginal rate of time preference is (roughly) equal to the marginal rate of productivity, both being equal to the MARR.

Second, Grant and others dealt extensively but confusingly with the matter of reinvestment (1, p. 563). In essence, they deal with the calculation and interpretation of an adjusted- or effective-rate-of-return figure.

They argue that such a method of computing an adjusted rate of return (or R_x) is fallacious, saying in part (1, pp. 563-565):

> Sometimes an analyst uses two or more interest rates because this method of analysis is required by company policy. Or he may mistakenly believe that this technique will give him useful conclusions. In either case, one aspect of his computational procedure will be the assumption of reinvestment at some stipulated interest rate. Various weaknesses in the reinvestment assumption are brought out in [the Table 4 example, to be discussed] and in several of the problems at the end of this appendix. . . .
>
> The fallacy in this type of analysis [i.e., that in which an adjusted or effective rate of return is calculated] may be even more evident if we apply the [adjusted-rate-of-return] method to the following estimates for another investment proposal [shown in Table 4].

For the cash flows shown in Table 4 (1, p. 565), the internal rate of return is 0 percent, thus indicating that the NPV is zero at 0 percent. Moreover, for any positive discount rate, the NPV is negative, thus indicating that the investment proposal is financially unattractive. In turn, Grant and others state that 10 percent is the "rate that the company is expected to make on other investments," thus indicating that 10 percent is the MARR and, therefore, the appropriate reinvestment rate for early-year gains (1, p. 565). Accordingly, they calculated the adjusted rate of return (R) and found it to be 6 percent. (Although I have some reservations about the way in which Grant and others calculated the adjusted rate of return, I will withhold discussion of those points until later.) In turn, they conclude (1, p. 565): "In effect, the investment proposal yielding 0 percent has been combined with the 10 percent assumed to be earned elsewhere in the enterprise to give the misleading conclusion that the proposal will yield 6 percent." To the contrary, the 6 percent is a true indicator of the effective yield to be anticipated from the cash flows shown in Table 4 since it properly reflects the reinvestment earnings of the positive cash flows. Moreover, the unattractiveness of the project is reflected in the fact that the overall yield is less than 10 percent, the effective yield that would be anticipated if the year 0 and year 1 funds were invested in other foregone opportunities. That is, if we were to invest in this plan for five years, then profits would be lost relative to other investment possibilities but not in an absolute sense.

Also, the folly of arguing that reinvestment should not be considered in calculation of the effective yield is obvious when we consider the circumstances for the cash flows as altered in Table 5. The cash flow for investment is the same for the Table 4 and 5 investments, and the internal rate of return is identical for the two (i.e., 0 percent). Even so, it should be apparent that the plan in Table 5 is much less attractive financially than that in Table 4. That is, we surely do care about when the earnings are received (i.e., early versus later) and what we do with them. The internal-rate-of-return figure of 0 percent would not reflect that fact, but the calculation of a modified or reinvestment-corrected rate of return would vividly demonstrate it. To be specific, and using the Grant procedure, the effective rate of return for the plan in Table 5 would be about 2 percent, thus indicating that the effective yield of this plan is far less than that for the plan in Table 4.

Also, the internal rate of return will be equal to the effective yield to be expected from a project in only two circumstances (both of which must be considered highly unlikely): (a) when all of the project earnings are accumulated solely at the end of the analysis period or (b) when the MARR is exactly equal to the internal rate of return. Only in these two cases will the NPV and internal-rate-of-return methods incorporate identical assumptions with respect to reinvestment and thus always provide identical decisions about acceptability and ranking.

Finally, note that the procedure to be used for calculating the adjusted or effective rate of return must be geared to the assumed financing plan and reinvestment strategy. Since either can vary, the procedure is somewhat arbitrary and the results will change accordingly. In the Table 4 example, Grant and others first discount the investment costs for years 0 and 1 to their present value at 10 percent; in effect this implies that $427 300 was borrowed in year 0 and that the balance between $427 300 and the $200 000 needed in year 0 was invested for one year at 10 percent and then used to pay the year 1 investment costs. This assumed financing policy, along with reinvestment of the $10 000 year 1 earnings, resulted in the maximum adjusted-rate-of-return value that could be achieved with a MARR of 10 percent. Alternatively, we could have simply adjusted the cash flows in the right- and left-hand columns, borrowed only $200 000 in year 0 and $120 000 in year 1, and then calculated the adjusted rate of return. For the latter set of assumptions, the adjusted-rate-of-return value for the Table 4 example would be 3.9 percent and that for Table 5 would be 0.85 percent. These effective yield values are lower than those obtained for the different financing and reinvestment strategies used by Grant and others, thus emphasizing that the yield is dependent on the reinvestment and financing plans and that it is necessary to be explicit about both.

Table 4. Five-year cash flows for an investment proposal.

Year	Cash Flow for Investment ($000s)	Cash Flow from Excess of Operating Receipts over Disbursements ($000s)
0	-200	
1	-250	+130
2		+110
3		+90
4		+70
5		+50
Total[a]	-450	+450

[a]Present worth at 0 percent.

Table 5. Five-year cash flows for an altered investment proposal.

Year	Cash Flow for Investment ($000s)	Cash Flow from Excess of Operating Receipts over Disbursements ($000s)
0	-200	
1	-250	+10
2		+20
3		+30
4		+40
5		+350
Total[a]	-450	+450

[a]Present worth at 0 percent.

SUMMARY AND CONCLUSIONS

With both public and private investment projects, multiple solutions can occur and thus can lead to ambiguous or incorrect investment decisions if one uses the internal-rate-of-return method. Even if this occurrence is rare (a fact that has yet to be established), its possibility alone should discourage even the most serious advocate of the internal-rate-of-return method. This method lacks generality.

ACKNOWLEDGMENT

I am grateful for the painstaking reviews and helpful criticism afforded by my colleagues Tung Au and Chris Hendrickson and by Robley Winfrey. The manuscript was markedly improved as a result of their collective effort, though I do not hold them accountable for any errors of fact or judgment, nor do I imply agreement on their part with my views.

REFERENCES

1. E. Grant, W. Ireson, and R. Leavenworth. Principles of Engineering Economy, 6th Ed. The Ronald Press Co., New York, 1976.
2. R. Winfrey. Economic Analysis for Highways. International Textbook Company, Scranton, PA, 1969.
3. D. Newnan. Engineering Economic Analysis. Engineering Press, Inc., San Jose, CA, 1977.
4. R. Winfrey. Misunderstandings of Cost-Benefit Analysis as Applied to Highway Transportation Investments. TRB, Transportation Research Record 635, 1977, pp. 37-41.
5. E. Solomon, ed. The Management of Corporate Capital. Free Press of Glencoe, New York, 1959.
6. E. J. Mishan. Cost-Benefit Analysis. Praeger Publishers, New York, 1976.

Discussion

Robley Winfrey, Consulting Civil Engineer, Arlington, Virginia

Wohl has proved nothing. He does not tell the whole story. He uses misleading illustrations. He does not point out the shortcomings of the NPV method. He does not prove the reinvestment assumption. He does not mention that industrial officials want to know the rate of earning and not their net dollar sum of present worth. He does not recognize that transportation officials (highway, mainly) prefer the benefit/cost ratio or the rate of return. He does not mention that the rate-of-return method has been used for years without the disastrous results he says could happen.

In the rate-of-return method, under conditions of two or more reversals of sign in the combined accumulated negative and positive cash flows, it is acknowledged that two or more rates of return may be found. But Wohl does not mention that the NPV method may also give two or more solutions. Which alternative gives the highest NPV depends on the choice of discount rate (MARR). For every rate of return isolated by the rate-of-return method in a multiple-rate example, there is likewise an NPV of zero. The in-between NPV solutions will be net negative or net positive. A data set can easily be formed in which a discount rate of 8 percent would indicate an NPV to support alternative A; however, a discount rate of 9 percent would support alternative B.

Wohl's statement and illustration that many roadways and bridges provide examples in which multiple rates would occur are far from reality and practice. His example shows that an author can find a set of figures to prove a hypothesis, no matter how unrealistic the figures used may be.

A highway engineer would not invest 50 and 55 units of cost at ages 0 and 1, and then at age 10 invest 705 units, followed in successive years by costs of 610 and 495 units. This cost cash flow is bad enough, but look at the annual benefits. They follow a straight-line increase of two units per year. There is no change in this benefit flow after the initial investment to reduce the "eminent danger of collapse" or after the three years of heavy investments. If these years of investment would not reduce costs or increase the benefits, then why make the capital expenditures? Another unrealistic feature of Table 1 is that the straight-line flow of benefits continues to age 30 years. Traffic over the bridge just could not produce such a constant rate of cost reduction.

It is useless and misleading in economic analysis to bring into the solution for today's choice of alternative those far-into-the-future needs for maintenance and repairs under conditions that, as of today, cannot be estimated under any acceptable probability of actual occurrence. Any such estimates destroy confidence in the analysis.

Engineers and the engineering economists agree that the objective of the analysis for the economy of proposed investments is to determine which alternative proposal will maximize the net dollar return—on a discounted basis, of course. We also take the position that all methods of analysis, when properly calculated, will arrive at the same alternative as the choice. Wohl's paper does not disprove this statement. Let him try some realistic situations in place of his custom-built unrealistic illustrations in Tables 1-3 and solve them by the methods of (a) equivalent uniform annual cost, (b) benefit/cost ratio, (c) rate of return, and (d) NPV. He can even try some realistic multiple-rate situations and, when properly handled, find agreement within the methods. Note that Tables 1-3 each have two or more solutions in the rate-of-return and NPV methods.

The rate-of-return solution could involve more calculations than would the NPV solution, but not extensively so when applied to realistic alternatives and when computers are used. But remember that the NPV solution should be made for a range of discount rates—particularly when there is evidence of two or more reversals of sign in the cash-flow sequence or when the NPV sums for a pair of alternatives have but a small difference.

The rate-of-return method is not ambiguous and inconclusive. It is the NPV and benefit/cost ratio methods that have these characteristics. In the example of Table 1 and Figure 1, the three rate-of-return answers are correct and the whole truth. Getting only one answer from the NPV and benefit/cost ratio methods leaves us in ignorance of the whole truth. For example, had a MARR of 8 percent or less been chosen, the project would have had a positive NPV. Also, a positive NPV would have been found at any MARR between 18.66 and 73.57 percent. Are these answers not facts the analyst should give to the decision maker?

The decision maker wants to know what each alternative will produce on its own: What is its rate of earning? What are the comparative earnings by pairs of the multiple alternatives under analysis (the differential solutions)? The analysis for economy in no way deals with the handling of the paid-back income. This matter is for the decision maker to evaluate on the basis of what is expected to be the future situation for investment when the project generates the incomes forecast.

Throughout the paper the economic analysis is confused with the decision on choice of alternative by the decision maker. These two items are distinctly separate. The findings of economic analysis are only guides to use in the decision-making process.

Under certain conditions, a choice of alternatives may be different, depending on what MARRs are used and whether there are two or more changes in sign. But such difference is not due to the reinvestment assumption. Any difference is due wholly to the results of the combinations of the three variable factors within the solution equations. The three factors are (a) the discount rate, (b) dollar amount of each cash flow, and (c) yearly time spacing of the cash flows. Thus, when the NPV method is used at a discount rate of 15 percent, both alternatives 1 and 2 in Table 2 would be favored. The basic characteristic of exponential mathematics results in multiple solutions whenever there are two or more reversals of sign. The factor $(1 + r)^n$ produces strange results with changes in the magnitude of the dollar sums in the cash flow.

Regardless of the reinvestment assumption and other factors, Tables 2 and 3 under the rate-of-return solution present the true rate of return for each alternative. Further, the NPV answers are fact for each alternative for the single discount rate used in the calculations.

Tables 1-3 each contains two or more reversals of sign, either in the prime statement of cash flows or in the differences in flow between the two alternatives. Could it be that it is only under conditions of two or more reversals of sign that the difference in choice of alternative is likely to differ between the rate-of-return and NPV methods? The paper does not answer this question.

Wohl has proved nothing about the reinvestment assumption. Wohl says that his calculations prove that the returns were reinvested. The basis for my statement is found in the mathematical equation used in calculating the rate-of-return solution. The equation can be expressed in terms of present worth or in terms of compounded amounts. The present-worth solution equation may be written for plan 1, Table 3:

$$0 = -70 + 75(\text{PW} - x\% - 1 \text{ year}) + 70(\text{PW} - x\% - 10 \text{ years}) \quad (6)$$

When this is solved by trial, the rate (x) will be found to be 22.84 percent. The proof solution is

$$0 = -70 + 75(0.814\,067) + 70(0.127\,821) = -70 + 61.05 + 8.95 = 0 \quad (7)$$

The compound-amount solution equation for the rate of return may be written

$$0 = -70(\text{CA} - x\% - 10 \text{ years}) + 75(\text{CA} - x\% - 9 \text{ years}) + 70 \quad (8)$$

Again, solving by trial produces 22.84 percent for the rate (x). The proof calculation is

$$0 = -70(7.823\,446) + 75(6.368\,810) + 70 \quad (9)$$

$$0 = (-547.66 + 477.66)/70 = 0 \quad (10)$$

Proof Equations 9 and 10, in terms of compounding, are what Wohl says prove the rate-of-return solution; Equation 6 assumes the reinvestment of the 75 payback at age 1. There is absolutely nothing in any of the above equations that supports the assumption of the reinvestment of the 75 payback. In fact, all that Wohl has proved is that, if $P(1 + r)^n = CA$, then $P = CA\,[1/(1 + r)]$ or that CA times the present-worth factor equals P.

Equation 6 simply finds the present worth of the cash-flow items that enter the equation as isolated items pertaining to the situation under analysis. They represent both outgo and income with no reference as to future disposition of the paybacks of the initial sum with interest earnings. Certainly, Equation 6 does not handle the 75 factor twice.

When Table 3 is analyzed by differences in cash flow, plan 1 is preferred at a MARR of 15 percent (NPV is -1.29), but plan 2 is preferred at a MARR of 17 percent (NPV is +0.67). The break-even MARR is 16.26 percent. This example is evidence that two or more solutions are possible with the NPV method, a fact that is not mentioned in the paper.

There is some logic to the fact that compound interest factors and their reciprocals (the present-worth factors) retain in their calculations the interest earning for each time period. This truth is well known and is inherently involved in all compound-interest calculations. But this is far different from stating that the compound-interest factors assume the reinvestment or retention of principal repayments as well as the interest earnings from their data of cash flow.

Obviously, under the rate-of-return procedure the project must earn at the rate solved for. That answer is what the rate-of-return method is supposed to produce. As compared to NPV, the important fact is that, when the MARR is less than the rate-of-return solution, the project earnings above MARR (the NPV dollars) have to be earned at a rate above MARR. That rate of earning is given by the rate-of-return solution. Thus, the two methods are consistent. All three methods (including the benefit/cost ratio) use the identical input data and the same compound-interest theory. Their answers must agree in result or be convertible to each other.

Perhaps Wohl is not referring to this retention and compounding of interest earnings within the mathematical system. But what does he have in mind about reinvestment? His paper does not prove the correctness of the reinvestment assumption. If this reinvestment theory is true, then such reinvestment assumption applies to every possible application of compound-interest mathematics that involves outgo and income cash flows.

Wohl states that the rate-of-return method assumes a reinvestment rate equal to the rate given by the solution and that the NPV method assumes a reinvestment rate equal to the MARR rate used in calculating the NPV. My conclusion is that neither method assumes a reinvestment of payback sums, and nowhere in the total use of compound interest can it be found that such reinvestment is included.

Acceptable managerial procedure is to give consideration to the timing and dollar amount of each payback cash flow. This consideration is not and should not be a part of the calculations to determine which of a pair of proposals has the highest rate of return or NPV. All such considerations for disposal of paybacks reflect judgments based on current positions of the owner and of the community.

Author's Closure

Winfrey refuses to recognize the problems created by use of the internal-rate-of-return method and by failure to account for reinvestment of any early-year gains that are accrued by a project. [Fortunately, other engineering economists understand these points and do take account of reinvestment in their engineering economy texts; among them would be Newnan (3, Appendix 7-A) and White and others (7).] Space does not permit me to retrace all the arguments that underlie these concepts or to respond to all of Winfrey's discussion. Rather, I will restrict my closure to the following points.

Winfrey begins by noting my failure to mention that industrial officials want to know the rate of earnings and that transportation officials prefer the benefit/cost ratio or internal-rate-of-return method. Sadly enough, this may be true since they probably were incorrectly taught to believe in the sanctity of the internal-rate-of-return method. More importantly, Winfrey proclaims that "the [internal] rate-of-return method has been used for years without the disastrous results [that Wohl said] could happen." But how can a project, once it is built, possibly provide any information to suggest or prove that some other rejected alternative was more preferable? After all, no project will automatically signal that the wrong alternative was chosen. Only the analyst can prevent that, before the fact.

Winfrey says, "In the rate-of-return method... it is acknowledged that two or more rates of return may be found. But Wohl does not mention that the NPV method may also give two or more solutions." On the contrary, for any given MARR value, the NPV method will give only one solution, but the rate-of-return method can easily give more than one solution. To indicate otherwise is to be misleading. (Uncertainty with respect to determining which MARR should be used is a very different matter and applies equally to all methods, not just to the NPV method.)

Winfrey is obviously disturbed by my example in Table 1, which assumes heavy capital outlays in future years and indicates a straight-line growth in benefits (net of operating costs) over a 30-year period. To make matters worse, Winfrey says, "It is useless and misleading in economic analysis to bring into the solution for today's choice of alternative those far-into-the-future needs for maintenance and repairs under conditions that, as of today, cannot be estimated under any acceptable probability of actual occurrence." Foolish or not, traffic and transportation engineers commonly use a 25- to 35-year analysis period and they commonly assume a straight-line growth in net revenues or benefits. As one pertinent example, only 3 years ago De Leuw, Cather and Company conducted a 30-year economic analysis of four transit alternatives for Pittsburgh (8, Chapter X). They estimated the federal share (or 80 percent) of the year-by-year capital outlays for the four alternatives [transit expressway revenue line (TERL), light rail transit (LRT), rapid rail transit (RRT), and express bus transit (EBT)] as shown in Table 6 (8, p. XI-18); it is obvious that both the LRT and EBT alternatives will have heavy capital outlays some considerable years in the future. Moreover, the consultants assumed that both the operating costs and benefits would grow linearly up to year 2005 (8, pp. X-7, X-12, X-17, and X-18). Thus, here is a recent and actual example of exactly the situation that Winfrey feels is unrealistic. Fortunately, the consultants did not use the internal-rate-of-return method to select the best alternative because, for the capital outlay patterns shown in Table 6 and for the assumed conditions for benefits and operating costs, there is a wide range of benefit levels that will produce multiple rates of return.

Winfrey comments: "Under certain conditions, a choice of alternatives may be different, depending on what MARRs are used and whether there are two or more changes in sign. But such difference is not due to the reinvestment assumption.... Regardless of the reinvestment assumption and other factors, Tables 2 and 3 under the rate-of-return solution present the true rate of return for each alternative." While Winfrey admits that problems can arise with multiple sign changes, he totally confuses the issue by incorrectly stating that the differences are not due to different reinvestment assumptions and that the internal rate of return is the true rate of return. Consider the example in Table 7. If Winfrey were correct in saying that the internal rate of return was the true rate of return, then presumably he would be indifferent between alternatives 1 and 2 since (according to him) the true yield of each is 20 percent. In fact, for a MARR value below 20 percent, alternative 1 is clearly better than alternative 2 and will have a higher true yield than will alternative 2. Simply stated, alternative 2 will have a true yield of 20 percent if and only if the $10 000 earned at the end of year 1 can be reinvested for the remaining year at a rate of exactly 20 percent. But if the $10 000 is reinvested at any other rate, the true yield for alternative 2 will not be equal to its internal rate of return. However, the internal rate of return for alternative 1 is equal to its true yield since its earnings are accrued entirely at the end of the two-year analysis period.

As cited before, Winfrey admits that multiple internal rates of return could lead to different economic choices than would result from the NPV method, but he insists that "such difference is not due to the reinvestment assumption." The example in Table 7 is instructive. First, with the internal-rate-of-return method and for a MARR less than 20 percent, alternative 2 would be selected as acceptable and the best (or the choice would be ambiguous because of the multiple incremental internal rates of return). By contrast, the NPV or benefit/cost ratio methods would both show that, for a MARR below 20 percent, alternative 1 is unambiguously acceptable and the best choice. Second, and despite Winfrey's assertion that the above economic choice difference is not due to the reinvestment assumption, there can be nothing other than a difference with respect to reinvestment that can result in alternative 2 being wrongly chosen by the internal-rate-of-return method. That is, the economic yield from alternative 1 is in no way affected by reinvestment since all the earnings are accrued at the end of the two-year analysis period. But with alternative 2, the issue must be, What happens to the $10 000 that is earned at the end of the first year? Certainly, these first-year earnings will not be ignored or placed in a safety deposit box. Rather, these funds either will be spent on consumption and thus provide extra enjoyment for the remaining year (i.e., they will be reinvested on early- rather than later-year enjoyment) or they will be reinvested in the best foregone investment and thus provide some yield for the last year. If we assume (as is normally done) that (a) people's rate of time preference is equal to the rate of productivity, (b) both are equal to the MARR, and (c) the borrowing rate is less than or equal to the MARR, then the only reinvestment rate for the first-year earnings that will result in the effective yield being equal to the internal rate of return for the two-year period is 20 percent. That is, if r is the reinvestment rate and r_2 is the internal rate of return for alternative 2, then r is the discount rate that satisfies the following identity:

$$0 = -101\,000 + [10\,000/(1 + r_2)] - [10\,000/(1 + r_2)] + [10\,000(1 + r)/(1 + r_2)^2] + [133\,440/(1 + r_2)^2] \quad (11)$$

If r_2 is 20 percent (or 0.20), then the only reinvestment rate (r) that can satisfy this identity is also 20 percent.

All in all, Winfrey seems to think that there is something inherently different about being explicit with respect to reinvestment as opposed to being implicit. For instance, in Winfrey's discussion of loan plan 1 for the example in Table 3, he says that the internal rate of return is the interest rate (or x percent) that satisfies Equation 8, or

Table 6. Annual federal shares of total system capital costs for four transit alternatives.

Year	Alternative 1, TERL ($000 000s)	Alternative 2, LRT ($000 000s)	Alternative 3, RRT ($000 000s)	Alternative 4, EBT ($000 000s)
1976	-	-	-	-
1977	72.5	16.8	22.2	6.7
1978	77.7	73.5	76.7	8.1
1979	75.6	64.2	76.4	57.1
1980	73.6	65.7	72.0	62.2
1981	77.6	69.1	90.9	63.2
1982	1.7	1.5	1.6	3.4
1983	1.8	1.6	1.8	3.7
1984	1.9	1.7	1.9	4.2
1985	2.0	1.8	2.0	4.2
1986	2.1	1.9	2.1	4.4
1987	2.3	2.0	2.3	10.3
1988	2.4	2.2	2.4	11.2
1989	2.6	2.3	2.6	18.1
1990	23.2	21.2	22.0	22.1
1991	2.9	4.0	2.9	5.8
1992	3.0	13.4	3.0	6.2
1993	3.2	28.7	3.2	6.5
1994	3.4	30.7	3.4	6.9
1995	3.6	3.3	3.6	7.4
1996	3.8	3.5	3.8	7.8
1997	4.1	3.7	4.1	8.2
1998	4.3	3.9	4.3	8.8
1999	4.6	4.1	4.6	9.3
2000	4.8	4.4	4.8	9.8
Total	454.7	452.2	414.6	355.7

Table 7. Net annual cash flows for two alternatives.

Year	Alternative 1 ($000s)	Alternative 2 ($000s)	Δ (Alternative 2 - Alternative 1) ($000s)
0	-100	-101	-1.00
1	0	+10	+10.00
2	+144	+133	-10.56

Note: r_1 = 20 percent; r_2 = 20 percent; and Δr for alternatives 1 and 2 = 20 and 780 percent.

$$0 = -70(1 + 0.01x)^{10} + 75(1 + 0.01x)^9 + 70 \qquad (12)$$

In turn, x = 22.84 percent. But Winfrey incorrectly says, "There is absolutely nothing in any of the above equations that supports the assumption of the reinvestment of the 75 payback." Winfrey has compounded the $75 year 1 payback amount for nine years at 22.84 percent and thus has implicitly assumed reinvestment at that amount for the nine remaining years. After all, what other explanation can there be?

Winfrey asks, "Could it be that it is only under conditions of two or more reversals of sign that the difference in choice of alternative is likely to differ between the rate-of-return and NPV methods?" The answer to this question is usually. That is, problems can also arise (even when there is a unique internal rate of return) if the order in which mutually exclusive alternatives are ranked is changed. This could occur when two or more alternatives have identical annual cash flows during the initial time period.

REFERENCES

7. J. A. White, M. H. Agee, and K. E. Case. Principles of Engineering Economic Analysis. Wiley, New York, 1977.
8. De Leuw, Cather and Company. Comparative Analysis of Alternative Transit Systems for the South Hills Corridor. Port Authority of Allegheny County, PA, March 1976.

Discussion

Richard S. Leavenworth, Department of Industrial and Systems Engineering, University of Florida, Gainesville

My intention is neither to defend nor to condemn the use of any of the accepted and proven valid methods for analyzing investment proposals. In our discussion, Reversals of Sign in Cash Flow Series (1, p. 556), we discuss the "cult of NPV" in some detail. Surely Wohl falls into this category of economist.

In teaching students, however, it is necessary that the major techniques actually used be presented in such a fashion that they will not be used incorrectly when the student applies them in the real world of decision making. Further, the results of a survey of the 1971 Fortune 500 firms (9) indicate that some 39-43 percent of the firms that responded use the rate-of-return criterion as their primary evaluation technique. Present worth ranked about second, along with payback period, as the primary techniques. Hence, a student should be taught the strengths and weaknesses of each of the major techniques of capital expenditure analysis.

Wohl is so intent on convincing us of the superiority of the present-worth criterion that, in the latter pages of his paper, he uses the future-worth criterion to demonstrate the validity of his arguments. Other authorities, who argue the superiority of future worth as the criterion, would appreciate his support.

Since the bulk of the discussion revolves around the multiple rate problem (what my coauthors and I call the problem of reversals of sign in cash-flow series), this is the specific point that I will address. Wohl makes the argument that each of the solving rates of return is exactly correct. This statement, however, requires a very specific qualification. Each is correct if the MARR that a firm requires from its investments equals the interest rate at which it acquires investment funds (the so-called borrowing rate). If this is in fact the case, then the firm is acting primarily as a money changer. There is little, if any, allowance in its financing-investment structure for productivity improvement.

Most firms other than regulated utilities have many more profitable ventures from among which to choose than their available capital will support. As the result, the MARR is usually substantially greater than the interest cost of capital. Once the constraint that cost of capital must equal the MARR is removed, it becomes necessary to interpret the meaning of the cash flows.

The two basic series of cash flows are pure investment series and pure borrowing series. The cash-flow pattern in Figure 4 represents an investment. At time 0, $1000 is expended to produce positive cash flows of $500 at the end of each of three periods. A plot of NPV is a function of interest rate (i) for a pure investment situation and will always be positive and equal to the algebraic sum of the cash-flow series (+$500) when i = 0, sweep in a downward direction to the right, and become asymptotic to the initial investment amount (-$1000) as i increases without bound. The value of i at which NPV(i) intersects the zero axis on the NPV scale is the rate of return (i*) on the particular cash flow. The NPV(i) plot for a pure investment situation will always exhibit this appearance. If the value of i* is greater than or equal to the MARR, the cash-flow series is preferable to alternative investment at the MARR.

The cash-flow series in Figure 5 clearly represents borrowing. At time 0, $1000 is received for which payments of $500 are made at the end of each of three periods. A plot of the NPV as a function of i for a pure borrowing situation will always be negative and equal to the algebraic sum of the cash-flow series (-$500) when i = 0, sweep in an upward direction to the right, and become asymptotic to the initial amount borrowed ($1000) as i increases without bound. The value of i at which NPV(i) intersects the zero axis on the NPV scale is the interest rate (i*) paid on the particular cash flow. The NPV(i) plot for a borrowing situation will always exhibit this appearance. If the value of i* is less than or equal to the firm's cost of capital from other sources, then this cash-flow borrowing series is preferable to acquiring capital from those other sources.

With these fundamental characteristics of pure investment situations and pure borrowing situations in mind, let us now turn our attention to Wohl's examples. In the interest of brevity, I will concentrate on the simplest cash-flow series of his several examples, Table 2. The solving rates of return for alternatives 1 and 2 compared to doing nothing are 25.69 and 26.16 percent, respectively. Plots of NPV(i) for each of these independent alternatives are shown in Figure 2. The independent cash flows are given in Table 8.

Having calculated the i*'s for each alternative independently and found alternative 2 to have a higher i* than alternative 1, casual observation of the cash flows suggests that there must be an unusual difference between them. Alternative 2 requires an additional $10 000 investment and nets $5 000 less return.

If alternatives 1 and 2 are mutually exclusive, it then becomes necessary to analyze this difference by first looking at and interpreting the difference between the two cash-flow series. This is indicated in Table 8 as Δ(Alternative 2 - Alternative 1).

The cash-flow series and its attendant NPV(i) plot in Figure 6 clearly do not match either the pure investment or pure borrowing situations. It must, therefore, be a mix of the two. Since the sum of the cash flows is negative (-$5000) it looks more like borrowing than anything else. However, the NPV(i) plot bends back down and becomes asymptotic to -$10 000, as would be the case for an investment project. The zero NPV axis is intersected twice.

The -$10 000 at time 0 clearly is an investment that can only be recovered from the +$45 000 at time 1. The -$40 000 at time 2 must be the repayment of a loan that can only come from the +$45 000 at time 1. Ignoring the firm's MARR, assume that the firm can acquire investment capital at a cost of 10 percent/period. Then the implied amount of the loan at time 1 would be $40 000 (1/1.10) = $36 360. This would leave ($45 000 - $36 360) = +$8640 to return the -$10 000 investment at time 0. This may be illustrated as in Table 9 (columns 1-3).

Since the initial investment of -$10 000 is not returned, the rate of return on the investment is negative and the present worth at any positive interest rate is negative. Thus, by either the rate-of-return or present-worth methods, the conclusion would be reached that alternative 2 is not preferable to alternative 1.

Assume, now, that the firm has a MARR of 25 percent. We could then evaluate the remaining balance of the time 1 cash flow by representing the loan as $45 000 - $10 000(1.25) = $32 500. The breakdown of the cash-flow series would then be as given in Table 9, columns 1, 4, and 5. The loan cost, in this case, is 23.08 percent. If the firm can borrow from any source at less than this figure, then alternative 2 is not preferable to alternative 1.

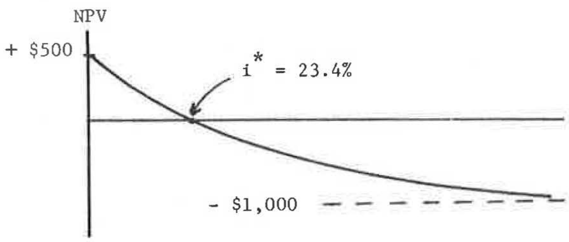
Suppose, however, that the analyst used the 25 percent MARR to evaluate the NPV of the entire cash-flow series Δ(Alternative 2 - Alternative 1) as Wohl suggests. This NPV is +$0.4 and alternative 2 is preferable to alternative 1. (In fact, for any chosen MARR between 21.92 and 228.08 percent the NPV will be positive.) This solution assumes that the cost of investment funds to the firm (the borrowing rate) is 25 percent and equals the rate required from investments (the MARR). If the borrowing rate does equal the MARR, then, of course, this result is correct and both the NPV and rate-of-return methods, properly applied, will show it (Table 9, columns 1, 6, and 7). The rate of return on the investment portion is 30 percent, in excess of the required MARR of 25 percent. Correspondingly, the present worth at 25 percent is positive and equal to 0.4.

Three basic points have been demonstrated:

1. The claim that present worth always leads to the correct choice requires the assumption that the borrowing rate to the firm always equals the MARR at which it invests funds. This is not the case with the vast majority of firms.

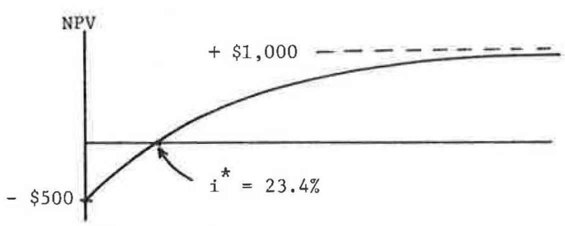
2. When a cash-flow series does not match the typical pure investment or pure borrowing models, that is, when it is a mix of borrowing and investing, it is necessary to divide the series into its investment and borrowing components.

Figure 4. Cash flow and NPV as a function of i, pure investment.



0	- $1,000
1	+ 500
2	+ 500
3	+ 500
∑	+ $ 500

Figure 5. Cash flow and NPV as a function of i, pure borrowing.



0	+ $1,000
1	- 500
2	- 500
3	- 500
∑	- $ 500

Table 8. Independent cash flows for the Table 2 example.

Year	Alternative 1 ($000s)	Alternative 2 ($000s)	Δ (Alternative 2 - Alternative 1) ($000s)
0	-100	-110	-10
1	+70	+115	+45
2	+70	+30	-40
Total	+40	+35	-5

Figure 6. Cash flow and NPV as a function of i, mixed borrowing and investment.

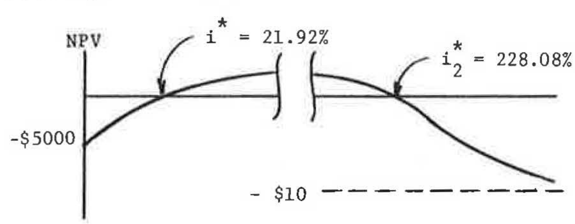


Table 9. The cash-flow series for the Table 2 example at different MARRs.

Year	Δ (Alternative 2 - Alternative 1) ($000s)	Loan Portion, 10% ($000s)	Investment Portion ($000s)	Loan Portion ($000s)	Investment Portion, 25% ($000s)	Loan Portion, 25% ($000s)	Investment Portion ($000s)
0	-10	0	-10.00	0	-10.0	0	-10
1	+45	+36.36	+8.64	+32.5	+12.5	+32	+13
2	-40	-40.00		-40.0	0	-40	0
Total	-5	-3.64	-1.36	-7.5	+2.5	-8	+3

3. Both the rate-of-return and present-worth methods will lead to correct choices when applied correctly. And both methods can lead to incorrect choices if applied incorrectly.

One additional caution should be added at this point. At no time have I used two interest rates that span the same time periods. One rate was used for cash flows in periods 0 and 1 and the other for periods 1 and 2. At no point are both interest rates used to transform cash flows or portions of cash flows over periods 0, 1, and 2. It is possible to show that $100 today is equivalent to $200 today if two interest rates are used over the same time periods. As the consequence, any method that allows for such equivalence transformations can lead to very strange results indeed.

REFERENCE

9. J.W. Petty, D.F. Scott, Jr., and M.M. Bird. The Capital Expenditure Decision-Making Process of Large Corporations. The Engineering Economist, Vol. 20, No. 3, Spring 1975, pp. 159-173.

Author's Closure

Leavenworth's discussion has confused the issues and introduced a confusing terminology. Let me attempt to unravel his complicated machinations.

Leavenworth says that "Wohl is so intent on convincing us of the superiority of the present-worth criterion, that, in the latter pages of his paper, he uses the future-worth criterion to demonstrate the validity of his arguments." Of course—any student of cash-flow equivalence knows that present worth and future worth are interrelated by a single-payment compound factor (or its inverse), that if one value is positive so is the other, and that, if the present worth for project A is larger than that for project B, then the future worth for project A will always be larger than that for project B. (This point was made in the paper.) In short, present worth is not a superior criterion to future worth, and vice versa; they are identical. Both are correct and consistent, and both are superior to using the internal-rate-of-return method. (Also, the benefit/cost ratio method is as valid as either present-worth or future-worth methods.)

Leavenworth then launches into lecture 1 of engineering economics, carefully explaining the mechanics of pure borrowing and pure investment situations. Then, turning to my Table 2 example, he shows that the net cash flows between alternatives 2 and 1 lead to incremental cash flows of - $10 000 in year 0, + $45 000 in year 1, and - $40 000 in year 2. Leavenworth then adds: "The cash-flow series and its attendant NPV(i) plot clearly does not match either the pure investment or pure borrowing situation. It must, therefore, be a mix of the two." Such an interpretation is outrageous. The situation is simple. The firm (in this example) has two pure investment choices (as well as the do-nothing alternative): to invest $100 000 in alternative 1 or to invest $110 000 in alternative 2. Or does it invest in neither? Moreover, if it invests in alternative 2 instead of 1, it simply must invest an extra $10 000 at time 0, then as a result earn $45 000 more at time 1 but $40 000 less at time 2. Thus, when deciding between alternatives 1 and 2 (and aside from the question of the acceptability of either), the question is whether the time 1 and time 2 amounts merit the extra investment of $10 000 at time 0. That is a pure investment question. By contrast, Leavenworth says, "Ignoring the firm's MARR, assume that the firm can acquire investment capital [i.e., borrow] at a cost of 10 percent/period. Then the implied amount of the loan at time 1 would be $40 000 (1/1.10) = $36 360." On the contrary, no loan is implied and no money needs to be borrowed at time 1, but only at time 0; that is, we must acquire the capital to make the extra $10 000 investment at time 0 if, in turn, we are to receive $45 000 more at time 1 but $40 000 less one year later. The real question is, Can I reinvest the $45 000 extra accrued at year 1 so as to recoup enough to offset both the $40 000 loss at time 2 as well as the foregone earnings on the initial investment? Moreover, if we ignore the firm's MARR (as Leavenworth suggested) and only consider the borrowing rate, then we overlook the most basic principle in engineering economics—to guarantee that the yield from additional investment is at least as large as that of our foregone opportunities.

Leavenworth's next example is also absurd. Therein, he assumes that the firm has a MARR of 25 percent and, thus, that the extra $10 000 in time 0 can be invested for one year, accumulating a total of $12 500 and thereby reducing the amount needed for a loan at time 1 to $32 500. Again, such calculations are mere fiction and fantasy. First and foremost, $10 000 is not available at time 0 for investment at the MARR of 25 percent. Rather, that amount of investment is required in order to gain an extra $45 000 at time 1 and $40 000 less at time 2; nor is a loan required or implied at time 1.

Last, it is not necessary (for his assumptions) to break the cash-flow stream into investment and loan portions and to carry out such arduous calculations. A simple NPV or net future worth calculation at the appropriate MARR will suffice. Of more importance, such machinations bear no resemblance to the internal-rate-of-return method, but are the heart of the ad hoc rate-of-return procedure of Grant, Ireson, and Leavenworth (1, Appendix B) to avoid the problems created by the internal-rate-of-return method.

My paper explored the use of the internal-rate-of-return and NPV methods under a rather strict and explicit set of assumptions about the borrowing and lending rates and about the time value of early (rather than late) consumption by people. Moreover, I explicitly noted that the use of different rates, consumption preferences, or different financing and investment plans could affect the effective yield that will be forthcoming from one project or another. In a roundabout fashion, Leavenworth is compelled to agree (as does Bergmann in a later discussion) that my conclusions are entirely correct for the assumptions that I made. His discussion then turned to the circumstances when the borrowing rate is different from the MARR. Unfortunately, his first two examples (which dealt with the incremental cash flows for the Table 2 example) were explicit only about one of these two rates (the borrowing rate in the first and the MARR in the second) but not about both. Then, later in his discussion, he explicitly assumes the borrowing rate and MARR to be 25 percent and concludes that "the rate of return on the investment portion is 30 percent, in excess of the required MARR of 25 percent." Presumably, then, an investor would conclude from Leavenworth's remarks that, all things considered, the effective annual yield (from investing in alternative 2 rather than 1) will be 30 percent for the two-year period, as compared with the MARR of 25 percent. This would suggest that, relative to investing in some foregone investment opportunity, the investor would accrue a profit of $1275 by the end of two years if the extra $10 000 were invested in alternative 2 rather than 1. But such a conclusion would be incorrect. Rather, for any borrowing and investment plan that can be devised (given Leavenworth's assumed borrowing rate and MARR), it can easily be shown that only a $625 profit and not $1275 can be accrued by the end of two years. In short, the effective yield from the two-year investment will be 27.48 percent (rounded off) rather than 30 percent, as indicated by Leavenworth.

The more important issue is, How should the analysis be done (especially for cash flows with two or more sign changes, as in the previous example) when the borrowing rate and MARR are not equal? First, we usually assume that the MARR is at least as large as the borrowing rate. Second, it also should be apparent that for an NPV analysis we do not apply the borrowing rate to any capital loans and then use the MARR value to discount the cash flows. Rather, we simply apply the MARR to determine the NPV. Remember, the choice is a simple one: Either acquire the investment capital for the project being analyzed and then accrue any and all future-year cash flows as a result thereof or acquire the same amount of investment capital and instead invest it in the best (otherwise) foregone investment opportunity that, by definition, will yield the MARR. The finance charges will be identical for both of the investment choices and thus can be disregarded in deciding whether to accept some project or do nothing (that is, or invest in the best foregone alternative). In short, it is neither necessary nor desirable to carry out the ad hoc rate-of-return procedure described by Leavenworth and others, at least not for these assumptions.

In the next to last paragraph of his discussion, Leavenworth claims to have demonstrated that "both the rate-of-return and NPV methods will lead to correct choices when applied correctly." On the contrary, what he did demonstrate was that the internal-rate-of-return method does not apply when multiple rates occur. He showed that an ad hoc rate-of-return method must be used in order to get around the problems created by multiple internal rates of return.

Discussion

Henry Malcolm Steiner, Department of Engineering Administration, George Washington University, Washington, D.C.

Wohl has taken a number of positions in his paper with which I disagree. I propose, however, to discuss only one area of disagreement—the reinvestment issue.

The question of the reinvestment assumption in the use of the internal-rate-of-return criteria for project analysis has become an issue of major importance. Some authors side with Wohl [White and others, for example (10, pp. 148-149)]; some do not [such as Grant and others (1, pp. 563-571)]. The issue is one of fundamental economic theory and definitions.

To illustrate the economic theory and resulting definition behind the concept of rate of return, I refer to Samuelson (11, pp. 599-601). Consider a perpetual investment (P) from which an amount (A) will be thrown off each year at interest rate (i). The interest rate is defined then as

$$i = A/P \tag{13}$$

The NPV of such an investment is P since

$$P = A/i \tag{14}$$

The internal rate of return of such an investment is i, since P = the discounted present worth of the costs, A(1/i) = the discounted present worth of the benefits, and Equation 14 shows them as equal. This is the definition of interest, present worth, and internal rate of return.

Under such a definition, is it necessary to invest A at i for Equation 14 to be true? The answer is obviously no. Yet the belief that such is necessary forms the basis of Wohl's contention that reinvestment of the benefits of an investment must be made, or be capable of being made, for internal-rate-of-return analysis to be valid.

The example of a perpetual investment illustrates this clearly. But what is at issue here is actually the fundamental definition of interest—in fact, of what return on investment means in the most basic sense. Investment returns may be invested (the reinvestment of Wohl) or consumed. Thus a definition of internal rate of return that requires reinvestment of the returns violates a fundamental economic concept, recognized by authorities such as Keynes (12, pp. 135-146) and Samuelson (11, pp. 599-601). A moment's thought brings to mind many examples where reinvestment of total product would be disastrous for society—agricultural production for example. If all the returns of agriculture were invested as seed, following the usual economic example of consumption and investment, nothing would be left for food.

I am not suggesting that adoption of Wohl's startling definition of internal rate of return is likely to lead the world to starvation. The world will continue to follow economic rules in spite of what academicians say. However, academicians should remain as much in touch with economic reality as they possibly can. Therefore, I urge Wohl to return to the fundamentals from which he has strayed so far.

The table that follows illustrates the induction of the formula for the future worth of an investment. A close examination of it may help to dispel the doubts of those who still cling to the reinvestment fallacy.

N	P ($)	Payment at Interest i ($ at 10%)	F_N($)
0	1000		
1		100	1100
2		110	1210
3		121	1331

$$F_1 = P + Pi = P(1 + i)^1 \tag{15}$$

$$F_2 = P(1 + i)^1 + iP(1 + i)^1 = P(1 + i)^2 \tag{16}$$

$$F_3 = P(1 + i)^2 + iP(1 + i)^2 = P(1 + i)^3 \tag{17}$$

Note that it was not necessary to invest the sum at the end of the third year ($1331) in order to induce our formula $F_N = P(1 + i)^N$. We merely had to reinvest the original investment itself and its interest. But this is the very meaning of investment. By extension, any cash flow of benefits that results from any investment may be considered single B_Ns thrown off because of each one's part of the original investment. Of course, future worth in all formulas implies reinvestment of the positive cash flows. But this is no more and no less than the meaning of future worth.

It is possible to specify that all returns of a given alternative must be reinvested at the opportunity cost of capital or at any other interest rate. But this reinvestment is merely a characteristic of the alternative and does not imply anything about the definition of the internal rate of return. Accordingly, I urge the logic of the discussion, so often mentioned by Wohl, of Grant and others (1, pp. 563-571).

Many authors also make Wohl's error. This has caused efforts to reconcile the internal-rate-of-return method with the other methods when, in fact, they were already reconciled. What is at stake in this controversy is Wohl's unacceptable definition of interest or internal rate of return. Return on capital may be consumed or reinvested, in whole or in part, but whether or not it is is a separate decision.

REFERENCES

10. J. A. White, M. H. Agee, and K. E. Case. Principles of Engineering Economic Analysis. Wiley, New York, 1977.
11. P. A. Samuelson. Economics, 10th Ed. McGraw-Hill, New York, 1976.
12. J. M. Keynes. The General Theory of Employment, Interest, and Money. Harcourt, Brace and Company, New York, 1936.

Author's Closure

Steiner has distorted the intent of my paper. To begin, Steiner correctly shows that the internal rate of return is equal to the interest rate at which the discounted benefits are equal to the discounted costs; this definition is identical to that which I incorporated in my paper and in Equation 1. (There are, of course, other correct ways of defining the same internal rate of return; e.g., it is the interest rate at which the future worth of the benefits is just equal to the future worth of the costs, or it is the interest rate at which the equivalent uniform annual benefits are just equal to the equivalent uniform annual costs.) But, then, he thoroughly distorts the matter by insisting (to paraphrase his words) that my startling definition of internal rate of return requires reinvestment of the returns and violates a fundamental economic concept, recognized by authorities such as Keynes and Samuelson. By contrast, examination of Equation 1 and the remarks that immediately follow will show that my definition of the internal rate of return is consistent and identical with those of every author cited by Steiner. In no way have either I or others implied or said that the definition of the internal rate of return required reinvestment of the returns during the analysis period. Rather, what I and many others have maintained is that use of the internal rate of return as the sole indicator of the effective yield from some investment program over the planning horizon or as the criterion for making economic choices can lead to incorrect or ambiguous answers. Also, to use the internal rate of return as the sole indicator of the yield to be anticipated from a project implies that the yield from either reinvestment or consumption of any early-year returns is equal to the internal rate of return.

Moreover, Steiner even agrees to the above conclusion by saying (in the last sentence of his discussion): "Return on capital may be consumed or reinvested, in whole or in part, but whether or not it is is a separate decision." Even so, Steiner misses the essence of the problem. And that is, whether a separate decision or not, the early-year returns will clearly be consumed or reinvested in something during the remaining years and thus the rate of reinvestment of earnings can and will affect the overall earnings, effective yield, and decision for projects. As a consequence, to simply use the internal rate of return as a guide to economic decision making is to open the door to incorrect or ambiguous economic choices by implying that the reinvestment rate is equal to the internal rate of return.

In my paper I indicated how we can compute the effective rate of return for a project, a yield figure that will properly reflect not only the internal earnings but also the external gains from consumption or reinvestment of early-year gains. I also later indicated that the effective rate of return will be equal to the internal rate of return only when all earnings are accumulated solely at the end of the analysis period or when the MARR (or reinvestment rate) is exactly equal to the internal rate of return.

Discussion

Dietrich R. Bergmann, St. Clair Shores, Michigan

The paper is a partial repetition and extension of Wohl's past criticisms of the internal-rate-of-return method for comparing mutually exclusive alternatives. His initial critique (13) used illustrations in which the solution for the internal rate of return on either the base or incremental investment was unique. Those criticisms and similar criticisms of other authors were discussed by me in 1973 (14).

Wohl's 1975 paper (15) continued his criticism of the internal-rate-of-return method and devoted a part of it to situations where the solution for the internal rate of return is not unique. Many of the points in his 1975 paper are repeated in the present paper. Consequently much of my reaction to the paper corresponds to the viewpoints expressed in my discussion (16), which was published with that paper. In the interest of brevity, I refer the reader to that discussion.

Several additional points are appropriate on this occasion. Wohl states that "...the MARR can be regarded as the opportunity cost of capital for both borrowing and lending situations." Such could be the case and would be true when there is no limit on an organization's borrowing; however, more often than not the interest rate that applies to an organization's borrowing will differ from the MARR for that organization's investment opportunities. His analysis of the illustrations presented in Tables 1-3 implicitly assumes that these two rates are identical to each other. For the instances that they are, I have no quarrel with Wohl's designation of the preferred project.

Consider, though, the general case where the rate of interest on money borrowed by the investing organization is different from the investing organization's MARR. Teichroew, Robichek, and Montalbano (17) have shown that, for this more general case, the NPV and the internal-rate-of-return methods must both be refined if the projects under consideration are mixed projects (i.e., projects that combine the features of both investment and lending situations). In each of the three illustrations summarized in Tables 1-3, either the base or incremental cash-benefit flow falls into the mixed project category as defined by Teichroew and others (17).

The specific approach to be used in analyzing a mixed project is dependent on the investment situation, of which the following three are envisioned:

1. The investing organization is also the organization that receives the benefits that accrue from the investment;
2. The investing organization is not the recipient of the benefits that accrue from the investment; instead, the benefits accrue directly and entirely to the public; and
3. The benefits generated by the investment accrue partly to the investing organization and partly to the general public.

Wohl's illustrations summarized in Tables 2 and 3 typify investment situation 1. His remaining illustration (summarized in Table 1) appears to be an example of investment situation 2.

When analyzing mixed projects that reflect investment situation 1, the cash-flow stream must be divided into a borrowing portion, for which the interest rate paid on the borrowing is specified, and an investment portion, for which the MARR is specified. After this division is completed, the analyst must evaluate the investment portion of the cash-flow stream. The project is deemed to be acceptable either if the internal rate of return for the investment portion of the cash-flow stream is in excess of the MARR or if the NPV of the investment portion is greater than zero.

Wohl and I would probably agree that the analysis of mixed projects, reflecting investment situation 2, has been inadequately treated in the literature. There are at least two unique features here, the first of which is that there is no borrowing by the investing organization from the project's beneficiaries. The second is that the benefited members of the public do not reinvest their benefits at the agency's MARR. Instead, those benefits are either consumed or reinvested at a rate that economists occasionally have referred to as the social rate of discount. Consequently the analytical problem presented by investment situation 2 is as evident in applying the customary NPV method as it is when applying Wohl's version of the internal-rate-of-return method.

One way to resolve the analytical problems posed by mixed projects involving investment situation 2 would be to apply the social rate of discount to transform all the project's public benefits to their future value at the end of the planning horizon and to accept the project if, for the modified cash-benefit flow stream, the present worth exceeds zero or the internal rate of return exceeds the MARR. I might add that, even for pure investment projects characterized by investment situation 1, there is a good basis to support the contention that public benefits should be transformed to their future value at the planning horizon by using the social rate of discount before the project's NPV or internal rate of return is calculated for the transformed cash-benefit flow stream.

Obviously, mixed projects, which are characterized by investment situation 3, require an analytical approach that combines the features described above for investment situations 1 and 2.

In conclusion,

1. It appears to me that Wohl has endorsed the NPV approach for evaluating mutually exclusive alternatives chiefly because, using his words, it is "conclusive and unambiguous." I agree that the NPV method is conclusive and unambiguous; however, I must add that application of the method in the way Wohl suggests makes invisible to the analyst those investment situations where his version of the method should be refined, as suggested in the preceding paragraphs.

2. It appears that Wohl would have us never calculate a rate of return on an investment proposal. If so, how are we to determine a value for the MARR when the investment opportunities are far in excess of the available budget? I can think of one way for doing that, but it is no less computationally tedious than calculating a rate of return.

3. Use of the internal-rate-of-return method does not imply, as Wohl and others suggest it does, that the positive cash flows from an investment alternative are reinvested at the internal rate of return. When we calculate the internal rate of return we are merely developing an ordinal ranking statistic that is compared with the MARR, just as the value of the NPV is compared with zero in order to identify the preferred project. No more and no less is either explicitly stipulated or implied by the two methods.

4. The problems that Wohl describes result partly from a disagreement about the definitions of the NPV and internal-rate-of-return methods for evaluating mutually exclusive investment alternatives. He is critical of the internal-rate-of-return method because he has defined it in a manner that involves the shortcomings he alleges it to have. Hopefully, future papers on this subject will strive to identify commonly accepted definitions for both methods.

ACKNOWLEDGMENT

This discussion has benefited from conversations with Henry M. Steiner and Richard S. Leavenworth. However, the views expressed are my own; they do not necessarily represent those of other individuals, my employer, or other organizations with which I am affiliated.

REFERENCES

13. M. Wohl and B. V. Martin. Evaluation of Mutually Exclusive Design Projects. Highway Research Board Special Rept. 92, 1967, 106 pp.
14. D. R. Bergmann. Evaluating Mutually Exclusive Investment Alternatives: Rate of Return Methodology Reconciled with Net Present Worth. HRB, Highway Research Record 437, 1973, pp. 75-82.
15. M. Wohl. Clarifying the Ambiguities of Internal Rate of Return Method Versus Net Present Value Method for Analyzing Mutually Exclusive Alternatives. TRB, Transportation Research Record 550, 1975, pp. 48-59, 61-62.
16. D. R. Bergmann. Discussion of paper, Clarifying the Ambiguities of Internal Rate of Return Method Versus Net Present Value Method for Analyzing Mutually Exclusive Alternatives. TRB, Transportation Research Record 550, 1975, pp. 59-61.
17. D. Teichroew, A. A. Robichek, and M. Montalbano. An Analysis of Criteria for Investment and Financing Decisions Under Certainty. Management Science, Vol. 12, No. 3, Nov. 1965, pp. 151-179.

Author's Closure

Bergmann concedes that he has no quarrel with my designation of the preferred project for the situation explored in my paper, in which the MARR can be regarded as the opportunity cost of capital for both borrowing and lending situations. However, his arguments deal mainly with an entirely different situation, that in which the borrowing rate and lending rate (that is, the yield of available investments outside the firm or MARR) are unequal. Accordingly, let me also address this different set of conditions as they relate to use of the NPV method, or such others as may be appropriate.

Bergmann argues that mixed projects are those that combine the features of both investment and lending situations; the examples in Tables 1-3 of my paper fall into the mixed-project category; and whenever the borrowing rate and MARR are equal, "the NPV and internal-rate-of-return methods must both be refined if the projects under consideration are mixed projects." Also, he says that, for projects such as those in Tables 2 and 3, "...the cash-flow stream must be divided into a borrowing portion, for which the interest rate paid on the borrowing is specified, and an investment portion, for which the MARR is specified." To the contrary, and notwithstanding the work of Teichroew (17) and Grant (1), it seldom is necessary to overcomplicate the analysis in this fashion. Put simply, the mere existence of a mixed project and of unequal borrowing and lending rates is not a sufficient condition for abandoning a simple NPV or net future value calculation.

Let me demonstrate this conclusion under the following conditions: (a) the lending rate (or rate at which an agency or firm is willing to loan or invest its assets) is its MARR, (b) the borrowing rate (BORR or rate at which investment capital can be acquired from outside sources) is unequal to the MARR, and (c) for public projects people's rate of time preference can be regarded as being equal to the MARR (more will be said about this assumption later). Under these conditions, an agency or firm when evaluating any project (or the difference between two competing projects) is faced with three options:

1. Acquire no capital and invest in none (i.e., the null alternative);
2. Acquire capital from liquidated assets, working capital, or borrowing and invest it in the best (otherwise) foregone investment opportunity; and
3. Acquire capital and invest it in the proposed project.

How, then, do we evaluate these options when MARR > BORR or MARR < BORR?

CASE 1: LET MARR > BORR

In this case, option 2 is always preferable to doing nothing. Thus, the choice is simply between a foregone alternative and the project in question. The investment capital to be acquired for each is identical, as will be the financing costs; therefore, we need only to compute the NPV (or, equivalently, the net future worth) of each proposed project with MARR as the discount rate. If the NPV of a proposed project is positive, it is acceptable; otherwise, invest in our best foregone investment alternative. Among a set of competing projects, the one having the highest positive NPV is the best. No other analysis is necessary for this case (nor for that in which MARR = BORR), and the complicated procedures of Teichroew and Grant can be ignored. Turning to the example in the table below, or to those in Tables 1-3, a straightforward calculation of the NPV at any given MARR value will always correctly identify a project's acceptability, as well as the best one when there is a competing choice. (Such calculations, if carried out for the example in the table below, will show that project X is unambiguously the best option whenever the MARR is between 30 and 40 percent. Also, the internal rate of return in the table is 0, 30, and 40 percent.)

Year	Net Annual Cash Flows ($)
0	-1000
1	+3700
2	-4520
3	+1820

CASE 2: MARR < BORR

At first glance, it may seem that, since the borrowing rate is greater than the MARR, a firm would always find option 2 unattractive. But that would not necessarily be correct. Simply stated, whenever a firm can acquire the necessary capital for investment in either the proposed project or a foregone alternative by liquidating some of its assets or by using working capital rather than by borrowing from outside sources, the financing costs would be represented by the MARR rather than by the BORR. Clearly a firm faced with these borrowing and lending rate conditions would avoid, so far as possible, borrowing from outside sources. Accordingly, if all the required capital can be acquired by using the agency or firm's assets, the appropriate discount rate is the MARR and the economic choice boils down to one between options 2 and 3, as before. And again, in this instance, a simple NPV or net future worth calculation at the MARR will indicate which choice is superior.

Second, whenever a firm or agency can act not only as an investor but also as a lender (a situation in which numerous private firms find themselves), and whenever the borrowing rate is larger than MARR, a firm would then regard the borrowing rate as its MARR and, accordingly, would simply carry out a straightforward NPV or net future worth analysis with the BORR (rather than MARR) as the appropriate discount rate. So once again, a complicated analysis is not necessary, even though the BORR > MARR.

Third, if and only if a firm can acquire the necessary investment capital only from outside sources, cannot become a lender (as well as investor), and the BORR > MARR, only then must we refine our techniques for evaluating mixed projects. (Frankly, I suspect that we seldom confront this rather special situation.) But even in this case, a simple and straightforward procedure will suffice to properly evaluate any proposal. [The one to be described is akin to that outlined in Teichroew (17); also, it is much simpler than that described by Grant(1).]

The underlying principle is to maximize the net future value and thus to minimize the amount of borrowing over the years. In short, borrow as little as possible and pay back borrowed funds as soon as possible. [Also, it should be evident that for this special case (i.e., capital can be acquired only by borrowing and the firm cannot be a lender) doing nothing is always preferable to acquiring capital and investing in a foregone alternative. Thus, the economic choice is simply between doing nothing and investing in a proposed project.] Accordingly, the procedure is as follows: Carry out a net future value analysis, compounding year by year over the n-year planning horizon. Whenever the accumulated net future value (in years 0 through n - 1) is negative, compound the balance to the next year at the BORR; if the accumulated balance is positive, compound the balance to the next year at the MARR. If the net future value at the end of n years is positive, the project will be acceptable; if not, the project is rejected. For a set of mutually exclusive projects, the one having the highest positive net future value (over the same planning horizon) will be the best.

As an example of the above procedure, let us carry out the net future value analysis for the cash flows shown in the preceding table, for a BORR of 35 percent and for a MARR of 32 percent (see Table 10). Since the net future value at the end of three years is negative (or -94.30), project X will be unacceptable and thus doing nothing will be best.

In summary, a simple NPV analysis with a discount rate of the MARR will always suffice so long as the BORR is equal to or less than the MARR. When the BORR is larger than the MARR, a simple NPV analysis with a discount rate of MARR will also lead to the correct economic choice if the firm can use its own assets to acquire the necessary capital. When the required capital can only be acquired from outside sources, and the firm can be a lender as well as investor, a simple NPV analysis with a discount rate of BORR will lead to correct economic choices. Finally, a more complicated type of

Table 10. Net future value analysis.

End of Year t	Cash Flow in Year t($)	Prior-Year Net Future Value Compounded to Year t ($)	Net Future Value at End of Year t ($)
0	-1000	-	-1000, compound at 35 percent
1	+3700	-1000(1.35) = -1350	+2350, compound at 32 percent
2	-4520	+2350(1.32) = +3102	-1418, compound at 35 percent
3	+1820	-1418(1.35) = -1914.3	-94.30

analysis (of the sort alluded to by Bergmann and described in the previous two paragraphs) must be used only when a firm cannot be a lender, cannot acquire capital except from outside sources, and has a borrowing rate greater than its MARR.

Bergmann also maintains that any benefits received by the public are consumed or reinvested at the social rate of discount. Many other economists argue differently and with persuasion. Not only is this a murky subject, but it also is one that is fraught with difficulty when we attempt to determine the appropriate social rate of discount. In sum, there is only an arbitrary basis for deciding on the propriety of its use as well as its value. [See Margolis's (18) review of this subject.]

Two final comments: One, Bergman is wrong in saying "...Wohl would have us never calculate a rate of return on an investment proposal." To the contrary, and as noted in my paper, if a rate-of-return figure is necessary (say, because of budget constraints) before the fact, then an effective rate of return should be calculated rather than the internal rate of return. Moreover, it is perfectly obvious that we need the actual effective yield that is being obtained from other ongoing (or past) projects. Two, Bergmann is clearly wrong in saying that: "Use of the internal-rate-of-return method does not imply, as Wohl and others suggest it does, that the positive cash flows from an investment alternative are reinvested at the internal rate of return." On the contrary, if the internal rates of return are used as the sole guide to make economic choices among mutually exclusive choices, then certainly he is wrong. Assume for the Table 7 example that the MARR is 20 percent (and equal to the borrowing rate). Then, what else other than an assumed reinvestment of the $10 000 in year 1 at 20 percent could have caused the two alternatives to be exactly equivalent for the two-year period? In a word, nothing.

REFERENCE

18. J. Margolis. The Economic Evaluation of Federal Water Resource Development. The American Economic Review, Vol. 49, No. 1, March 1959, pp. 102 ff.

Publication of this paper sponsored by Task Force on Economic Analysis.

Potential of Pricing Solutions for Urban Transportation Problems: An Empirical Assessment

John Pucher, Department of Urban Planning, Rutgers University, New Brunswick, New Jersey
Jerome Rothenberg, Department of Economics, Massachusetts Institute of Technology, Cambridge

This paper surveys the available empirical evidence on the elasticity of travel demand to assess the potential of pricing policies to alter travel behavior and thereby to solve various urban transportation problems. The first set of studies considers the responsiveness of fuel use to changes in gasoline price. The second set, econometric models of urban travel demand, estimates the direct and cross-price elasticities of the use of different modes with respect to different components of trip cost. The third set of evidence is composed of arc elasticity estimates based on the impacts on travel behavior of actual changes in the levels of roadway user charges and transit fares. For each study dealt with, the paper briefly summarizes its methodology, data base, and findings and subjects these to critical evaluation. The paper concludes with an evaluation of the body of results for the usefulness of pricing policies in urban transportation.

Economists have criticized perverse pricing as the crux of the urban transportation problem and, thus, have regarded corrective pricing policy as the key to the solution. The theoretical basis for such alternative pricing involves the need to internalize the often significant external social costs associated with urban travel (such as congestion, air pollution, and noise), particularly as these vary with respect to time of day, route, and mode of travel.

The objective of this paper is to determine how responsive urban travel behavior would be to various corrective pricing strategies:

1. To what extent, for example, could peak-period pricing alleviate congestion by diverting automobile drivers to other modes or to use of their automobiles at less-congested times or on less-congested routes?
2. To what degree might higher gasoline prices encourage motorists to economize on fuel by driving fewer kilometers or by purchasing smaller, more fuel-efficient automobiles? and

3. How much would transit ridership increase as a result of surcharges on automobile use or reductions in transit fares?

Elasticity of demand is a useful measure for answering these questions. It is defined as the ratio of the percentage change in demanded use to the percentage change in the relevant exogenous variable. This dimensionless measure indicates the relative sensitivity of the demand for a specific type of travel to the cost of such travel, the cost of competitive or complementary travel, and to performance levels or other aspects of the quality of travel alternatives. The degree of price elasticity is important for public policy purposes because, in general, the more elastic is the demand, the greater is the potential leverage of pricing policy in altering urban travel behavior and thus in helping to solve various transportation problems.

This paper assembles information about the size of relevant demand elasticities from a variety of recent empirical studies. Our first set of studies considers the responsiveness of fuel use to changes in gasoline prices. Next, we examine a number of urban travel-demand models, which yield estimates of direct and cross-price responsiveness for different urban travel modes and for different components of travel cost. Finally, we note the wide range of arc elasticity estimates based on impacts of actual changes in the levels of roadway user charges and transit fares. For each study dealt with, we briefly summarize its methodology, data base, and findings and subject these to critical evaluation. We compare results of studies based on different methodologies to test for robustness of elasticity estimates. The paper concludes with an evaluation of the body of results for the use of pricing in public policy.

PRICING GASOLINE TO CONSERVE FUEL

Public interest in fuel conservation peaked in intensity during the energy crisis of 1973-1974 and has remained an important consideration of federal policy in the years since. Because the highway use of gasoline has been a large percentage of the total demand for petroleum products, such transportation fuel usage, particularly by automobiles, has been a prime target for energy conservation. In the short run, gasoline savings can be achieved by (a) the elimination of nonessential trips, (b) a shift to modes less energy-intensive than single-occupant automobile use (such as transit or carpooling), or (c) a reduction in driving speeds to increase kilometers per liter of fuel. Additional means of conservation include a reduction in the level of automobile ownership, technical development and market penetration of smaller, more fuel-efficient automobiles, and a transformation of urban land-use patterns toward decreased spatial diffuseness.

Each of these fuel-conservation measures is to some extent automatically encouraged by higher gasoline prices, which signal to consumers the increased scarcity of energy resources and force them either to conserve or pay the price for not doing so. Without government intervention, fuel prices would be expected to rise as a result of the interaction of diminished fuel supply and increased demand and thus to induce some amount of conservation, but policy intervention can mitigate or even accentuate the rise.

In addition to leverage for overall energy use, gasoline prices can be the instrument for influencing the absolute and relative use of automobiles in urban transportation. Whatever its policy goal, however, its potential effectiveness depends on the demand responsiveness to its use. A useful approximation to this demand sensitivity is the own-price elasticity of demand for gasoline. The following section reviews several different types of econometric models of gasoline demand that yield estimates of this price elasticity measure.

Econometric Models

The econometric models described below are roughly of two broad types. Flow-adjustment models generally express the demand for gasoline in the current period as a function of the consumption of gasoline in the previous period, the real price of gasoline in the current period, real personal income per capita, and a variety of other variables thought to influence the level of fuel use (see Table 1, studies 1, 2, and 5) (1-5). The identifying characteristic of this approach is the use of last period's gasoline consumption to predict the current period's consumption.

It is assumed that the degree of consumer response to any change in price is a function of the length of time allowed for the response to occur. For example, in the long run, extensive adjustments in travel behavior are possible through changes in the level of automobile ownership, vehicle efficiency, and residential and workplace locations. Lags in the adjustment process are not explicitly represented in flow-adjustment models. Rather, to proxy for these complex initial influences and the time-phased adjustment, past levels of the dependent variable are used to explain current levels of that variable.

The more significant the lagged variable is in explaining current levels of gasoline demand, the greater the friction of adjustment and the longer the implied period of time requisite to a full adjustment. The actual importance of such momentum can be calculated in these models. Through the estimated coefficient of the lagged dependent variable, the time period considered to be the long run is mathematically inferrable from the estimated equation. Both short- and long-term impacts can be measured. The former is the estimated parameter weight on each explanatory variable. The latter are the weights of the steady-state equilibrium of the system.

Unfortunately, flow-adjustment models provide no specific information on the nature of the adjustment process, only on the implicit timing and overall size. Yet, for policy purposes, it is important to know to what degree a price-induced reduction in gasoline use stems differentially from changes in the number, length, timing, or location of automobile trips; changes in gasoline efficiency; changes in automobile ownership; or shift in modal choice. The specific nature of the adjustment process can influence significantly the relative attractiveness of different fuel-conservation policies because they differ with respect to other public policy objectives.

A variant on the flow-adjustment models substitutes measures of vehicle fleet size and gasoline efficiency for the lagged dependent variable. Thus, the drag on adjustments in travel behavior to gasoline price attributable to these two long-term considerations is estimated directly. The short-term models of the Rand Corporation and Charles River Associates are of this type (see Table 1, studies 3 and 4).

The Rand Corporation's multiequation recursive model distills the long-term adjustment process into an assumed sequence of component stages (Table 1, study 6). In contrast to the flow-adjustment models, each aspect of the long-term impact is explicitly represented.

The effect of gasoline price on the level of fuel use is estimated both directly (through its impact on vehicle kilometers traveled by a given automobile stock with a fixed average vehicle efficiency) as well as indirectly (through its influence on new and used automobile ownership and vehicle fuel efficiency). By breaking down the long-term adjustment process into its components, the relative importance of each is determined as well as the total impact. Unfortunately, however, this approach fails to represent explicitly the time phasing of adjustment, the main advantage of the flow-adjustment models.

Table 1. Aggregate econometric models of gasoline demand and vehicle kilometers of travel.

Study	Dependent Variable	Independent Variables	Functional Form	Sample	Estimation Technique	Short-Range Elasticity*	Long-Term Elasticity
Data Resources (1)	Highway motor fuel use per capita	Real gasoline price by state, real disposable income per capita, lagged per capita highway gasoline consumption	Log linear	Quarterly data, 1963-1972, 48 states and DC	Error components	-0.23 to -0.30	-0.32 to -0.54
McGillivray-Urban Institute (2)	Automobile gasoline use per capita	Real gasoline price, new passenger automobile registration per capita, average gasoline consumption per automobile, lagged per capita automobile gasoline consumption	Linear	National total, United States, ordinary least squares 1951-1969, annual	Ordinary least squares	-0.23	-0.76
Rand I (single-equation, short run) (3)	Highway motor fuel use per capita	Real gasoline price, disposable (per capita) personal income, total vehicle registrations per capita, fuel efficiency of vehicle fleet by state, trucks as percentage of total vehicles	Log	Annual, pooled time-series cross-section, 1955-1970 for 48 states	Ordinary least squares	-0.26	
	Automobile gasoline use per capita	Real gasoline price, vehicle registrations per capita, dummy variable for safety and emissions standards	Log	Annual national total, 1950-1972	Ordinary least squares	-0.38 to -0.43	
	Vehicle kilometers traveled per capita	Real gasoline price, vehicle registrations per capita, dummy variables for safety and emissions standards	Log	Annual national total, 1950-1972	Ordinary least squares	-0.37 to -0.39	
Charles River Associates (short run) (4)	Highway gasoline consumption per licensed driver	Real gasoline price, stock of automobiles per licensed driver, stock of trucks per licensed driver, augmented fuel efficiency of automobile stock, number of post-1968 registered vehicles on road in each year per licensed driver	Linear	Annual, pooled time-series cross-section, 1950-1971, annual, 48 states and DC	Two-stage least squares	-0.18	
Charles River Associates (long run) (5)	Highway gasoline consumption per licensed driver	Real gasoline price, real disposable income per capita, lagged highway gasoline use per licensed driver, six dummy variables for each of six different types of gasoline consumption characteristics of state	Log	1951-1971, annual pooled, 48 states and DC	Two-stage least squares	-0.29	-1.37
Rand II (5-equation, recursive, long run) (6)	Used automobile price	New automobile price, real gasoline price, disposable income, lagged automobile stock, automobile strike dummy variable	Linear	Annual, national total, United States, 1950-1972	Two-stage least squares, ordinary least squares		
	New automobile sales	New automobile price, used automobile price, growth in disposable income, automobile strike dummy variable	Linear		Two-stage least squares, ordinary least squares		
	Used automobile stock	Used automobile price, new automobile price, real gasoline price, disposable income, strike dummy variable	Linear		Two-stage least squares, ordinary least squares		
	Average kilometers per liter	Real gasoline price, dummy variable for safety and emissions standards	Log		Ordinary least squares		
	Automobile kilometers driven per capita	Automobile stock, real gasoline price, safety and emissions dummy variables	Log		Ordinary-least squares		
	Automobile gasoline consumption per capita	Average km/L × automobile-km driven per capita					-0.64 to -0.68
Chase Econometrics (7)	Total vehicle kilometers traveled	Total automobile ownership, relative price of gasoline and oil, change in consumer price index, average price of new automobiles lagged two periods, growth in wages and salaries	Linear	Annual, national total, United States, 1956-1972	Ordinary least squares	-0.5	-0.72
Federal Energy Administration (8)	Vehicle kilometers traveled per capita	Automobile operating costs per kilometer, real disposable income per capita, unemployment rate, lagged vehicle kilometers of travel	Linear	Annual, national total, United States, 1950-1972	Nonlinear least squares with first-order automobile-regressive transformation	-0.12	-0.72

*Short range means no more than one year.

Each of these three types of model yields a different type of elasticity estimate. The short-term elasticity derived from lag-adjustment models is calculated directly from the estimated equation by specification of an arbitrary short-term period (set here to one year for comparability). Short-term elasticities derived from models other than the flow-adjustment type are the estimated effects of gasoline price on fuel consumption, holding automobile ownership levels and vehicle efficiency constant. So in the former model type the short term is an arbitrary time period; in the latter it is the period for which automobile ownership and fleet efficiency are fixed.

Similarly, for the second and third types of models, the long-term period is implicitly that length of time after which price will also have affected gasoline consumption indirectly through changes in automobile ownership and vehicle efficiency. One cannot infer from these models the actual number of years required for such adjustment. The length of the long run can be explicitly calculated, however, in lag-adjustment models. It is equal to the number of periods subsequent to some price change required for the difference equation to reach a steady-state equilibrium solution. The Data Resources study, for example, implies a long term of about 2.5 years; McGillivray's long-term period is 10 years, and the Charles River Associates long term slightly exceeds 12 years. Thus, the duration of the long term varies significantly from one lag-adjustment model to another. These varying interpretations of short and long term make comparisons of elasticity estimates from different models dubious.

Empirical Results

Despite a wide variety of specifications, estimation techniques, and data bases, most of the recent econometric studies of the demand for gasoline indicate that the short-term direct elasticity of gasoline use with respect to gasoline price falls within the range of about -0.2 to -0.3 (see Table 1). In contrast, long-term elasticity estimates from this same set of studies vary substantially, ranging from -0.32 to -1.37. The duration of the long-term ranges (implicitly from 2.5 to 12 years) makes comparison of the results especially difficult because the specific durations are not deliberately chosen but rather are the statistical outcome of the type of demand modeling employed. Estimates of the elasticity of new automobile sales with respect to gasoline price range from -0.7 to -1.0 (3, 5). [Corresponding estimates of the elasticity of new automobile sales with respect to new automobile price ranged from -0.88 to -1.6 (3, 5).] Especially in light of the extremely important effect of automobile ownership on urban travel behavior, these figures suggest that public policies that affect the price of gasoline may substantially alter travel behavior. The short-term, direct effect of gasoline price is reinforced by the indirect, long-term effect of price on automobile ownership and, in turn, on travel behavior.

Most of the studies reviewed here are beset by statistical problems of estimation—multicollinearity and simultaneous-equations bias being the most important. [For a detailed analysis of the methodology, statistical limitations, and data bases of each of the studies, consult the background report on which this paper is based (6). This report also contains a lengthier discussion of the empirical results and their policy implications.] The convergence of the short-term estimates lends credence to a -0.2 to -0.3 range of elasticity values; however, despite divergence, the long-term elasticity estimates uniformly increase over longer periods, especially as turnover occurs in the automobile fleet. None of the studies explicitly examines the possibility that large, long-lasting changes in fuel price might also significantly affect locational decisions over time, thereby further increasing the long-term price elasticity.

Because the gasoline studies differentiate neither between urban and rural demand nor (more importantly) among regions of the country, income classes, trip purposes, or trip routes, they are of limited usefulness in the determination of the potential leverage of pricing to alter specifically urban travel behavior. Yet they do suggest considerable aggregate gasoline conservation in response to higher gasoline prices, especially in the long run.

SENSITIVITY OF URBAN TRAVEL DEMAND

Three types of studies have examined the elasticity of demand for urban travel. Direct travel-demand models estimate total zone-to-zone traffic volumes, by mode and trip purpose, as a function of various cost and performance variables. In contrast, disaggregate travel-demand models use individual household observations to determine the independent influence of each of a number of cost, performance, and socioeconomic variables on the probability that individuals with specific socioeconomic characteristics will select particular origin-destination pairs, travel times, modes, and trip frequencies. Finally, highly fragmentary evidence of the effect on travel behavior of changes in transit fares, bridge and tunnel tolls, and parking charges has been used to calculate arc elasticities, measures of demand sensitivity that are based on only the two price-quantity observations of a single price change.

Evidence from Direct Demand Models

The aggregate version of the behavioral travel-demand model estimates the total number of round trips between each origin and destination zone in an urban area as a function of the travel times and money costs of each of the available alternative modes, the socioeconomic characteristics of the origin zones (usually income per household and per capita automobile ownership), and the employment density of destination zones. A separate equation is estimated for each mode and trip purpose combination on the basis of aggregate, not individual, interzonal variations in the variables. Simultaneous determination of mode choice, trip frequency, trip distribution, and trip generation is assumed.

The earliest, most fully developed, and best-known of the aggregate models is the Charles River Associates direct-demand model designed to assess the impact of free transit in Boston (7). Four separate mode and trip purpose demand equations are estimated: automobile work trip, automobile shopping trip, transit work trip, and transit shopping trip. For each, the dependent variable is the total number of round trips taken between each origin and destination zone by the specified mode. The explanatory variables are socioeconomic characteristics of each origin (reflecting trip generation and modal preferences), land use characteristics of each destination (reflecting the relative trip attractiveness of these), and performance and cost measures for both automobile and transit on each specific origin-destination link (reflecting the relative desirability of alternative destinations and modes). All independent variables were included in both linear and logarithmic forms to capture the effect on travel behavior of both absolute and relative changes in ex-

Table 2. Elasticities from direct demand models.

Independent Variable	Automobile Work Trips: Charles River Associates	Automobile Work Trips: Fulkerson	Automobile Shopping Trips: Charles River Associates	Transit Work Trips: Charles River Associates	Transit Work Trips: Talvitie: Rail	Transit Work Trips: Talvitie: Bus	Transit Work Trips: Talvitie: Pooled Transit	Transit Work Trips: Fulkerson	Transit Shopping Trips: Charles River Associates
Automobile									
In-vehicle time	-0.82	-0.39	-1.02	0	0.84	0	0.37		0
Access time	-1.44		-1.44	0	0	0	0		0
Line-haul operation cost	-0.49	-0.12	-0.88	0	1.34	0.36	0.80		0
Out-of-pocket cost	-0.07		-1.65	0	0	0	0		0
Transit									
Line-haul cost	0.14	0.15	0	-0.09			-0.38	-0.40	
Access cost	0		0	-0.10			-0.08		-0.32
Line-haul time	0		0.10	-0.39			-0.20	-0.19	
Access time	0.37		0	-0.71			-0.69	-0.38	-0.59
Bus									
Out-of-vehicle time					0	-1.84			
In-vehicle time					0.23	-1.10			
Access cost					0.08	0			
Line-haul cost					0	-0.51			
Rail									
Out-of-vehicle time					-1.74	1.00			
Access-in-vehicle time					-0.55	0.25			
Total access time					-2.06	1.15			
Line-haul time					-0.80	1.02			
Access cost					-0.30	0.28			
Line-haul cost					-1.80	0			

planatory variables. To counter the inevitable multicollinearity among explanatory variables, which necessarily accompanies such a dual specification, the ordinary least-squares estimation procedure was modified to constrain direct elasticities to be nonpositive and cross-elasticities to be nonnegative. Almost half of the coefficients in the four demand equations were set equal to zero because the imposed constraints were binding. Explanatory variables (primarily cross-price effects) whose logarithmic- and linear-form parameters were both set equal to zero for this reason had their predicted signs rejected but were not actually estimated as zero.

The statistical reliability of the four estimated equations is dubious. Multicollinearity is an obvious problem for several reasons: (a) the log and linear specification of each variable; (b) the pervasive effect of income on automobile ownership, on the degree of disutility represented by any measured level of time or cost, and on modal choice and the total demand for travel; and (c) the inevitable correlation between travel times on different modes between the same two zones. However, no analysis of the effect of the multicollinearity problem was made in the Charles River Associates study. Although the constrained estimation process is employed specifically to mitigate collinearity problems of the first type, it may have exacerbated the estimation bias induced by the relationship across modes between travel times and costs or the multifaceted impact of income on the other explanatory variables. Moreover, it is likely that the partial effects of travel time and cost on travel demand depend importantly on consumer income levels; a multiplicative specification would probably have been more appropriate to capture this influence.

The form in which the model is estimated assumes that each of the right-hand explanatory variables is truly exogenous and not mutually a function of the dependent variable (traffic volume). Yet travel times, and thus also gasoline operating costs, are dependent on traffic volumes via the level of traffic congestion. They vary nonlinearly as such volumes change. The failure to account for this joint determination may have introduced serious simultaneous-equation bias into the estimation process—another reason to regard the coefficient estimates with suspicion. The direction of multicollinearity-induced bias is ambiguous, but one consideration suggests that it understates elasticities. Nominal trip costs omit congestion (time) costs. So, heavily used origin-destination pairs have higher real costs than are measured; real costs for lightly used pairs are as measured. Large differences in levels of route use are associated generally with smaller real cost differentials than are measured; estimated elasticities are thereby understated.

Another problem is that the mode alternative to automobiles was taken as an assumed homogeneous transit mode—simply the sum of all nonautomobile trips. In fact, these nonautomobile trips are quite heterogeneous. The described treatment may impart another bias toward understating the potential substitutability of automobile with nonautomobile modes, since actual changes in modal-split behavior will be influenced by the most substitutive of alternative modes, not the average of such modes. The behavior of such closer substitutions is hidden (indeed, distorted) by the present treatment: Aggregations probably understate behavioral sensitivity toward actual intermodal competition.

The elasticity results given in Table 2 were computed from estimated regression parameters. (Unfortunately, the statistical significance of the latter was not reported; hence, there is no way to gauge the reliability of the derived elasticity values.) These elasticities indicate that automobile travel demand is moderately responsive to the time components of real trip cost, more for out-of-vehicle time (walking) than for in-vehicle time. Transit demand is less sensitive to time, but only the elasticity with respect to line-haul time is really low. Use of both modes is less responsive to money costs than to travel time. Automobile demand is more responsive than transit demand—the latter shows a near zero sensitivity. The few estimated cross-elasticities are extremely low. In terms of trip purpose, work trips are generally less responsive to cost components than are shopping trips, an expected result given the largely fixed number and geographic pattern of commuting trips in the short run.

Overall, the Charles River Associates direct-demand results indicate that, although transport prices affect travel demand in the manner predicted, several effects tend to be quite small. Socioeconomic factors are overwhelmingly more important than cost or performance in determining travel behavior, especially modal choice. These results have been the basis of pessimistic appraisals of the usefulness of pricing in urban transportation policy; however, such appraisals may not, in fact, be justified.

The econometric problems of multicollinearity, misspecification, simultaneous-equations bias, and overly aggregative treatment of nonautomobile trips may have understated sensitivity to cost. There are other grounds as well for suspecting downward bias of the demand elasticities. Time of day and specific route of travel are important policy issues. Neither of these is treated in the Charles River Associates model. Differentiation by these trip dimensions would expose a wider variety of travel substitutes and, thus, reveal inherently higher elasticities with respect to crucial aspects of travel. Moreover, the measured elasticities represent an average responsiveness of travel among all origin-destination zone pairs. Insofar as pricing policy can be designed to differentiate both by route and time of day, demand responsiveness will be greater than suggested by the Charles River Associates estimates. This omission of specific route and time-of-day aspects of trips may be a serious bias because travel externalities vary significantly with time of day and specific route; optimal pricing policy would presumably differentiate according to these dimensions. The Charles River Associates estimates are probably lower bounds of demand responsiveness to the variety of cost-impacting policies that might reasonably be employed to achieve public policy goals.

Almost identical in structure to the Charles River Associates model, Talvitie's direct demand model of transit work trips in Chicago is subject to virtually all the criticisms levied at the Charles River Associates model (8). As Table 2 indicates, however, disaggregation of the transit mode into bus and rail results in substantially larger elasticities. Even in his pooled transit version, Talvitie estimates cross-elasticities of transit demand with respect to automobile in-vehicle time and automobile operating cost of 0.37 and 0.80, respectively—values large enough to cast further doubt on the Charles River Associates estimates obtained by constraining these cross-elasticities to zero.

The Moses and Williamson study of modal choice by Chicago commuters also found a considerable elasticity of transit demand with respect to the level of automobile operating costs (9). These survey results indicated that lowering transit fares to zero would have diverted to the bus only 13 percent of the automobile drivers who reported that this was their next most preferred mode; however, the imposition of a $1.00 round-trip surcharge on automobile commuting for work trips would have shifted to the bus 47 percent of automobile drivers who indicated that the bus was their second-best mode. The reliability of the survey data on which these estimates are based is questionable, and the shift estimates are upper bounds. Nevertheless, it is noteworthy that the cross-price effect of automobile costs on transit demand not only appears to be significant, but even exceeds the own-price effect of fares on transit demand.

The direct-demand model of work trips that Fulkerson estimated for Louisville is of interest primarily because, although it employs most of the same explanatory variables, it exposes several of the deficiencies of the aggregate model that the Charles River Associates and Talvitie versions conceal (10). The avoidance of the dual log and linear specification of each explanatory variable and the straightforward use of ordinary least squares without a priori constraints facilitate the isolation of multicollinearity problems other than those obviously associated with the dual variable specification. Not surprisingly, an analysis of the correlations among explanatory variables indicated that system performance and cost variables (both within and across modes) are highly interrelated and that this was particularly serious in light of the low coefficient of determination of the estimated direct-demand equations. Due to the fragmentary reporting of the Charles River Associates and Talvitie models, one can only surmise the degree to which the estimated demand equations were distorted, but the Fulkerson results suggest that the problem is not minor.

Fulkerson facilitates interpretation of the results by reporting t-statistics. The uniformly low magnitude of the estimated coefficients of travel cost and time variables (as well as the insignificant t-statistics) may be at least partly the result of the reported multicollinearity among variables. Interestingly, by far the variable that most affects both automobile and transit travel demand is automobile ownership. Not only are the parameter coefficients of high statistical significance, but the associated demand elasticities exceed 1.0 in absolute value. For automobiles, in-vehicle time and money operating cost elasticities are considerably smaller than the Charles River Associates estimates (see Table 2). The cross-elasticity with respect to transit line-haul money cost is essentially the same as in the Charles River Associates study. For transit work trips, transit line-haul time and money cost have the same elasticities as in Talvitie's study. But the travel-time elasticity is considerably less and the cost elasticity is greater than the comparable Charles River Associates values for Boston. The elasticity for transit access time in the Fulkerson model is only half that estimated in the other two studies.

Fulkerson attributes her generally lower elasticities to the much less extensive and varied transit network in Louisville relative to those in Boston and Chicago. Despite the variables that express performance characteristics of the different modes, the variety and availability of modes alternative to automobiles is not captured, especially where observations are zoned aggregates. Thus, in the three cities being compared, a different spatial conformation in conjunction with a different availability of nonautomobile modes in each will result in different automobile driver and transit rider populations in each city. Since they will typically have different trade-off valuations, measured sensitivity to the various trip price components can be expected to differ. In effect, the closer and more plentiful are the substitutes to a given mode, the greater will be the price sensitivity. Talvitie's disaggregation of the transit mode into rail and bus reflects this.

Evidence from Disaggregate Demand Models

Disaggregate demand models have been developed to overcome many of the deficiencies of direct travel-demand models. They do this by basing their predictions of group behavior on the choice situation of the individual. The probability of making each kind of trip (defined in terms of mode, frequency, time of day, destination, and purpose) is formulated as a conditional logit model, where the equation is a ratio of the exponent of the household's utility level for the specific travel choice to that of the sum of the exponentials for all alternative travel decisions. The exponents in this expression are weighted linear combinations of the explanatory variables. The coefficients, which represent the independent effect of each variable on the utility level of household, are estimated by maximum likelihood. The explanatory variables fall into three classes: socioeconomic characteristics of the individual decision maker that might influence travel choice, comparative characteristics of alternative travel options, and comparative characteristics of the alternative destinations that relate to their relative attractiveness. They can be summarized crudely as household income levels, travel times and costs, and destination employment levels, respectively. [For a more detailed discussion of the disaggregate modeling technique, see the full report (6) or McFadden and Domencich (11).]

Disaggregate work-trip models restrict the range of travel choices to mode only, assuming that trip frequency, destination, and time of day remain unchanged for work commuters. To the extent that the journey to work is subject to time, frequency, or destination variation, this assumption imparts a downward bias to be the estimated responsiveness of work travel to pricing policy. Disaggregate shopping models allow for a fuller range of variation in travel behavior, although to varying degrees and employing different assumptions on the structure of travel decisions. Destination and frequency of shopping trips (as well as mode choice) are usually allowed to vary. In addition, one allows a very limited but interesting choice between peak and off-peak travel.

Two basic equation structures have been employed in the estimation of the shopping model. Charles River Associates estimates a set of up to four conditional probability equations: (a) modal split, (b) time of day, (c) destination, and (d) trip frequency (11). This assumed separability of shopping travel demand requires that mode choice be independent of the time of day, destination, and frequency of trip making and be a function only of the relative travel time and cost characteristics of the available modes. Also, it is assumed that destination, frequency, and time of day of travel are similarly conditionally layered in a sequence of separable decisions. Through the successive estimation of these four equations, the probability that an individual will make a shopping trip with a given frequency, to a certain destination, and via a particular mode is a function of the costs and travel time of all available modes, the relative attractiveness of alternative destinations, and a set of socioeconomic traits of the individual.

The assumption that shopping travel demand is separable facilitates estimation; however, it is dubious. Characteristics and availability of each mode are not independent of destination and time of day. So modal choice, for example, will be influenced by trip destination. Travelers are more likely to use the automobile for a circumferential suburban trip and transit for a radial central business district (CBD)-oriented trip. Moreover, even for shopping trips, transit is more likely to be chosen during peak hours of travel.

To avoid this weakness, Ben-Akiva and Adler estimated what they term a joint-probability shopping model. In this model, the probability that an individual will make a shopping trip with a certain frequency, by a given mode, and to a particular destination is estimated in a single equation, which includes all explanatory variables that influence one or more of the identifying aspects (e.g., mode choice, frequency) of the trip (12). Although it avoids separability, the combination of all variables into one equation introduces multicollinearity and simultaneous-equations bias.

Elasticity Results from Work Trip Models

Demand elasticity information provided by disaggregate models is obtained either through simulation (with real data) or by the substitution of hypothetical variable values, by using the probability equations estimated econometrically. Alternative values of policy variables are fed into the estimated equations. The price elasticity is then calculated by comparing the size of the resulting predicted change in travel behavior with the various initiating cost changes. It is considerably less straightforward to evaluate the reliability of the elasticity estimates so derived. However, the statistical reliability of the underlying disaggregate probability equations from which the simulations are generated can be assessed. Most of the studies report indices of statistical reliability that indicate an impressively good fit for whole equations and significance for many coefficients. The simulation estimations are, however, separated from these equations by an additional stochastic step that blurs the reliability of the final results.

Simulations of the Massachusetts Institute of Technology (MIT) - Cambridge Systematics disaggregate demand model of work trips in Washington, D.C., suggest that modal split sensitivity to trip costs is quite low (13). Even large increases in the money cost of automobile use are predicted to only slightly affect the overall distribution of work trips among modes in the metropolitan area. For example, a quadrupling of the price of gasoline would presumably reduce the drive-alone share by less than five percent. Similarly, large parking surcharges are predicted to have only a minor impact. Adding \$3/day to the cost of a downtown parking place would only diminish the drive-alone share of work trips by 6.5 percent. This extreme insensitivity to price suggests that price would not be an efficient lever to help relieve rush-hour congestion or reduce fuel use by diverting automobile drivers to transit or carpools. However, these results are highly suspect. They involve extrapolation far outside the range of observed data. Although there is no reason to believe that demand responsiveness remains constant at all price levels, the Cambridge Systematics policy simulations are necessarily based on this assumption.

In contrast to the low elasticity implicit in the Cambridge Systematics results, the McFadden model of travel demand in the Oakland-Berkeley area suggests considerable responsiveness of travel to price (14). It reveals interesting differences in elasticities with respect both to different components of real trip cost and to different components of alternative mode characteristics. Time and money cost sensitivity differ, as do different components within each of these broader categories. Cross-mode effects differ as well. Strongly confirmed is the Moses and Williamson survey finding that transit demand is more sensitive to the cost of automobile use than it is to its own price (9). For example, in the two-mode case, although the elasticity of bus trips with respect to bus fare is only -0.45, the cross-elasticity of bus trips with respect to automobile operating and parking costs is more than twice as great—nearly 1.0. It is greater, in fact, than either the cross-mode or the own-mode effect of travel time. This result is not quite as striking in the three-mode case; nevertheless, the bus trip cross-elasticity with respect to automobile costs is 0.81 (considerably greater than the own-elasticity with respect to bus fare of -0.58, which in itself is substantial) and the rapid transit trip cross-elasticity with respect to automobile costs is 0.82 (only slightly less than the own-elasticity of -0.86 with respect to rapid transit fare). This is an important result. It appears to contradict the Cambridge Systematics results by indicating the possibility of significant leverage for public policy use of the pricing instrument to change modal split. The suggested diversion to transit effected by the increased cost of work-trip automobile use might, for example, significantly alleviate rush-hour road congestion.

On closer examination, the discrepancy between the two studies is hardly surprising and is, in fact, quite illuminating. McFadden measures modal-split elasticity only in corridors that have extensive transit service—that is, where transit is a feasible alternative to the use of the automobile for the work trip. In contrast, the Cambridge Systematics model estimates modal shifts for the metropolitan area as a whole. Thus, its sample includes travelers who reside in areas where residents could not possibly opt for transit over the automobile in the short run, regardless of the magnitude of the economic incentives. Since, in fact, this infeasibility of the transit choice applies for most suburban portions of American urban areas, whatever modal shift occurs in those corridors where transit is available is greatly diluted by the inevitable lack of responsiveness in the suburbs, where transit is not an available alternative. Although Washington, D.C., traffic corridors well served by transit might display the same degree of work-trip sensitivity as that estimated by the McFadden study, this sensitivity would not be evident in the Cambridge Systematics-type of aggregate, metropolitan-wide results. Thus, a seemingly low aggregate responsiveness may conceal important variations in responsiveness among the hundreds of individual traffic corridors that comprise the urban transportation network.

This does not deny the usefulness of the Cambridge Systematics results, however. For energy conservation questions, for example, the Cambridge Systematics results are clearly more appropriate, since the amount of gasoline saved is more a function of total automobile use than of distribution of automobile traffic among specific urban corridors. But, for urban transportation problems that arise from the externalities of automobile use, the spatial and temporal composition of automobile use determine the social costs associated with any level of automobile driving. The severity of congestion, in particular, is almost entirely dependent on the density of traffic, which varies greatly from one part of the metropolitan area to another.

Significantly, transit is most frequently available in those corridors that have the highest traffic density—precisely those routes where modal shift away from the private automobile would be most beneficial in terms of mitigating congestion and other externalities. Consequently, the McFadden results suggest that properly targeted pricing of automobile use would have maximum impact where it is most needed. That the impact of automobile pricing policies would be minimal in low-density suburbs is of little importance with regard to the potential effectiveness of public policy use of pricing to mitigate the social costs of the automobile. The problems associated with automobile use are least serious in these areas.

Price elasticity estimates can also be inferred from the Charles River Associates work-trip model originally estimated with Pittsburgh data and subsequently recalibrated with Los Angeles data to simulate the effects of various pricing policies in promoting energy conservation, pollution abatement, and road congestion reduction (15). Results indicate an elasticity of automobile work trips with respect to automobile line-haul costs of -0.27 and of automobile vehicle kilometers traveled with respect to line-haul costs of -0.38, which suggests only slightly less responsiveness to price than does the McFadden study.

Elasticity Results from Shopping Models

The shopping-trip model in the McFadden and Domencich Charles River Associates studies is especially rich. Unlike the work trip, where frequency, destination, and time of day are assumed to be constant, most dimensions of shopping trips are allowed to vary and are explained in a set of equations that represent a recursive, sequential decision structure. Two sets of shopping demand-responsiveness results have been assembled by Charles River Associates. The first stems from a four-equation conditional probability model, which uses Pittsburgh household survey data in conjunction with calculated trip time and cost values relevant to surveyed households. In this model, automobile line-haul cost has an elasticity of -0.17 with respect to modal choice and destination in automobile trips. Trip frequency, time of day of travel, and trip destination are significantly more responsive to shifts in relative trip times and money costs than is modal choice. Sample calculations for a hypothetical individual suggest, for example, that trip time of day is approximately twice as responsive as modal choice to variations in transportation cost and performance measures.

Even the elasticities of trip frequency and destination significantly exceed the modal-choice responsiveness when account is taken of the greater number of alternative destinations than modes. Because the problem of road congestion is basically one of timing, location, and frequency, these results indicate that a selective pricing of urban transport facilities targeted to influence the timing and destination of trips would be more successful in reducing congestion than would a pricing policy aimed at inducing a shift away from the use of the private automobile to other modes. The absolute size of each of the former adjustments is not large, but together they represent a substantial impact.

Of further relevance to our study is the finding that, although average modal-choice elasticity for work trips vis-à-vis cost is low, it is increased by the introduction of additional, closely substitutable modes. Thus, a modal-shift-oriented price policy for work trips, combined with a pricing policy aimed at influencing the frequency, destination, and time of day as well as mode of shopping trips, and the introduction of substitutive modes, might be a quite effective policy package to counter congestion and pollution.

The second set of Charles River Associates results is derived from a simulation of the Pittsburgh-type disaggregate shopping model by using Los Angeles data. The elasticity of vehicle kilometers of travel with respect to automobile line-haul costs is estimated to be -0.17 with mode and destination variable and -0.34 with mode, frequency, and destination variable.

EVIDENCE OF TRANSIT FARE ELASTICITY

The degree of sensitivity of transit use to fares is important to transit operators, who must set fare and service levels that will minimize the operating deficit but maintain ridership. An elastic transit demand implies revenue loss by operators if they raise fares but revenue gains if they lower them. Conversely, the more inelastic is transit demand, the more revenues can be increased by fare hikes.

Although the assembled estimates of fare elasticity range widely (from -0.09 to -1.8), most of the estimates fall substantially short of 1.0. The travel demand models suggest an average elasticity of around -0.4 or -0.5, although there is a great deal of variation [see

Table 3. Transit fare elasticities derived from travel demand models.

Model	City	Transit Mode	Short-Term Elasticity	Comments
Warner	Chicago	Aggregate transit	-0.96	Applies only to those travelers not restricted to a single mode (i.e., noncaptive)
Lisco	Chicago	Aggregate transit	-0.4	Noncaptive travelers
Lave	Chicago	Bus	-0.7	
Charles River	Boston	Aggregate transit	-0.09	
McGillivray	San Francisco	Aggregate transit	-0.11	All trip types
			-0.19	All trip types (noncaptive)
			-0.87	Work trips (noncaptive)
Talvitie	Chicago	Aggregate transit	-0.38	Noncaptive
		Bus	-0.51	Noncaptive
		Rail	-1.8	Noncaptive
McFadden	Berkeley-Oakland	Bus	-0.45	In absence of BART
		Bus	-0.58	In conjunction with BART
		Subway	-0.86	BART
Fulkerson	Louisville	Bus	-0.4	
Regional Plan Association	New York	Bus	-0.31	Time series analysis from 1950 to 1974
		Subway	-0.16	

Table 3 (7, 8, 10, 14, 16-20)]. Of note is the deviance of the Charles River Associates estimate of -0.09, which is exceeded significantly not only by the results of most of the other models listed in the table but also by most arc elasticity estimates that have been calculated. These latter, with a mean value of about -0.4 (but ranging widely), can at least be said not to contradict the formal model estimates (6).

Of course, the reliability of both types of estimates is limited to those conditions under which they were estimated. Extrapolation to other cities, other transportation environments, and other price-service levels is dangerous. This is especially so in the case of the arc elasticities, which do not control for nonfare variables. Moreover, since the models are calibrated on the specific transportation situations of particular cities, the estimated demand parameters are similarly constrained by this structural dependence on the specific data base.

The generally inelastic demand for transit use ensures that fare decreases will almost certainly increase the operating deficits of transit authorities in the short run and that fare increases will reduce deficits. A number of considerations may, however, dictate against the upward adjustment of fares simply to reduce deficits. Long-term transit fare elasticities, which have not been estimated, definitely exceed corresponding short-term elasticities because a wider range of transportation choices and residential as well as workplace locations become available over time. Indeed, long-term elasticities may actually exceed unity and thus make fare increases ultimately perverse. Moreover, the economies of scale in transit operation complicate the adverse effects of price increases. Not only does ridership decrease (albeit at a lesser percentage than that of the price increase), but operating costs per passenger rise and service levels usually fall (especially frequency and geographic coverage). This stimulates the need for additional fare increases and subsequent patronage declines. Finally, to whatever extent external social benefits accrue to the provision of transit service, these may justify the public subsidization of transit so that these services can be provided at below cost.

EVIDENCE OF AUTOMOBILE TRAVEL ELASTICITY

Traffic Response to Tolls

Especially in densely developed northeastern American cities, bridges and tunnels are important access links to the congested CBDs of metropolitan areas. Moreover, limited-access expressways are also significant feeder routes to the downtown areas of almost all metropolitan areas in the United States. In contrast to the great difficulty of imposing ideal marginal congestion cost charges on users of most city streets, the limited points of access to expressways, bridges, and tunnels improve the feasibility of appropriate congestion pricing of these arterial routes, which so exacerbate the congestion problems of central urban districts.

Although evidence on the effectiveness of expressway pricing is not available (most urban freeways are free to users), the limited experience analyzed for bridge and tunnel toll surcharges indicates an average arc elasticity of automobile traffic with respect to tolls of -0.13 for New York City and -0.17 elsewhere. Little can be inferred from these data, as no effort is made to control for other variables that might have altered traffic levels. An interesting result, however, is that bridges and tunnels that have a number of substitute routes exhibit the most elastic demand response, as economic theory would predict (21).

Traffic Response to Parking Charges

In lieu of ideal marginal social cost pricing of urban roads (technically difficult and politically infeasible), a less controversial and more easily implementable approximation to congestion pricing is the coordinated use of parking taxes, meters, and supply restrictions to parking availability, which is crucially complementary to automobile use. Parking charges may be especially effective in reducing traffic levels in congested downtown areas, for the number of parking spaces tends to be scarce relative to demand at peak hours and is sensitive to public regulation.

Hard evidence on the indirect effect of parking policies on traffic levels, as reviewed by Kulash, is virtually nonexistent (22). Even studies of the direct, and less policy-significant, effect of such charges on parking demand are seriously flawed. Most conclude that parking demand is inelastic with respect to own-price, averaging around -0.3 or -0.4. However, the extent of the direct influence on parking demand per se is of less importance than the indirect effect of these fees as a component of trip costs on overall automobile travel demand. The evidence on this issue of primary importance is of a hypothetical nature. Not only is it based on the dubious reliability of survey data, but it involves radical extrapolations to price levels many times higher than those that have actually been obtained in any American metropolitan area. It does suggest, however, the need for pricing policies to be comprehensive and coordinated in order to have any impact on the level and pattern of traffic and thus, indirectly, on the levels of road congestion and automobile pollution.

CONCLUSIONS

Some of the elasticity estimates we have examined have been interpreted by observers as testifying that price is an ineffective tool for helping to cope with urban transportation problems. We do not share this interpretation. It reflects an incorrect view of the role that pricing policy should perform. The objective of public pricing policies for urban transport is not simply to reduce aggregate automobile use. Specific program goals may be formulated that require such decreases, and pricing may be considered a tool for helping to bring them about. The natural role for public pricing is to force users of any scarce resource to confront the full social costs and benefits to the community of their use of that resource, so that the resource will be allocated most beneficially with respect to the overall amount and distribution of use.

Urban travel, especially via private automobile, involves several kinds of indirect effects on users and nonusers, which do not now enter into private travel decisions. The distinctive goal of pricing is to incorporate these externalities (e.g., congestion, air pollution, and noise) into the decentralized urban travel decisions of the population. Elasticity values indicate how much alteration in urban travel behavior would occur under this form of price policy. They suggest how serious a distortion of resource use occurs because of externalities, although they shed no light on the magnitude of these social costs. In this sense, the smaller the elasticity numbers, the less serious is the resource distortion that pricing policy is called on to rectify. But this is not the same as the usefulness of pricing policy.

The magnitude of the transport externality problem varies crucially over times of day and at different locations within the metropolitan area. Therefore, improved public pricing policy should impose significantly different charges, depending on time, location, and the social as well as natural environmental context. Studies that differentiate the demand response to pricing by changes in trip frequency, route, and time of day as well as modal choice tend to find much greater responsiveness to price than do those studies that assume travel demand to be homogeneous. The particularly keen leverage of pricing in altering the composition of automobile use makes pricing an especially attractive policy instrument. The social costs of the automobile involve less the aggregate level of use than the nature and composition of this use. Thus, improved pricing might, for example, decrease urban congestion problems significantly without reducing the total number of kilometers driven in the metropolitan area.

Perfect pricing of automobile use (sensitive to every variation in social cost) is impossible; however, some differentiation by type of automobile use can probably be achieved through the appropriate orchestration of the various pricing instruments that are available. For example, parking charges could be made to vary with location, gasoline taxes could be increased, and tolls could be imposed at entrances to key congested links. A suitable degree of differentiation among automobile uses should result in the imposition of charges on some uses but not on others. Therefore, the specific automobile uses for which a surcharge is levied will inevitably have a set of closer substitutes than would aggregate automobile use. As confirmed by the disaggregate studies, the demand responsiveness to price will be greater for these specific uses than aggregate use elasticities indicate.

Almost all of the estimates of elasticity are dependent, at least to some extent, on the specific city from which the data are assembled. Urban transportation demand depends importantly on city-specific characteristics (such as density, area, urban land-use patterns) and the extensiveness and quality of the existing road system and transit network. Elasticity estimates derived from demand studies that have not been sensitive to these interurban differences cannot be reliably extrapolated to predict the travel response of pricing policies in cities other than the one on which they were calibrated. An advantage of the disaggregate modeling approach is that this limitation has been mitigated, at least to some degree.

There are other aspects of the extrapolation problem, however. Elasticity estimates of even the best demand studies are less reliable outside the observed range of cost and performance characteristics of the transportation system. Forecasts of the policy impact of transport prices that are either substantially higher or more differentiated than past or present levels on the basis of historical elasticities are of dubious validity. Not only may the magnitude of the effect change, but the nature of the impact may also change. For example, as travel prices increase to levels previously unobserved, responsiveness may increase at some nonlinear rate, or travel choices might shift radically on reaching some crucial price threshold.

Furthermore, it is important to note the short-term nature of the urban travel-demand studies. Although the effect of automobile ownership levels on travel behavior has been estimated to be considerable, in no model is the indirect effect of price on travel behavior through changes in automobile ownership taken into account. Yet the Chase Econometrics and RAND models estimate equations of automobile demand that suggest an extremely strong impact of gasoline price both on the level of automobile sales and on the composition of the automobile stock by size, class, and fuel-efficiency characteristics. Because urban travel-demand models have not incorporated this indirect response to pricing, the calculated elasticities almost certainly underestimate the longer-range impact of pricing.

A potentially even more important, indirect effect of price on travel demand is that of a changing structure of transportation prices on the pattern of land use in metropolitan areas. This avenue of change has been given almost no serious attention in empirical demand studies. The ramifications of the land-use impact are extremely varied and complex. The difficulties of modeling the effect are, therefore, considerable and empirical estimation has so far proved infeasible. Nevertheless, the exclusion of this indirect long-term locational effect of price on travel demand certainly understates the total responsiveness induced by price changes. It would seem that an alternative evolution or redirection of urban development patterns would represent the most fundamental solution to many urban transportation problems.

In sum, even on the basis of the underestimates of elasticity values that have been calculated, the potential of corrective pricing policies to aid in the solution of urban transportation problems is significant. We would argue, however, that this potential is, in fact, considerably greater than can be inferred from the existing empirical work.

ACKNOWLEDGMENT

This paper represents an abridgement of a report prepared for the U.S. Department of Transportation. The background report contains detailed analysis that was necessarily omitted in this shortened version.

REFERENCES

1. A Study of the Quarterly Demand for Gasoline and Impacts of Alternative Taxes. Data Resources, Inc., Lexington, MA, Dec. 1973.
2. R. G. McGillivray. Gasoline Use by Automobiles. The Urban Institute, Washington, DC, Aug. 1974.
3. B. Burright and J. Enns. Econometric Models of the Demand for Motor Fuel. Rand Corporation, Santa Monica, CA, 1975.
4. The Effect of Automotive Fuel Conservation Measures on Automotive Air Pollution. Charles River Associates, Cambridge, MA, 1975.
5. The Effect of Tax and Regulatory Alternatives on Car Sales and Gasoline Consumption. Chase Econometrics, New York, 1974.
6. J. R. Pucher and J. Rothenberg. Pricing in Urban Transportation: A Survey of Empirical Evidence on the Elasticity of Travel Demand. Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, 1976.
7. T. Domencich, G. Kraft, and J. P. Valette. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. HRB, Highway Research Record 238, 1968, pp. 64-78.
8. A. Talvitie. A Direct Demand Model for Downtown Worktrips. Transportation, Vol. 2, No. 2, July 1973, pp. 121-152.
9. L. Moses and H. Williamson. Value of Time, Choice of Mode, and the Subsidy Issue in Urban Transportation. Journal of Political Economy, Vol. 71, June 1963, pp. 247-264.
10. H. Fulkerson. A Direct Demand Model of Work Trips in Louisville, Kentucky. Transportation Systems Division, Massachusetts Institute of Technology, Cambridge, 1975.
11. T. A. Domencich and D. McFadden. Urban Travel Demand: A Behavioral Analysis. American Elsevier-North Holland, New York, 1975.
12. M. E. Ben-Akiva and T. Adler. A Joint Frequency, Destination, and Modal Choice Model for Shopping Trips. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, 1975.
13. T. Atherton, J. H. Suhrbier, and W. A. Jessiman; Cambridge Systematics. The Use of Disaggregate Travel Demand Models to Analyze Carpooling Policy Incentives. Federal Energy Administration, Draft Rept., Oct. 1975.
14. D. McFadden. The Measurement of Urban Travel Demand. Journal of Public Economics, Vol. 3, 1974, pp. 303-328.
15. Economic Analysis of Policies for Controlling Automotive Air Pollution in the Los Angeles Region. Charles River Associates, Cambridge, MA, 1975.
16. S. L. Warner. Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice. Northwestern Univ. Press, Chicago, 1962.
17. T. E. Lisco. The Value of Commuters' Travel Time. Univ. of Chicago, Ph.D. thesis, 1967.
18. C. A. Lave. Modal Choice in Urban Transportation. Stanford Univ., Stanford, CA, Ph.D. thesis, 1968.
19. R. G. McGillivray. Binary Choice of Transport Mode in the San Francisco Bay Area. Univ. of California, Ph.D. thesis, 1969.
20. Regional Plan Association. Financing Public Transportation. Regional Plan News, March 1976.
21. R. Kirby. Road User Charges: Some Practical Considerations. The Urban Institute, Washington, DC, Paper 1212-5, 1974.
22. D. Kulash. Parking Taxes for Congestion Relief: A Survey of Related Experience. The Urban Institute, Washington, DC, Paper 1212-1, 1974.

Publication of this paper sponsored by Task Force on Economic Analysis.

Abridgment

Gasoline Rationing Based on Licensed Drivers or Vehicles: Potential for Coupon Sales Between Income Groups in Michigan

Martin E. Lee, Research and Evaluation Division, Michigan Department of State, Lansing

In a proposed standby gasoline rationing plan released for public comment in June 1978, the U.S. Department of Energy (DOE) proposed that the unit of allocation for gasoline be registered vehicles rather than licensed drivers. It was asserted that this would make rationing quicker to implement and be a more realistic response to existing use than driver-based allocation (1). The plan also emphasized the value of a "white market" for the unrestricted exchange of rationing rights at uncontrolled prices. The vehicle-based allocation and white-market proposals raise controversial questions about the impact of rationing on different income groups. These questions can only be addressed by using detailed data on the trip-making characteristics of drivers and vehicles. This paper summarizes analyses of a micro-data base on vehicle ownership and use to compare the impact of rationing on Michigan drivers grouped by location, income, and vehicle size.

ORIGIN OF THE DATA

The data base, known as the Michigan driving experience survey (MDES), was created to investigate public policy issues in driver and vehicle licensing, traffic safety, driver behavior, and energy conservation. It is based on 7581 interviews of applicants for renewal of driver's licenses, which were conducted throughout the state during 1976. It used a controlled selection procedure to randomly select 30 sites and, because of the paucity of rural trip-making data, rural areas were deliberately oversampled. Within the sites, a random number system was used to select seven or eight interviewees per office per week. Overall, this provided an excellent random sample of the Michigan driver population and a very high response rate. Note, however, that drivers under the age of 19 are not included because they are not old enough to have reached first renewal.

The interviews were conducted by the managers of the local license offices. The emphasis of the survey was on the careful reconstruction of a recent trip day (usually the previous day) and on the complete set of vehicles to which the respondent had access, together with basic socioeconomic information on the respondent and his or her household.

THE STANDBY RATIONING PROPOSAL

The 1978 standby rationing plan provided for a fixed allocation of gasoline to all registered vehicles. All privately owned vehicles under 4535 kg (10 000 lb) would receive an allocation based on the national average consumption for an automobile [estimated to be 2830 L/year (748 gal/year) (1)], less a percentage necessary to respond to the predicted shortfall in supply. The allocations would be made for periods of about 90 days. Ration rights would be distributed directly to the public, with a small amount of additional rights (perhaps 3 percent of the total) distributed through state government agencies to provide relief for hardship cases. Ration rights could be traded legally at uncontrolled prices, and much is claimed for the value of this white market in the redistribution of income to offset general inflationary effects and benefit poorer households.

Methods

For our purposes, analyses are confined to the 98.7 percent of the respondents who stated that the vehicle they drive most often is a private automobile, van, pickup, or utility.

In order to compare driver- and vehicle-based allocation methods, the detailed respondent trip-making information available in MDES was analyzed together with a surrogate for per-vehicle trip-making, obtained as follows:

$$TV = (TP \times DH/VH) \quad (1)$$

where

DH = number of drivers in household,
VH = number of vehicles in household,
TP = respondent trip making (km, min), and
TV = vehicle trip making (km, min).

Analyses of trip making were performed by using three subgroupings of the respondents:

1. Income group—self-reported household income;
2. Vehicle size—size of vehicle most often driven, classified from make and model; and
3. Location of residence—five strata of counties ranked by population density.

In the analysis of costs, white-market cash exchanges are, of course, subject to widely differing assumptions. However, because the value of the MDES data is in providing trip-making data, some simplistic scenarios are postulated to examine the distributional effects of a hypothetical 25 percent shortfall in gasoline supply. Costs are estimated here on the artificial basis that, if all drivers reduced their travel by the same percentage as the shortfall (25 percent in this instance), a fixed allocation of gasoline based on a similar reduction in supply would be oversufficient for some and insufficient for others. The average cost in dollars per month is calculated by the formula

$$\text{Dollars per month} = [(KMD \times 0.75) - KMR] \times (365/12) \times [WM/(KM/L)] \quad (2)$$

where

KMD = kilometers per day (per driver or vehicle),
KM/L = kilometers per liter of vehicle used,
KMR = kilometers allowed by ration, and
WM = white market cost per liter.

For analyses by vehicle size, KMR is adjusted to the average fuel economy of the vehicle class analyzed, and KM/L is set to that figure. The fuel economy constants are mostly based on U.S. Department of Transportation (DOT) standards for 1972 vehicles (2), and range from 9.35 km/L (22 miles/gal) for subcompacts to 4.68 km/L (11 miles/gal) for vans and pickups. (The median year of vehicles in the survey is 1972.)

The formula used for KMR is

$$KMR = [\overline{KMD} \times 0.75 \times (KM/L)]/(\overline{KM/L}) \quad (3)$$

where

$\overline{KMD}$ = grand sample mean of kilometers per day (per driver or vehicle),
$\underline{KM/L}$ = kilometers per liter of vehicle used, and
$\overline{KM/L}$ = average kilometers per liter.

For analyses in which vehicle size is not differentiated, the KM/L constant used is the same as that quoted in the DOE plan, namely 5.74 km/L (13.5 miles/gal).

Results

Two fundamental findings of the MDES data are the similarity between income groups in the number of vehicles per driver and the major increase in daily kilometers driven with increasing household income (Table 1). Therefore, the potential for a white market to operate between high- and low-income groups is considerable; it is slightly greater for a per-driver than for a per-vehicle allocation basis. Other analyses showed that about 54 percent of drivers in the top two income groups would have enough gasoline for all of the driving they now do if a 25 percent shortfall occurred, compared to about 78 percent of drivers in the bottom two income groups.

There are also significant differences in the amount of daily travel as a function of the vehicle size most often used. The table below shows that use of smaller automobiles is associated with increased driving, and other MDES analyses have revealed that this effect generally holds true, regardless of the age of the driver. The higher averages for those driving vans and pickups reflect some degree of rural bias in the location of these vehicles, with associated longer trip lengths. The full-sized vehicle class includes luxury automobiles; full-sized vehicles alone have lower average travel (1 km = 0.62 mile).

Vehicle	Average Daily Kilometers per Driver	Number of Respondents Using Size
Automobile		
Subcompact	45.1	644
Compact	43.9	1042
Intermediate	41.6	1327
Full-sized	41.2	2700
Van, recreational vehicle, and pickup	55.2	892
Motorcycle	39.1	16
Truck and bus	182.8	77
Total		6698

The effect of vehicle size on rationing is also influenced by the distribution of vehicle classes within each income group. In general, higher-income groups opt for more large vehicles than do lower-income groups; vans and pickups are a middle-income phenomenon. The popularity of the smallest automobiles, once a higher-income speciality item, is now growing in the lower-income groups.

The cost analyses were performed by using mean daily travel for the various population subgroups. Summaries of the costs to each income group of a 25 percent shortfall and postulation of a 25 percent reduction in travel by all are shown in Figures 1 and 2. Figure 1 examines differences by geographical location; Figure 2 gives results by the vehicle size most often used. Both figures compare the costs on a per-driver basis (upper graph) and a per-household-vehicle basis (lower graph). The hypothetical average coupon price of $0.24 [predicted by the DOT plan (3) for a 20 percent shortfall] is used as the basis for the tentative costs shown. The results for Figure 1 assume that all household vehicles have similar fuel economy.

Figure 1. Distribution of potential ration coupon exchange costs by income group and residence location for a 25 percent shortfall (truck, bus, and motorcycle users excluded).

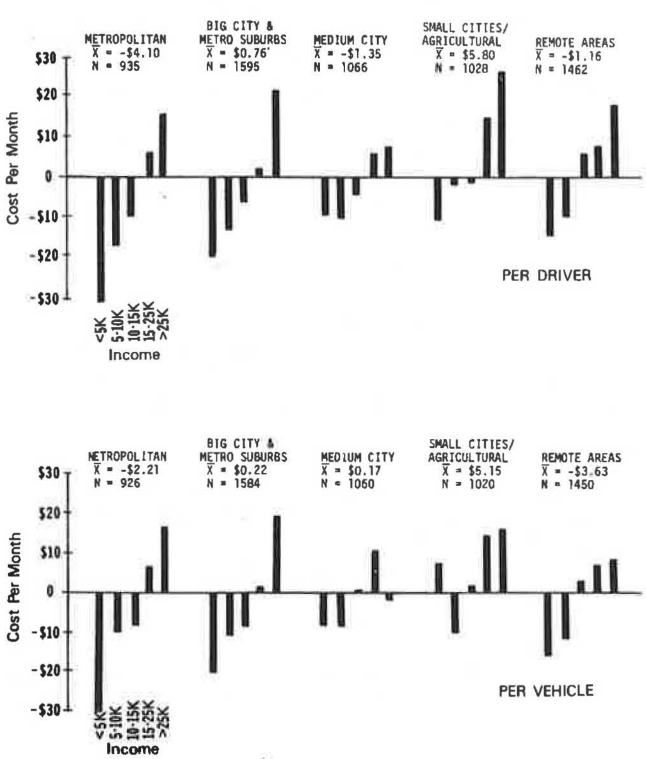


Table 1. Driver-vehicle ratio and mean daily travel by income group.

Household Income ($000s)	Ratio in Household (Drivers:Vehicles)	Mean Daily Travel: Kilometers per Driver	Mean Daily Travel: Kilometers per Vehicle in Household
Under 5	1.07:1	24.8	27.8
5-10	1.01:1	31.7	32.5
10-15	0.99:1	39.4	41.2
15-25	0.94:1	50.5	50.7
Over 25	0.95:1	63.1	58.6

Note: 1 km = 0.62 mile; respondents who normally drive trucks, buses, or motorcycles are excluded (N = 6605).

Figure 2. Distribution of potential ration coupon exchange costs by income group and vehicle size most used for a 25 percent shortfall (truck, bus, and motorcycle users excluded).

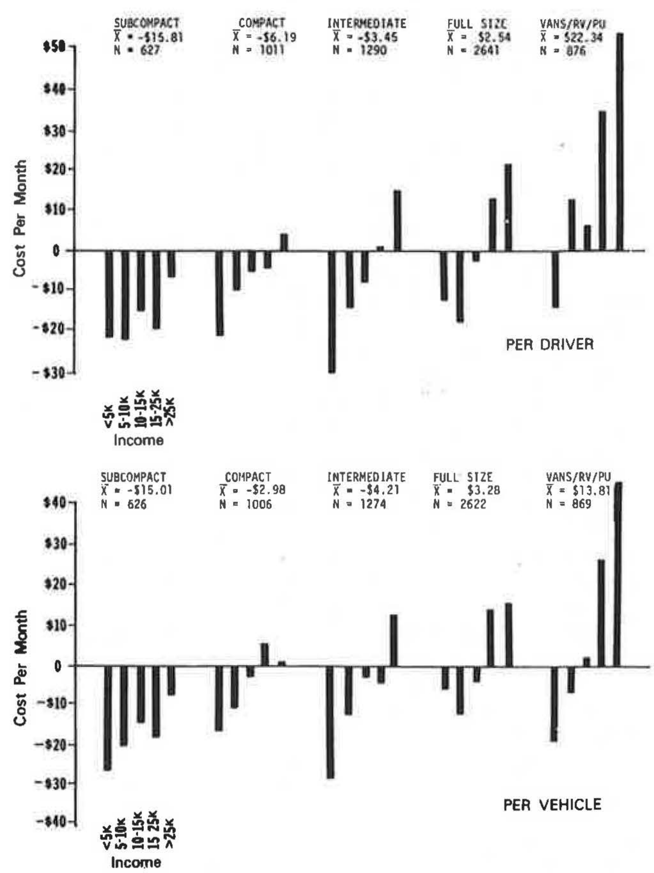
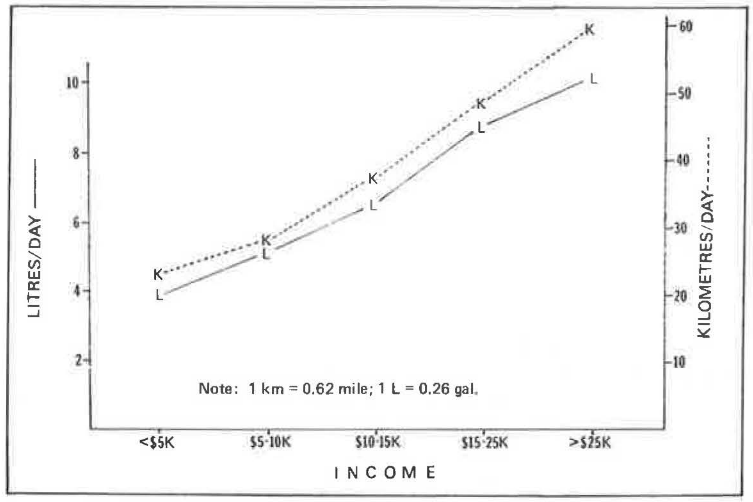


Figure 3. Average gasoline consumption and kilometers driven per day by household income of respondent.



Of major importance are the relative effects of this scenario on different income groups. It is clear from this perspective that the per-vehicle and per-driver schemes are similar in impact. Although this might be expected, given the almost one-to-one ratio of drivers to vehicles in Michigan regardless of income, it is not necessarily true that the amount of driving done is independent of varying driver-per-vehicle configurations found in households.

The differences between geographical locations are of particular interest in that the most remote locations do not seem to carry the penalty of increased driving, which is conventionally assumed. One possible explanation is the concentration of retirees in the more remote parts of Michigan. By far the heaviest average use of gasoline is in the agricultural centers. Metropolitan drivers (Detroit area) would have the highest income potential on a per-person basis and the second highest on a per-vehicle basis. This provides an estimate of the penalty associated with the lower levels of vehicle ownership in metropolitan areas under a per-vehicle rationing plan. However, this study does not reveal anything about those people in central cities and elsewhere who would not appear in the system at all because they have no access to private automobiles.

The data on vehicle size in Figure 2 reveal that those who have the smallest automobiles could be in a positive cash-flow situation in this scenario, regardless of income group. The three lowest-income groups could sell some ration rights; however, those who use larger vehicles do not appear to reduce their driving enough to compensate for the poorer fuel economy. The higher average travel of those who drive vans and pickups shows up clearly, and those in the higher-income groups could spend over $600/year more to maintain 75 percent of their previous driving activity.

A more accurate calculation of gasoline consumption is supported in MDES by the data on vehicles actually driven during the trip days. Figure 3 shows consumption by income group based on the average kilometer-per-liter estimates for the vehicles actually driven. For comparison, the average kilometers driven within each income group is plotted against the right-hand scale. A comparison of the shape of the curves suggests that there is a slight trend for the higher travel of the $25 000+ group to be associated with more fuel-efficient vehicles. Considerably more adaptation to fuel-efficient vehicles by high-kilometer drivers than these analyses reveal will be necessary if this method of gasoline rationing is to encourage conservation. It should be noted that average fuel economy has improved since 1972, the year for which estimates were used and the median year of vehicles in this sample.

CONCLUSION

This is a manipulation of personal travel data to examine who might be able to benefit from a white market in ration rights. It poses the highly improbable, zero-sum, scenario that, under rationing, all drivers would reduce their travel by the same proportion. Taken at face value, it does appear that wealthier vehicle users would be likely to buy available ration rights from the spare capacity of lower-income groups. Before assuming that the operation of such a market contributes to the general welfare, it should be asked whether it is right in a shortage situation to assent to a system that reinforces existing demand patterns. Wealthier people would, in all probability, pay the white-market price; and, at an extra $0.24/L ($0.90/gal), the inhibition of their driving would probably be minimal, given the inelasticity of demand for gasoline. A shortfall situation would seem to be an opportunity to reward conservation more specifically than through a white market. The travel needs of lower-income groups should be examined in more detail to establish the price of inhibited travel in terms of quality of life, not just in terms of (uncertain) cash flow or procrustean ideas of existing nonessential travel demand.

ACKNOWLEDGMENT

This study is based on data that were sponsored, in part, by the National Highway Traffic Safety Administration, U.S. Department of Transportation, through the Michigan Office of Highway Safety Planning.

REFERENCES

1. Proposed Standby Gasoline Rationing Plan. U.S. Department of Energy; Federal Register, Vol. 43, No. 125, June 28, 1978.
2. G. Maring. Highway Travel Forecasts. Office of Highway Planning, Federal Highway Administration, Nov. 1974. NTIS: PB 239 169/6SL.
3. Proposed Standby Gasoline Rationing Plan: Economic and Regulatory Analysis Draft. Office of Regulations and Emergency Planning, U.S. Department of Energy, DOE/ERA 0009, June 1978.

Publication of this paper sponsored by Task Force on Economic Analysis.

Framework for Involving Local Citizens in a Small Urban Area's Transportation Planning Process

John Lichtenheld, Straam Engineers, Inc., Portland, Oregon
James Etmanczyk, Wisconsin Department of Transportation, Madison

This paper presents a framework developed for involving citizens in transportation planning efforts in a small area. The paper describes the steps involved in the citizen participation process, provides a critique of the process, and outlines recommendations based on the results of the process.

The importance of involving the users of a system in planning the design of that system is well documented (1-4). Although effective citizen involvement does not always guarantee public acceptance of a project or plan, the absence of any citizen involvement in the planning process is an almost sure guarantee that a plan or project will meet strong citizen opposition and may, as a result, be difficult to implement.

The basic problem for most planning agencies today is how to effectively involve citizens in the planning process. There is no universally accepted method of citizen participation but rather a variety of individually tailored formulas, ranging from participatory tokenism to computerized sophistication (5, 6). Sophisticated citizen participation techniques, such as the Delphi method and interactive graphics, are a problem for small cities because they lack the necessary staff to implement these more costly and time-consuming techniques.

PURPOSE

This paper outlines a framework for incorporating citizen participation in the planning process based on one small community's trial-and-error attempts. This framework could have widespread applicability in planning improvements to any aspects of a community's environment where manpower and financial resources are limited. If public involvement is to work, it must allow a two-way communication between the planner and the public. It must serve to

1. Educate the public as to the complexity of the planning process and make the citizen familiar with existing conditions and future projections. If nothing else, the citizen participation effort should make the citizen aware that a coherent process is involved in planning future improvements and that decisions are not completely ad hoc.
2. Act as a sounding board for new ideas and alert the planner to possible problems at a stage where plans can be modified.
3. Allow free flow of information and comment between the planner and citizen by providing a forum for discussion; this would allow the presentation of new ideas from the citizens, which could be further evaluated by the planner.

BACKGROUND

The state line area transportation study (SLATS) was initiated in the fall of 1974. The purpose of this study was to develop a long-range transportation plan. The study area is located in south-central Wisconsin on the state line between Wisconsin and Illinois. The principal cities involved in the study are Beloit, Wisconsin, and South Beloit, Illinois. The combined population of the area is 61 400 with an employment of 22 300; the land area is 368 km^2 (142 $miles^2$).

SLATS was originally set up under two committees, a policy committee and a technical committee. The policy committee is composed of representative public officials from the county, city, and towns, as well as officials of state and federal transportation and planning organizations. The policy committee reviews and approves all recommendations made by the technical committee. The technical committee is composed of technical officials who are concerned with the area's transportation system. The technical committee is responsible for development and analysis of improvements and makes recommendations to the policy committee.

When SLATS was initiated, all citizen participation techniques applicable to a small urban area were reviewed (7). The most important criterion in this review was to define a process that would allow the study to remain in contact with the average citizen at all times. As a result, during the preliminary stages of the study a decision was made that, rather than establishing a committee composed of lay people, such as a citizens' advisory committee, a more effective means for involving local citizens in the study process would be a series of open-house and public information meetings. It was felt that a citizens' advisory committee soon loses its original intent because it changes a group of lay people into specialists, who then cease to represent the average citizen.

An initial policy on citizen participation adopted by the policy and technical committees led to the implementation of a program of public meetings, which was continuously modified and updated as the study progressed. The following schedule of open houses was established:

1. An initial open house to educate the public about SLATS and to display the preliminary improvement options,
2. A series of neighborhood open houses to address traffic problems and improvement options in particularly sensitive areas,
3. A second areawide open house for comments on the final group of four alternatives, and
4. A final open house or public hearing on the area's recommended transportation plan.

FORMAT

The open house or public information meeting concept used in the study grew out of a similar strategy used with some success by the Division of Highways, Wisconsin Department of Transportation. The public was invited to attend the open house at their own convenience during an established time frame that usually ran 6-7 h. No formal presentations were made during the meeting, and no official transcript was taken and published. Information on the study was displayed via graphics, photographs, handouts, and reports. Members of the technical committee were available to answer questions and discuss proposed improvement options. Feedback from citizens was documented through the use of questionnaires for incorporation into the further evaluation of the improvements.

INITIAL OPEN HOUSE

Prior to the first open house, citizen participation had been limited. Although all policy and technical committee meetings were open to the general public and publicized through the local media, only a few citizens attended the meetings. In addition, presentations were made at regularly scheduled public meetings of the city councils, plan commissions, and town boards. At these meetings each unit of government reviewed and approved the population and socioeconomic projections as well as the goals and objectives for the study area.

The purpose of the first open house was to inform the public of the area's transportation planning effort and to present a set of 24 possible improvements to the areas' street system over the next 24 years. Information was also provided on the projected growth patterns and future travel demand for the area.

A mailing list was compiled of public officials, civic groups, clubs, school boards, and news media. Individuals from each organization on the mailing list were sent an announcement of the open house two days prior to the event. A proposed mass mailing of announcements to a 5 percent sample of the population was considered but rejected because it was decided that more exposure could be obtained by contacting group leaders, who would in turn pass the information on to group members (1, pp. 71-89). An attempt was also made to get the local newspapers to run articles on the study during the week immediately prior to the open house. Unfortunately, the press did not give the event comprehensive coverage.

Procedure

The open house was held at a centrally located hotel. It was held on a weekday between the hours of 1:30 and 8:00 p.m. When individuals entered, they were asked to sign a guest book and were handed a questionnaire and a brief summary of the background and purpose of the study.

Graphics were displayed around the room so that individuals could progressively follow the major elements of the study on their own as they circled the room. The graphics depicted were in the following order:

1. The projected areas of major growth (i.e., population and dwelling units),
2. The number of trips in and out of the study area along the major arterials during the base year,
3. The projected internal trip demands for major areas [i.e., shopping center, industrial park, and central business district (CBD)],
4. The current traffic volumes on the street and highway network,
5. The future traffic volume projections on the street and highway network,
6. The network problem areas (as defined by current traffic accident data and future volume-to-capacity ratios),
7. A map of proposed improvement options,
8. Photographs of selected existing street sections that showed how far a widened or newly constructed facility would extend both in terms of its pavement and right-of-way width, and
9. A map of the proposed bus route system.

On the large map of all the improvement options, each improvement was numbered to correspond with a two- or three-page handout that described the improvement in greater detail. These handouts were placed nearby so that an individual interested in more details on a particular improvement could easily find that information. The handout described in detail each improvement's (a) location; (b) design; (c) volume of future traffic projected to be using the facility; (d) environmental, social, and economic impact; and (e) capital cost, including construction and right-of-way acquisition. Staff members were available to answer questions. In some cases small groups of people were taken through the presentation together.

An individual was expected to fill out a questionnaire after he or she had an opportunity to absorb the information and view the options. The questionnaire was modeled after a similar instrument used as part of the Dane County Area Transportation Study (8). It attempted to address people's preferences for future transportation policy direction as well as their comments on specific improvement options that would satisfy the area's future transportation needs.

Results

Approximately 75 people attended the open house, a somewhat lower turnout than anticipated. Possibly the most important result of the open house was that two alternate improvements were suggested by citizens in attendance. These proposals were incorporated into the study for further analysis. Some of the other more notable results follow.

Most of the attendance was during the first few hours. Attendance tapered off after 4:00 p.m. and then picked up somewhat after 7:00 p.m. The majority of those in attendance were from an area where an improvement was being considered.

Only 20 questionnaires were returned—a disappointing 27 percent rate of return. This low response could possibly be attributed to the length and detail of the questionnaire. Of those returned, very few were filled out completely (particularly those questions that were more policy oriented at the beginning of the instrument). Another possible reason for the low rate of response was our encouragement for people to complete the questionnaire at home. Once away from the open house, people's interest waned. A number of people took more than one questionnaire home for friends and family, which made it difficult to determine whether the survey results were representative of those in attendance, much less representative of the population as a whole.

A large amount of information was presented—too much for one person to absorb in a short period of time. The majority of people were most interested in the map display that showed all the improvement options being considered. The photographs of selected streets were an effective means of illustrating the potential extent of proposed improvements. The photographs were used in place of the usual cross-section drawings.

SUBAREA OPEN HOUSES

Sections of the study area were sensitive or controversial prior to the inception of the study or else became so when alternative solutions to the traffic problems were developed and presented. These sections were identified and designated as special subareas of the study. Special emphasis was placed on the development of alternative options for improving traffic flow in these areas and involving local citizens in the development of these improvement options.

Two special subareas were identified, one residential and one commercial. One subarea was an older, established residential neighborhood in which widening had been proposed for two of the major streets. The other subarea involved streets within and around the central commercial district, an area that is declining as a commercial center. Each area was unique and a separate program was set up for each to involve its residents in the selection of improvements for that particular area.

Neighborhood

Traffic-flow patterns in this neighborhood have been the subject of controversy for many years. Previous debate had been over whether or not to widen two of the major streets within the neighborhood. As a result of suggestions from citizens in the first open house, two additional options for improving traffic flow were developed. The purpose of the neighborhood meeting was to provide citizens with the opportunity to review the options that had been developed and to provide them with the opportunity to make comments and suggestions.

Only groups and citizens on the mailing list who lived within the immediate area of the neighborhood were notified of the meeting. This time meeting notices were distributed two weeks in advance of the meeting. Neighborhood leaders were also contacted, and articles appeared in the local newspapers about the neighborhood meeting.

Procedure

The meeting was held early in the evening at a public school within the target neighborhood. The meeting was more formally structured than was the first open house. A more formal procedure was adopted because improvements in this area were highly controversial and we wanted to make sure that everyone was exposed to the same information on the improvements. A presentation was made, followed by a question-and-answer period, and then the meeting was scheduled to break into smaller groups. Those in attendance at the meeting were asked to sign a guest book and encouraged to fill out and return a questionnaire before leaving the meeting.

The purpose and format of the meeting was explained. Background information on future growth and travel demand projections were presented via slides of graphics taken at the previous open house. These included projected growth patterns, future travel demand, and future problem areas. Next, the four alternative options for helping travel patterns in the area were detailed. Included for each option was

1. An explanation of the improvement's design,
2. The impact the improvement would have on projected traffic volumes in the neighborhood,
3. The capital cost of the project, and
4. The effect of the improvement on the local environment, economics, and neighborhood.

After each of the proposals was explained in some detail, the floor was to be opened up for questions. A transition to a more informal meeting was scheduled to result by inviting people to view the graphic displays more closely and encouraging them to break up into smaller groups for discussions with staff members.

Results

Approximately 60 people attended the open house, of whom all but a few were from the immediate neighborhood. Even before the meeting began, we realized that we were dealing with a very emotional issue and that the feelings of many of those in attendance were already running high. Unfortunately, the meeting did not proceed exactly as planned due somewhat to failures on our part.

The slide show of background information could have been eliminated. It took up much valuable time and answered too few of the questions in which those attending were interested. Because the slide presentation cut into our time, the meeting did not become productive until a late hour, after some of the people had left.

Our initial presentation may have reduced some tension and dispelled some preconceived notions if we had (a) assured people that we were looking at long-range problems and, therefore, long-range solutions (many were concerned that a decision had been made and that the improvements would be constructed immediately) and (b) emphasized that there were four options, all of which were feasible. We were able to show people that their suggestions and input did produce results. Two of the options we were reviewing for this area came from suggestions made by individuals at the first open house.

Since much of the meeting operated under a question-and-answer format, certain people in the audience tended to dominate the discussion. In some cases we found it necessary to benignly ignore certain individuals and give preference to those who had not been heard from previously.

Nearly everyone in attendance filled out and returned a questionnaire. This high rate of response could be attributed to our emphasis on completion of the questionnaire before leaving the open house and the brevity of the quesionnnaire.

Downtown

The downtown was one of the last subareas addressed. The Beloit downtown area is declining as a commercial center. Proposals for improving traffic patterns in the area needed to be flexible because the downtown merchants were also considering changes to make the area more attractive for shoppers.

Procedure

Two major options were developed for improving traffic flow in the central urban core. Since no single organization represented the interests of the downtown merchants, the proposals were presented individually and collectively to a cross-section of merchants for comment. Businesses represented included banks, automobile dealers, service stations, restaurants, and a downtown merchants' organization. In most cases an appointment was set up at the individual's place of business. At these meetings little background information was given on the area's transportation study; the discussion focused primarily on the options available in the central area. The impacts of each option were articulated in terms of how they would affect

1. Traffic patterns within the central core;
2. Safety, pedestrians, and parking within the central core;
3. Traffic patterns outside the urban core; and
4. Project costs.

Results

As a whole, the business community was opposed to any major changes in the downtown area. Since such a diverse group of merchants was represented, their perceived needs conflicted. Shop owners were more in favor of a pedestrian-oriented plan; however, service stations, automobile dealers, and restaurants preferred an automobile-oriented plan. A number of strategic errors were made in the presentation of the proposals, which may have had some bearing on how well they were received.

An article in the local press on one of the two options had inaccurately stated that it had been recommended. A week after the article appeared, we were finally able to schedule a meeting with the downtown merchants. This gave the merchants plenty of time to organize against the proposal before we felt they had heard all the facts. Some of the individuals were not interested in hearing about the proposals. They stated emphatically that they were against any improvement and gave us little or no opportunity to present the options completely.

We had no strong merchants' organization with which to work in developing the improvement options. Even if we had been able to get their endorsement, there was no guarantee that the individual members would give their support.

Although many of the fears expressed by the merchants in opposition to the improvements were not substantiated by the facts, we realized that the improvements would fail unless they had the support of the merchants. Furthermore, if an unacceptable improvement were instituted, it would merely serve as a scapegoat for future problems in the area.

SECOND OPEN HOUSE

Once the policy and technical committees had developed a final set of four alternatives, a second areawide open house was held. The objective of this meeting was to inform citizens of the progress of the study and to obtain their comments on their choice among the final four alternative future transportation systems. The results of the citizen input from this open house would be used by the technical and policy committees in the recommendation of one of the four alternatives to the local units of government.

Procedure

The format of this open house was similar to that of the first. The meeting was held at the same place on a weekday between the hours of 1:30 and 8:00 p.m. An announcement, similar to the first, was sent out two weeks in advance of the meeting. By this time our mailing list had grown to more than 250 individuals and organizations in addition to the news media. Press coverage of the open house was much more extensive than that given to the first open house.

Graphics consisted of a map of each of the final four alternatives and the improvements unique to that alternative. Each of the four maps contained the year-2000 projected traffic volumes, both with and without the improvements for that alternative. In addition, a one-page summary sheet of each alternative was available. These summary sheets outlined street, transit, bicycle, and rail improvements.

Street and Highway Improvements	Other Transportation Improvements
Number of kilometers widened and number of kilometers of new facility	Increased transit ridership
Areas of major change in traffic volume (either increases or decreases)	Provision of better bicycle routes
Primary environmental, social, and economic benefits and negative impacts	Reduced rail-automobile conflicts
	Increased vehicle occupancy

At the entrance, individuals were asked to sign a guest book and given a questionnaire, which they were asked to fill out and return before leaving. Tables and chairs were set up off to the side so that people had a comfortable place to complete the questionnaires. Coffee and pencils were provided.

The four large maps (one of each alternative) were displayed as the focal point of the meeting. Handouts that summarized the alternatives were available next to the maps. Technical committee members were present to answer questions, provide information, and explain the alternatives in detail.

Results

Approximately 70 people attended this open house, about the same number as had attended the first open house, although more extensive publicity was given to the second areawide meeting. Attendance was sporadic and slow during the afternoon. It fell off almost completely during the supper hour (4:00-6:00 p.m.), but almost half those attending arrived between 6:00 and 8:00 p.m.

More than half the people attending the open house lived within the subarea addressed at the neighborhood open house. Most of the others were also from an area affected by the study. Although there were some new faces at this open house, most of the people in attendance had also shown up at the previous two open houses.

Fifty-two questionnaires were returned, a 73 percent rate of return. Almost all of the questionnaires returned were completely filled out. The provision of a place for people to return their questionnaires before leaving, as well as tables, chairs, coffee, and pencils, encouraged them to fill out and return the questionnaires.

CONCLUSION

The framework outlined for including local citizens has proved most effective as a means of incorporating citizens' input during the development phase of the planning process. This method was successful in involving the segments of the population that were directly affected. It did not involve a large segment of the indirectly affected public. We could have surveyed areawide opinions to obtain this segment's input, but due to fiscal and time constraints a survey was not possible.

The citizen participation process outlined here and used in the development of the alternatives

1. Opened the process to public scrutiny,
2. Disseminated information on the process, and
3. Incorporated citizens' suggestions and proposals.

RECOMMENDATIONS

The open-house procedure is most effective at the developmental phase of planning. The procedure provided us with a means of disseminating information, sounding out proposals, and developing new proposals. It is important that citizen input be obtained before major decisions are made. People react as much to the decision-making process as to the decision. If citizen participation is to work, it must be included throughout the study, not just at the end. Credibility is increased when you can show that citizen feedback does produce results. Two options that we had not previously considered were incorporated into the study after they had been suggested by citizens attending the first open house.

The most effective way to disseminate information about a meeting is by word of mouth. Contacting neighborhood leaders, local officials, and civic organizations brought out more people than all the newspaper notices and articles combined. Develop a mailing list of interested parties, civic leaders, organizations, and officials. People who express an interest in the study should be added to the list. Providing a place for people to write their name and address at the bottom of the questionnaire was one of the most effective means we found in expanding our mailing list. A guest book for people attending the open house to sign their name and address provided a useful record for documenting attendance and for analyzing the areas represented at the meeting.

We found large photographs of street sections to be more visually effective than engineering cross-sections in showing people what impact a proposed improvement would have on an area. The more information that can be reduced to graphics, the better. An open house is probably the best means of conveying the information in an easily understandable form.

Presentations to a group should be short and direct. Breaking a study into subareas or groups means presenting information that only pertains to that particular area of the study.

Questionnaires are a useful means of recording citizen feedback, but it is important that they be short yet open ended enough to allow suggestions. People are most likely to take the time to fill out and return a questionnaire if tables, chairs, pencils, and coffee are provided in a quiet corner. Results of the open house questionnaires could possibly be biased due to the limited number of people attending and their proximity to an improvement.

REFERENCES

1. Citizen Participation in Transportation Planning. HRB, Special Rept. 142, 1973, 142 pp.
2. Citizen's Role in Transportation Planning. TRB, Transportation Research Record 555, 1975, 52 pp.
3. Social, Economic, and Environmental Factors of Transportation. HRB, Highway Research Record 356, 1971, 174 pp.
4. M. C. Perfater. Citizen Participation and the Role of the Public Hearing. Virginia Highway and Transportation Research Council, Charlottesville, 1975.
5. D. Jordan and others. Effective Citizen Participation in Transportation Planning: Volume 2—A Catalog of Techniques. Arthur D. Little, Inc., San Francisco, 1976.
6. Application of Interactive Graphics in Citizen Participation. TRB, Transportation Research Record 553, 1975, 49 pp.
7. R. Yukubouski. Citizen Participation in Transportation Planning—A Selected Biography. Planning Division, New York State Department of Transportation, Albany, 1973.
8. Toward a New Transportation Plan. Dane County Regional Planning Commission, Madison, WI, 1976.

Publication of this paper sponsored by Committee on Citizen Participation in Transportation Planning.

Pricing in the Planning of Transportation Facilities

William Vickrey, Department of Economics, Columbia University, New York

If proper pricing according to marginal cost principles can be implemented, important improvements in the planning of investment are possible. In cases of daily or seasonal fluctuations in demand, less capacity will often be necessary. In some cases the optimum time for making an investment will be advanced (chiefly in new projects), but in others (particularly in additions to capacity) investment may be deferred. When future demand is uncertain, reductions in planned capacity may result by reason of the better adaptation to alternative developments that are possible with pricing. Where there are externalities, proper pricing methods or effluent charges applied according to eventual results can provide better planning of investment in pollution-abatement facilities or modifications than are likely to result from the imposition of standards.

Ideally, the development of an efficient transportation system (or any system that has large fixed costs) calls for an initial decision about the pricing policy to be followed after the facility is in place. If one does not know what pricing policy is to be applied after the facility is ready for service, and hence what the pattern of use is likely to be, an efficient facility cannot be designed. This does not mean that the actual level or pattern of prices must be decided in advance but, rather, what methods are going to be used to determine the prices that will produce the best results. The policy may dictate a pattern and level of pricing considerably different from what might have been chosen had a firm commitment been required at the time the project was started. The nature of the project should take into consideration alternative possible states of the market for the service, together with their probabilities, and the corresponding pricing and utilization patterns that would result from the application of the prescribed policy to these various possible future circumstances.

IMPLICATIONS OF A SHORT-RUN MARGINAL-COST-PRICING POLICY FOR PLANNING

In this paper, I will maintain that the appropriate pricing policy calls for pricing related to a relevant marginal cost. In some cases this will be a short-run marginal cost, as in cases where the price can be varied on short notice. Where a capability for varying prices from moment to moment would either be expensive or would be ineffective in influencing the pattern of usage, it may be appropriate to vary prices only at substantial intervals. Where, for whatever reason, prices are to be kept fixed for some time, the appropriate marginal cost is one calculated with respect to a time horizon that corresponds to the period during which the price under con-

Figure 1. Economics of scale from the demand side.

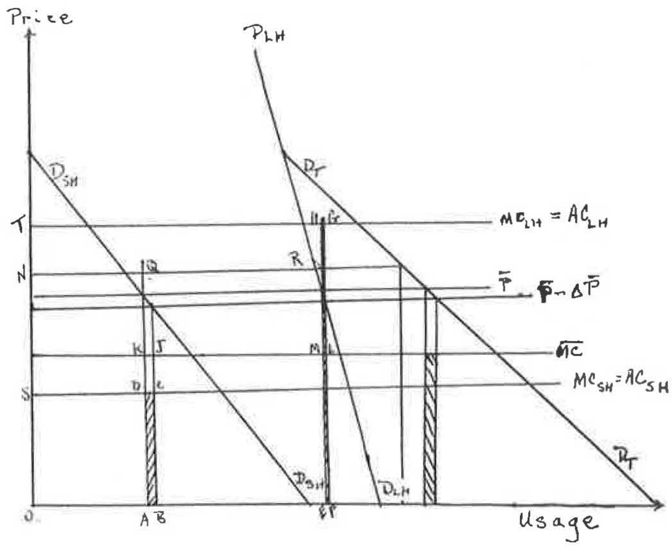
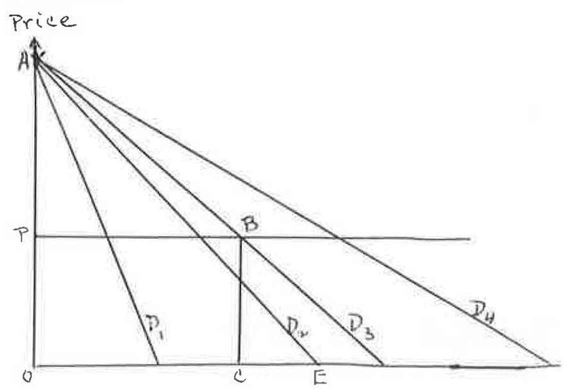


Figure 2. Optimum construction times with fixed and marginal cost pricing.



sideration will not be subject to reconsideration. In any case, the marginal cost that is relevant for a decision about the level at which to set a price is the amount of cost of that service that will be added to or subtracted from the load on the system by a minimum feasible decrement or increment in the price under consideration.

Thus, for example, it is decided that no differentiation can be made in the fare between short-haul and long-haul riders, and if variations in the fare will have relatively little effect on long-haul ridership within the range of fares under consideration (as in Figure 1), then the appropriate marginal cost is one in which the marginal cost of the short-haul riders is weighted more heavily in proportion to this greater responsiveness to fare changes. In Figure 1, if the short- and long-haul marginal costs and demands are MC_{SH}, MC_{LH}, D_{SH}, and D_{LH}, respectively, and if the uniform price is lowered from $\overline{P}$ to $\overline{P} - \Delta\overline{P}$, then the increases in cost will be ΔC = ABCD + EFGH. This will produce an effective marginal cost $\overline{MC}$ such that ABJK + EFLM = ΔC = ABCD + EFGH. In the extreme, where the long-haul ridership is totally insensitive to fare changes within the range being considered, the marginal cost of short-haul riders would be the ruling factor.

If, for the sake of argument, we could have a cost function where costs are strictly proportional to passenger kilometers, we could nevertheless come out with a marginal cost per passenger that is less than the average cost per passenger. In terms of the measure of output used as the basis for pricing, an economy of scale arises here from the demand side of the market rather than from the supply side: Reduction in the price results in a pattern of usage that produces a lower cost per ride, even though, from the strict technological viewpoint, there are constant returns to scale. Thus, in Figure 1 if MC_{LH} and MC_{SH} are both constant and there are no fixed costs, the separate marginal and average costs are equal. Nevertheless, the combined average cost $\overline{AC}$, such that ONQA + ONRE = OSDA + OTHE, exceeds the effective marginal cost $\overline{MC}$. This, in turn, implies that optimum exploitation with price at marginal cost would require a subsidy, even where, in terms of the production technology, there may be constant returns to scale.

Such an appropriate pricing policy would have important implications for the planning of investment facilities in a number of situations. Among the more important ones are the timing of lumpy investments, allowances for errors in forecasting, the design of substitute facilities, and the control of use during temporary conditions, such as during construction.

TIMING OF LUMPY INVESTMENTS

The traditional approach to the timing of lumpy investment in facilities for which a price is charged, such as a toll bridge, usually presupposes that a constant price will be charged, which will amortize the cost of the facility over some reasonable period of time. The project may be embarked on at a time determined (a) somewhat arbitrarily by engineering and financial considerations or (b) by the earliest date at which a favorable cost/benefit ratio can be demonstrated. A more rational approach, even under a constant price regime, would be to time the construction to the date at which the current annual benefits from the traffic flowing at the fixed price would begin to exceed the annual costs for interest, operation, and maintenance, exclusive of amortization (possibly with some allowance for a habituation factor).

A marginal-cost-pricing approach would entail the operation of the facility as a free facility or at a relatively low toll as long as it is congestion free and the embarkation on the facility according to the same rule (i.e., when the aggregate benefit exceeds the annual cost other than amortization). Now this same criterion would be satisfied earlier, given the higher level of traffic carried at the lower price. Thus in Figure 2, if we think of the demand curve for traffic as growing over time, becoming successively D_1, D_2, D_3, D_4, . . ., then under the constant-price arrangement the facility should be built when the demand curve reaches D_3, where the annual user benefit (measured by OABC) is equal to the costs associated with building the facility one year earlier. Under marginal cost pricing the project could optimally be undertaken at the earlier time when the demand reaches D_2, where the benefit measured by OAE will cover the incremental costs of the initial year of operation.

The situation may be reversed when it comes to expansion of the facility. In Figure 3, under a policy of maintaining a fixed price (P), the capacity of the original unit is indicated by F_1 and the addition of another unit by F_2.

When the demand level reaches D_4 at the fixed price P, congestion begins to be felt. Demand for enlarging the facility may become especially persuasive if the revenue produced by the traffic that uses the second unit, half of the total (given by MNHF$_1$), is sufficient to cover the incremental costs of the initial year of the new unit. Actually, even at the fixed price P, the optimal time to bring the new unit into service would be when the

Figure 3. Duplication of capacity under fixed and marginal-cost-pricing policies.

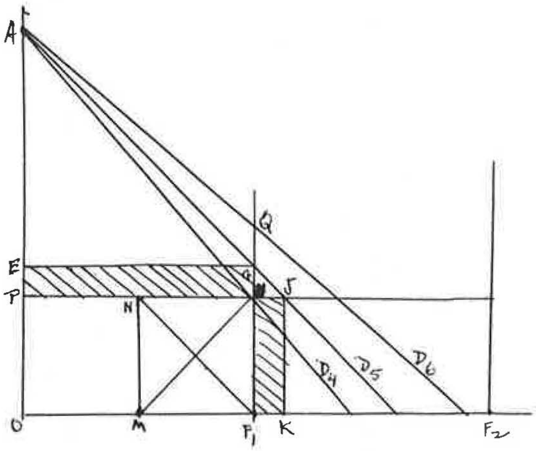
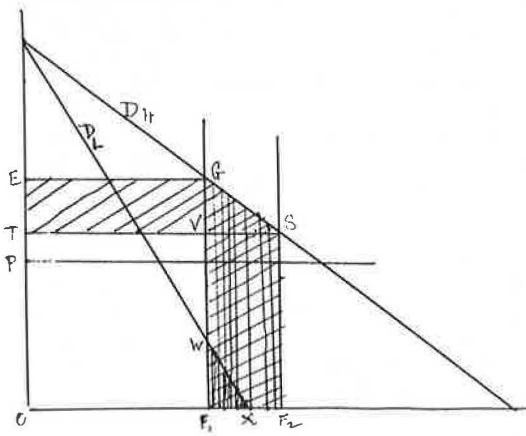


Figure 4. Investment planning for uncertain future demand with fixed and variable pricing policies.



demand curve reaches D_5, where the sum of the gain to the new traffic (F_1GJK) plus the gain from the reduction in congestion to the old traffic (PLGH) covers the incremental first-year cost.

With marginal cost pricing, price would be held at zero until the time at which demand at price zero reaches capacity. Price would then be increased so as to keep demand at capacity without congestion, until demand reaches D_6 at the price of F_1Q, at which time the benefit from the new unit with the price cut back to zero will be QRF_1, which is sufficient to cover the initial annual cost. Thus, although marginal cost pricing will justify bringing the first unit in earlier than would the constant-price policy, it will result in the deferral of the optimal time for the installation of subsequent units. (In the above analysis it is assumed that when the price remains at P, congestion increases as the demand increases from D_4 to D_5 so that, by adding the congestion costs to the price, a combined cost of travel is produced that just restrains demand to the capacity of the facility. Were demand to exceed capacity, queues would build up to increase the congestion cost until the equilibrium is achieved.)

In practice, the situation is often worse than this in that, just as traffic reaches levels that would justify a positive price to keep congestion at an optimum level, a demand arises for the reduction or elimination of the toll on the grounds that the facility has been fully amortized and should be made free. Equity, as popularly perceived, is often in sharp conflict with efficiency.

Errors in Forecasting

The forecasting of demand several years into the future is always fraught with considerable ranges of error. In the absence of pricing, or with an inflexible pricing policy, it is necessary to balance the risk of wasting resources on an excessively early or excessively large-scale development of a facility against the risk of losses due to the development of congestion on an inadequate facility. In Figure 4, with a fixed price (P) the advantage of having built capacity F_2 rather than F_1 will be nil if demand is D_1, $EGSF_2VT$ if demand is D_H, and F_1F_2W if demand is D_L. It will be seen that, in cases such as this, the optimal capacity is likely to be greater under a fixed-price regime than under a flexible-pricing one. This tendency will be the greater if, as is likely, planners will be subject to greater criticism if their scale is inadequate than if it is somewhat excessive, and also if, as is likely, there are "hypercongestion" phenomena in which queues backed up from one point of constricted capacity interfere with traffic that does not itself pass through the bottleneck.

This tendency of planners, when operating under a relatively inflexible price structure, to overbuild relative to the optimal scale when flexible pricing is contemplated may be in part responsible for the notion that short-run marginal cost tends to be lower than long-run marginal cost, at least in those cases (mostly involving a single product without joint costs with other products) where a long-run marginal cost can be adequately defined.

Substitute Facilities

The influence of pricing on planning may be even greater where there are substitute facilities or where substitution between peak and off-peak travel can occur. For example, the building of the third tube for the Lincoln Tunnel or the construction of the Throg's Neck Bridge may have been necessary at the time they were constructed, given a commitment to a uniform toll over a number of substitute facilities and over time. However, had flexibility of toll rates over time and among facilities been available as a policy, construction of one or both of these facilities could have been postponed or even deferred indefinitely.

Particularly in the case of the Lincoln Tunnel, through appropriate toll variation, traffic could have been diverted to other facilities and away from peak hours. Even today, the construction of the third tube might turn out not to be cost effective if the pricing alternative were considered an acceptable one. This involves not only the more obvious shifts to the Holland Tunnel and the George Washington Bridge but also shifts to the Verrazano Bridge on the south and the Tappan Zee and Beacon-Newburgh Bridges on the north. In the case of the Throg's Neck Bridge, during peak hours the possibilities for shifting traffic to the Whitestone Bridge or routes farther south may be quite limited. However, if the Throg's Neck Bridge were credited with only those toll revenues that correspond to traffic that could not be shifted elsewhere (either at the same or at a different time), the revenues would fall considerably short of covering the costs, and the consumer surplus element in the benefit would be comparatively small. There are few trips for which the running time, for passenger automobiles at least, would be more than a minute or two longer via the Whitestone Bridge.

Variations in the toll by time of day are not all that difficult. To be sure, one would not want to have large jumps in the toll at times of heavy traffic lest this create minipeaks and other disturbances to the smooth flow of traffic. The toll would have to vary, for most vehicles at least, in a reasonably smooth fashion from one moment to the next. Several methods of doing this conveniently are available. For frequent users, a credit card or an automatic vehicle identifier (AVI) unit could be the basis for charging the vehicle by billing by mail at suitable intervals. Another method would be to have two levels of cash toll, an off-peak toll of, for example, $0.25 with no refund and a peak cash toll of, for example, $3.00, in return for which the driver would receive a voucher equal in value to the difference between the actual toll in effect at the moment and the $3.00 paid. The value of the voucher would be indicated on it, and this value could be realized at the option of the driver in one of three ways: (a) by cashing in at a gasoline station at the next stop for fuel; (b) by rubbing off a black panel that, with appropriate probability, would make the voucher either worthless or valid for the $3.00 toll for a subsequent rush-hour trip; or (c) by rubbing off a second black panel that, with appropriate probability, would produce a winning lottery ticket worth, for example, $100.00. The latter two alternatives might be given a somewhat higher expected value than the flat cash option to minimize the cost of the cashing transactions. This might also enable service stations that cash the vouchers to turn them into lottery tickets, again reducing the cost of handling to them and providing a slight profit, on the average, for their trouble.

Flexible pricing would also largely eliminate any need for reserved lanes or bypass facilities for buses and high-occupancy and priority vehicles. If pricing is set at appropriate levels, there would be no significant congestion to be bypassed, in most cases; the high peak-hour toll would furnish the appropriate incentive for carpooling and other means of increasing occupancy. Thus, the costs of setting up the reverse bus-only lane for inbound buses in the morning on the approach to the Lincoln Tunnel could have been avoided had flexible pricing been adopted. The costly reserved express lanes on Shirley Highway would show up as wasteful extravagances if a free flow of traffic on the older lanes were maintained by appropriate pricing, which would require installation of some new pricing system, such as an AVI system for electronic identification of vehicles at suitably distributed scanning points. The results would be used to bill vehicle owners. The extra lanes on Shirley Highway do not represent any additional usable capacity, given the inherent limits on traffic flow across the Potomac.

The availability of pricing would also permit substantial economies to be made in construction costs where obstruction to traffic is involved. Often unnecessarily high costs are incurred, both in the delays to traffic and in the costs of "shoo-fly" arrangements put in place to facilitate the flow of traffic past the construction site. Pricing would permit the diversion of traffic in an efficient manner away from the affected site and permit the temporary arrangements at the site to be less costly. The tendency of traffic to avoid congested routes is insufficient to bring about an efficient allocation of traffic over alternative routes, given the wide disparity between the congestion experienced by the driver and his or her total contribution at the margin to overall congestion, the latter being typically two to five times the former.

The role of pricing in traffic diversion is not limited to motor vehicle traffic, but also concerns transit systems. At current levels of ridership the matter may well be moot, but in the 1940s, for example, conditions at Grand Central Station on the Lexington Avenue Line were a controlling factor in limiting the service on the entire line. Passengers transferring at that point between points in Queens and downtown contributed greatly to the congestion. If a suitable concession in the fare had been made for persons who would continue on to 6th Avenue or Times Square for their transfer, a significant reduction in congestion could have been obtained for relatively minor added inconvenience to the riders who would take advantage of this concession. Similar opportunities probably occur elsewhere.

One fairly common failing of even fairly sophisticated systems is to implement time-of-day pricing by schedules in which the maximum-fare period is so long that few riders at the height of the peak will have any interest in shifting the time of their riding, since the amount of the shift they would have to make would be so large relative to the saving in fare that it would not be worthwhile. To be at all effective in abating the peak proper, some incentive, even a relatively modest one, must be offered to those who will shift their time of travel by as little as 10 or 20 min away from the height of the peak. This is difficult to accomplish with a fare schedule that has only two levels. A time-of-day fare schedule needs to have several levels, maybe five or more; a peak fare that lasts no more than about 15 min; and fares that drop off at 10- or 20-min intervals. Since few riders would absolutely have to ride during the peak 15-min period, the fare at such times could be made quite high without giving rise to complaints that it is oppressive for low-income workers.

The manner in which the magnetic fare-card system has been implemented can also be considered rather inept. There is no reason for having the original fare-card-issuing machines outside the controls give change. This simply increases the cost of machines, causes delay to occasional riders when they attempt to pay for one ride at a time, increases the number of machines required to give satisfactory service, increases the cost of keeping the machines provided with change, and increases the bulk of the machines, which might cause space problems. The change delivery cups are inconveniently located and are so designed that considerable fumbling is needed to collect the change. Change giving could be limited to the add-fare machines, which must give change when wanted in any case. Travelers who wish to avoid being left with a balance on their fare card could obtain their change at these machines without having to guess at the fare, and travelers who expect to make further trips would be discouraged from getting their change by the need to use an additional machine. Even more important, add-fare machines should be arranged to provide the option of not giving change but of simply adding the amount deposited to the value of the fare card, either the same one or a replacement. There is no point in forcing the passenger to take this change, surrender the fare card to the exit turnstile, and then purchase a new fare card for the next trip.

Adaptation of a magnetic fare-card system to vary the fare according to the route taken, as in the routing via Times Square suggested above, is simple. All that is necessary is to provide registers at convenient points in the Times Square Station where passengers who wish to have the benefit of the reduced fare can insert their fare cards. The registers would produce a suitable coding on the card that would be recognized by the exit turnstile as entitling the passenger to the reduced fare via the Times Square route, otherwise the higher fare route would be charged.

PRICING AND POLLUTION

Perhaps the application of pricing principles to the control of pollution can also be considered a part of planning by the automobile manufacturing industry. So far only mediocre progress has been made in the reduction of automotive pollution through the attempt to impose fairly stringent standards on manufacturers, in terms of the performance of their vehicles as they leave the factory. The manufacturers have stalled, on the ground that the standards were impossible to meet, and the standards have been lowered. Manufacturers have elected to use the catalytic converter as the means to meet the standards, at least temporarily, which has resulted in automobile owners bypassing or discarding the converters or poisoning them with lead. Now a shortage of high-test unleaded gasoline is threatened. At best, manufacturers have been concerned with performance as measured at the factory gate, but what is important is the performance of automobiles on the road.

Ideally what might be called for is to periodically measure the pollution performance of each automobile on the road and levy a pollution tax, according to the results, on the owner. But such an approach would be inordinately costly to administer, especially in the light of the high proportion of total emissions accounted for by cold starts and the high cost in terms of inconvenience and otherwise of testing for this element. Moreover, important elements of the pollution control problem relate to design and quality control in a way that would be unlikely to affect manufacturers through the influence of taxes levied on owners or the demand for various models. Moreover, through the price or tax mechanism manufacturers could be made responsive to more than just the performance of the vehicle at the factory gate.

What suggests itself on this basis is as follows: As automobiles leave the factory they could be sampled, tested, and a pessimistic forecast made of the pollution likely to be emitted by automobiles over their lifetime. An appropriate tentative tax could be levied on the manufacturer, which, in principle, would represent the discounted present value of the marginal pollution damage attributable to this pessimistic forecast of pollution. Subsequently, at suitable intervals a random sample of these same automobiles could be selected in the field and tested for emission levels, and to the extent that the results are better than the pessimistic forecast, a rebate of the tax could be paid to the manufacturer. If the results should prove worse, a supplementary tax would be levied, though this might be difficult to collect if the manufacturer is no longer in business. Owners of vehicles selected for testing would be suitably compensated. One might economize in the testing by testing at two levels: a smaller subsample to be tested fully, including cold-start tests, the full sample to be tested fully warmed up. The sample would have to be properly stratified by time of year, climate, altitude, and environmental density. It would be appropriate to weight more heavily the sampling in areas of highest pollution levels.

Such a pricing approach to pollution would have several advantages over mandatory standards. Setting of mandatory standards cannot deal with the problem of providing an incentive for action to improve emission performance after a lapse of time, as distinct from performance when new. Another advantage is that the incentive would be applied where it will do the most good. More care will be given to automobiles likely to be heavily used in polluted areas and less to automobiles likely to be used chiefly in areas where pollution is of little consequence. Instead of concentrating on devices that may abate pollution of new automobiles, attention will be paid to methods of pollution abatement that are less vulnerable to neglect or abuse by owners, such as stratified charge engines rather than catalytic converters. Manufacturers would be given an incentive to make appropriate maintenance, repair, and retrofitting kits available to their service stations and parts distributors. Perhaps most important, manufacturers would not simply balk at making the desired changes: Failure to respond adequately would merely result in the levying of a tax, and the manufacturer would not be able to threaten to suspend production or simply offer nonconforming vehicles and engage in a confrontation of technological experts over the issue.

Such a program would have a higher administrative cost than simply an attempt to impose standards. But it is hard to see what alternative could achieve better results.

The problem with all of this analysis is that, ever since Thomas Aquinas, the use of pricing mechanisms to achieve economic efficiency has had a bad press. But, if injustices are produced by efficient prices, they can often be remedied by appropriate transfer devices; if inefficiencies are produced by just prices, there will seldom be any way of making good this loss. If enough efficiency is lost by insisting on just prices, everyone may wind up the loser.

Publication of this paper sponsored by Task Force on Economic Analysis.

Issues in Measuring the Costs of Railroad Accidents

G. A. Fleischer, University of Southern California, Los Angeles

The allocation of limited economic resources among competing investment proposals in the railroad industry, especially among projects and programs designed to reduce accident occurrence and severity, implies that a means exists for relating the costs of railroad accident-related deaths, injuries, and property damage. This paper addresses the principal associated issues and suggests courses of action to assist analysts and decision makers.

The principal focus of this paper is an assessment of the state of understanding and development of recommendations for an approach to the analysis of societal impacts of railroad accidents, including events involving fatalities, injuries, and property damage. Of particular interest is the investigation of alternative philosophies and methodologies for measuring societal impacts of death and injury in economic terms.

Each of nine separable but interrelated issues are discussed in this paper. For each, the relevant alternative positions are identified and discussed, and appropriate recommendations are provided.

SHALL THE COSTS OF MORTALITY, MORBIDITY, AND PROPERTY DAMAGE BE MEASURED IN ECONOMIC TERMS?

Arguments Against Measuring the Costs of Death and Injury in Economic Terms

Based on a review of the research activities in this area over the past 50 years, it is clear that there is no general consensus as to either monetary estimates of the costs of death and injury or an underlying philosophy and methodology. Estimates that are made are too uncertain to be relied on for decision making. Economic estimates, if used, can obscure other important issues, such as pain and suffering, that result from death and injury.

Rhoads and Singer argue (1), "It is demoralizing when society collectively and publicly places a value on life. It is especially so when a decision is made not to save an identifiable individual." To the lay public and their political representatives the attachment of monetary values to mortality and morbidity suggests a nonfeeling, noncaring indifference. There is an implied mechanical precision. Most public agencies avoid the issue operationally. For example, although the National Highway Traffic Safety Administration (NHTSA) has funded a substantial effort to identify the social costs of motor vehicle accidents, the resulting values are not currently being used by that agency.

Accident Cost Analysis

Accident costs can be defined in nonmonetary terms. In the evaluation of the relative efficacy of a proposed program or project having safety consequences, for example, the cost of the program or project could be stated in dollars and the effectiveness defined as reduction in risk, number of accidents, or number of casualties. However, no decision can be made as to the relative attractiveness of alternatives unless the decision maker knows something of the relation between the relative utility of the differences in effectiveness and the difference in costs. That is, the priority of alternatives cannot be determined unless effectiveness and cost can be expressed in a common dimension.

In the event that accident costs (i.e., deaths, injuries, and property damage) can be expressed in the same terms as the other consequences associated with a program or project, then a benefit/cost analysis can be performed to assist the decision maker in the allocation of limited capital resources.

Note that the actual implementation of benefit/cost analysis is not nearly as straightforward as outlined above. Complications arise, in large part, because of the stochastic character of principal estimates, chiefly the number of future occurrences of an event (e.g., fatality and injury) and the cost per occurrence. These are predictions of the future; they are not known with certainty.

Recommended Position

Inasmuch as accident costs must be described commensurately with other principal consequences in order to make informed decisions concerning capital allocation, and since the latter are normally described in monetary terms, accident costs, especially death and injury, should also be stated in monetary terms. Jones-Lee makes the point effectively (2):

> It is a fact of life that society confronts a problem of scarcity and must in consequence engage in continual allocative choices. Insofar as such choices occur in the public sector, it is desirable that those who make them should do so on the basis of more rather than less information concerning private preferences.

FROM WHOSE POINT OF VIEW SHALL IMPACTS BE ASSESSED?

Railroads generally assess costs from the point of view of the reporting railroad (e.g., loss or damage to freight, cost of clearing wrecks, and damage to railroad property). Railroad managements are concerned with effects that are directly reflected in their income statements. Federal agencies and the Association of American Railroads (AAR) have not required more extensive reporting, either because they see no particular reason to do so, the means are not at hand, or the costs would be prohibitive.

The bulk of public opinion on this matter concludes that, with respect to public decisions concerning the allocation of public funds, the appropriate perspective is that of the society at large. Jones-Lee summarizes this position (2):

> The resemblance between cost-benefit analysis and "commercial" project appraisal is, however, more apparent than real. This is hardly surprising since the ultimate objectives of public-sector decision makers are unlikely to bear much resemblance to the objectives of decision makers in the private sector. The essential difference is that the managers of a firm will probably be largely concerned with their own and their shareholders' interests while the public sector decision maker normally will be concerned with a more nebulous index of the welfare of society "as a whole".

HOW SHALL ECONOMIC VALUE BE DETERMINED WITH RESPECT TO MORTALITY?

A variety of viewpoints concerning the value of life have been expressed in published literature. Generally, they can be summarized as follows:

1. Willingness to pay (WTP);
2. Discounted future earnings (DFE) (gross) and DFE (net);
3. Societal costs, including DFE; and
4. Miscellaneous other (e.g., insurance premiums and court awards).

There are virtually no advocates of reason 4, for reasons that are rather obvious. The other views, however, do have their partisans among thoughtful scholars and practitioners.

WTP

One body of thought argues that the value of human life is best determined by the individuals involved, who can express their willingness to pay for certain risk-reduction options, either explicitly or implicitly. There are variations in this viewpoint. Jones-Lee (2-4) determines the functional relation between an individual's future income stream and the amount he or she will pay to reduce the probability of death. Mishan (5) requires, for each affected individual, the maximum amount he or she will pay rather than forgo a project that results in certain probabilities of death.

Some results reported by investigators using this approach follow (all figures are for the year reported):

Table 1. Societal cost summary discounted at 10 percent, 1975.

Cost Component	6 (Fatality)		Injury Severity 5		4		3		2		1		Property Damage Only	
	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent	Cost ($)	Percent
Production-consumption market	145 670	72.35	82 250	61.14	36 075	58.53	1645	20.35	865	19.86	66	3.01		
Home, family, and community	43 700	21.71	24 675	18.34	10 820	17.55	425	5.26	310	7.13	20	0.91		
Medical														
Hospital	275	0.14	5 750	4.27	2 250	3.65	1095	13.54	450	10.34	45	2.05		
Physician and others	160	0.08	5 520	4.10	2 160	3.50	525	6.49	165	3.80	55	2.51		
Coroner-medical examiner	130	0.06												
Rehabilitation			6 075		3 040									
Funeral	1 080	0.54												
Legal and court	2 190	1.08	1 645	1.22	1 090	1.77	770	9.52	150	3.45	140	6.39	7	1.35
Insurance administration	295	0.15	295	0.22	285	0.46	240	2.97	220	5.06	52	2.37	30	5.77
Accident investigation	80	0.04	80	0.06	70	0.11	45	0.56	35	0.81	28	1.28	6	1.15
Losses to others	3 685	1.83	4 180	3.11	1 830	2.97	260	3.21	130	2.98	32	1.46		
Vehicle damage	3 990	1.98	3 990	2.97	3 960	6.42	2920	36.12	1865	42.87	1595	72.83	315	60.58
Traffic delay	80	0.04	60	0.04	60	0.09	160	1.98	160	3.68	160	7.31	160	30.77
Total	201 335		134 520		61 640		8085		4350		2190		520	

Table 2. Costs of death, injury, and property damage per occurrence.

Source	Fatal($)[a]	Injury ($)[a]	Property Damage Only ($)[a]	Ratio of Cost of Fatality to Nonfatal Injury
Fromm (7), 1975[a]	287 175	3185	520	90
NSC (19), 1973	90 000	3700	570	24
NHTSA (20), 1971	43 000	2200	-	20
White House (15), 1972	140 000	2750	-	51
Helms (21), 1971[c]	83 200	1300	-	64
Niklas (22), 1970[c]	32 400	362	-	90
Reynolds (12), 1956[d]	5 580	1450 serious	-	4
		112 slight		50

[a] Values are expressed in dollars for the year of the study, not in constant dollars.
[b] Assuming a 7 percent discount rate.
[c] In 1971, \$1.00 = DM 3.70.
[d] In 1956, \$1.00 = £0.36.

(a) Carlson (6), \$200 000-\$1 000 000 (1963); (b) Fromm (7), \$210 000 (1965); (c) Thaler and Rosen (8) \$200 000 (1975); (d) Ghosh, Lees, and Seal (9), \$260 000 (1975); and (e) Blomquist (10), \$257 000 (1977).

DFE

Advocates of the DFE viewpoint argue that the value of an individual's life is measured by the wages that society is willing to pay for his or her future services. These are then discounted (by the social rate of discount). Most investigators take the view that the most appropriate measure is the net loss of output [i.e., future production less future consumption (11, 12)]. Others argue that the gross loss of output should be measured (13):

> The accidents that need to be costed are those that do not occur but which, without the introduction of some safety measure, would have occurred. The fact that on this basis the individual concerned is, indeed, still alive means that the individual consumption should not be deducted when assessing the benefits of preventing accidents, as he is alive and able to enjoy that consumption.

Some quantitative results are summarized as follows (all figures are for the year reported):

1. Gross—Dublin and Lotka (11), \$27 209 (1930); Bollay and Associates (14), \$250 000 (1963); White House (15), \$140 000 (1972); Usher (16), \$150 000 (1973); and Faigin (17), \$184 110 (at 7 percent discount) (1975).
2. Net—Dublin and Lotka (11), \$9802 (1930).

Societal Costs

Perhaps the most widely held view of the value of life is one that aggregates a number of societal cost components, including the forgone production of the individual, relevant medical costs, legal and court costs, and accident investigation costs (7, 12, 17, 18). In some variations the component that represents value of life to the individual is measured by WTP (2).

The most detailed and current effort based on this view is that of NHTSA during the early 1970s (17). The results for the principal cost elements are given in Table 1. [Results are given for five levels of severity of nonfatal accidents from critical (5) to minor (1)]. It is particularly interesting to note that, using a 10 percent discount rate, the production-consumption component (\$145 670) represents more than 72 percent of the total cost of a fatality (\$201 335). This element is determined by discounting forgone compensation, which is a proxy measure of societal valuation of production. The second largest component, home, family, and community services production losses, represents about 22 percent of the total. Thus, these two cost elements alone represent 95 percent of the total, a result that should influence the allocation of additional research resources.

A Sample of Results

Costs of fatalities, as well as injuries and property-damage-only accidents, are summarized in Table 2 for a number of sources. The data used by the National Safety Council (NSC) are shown for 1973; these most probably have been updated since 1973, but current values are not generally available. The principal reason for the substantial differences between NHTSA and NSC values for fatalities is that NHTSA discounts gross future earnings whereas the NSC discounts net future earnings.

HOW SHALL ECONOMIC VALUE BE DEFINED WITH RESPECT TO MORBIDITY? WHAT ARE THE COSTS OF INJURIES?

There are two fundamental questions closely related to this issue.

1. Shall a single cost per occurrence be established, irrespective of the severity of the injury, or shall separate costs be estimated with respect to separate levels, or classes, of injury severity?
2. Shall the cost of pain and suffering be included as a relevant component?

Table 3. Casualties of employees on duty on class 1 and 2 railroads, 1976.

Injury or Illness	Injuries	Workdays Lost	Average Days Lost per Injury
Nonfatal injuries			
Bruise—contusion	12 309	75 111	6.10
Sprain—strain	18 549	166 951	9.00
Cut or laceration—abrasion	10 018	33 701	3.36
Electrical burn or shock	261	1 546	5.92
Other burns	1 559	6 460	4.14
Dislocation	315	6 007	19.06
Fracture			
Arm or hand	722	18 840	26.09
Fingers	1 616	12 693	7.85
Leg or foot	1 011	40 062	39.62
Toes	667	6 806	10.20
Head or face	230	4 010	17.43
Torso	476	15 688	32.95
Other	165	2 326	14.09
Amputation			
Arm or hand	16	3 391	211.93
Fingers	151	4 040	26.75
Leg or foot	39	8 387	215.05
Toes	16	989	61.81
Other	1	365	365.00
Cinder or other foreign particle in eye	6 124	8 515	1.39
Hernia	377	12 655	33.56
Concussion	139	2 523	18.15
Nervous shock	35	337	9.62
Internal injuries	83	2 685	32.34
Loss of eye	7	632	90.28
Other	1 428	7 994	5.59
Total	56 314	442 714	7.86
Nonfatal occupational illnesses			
Occupational skin diseases or disorders	868	1 494	1.72
Dust diseases of the lungs	4	37	9.25
Respiratory conditions due to toxic agents	262	1 131	4.31
Poisoning	82	414	5.04
Disorders due to physical agents	144	308	2.13
Disorders due to repeated trauma	23	212	9.21
Other	192	719	3.74
Total	1 575	4 315	2.73
Fatalities	100	3 911	39.11
Total	57 989	450 940	7.77

The Cost of Pain and Suffering

With respect to the latter question, we note that nowhere in the literature is an attempt made to include pain and suffering as a cost component of morbidity. In part, this may be explained by the fact that the WTP approach has been associated historically only with the cost of fatalities, not injuries. And it is the WTP concept that provides a theoretical basis for estimating that amount that individuals would be willing to pay to avoid pain and suffering. An injury, moreover, is not the finite event that death is perceived to be. Pain and suffering are even less definable. Thus, even in using the WTP approach, it is probably infeasible to attempt to include pain and suffering as a societal cost element measured in monetary terms.

Injury Severity

As illustrated in Table 3 (23), injuries are currently described by reporting railroads in terms of the type (e.g., bruise or strain) rather than the severity. The existing classification scheme does not readily lend itself to costing. A contusion, for example, may result in very substantial costs or may require little or no medical aid or lost time. Moreover, as will be shown below, societal cost data developed by other investigators are related to injury classes other than that currently used by railroads.

The Abbreviated Injury Scale (AIS), published by the American Medical Association, Society of Automotive Engineers (SAE), and American Association for Automotive Medicine (AAAM) in 1971 and revised in 1976, is gaining increasing acceptance throughout the United States and abroad and is used almost exclusively in the classification of injuries from traffic accidents. The AIS is used in coding specific individual injuries. The codes are

0—no injury,
1—minor,
2—moderate,
3—severe (not life threatening),
4—serious (life threatening),
5—critical (survival uncertain), and
6—maximum (currently untreatable).

The overall AIS(OAIS) is used for coding multiple injuries (24): "Basically, the coder equates what in his judgment is the total effect of multiple injuries on a victim's body and systems with the effects on the body and systems of a single injury with a known AIS."

The advantages of using the AIS (or OAIS) for classifying injuries include (25) the following:

1. Single, comprehensive system for rating tissue damage,
2. Acceptable to both physicians and engineers,
3. Severity of injury can be rated in the AIS without regard to whether or not the victim dies,
4. Adopted by the multidisciplinary accident investigation (MDAI) teams established by the U.S. Department of Transportation (DOT) and by crash investigators worldwide, and
5. Good interrater reliability has been demonstrated.

Cost Estimates by Injury Severity

In Table 1 cost estimates were presented for various injury classes as reported by NHTSA (17). These values are given as point estimates. However, recent studies at the University of Michigan indicate that the variance of the cost distribution is quite large. Ranges of costs as actually experienced by a fairly small sample are reported as follows (18, 26):

OAIS Code	Costs ($) Range	Mean	NHTSA (1975 dollars)
1	0- 4 327	983	2 190
2	1775- 3 382	2 497	4 350
3	2569- 16 313	7 568	8 085
4	4457-217 979	46 924	61 640
5	4730-364 493	68 134	134 520

For comparison, costs derived by NHTSA (17) are shown in the last column of the above table. The reference year for both sets of data is 1975. Clearly the NHTSA-derived values are substantially higher than the means of costs actually experienced in the University of Michigan studies. Not too much should be made of this, however, inasmuch as the University of Michigan data are from a small regional sample.

AS CURRENTLY REPORTED BY RAILROADS, ARE INJURY DATA ADEQUATE FOR ASSESSING COSTS OF MORTALITY AND MORBIDITY?

Reportable Accidents and Incidents

Railroads are required to file monthly accident and incident reports with the Office of Safety, Federal Railroad Administration (FRA). These include

1. A monthly report of railroad accidents,
2. A rail equipment accident and incident report,
3. A rail-highway grade crossing accident and incident report, and
4. A railroad injury and illness summary.

Reportable accidents and incidents are defined as any impact between railroad on-track equipment and automobile, bus, truck, motorcycle, bicycle, farm vehicle, or pedestrian at a highway grade crossing; any collision, derailment, fire, explosion, or other event involving railroad on-track equipment that results in more than $2900 in damages to railroad on-track equipment, signals, track, track structures, and load bed; and any event arising from operation of the railroad that results in the death of one or more persons, injury to persons other than railroad employees requiring medical treatment, injury to employees (limited), and occupational illness of employee.

Major revisions in reporting requirements, effective January 1, 1975, are summarized in Table 4. Because of these changes, comparisons of the data from 1975 and later with data from previous years are virtually impossible.

Major Problems

The accident and incident data reporting system currently used by railroads is inadequate for an assessment of the qualitative and quantitative effects of safety impacts. Among the principal problems are the following:

1. Reporting is incomplete; minor injuries are unreported unless medical treatment beyond first aid is necessary.
2. The description of injury severity is imprecise and does not facilitate comparison with OAIS data developed in other contexts.
3. There appears to be little quality control; no mechanism exists for monitoring or ensuring that reporting procedures are uniform among the various railroads.
4. Objectivity of reporting is questionable; it is possible, indeed probable, that biases arise as the result of the reporting railroad's desire to avoid the appearance of negligence and inefficiency.
5. Currently, time lost due to injury is truncated at 365 days; if, for example, an injured person is expected to be incapacitated for 18 months, the cost of that lost time between 12 and 18 months is ignored. This is a systematic bias and tends to understate the true cost of time lost due to injuries.

ARE CURRENT PROCEDURES ADEQUATE FOR ASSESSING PROPERTY DAMAGE DUE TO RAILROAD ACCIDENTS?

A variety of data sources are used to identify property damage due to railroad accidents. The major sources are the Interstate Commerce Commission (ICC) (uniform system of accounts) and the FRA (yearly financial reports). Supplemental sources include the AAR and the internal accounting systems of the various railroads.

Damage to railroad property is reported to the FRA when the damage estimate exceeds $2900. All railroads are required to report damage to livestock on the right-of-way to the ICC in their annual reports. Costs include direct expenses and related employee salaries, expenses, office rent, and probable liability (i.e., the railroad's liability). Freight loss and damage is also reported annually to the ICC. The costs of clearing wrecks are not included in the damage costs reported to the FRA in the accident reports. Omitted are the costs of emergency services, which are borne

Table 4. Some changes made in reporting requirements.

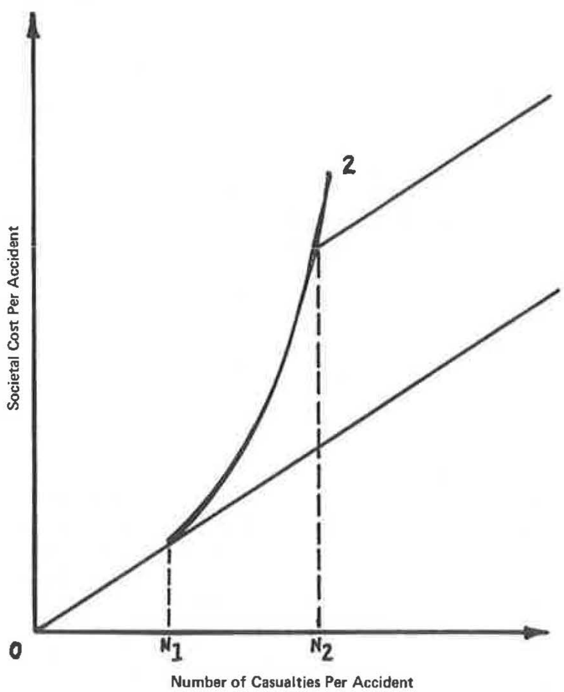
Reporting Requirement	Through 1974	1975 and Later
Damage threshold for reporting train accidents	$750	$1750 as of 1/1/75 $2300 as of 1/1/77 $2900 as of 1/1/79
Requirements for reporting rail crossing accidents and incidents	Only if reportable casualty or minimum of $750 damage to railroad equipment, track, or roadbed	Any
Reporting of fatalities	Death occurring more than 24 h after occurrence of injury reported as injury	Reported as fatality
Reporting of employee injuries	Only those injuries causing at least 2 days of lost or restricted time. Case remains active for 10 days	Those injuries that result in one or more lost or restricted workdays, medical treatment beyond first aid, transfer to another job, termination of employment, or lost consciousness. Case remains active for 365 days
Reporting of nonemployee injuries	Prevented from following vocation for more than 24 h during following day	Requires medical treatment beyond first aid

Table 5. Freight loss and damage costs reported to the AAR.

Cause	Total Freight Loss and Damage (%)*
Shortage, packed equipment	1.78
Shortage, bulk shipment	4.07
Defective or unfit equipment	3.40
Temperature failures	5.06
Delay	2.54
Robbery, theft, pilferage	5.26
Concealed damage	0.69
Error of employees	0.94
Vandalism	0.64
Fire, marine, and catastrophes	1.78
Train accident (lading only)	20.69
Miscellaneous	53.24

*1976 data.

Figure 1. Cost to society per accident as a function of the number of casualties per accident.



by local governments, and the costs of damage to structures owned by others.

The costs of freight loss and damage are especially interesting. As noted in Table 5 (27), more than one-half of the costs are attributable to miscellaneous causes. This categorization hinders the ability to relate damage to specific causal actions and also calls into question the reliability of the source data. Of the remaining cause categories, note that train accident is paramount.

GIVEN THAT COSTS OF MORTALITY, MORBIDITY, AND PROPERTY DAMAGE CAN BE EXPRESSED IN ECONOMIC TERMS, ARE THESE COSTS STRICTLY ADDITIVE?

This question can be rephrased: Does the whole equal the sum of the parts? or, Is the total cost of a specific accident equal to the sum of the costs of property damage, injuries, and fatalities associated with that accident? Current practice is to view a single accident that results in 100 deaths, for example, as just as costly as 100 separate accidents, each of which results in one death (everything else being equal).

Experience suggests that there may be an additional severity cost that is a function of the magnitude or gravity of the accident. Journalists seem to recognize this fact: Multiple-fatality accidents are much more likely to rate press coverage than single-fatality accidents, in part because their relative rarity makes them more newsworthy. The attention of legislators is also more likely to be drawn to perceived disasters and catastrophes even though aggregate losses may be no greater than that arising from a large number of relatively minor events.

This position is shown in somewhat simplified form in Figure 1. The line 0-1 represents the classical position: Total cost is the number of casualties (e.g., deaths and injuries) multiplied by the cost per casualty. The line 0-2 reflects the additional severity cost. A threshold (N_1) is indicated below which the severity cost is perceived to be negligible. Similarly, beyond N_2 the incremental severity cost is also perceived as negligible.

GIVEN THAT A SOCIETAL COSTS APPROACH TO THE VALUE OF LIFE IS INAPPROPRIATE, WHAT ARE THE RELEVANT COST ELEMENTS?

NHTSA Study

The cost elements included in the NHTSA 1975 societal cost study (18) are

1. Production losses—market, home, family, and community;
2. Medical—hospital, physician and other, coroner-medical examiner (fatalities), and rehabilitation;
3. Funeral;
4. Legal and court—tort action and accident citation;
5. Insurance administration;
6. Accident investigation;
7. Losses to others—employer and home care;
8. Vehicle damage; and
9. Traffic delay.

Jones-Lee Study

Jones-Lee, although much less specific, identifies these cost elements as follows (2):

1. Reduction in the individual's share of real resource costs occasioned by the death of others,
2. Reduction in the individual's share of the loss of net output due to the death of others, and
3. Reduction in the risk of his or her own death or that of anyone he or she cares about.

NSC Study

The position of the NSC is of special interest because of the relative influence of NSC figures among transportation planners. The NSC position was described in a recent paper (28):

> We have tried to measure the real dollars lost as the result of motor vehicle accidents. This includes: dollars that had to be spent as the result of the accident and dollar income that would not be received. This latter is seen as a reduction in contribution to the wealth of the nation using wages as a measure of the loss of productivity.

Specifically, the NSC cost elements are (a) net discounted value of future earnings and (b) medical costs (assuming 50 percent of fatals are dead on arrival), including hospital charges, doctor's costs, insurance (premiums less claims paid), and property damage (assumes one vehicle destroyed for every fatality).

The Conservation of Resources Approach

Winfrey proffered his own list of relevant cost elements (9):

1. Normal automobile use not incurred,
2. Costs and benefits of autopsies,
3. Costs and benefits of accident investigation,
4. Nonlegal expenses to fix accident responsibility,
5. Legal and court expense to fix accident responsibility,
6. Funeral costs (discounted),
7. Estate settlement,
8. Administration cost (overhead) of motor vehicle insurance in addition to cost of accident,
9. Work time lost and wages not continued,
10. Estimated future gross wage or salary income (discounted),
11. Future costs to maintain a worker in working status (discounted) (this is a reduction to cost),
12. Benefits of not working (union dues), and
13. Training of replacement employees (discounted).

IN WHAT WAYS, IF ANY, ARE THE COSTS OF RAILROAD ACCIDENTS DIFFERENT FROM THOSE EXPERIENCED ELSEWHERE, ESPECIALLY HIGHWAYS?

Employees Affected

Can the results of other investigators, working in other contexts, be applied directly to railroad accidents? The costs of fatalities and injuries described in other (nonrailroad) contexts universally assume that the individual affected is drawn from the general population. However, railroad employees represent about 6-8 percent of the total mortality and morbidity in railroad accidents and incidents (23). This proportion is not insignificant. The foregone earnings of railroad employees will influence the value of life, especially with respect to the DFE approach to valuation, because their earning patterns differ somewhat from those of the general population. Differences should also be noticeable with respect to injury valuation (e.g., medical costs and workdays lost).

Injury Severity

As illustrated in Table 6 (17), estimates of average cost per injury are dependent on the proportions of the total injured population that fall within each severity class. In the absence of comparable data for railroad injuries (i.e., the proportions of railroad injuries that fall within each severity class), it is questionable whether the same estimates are transferable to the railroad context. Railroad accidents, for example, may involve a larger (or smaller) proportion of less severe injuries than that experienced in the highway context.

Age Distribution

The DFE approach to the valuation of life depends on the age of the individual at the time of death. The distribution for motor vehicle and railroad casualties in 1976 is shown in Table 7 (30). The percentage distribution by age group is summarized in Table 8 along with comparable data from the 1975 NHTSA study.

Railroad employees killed and injured in rail accidents are clearly older than casualties in motor vehicle accidents. The total number involved in railroad accidents, employees as well as nonemployees, are also somewhat older than those involved in typical motor vehicle accidents. [Note the surprisingly high percentage (20.9 percent) of nonrail employees in the 0-4 age group who are in motor vehicles at the time of the railroad accident or incident. Only 3.9 percent of the other motor vehicle accidents are in the same age group. The difference may be explained, in part, by the large proportion of railroad accidents that occur at grade crossings in which very young children are passengers in the involved motor vehicles.]

Probability of Catastrophic Events

As discussed, there is an accident cost that is a function of the perceived overall gravity of the event. (Certainly this is evident with respect to commercial aviation accidents.) Catastrophic events are more likely to occur in rail accidents than in motor vehicle accidents. In terms of Figure 1, the critical threshold (N_1) is more likely to be surpassed in rail accidents. Thus the additional cost of severity becomes of interest.

Table 6. Frequency and cost of highway fatalities and injuries.

Injury Severity (AIS)	Number of Occurrences	Relative Frequency	Cost per Occurrence* ($)	Total Cost ($000 000 000s)
1	3 400 000	0.841	2 190	7.45
2	492 000	0.122	4 350	2.14
3	80 000	0.020	8 085	0.65
4	20 000	0.005	86 955	1.74
5	4 000	0.001	192 240	0.77
6 (fatality)	46 800	0.012	287 175	13.44
Total	4 042 800			26.19

*1975 dollars, 7 percent discount rate.

Table 7. Age distribution of highway and railroad casualties, 1976.

Age Group	Motor Vehicle Casualties*						Railroad Casualties							
									Nonrail Employees					
	Deaths		Injuries		Total		Rail Employees		In Motor Vehicle		Other		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
0-44	1 600	3.4	70 000	3.9	71 600	3.9	0	0.0	1179	20.9	1027	14.8	2 206	2.9
5-14	3 100	6.6	160 000	8.9	163 000	8.8	0	0.0	338	6.0	337	4.8	675	0.9
15-24	16 500	35.3	690 000	38.3	706 500	38.3	14 418	23.0	1632	28.9	1259	18.1	17 309	23.0
25-44	12 100	25.9	530 000	29.4	542 100	29.4	31 825	50.7	1392	24.7	1838	26.4	35 055	46.5
45-64	7 600	16.3	260 000	14.4	267 600	14.5	16 292	25.9	774	13.7	1712	24.6	18 778	24.9
65-74	3 100	6.6	60 000	3.3	63 100	3.4	90	0.1	209	3.7	493	7.1	792	1.1
75	2 700	5.3	30 000	1.7	32 700	1.8	27	0.0	114	2.0	287	4.1	428	0.6
Unknown	-	-	-	-	-	-	143	0.2	-	-	-	-	143	0.2
Total	46 700		1 800 000		1 846 700		62 795		5638		6953		75 386	

*Includes pedestrian and pedalcycle casualties.

Table 8. Percentage distribution for motor vehicle and railroad casualties by age group.

Age Group	Faigin Study (NHTSA) (%)	Motor Vehicle Casualties (%)	Railroad Casualties Employees Only (%)	Railroad Casualties Total (%)
0-14	19.5	12.7	0.0	3.8
15-24	21.0	38.3	23.0	23.0
25-44	41.3	29.4	50.7	46.5
45-64	18.6	14.5	25.9	24.9
≥65	0	5.2	0.3	1.9

SUMMARY

The costs of mortality, morbidity, and property damage should be measured in economic terms. The appropriate point of view is that of the general society with respect to those decisions requiring expenditure of public funds. However, it will be both useful and appropriate to identify also those costs to be incurred by the railroads and their employees.

Among the contending approaches to the valuation of human life, the societal costs approach appears most promising. Both the WTP and DFE perspectives should be explored over the near term to evaluate the most important cost component. Morbidity costs should be related to relative injury severity; the OAIS, as revised, shows greatest promise.

The quality of existing data bases is poor with respect to mortality and morbidity. Substantial improvements must be made before these data can be used with a reasonable degree of confidence. Current procedures for estimating the cost of property damage are poor. From a societal point of view the costs are understated.

The total cost of an accident is not equal simply to the sum of individual cost elements (i.e., mortality, morbidity, and property damage). An additional severity cost is a function (not necessarily linear) of the perceived magnitude of the accident. It is neither feasible nor desirable at this time to provide an exhaustive listing of societal cost elements. Nevertheless, as discussed in this report, an initial set of relevant components is available. It would appear that either WTP or DFE makes up the greatest part of the total.

There are are some important differences between costs of railroad and other accidents. Thus standard costs developed in other contexts should be used only with considerable care. Indeed, it would appear that railroad-specific standard costs should be developed.

ACKNOWLEDGMENT

Research leading to this paper was supported in part by a contract from Planning Systems International, Inc., Cambridge, Massuchusetts. That study, in turn, was directed by the Transportation Systems Center, U.S. Department of Transportation. This paper is an abridged version of Technical Report 79-1, Traffic Safety Center, University of Southern California, Los Angeles.

REFERENCES

1. S. E. Rhoads and M. Singer. What is Life Worth? The Public Interest, Vol. 51, Spring 1978, pp. 74-92.
2. M. Jones-Lee. The Value of Life: An Economic Analysis. Univ. of Chicago Press, Chicago, 1976.
3. M. Jones-Lee. Valuation of Reduction in Probability of Death by Road Accident. Journal of Transportation Economy and Policy, Vol. 3, Jan. 1969.
4. M. Jones-Lee. The Value of Changes in the Probability of Death or Injury. Journal of Political Economy, Vol. 82, July-Aug. 1974, pp. 835-949.
5. E. J. Mishan. Evaluation of Life and Limb: A Theoretical Approach. Journal of Political Economy, Vol. 70, July-Aug. 1971, pp. 687-705.
6. J. Carlson. Valuation of Life Saving. Howard Univ., Washington, DC, Ph.D. dissertation, 1963.
7. G. Fromm. Economic Criteria for FAA Expenditures. Federal Aviation Administration, Contract FAA BRD-355, June 1962.
8. R. Thaler and S. Rosen. The Value of Saving a Life. In Household Production and Consumption, Proc., Conference on Income and Wealth, National Bureau of Economic Research, 1975, pp. 265-298.
9. D. Ghosh, D. Lees, and W. Seal. Optimal Motorway Speed and Some Valuations of Time and Life. Manchester School of Economic and Social Studies, Vol. 43, June 1975, pp. 134-143.
10. G. Blomquist. Value of Life: Implications of Automobile Seat Belt Use. Univ. of Chicago, Ph.D. dissertation, March 1977.
11. L. Dublin and A. Lotka. The Money Value of a Man. Ronald Press, New York, 1930.
12. D. J. Reynolds. The Cost of Road Accidents. Journal of the Royal Statistical Society, Issue 119, 1956, pp. 393-408.
13. R. F. F. Dawson. Cost of Road Accidents in Great Britain. Road Research Laboratory, London, 1972.
14. E. Bollay Associates, Inc. Economic Impact of Weather Information on Aviation Operations, Volume 2. Federal Aviation Administration, Contract FAA BRD-410, Sept. 1962.
15. Cumulative Regulatory Effects on the Costs of Automotive Transportation. White House Office of Science and Technology, Feb. 1972.
16. D. Usher. An Imputation to the Measure of Economic Growth for Changes in Life Expectancy. In Measurement of Economic and Social Performance (M. Moss, ed.), National Bureau of Economic Research, 1973.
17. B. M. Faigin. 1975 Societal Costs of Motor Vehicle Accidents. National Highway Traffic Safety Administration, U.S. Department of Transportation, Dec. 1976.
18. J. D. Flora, J. Bailey, and J. O'Day. Financial Costs of Automobile Accidents. Highway Incident Texture Lab Repts. (Highway Safety Research Institute, Univ. of Michgan, Ann Arbor), Vol. 5, No. 10, June 1975.
19. Estimating the Costs of Accidents. National Safety Council, Traffic Safety Memorandum 113, July 1973.
20. Economic Analysis of the Occupant Crash Protection Standard. Office of System Analysis, National Highway Traffic Safety Administration, Staff Rept., April 1971.
21. E. Helms. Folgekosten der Strabenverkehrsunfälle 1968 nach Schadensarten. Zeitschrift für Verkehrssicherheit, Vol. 17, No. 4, 1971.
22. J. Niklas. Nutzen/Kosten-Analyse von Sicherheitsprogrammen im Beireich des Straben-verkehrs. Schriftenreihe des VDA, No. 7, 1970.
23. Federal Railroad Administration. Accident/Incident Bulletin 145. 1977.
24. The Abbreviated Injury Scale. Rev. Ed. American Medical Assn., Soc. of Automotive Engineers, and the American Assn. for Automotive Medicine, 1976.
25. E. J. Mackenzie, E. A. Garthe, and G. Gibson. Evaluating the Abbreviated Injury Scale. In

Proc., 22nd Conference of the American Association for Automotive Medicine and the 7th Conference of the International Association for Accident and Traffic Medicine, Ann Arbor, MI, July 10-14, 1978.
26. J. C. Marsh, R. J. Kaplan, and S. Kornfield. Financial Consequences of Injuries in Automotive Accidents. Highway Safety Research Institute Research Review, Vol. 9, No. 1, July-Aug. 1978.
27. An Evaluation of Railroad Safety. Office of Technology Assessment, U.S. Congress, May 1978.
28. J. L. Recht. Procedures of the National Safety Council for Estimating the Cost of a Highway Traffic Fatality. Paper presented at the 57th Annual Meeting, TRB, 1978.
29. R. Winfrey. Concepts, Principles, and Objectives of Economic Analysis Applicable to Traffic Accidents. TRB, Transportation Research Record 680, 1978, pp. 40-50.
30. Accident Facts. National Safety Council, Chicago, 1977.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Economics of a Unified Transportation Trust Fund

Gabriel Roth, Urban Projects Department, World Bank, Washington, D.C.

The paper describes the pricing and investment rules that might be appropriate to a unified transportation trust fund and suggests that they could be based on the same criteria of profitability that are used in the private sector. The consequences of applying such rules to the U.S. transport sector are explored, and it is concluded that rail passenger transport and some waterway transport could be lost, but bus transport and rail freight services could benefit. The effect on air transport would be to divert traffic from the more congested airports to less congested ones. The effect on road transport could be a substantial rise in fuel taxes, especially on diesel fuel, and in the annual registration fees payable by vehicles that impose heavy axle loads on the road system. It is concluded that, if suitable pricing and investment criteria are introduced, a unified transportation trust fund would be unnecessary; if they are not, a unified transportation trust fund could be wasteful.

The Highway Trust Fund is due to expire in 1979, and a number of proposals have been made for alternative financing mechanisms for highways and other transport modes. One such proposal is for the establishment of a unified transportation trust fund (UTTF) that would be used to finance all transport modes (1). The main purpose of this paper is to discuss the economics of a UTTF, particularly the rules that it might follow for pricing and investment.

CRITERIA FOR PRICING AND INVESTMENT RULES

A principal advantage of the UTTF, according to the Congressional Budget Office, is that it would "consolidate fiscal decisions for transportation as a whole and would permit better congressional coordination of modal financing" (1). It would also enable the U.S. Department of Transportation (DOT) "to better carry out the original purpose of integrating transportation programs" (2). Such integration implies that the same pricing and investment rules would apply to all modes supported by federal funds, so that the most economic solution can be developed for every need, irrespective of mode. Thus, a basic requirement of UTTF decision rules is that they can apply to all modes. A further requirement is that the rules should be applicable to transport activities in the private and public sectors. This is necessary to ensure that activities that can be carried out more economically in the public sector are not carried out by the private sector and vice versa.

PROFIT- AND BENEFIT-MAXIMIZING RULES

One of the main difficulties in the formulation of pricing and investment rules that would apply to all projects is that some modes, such as railroads, buses, and air carriers, provide services that are paid for by users, and investments in these modes can, in theory, be justified by the profits that they generate to the producers, without regard to the benefits enjoyed by the consumers. In a market economy, investments are typically justified in this way. On the other hand, facilities such as roads and waterways are generally regarded as free, and no charges are levied for use. Road and waterway projects are therefore generally assessed not by their profitability to their suppliers, but on the basis of cost/benefit analysis (CBA), which attempts to rank alternative schemes by comparing the benefits to society from each scheme with its costs to society. The private sector cannot function without profits and can only invest in projects that produce revenue in excess of expenditure. In contrast, the public sector can finance projects out of tax revenues and is not confined to revenue-raising projects. However, it should not be assumed that the benefits from revenue-producing projects go only to the suppliers: Laker's transatlantic air services produce substantial benefits to the consumer as well as profits to the airline.

Much of the effort that has gone into multimodal transportation planning has been directed at developing CBA to enable it to deal with revenue-producing, private sector projects, such as railroads, within the framework developed for the assessment of non-revenue-producing projects, such as roads. The method requires that total benefits to consumers, producers, and the general public be worked out for each project component and compared with the appropriate costs. The difficulty and ambiguity of such calculations enable poor projects to be justified on the basis of alleged social benefits. For example, according to Senator Domenici, the inability to measure the social demand for navigation projects leads to a "slipshod analysis of whether new projects are needed or not" (3).

This paper attempts to sketch an alternative approach: to use private-sector, profit-to-the-producer, criteria for the evaluation of social projects, such as roads. In fact, most of the problems that arise from this approach will apply to roads, which are the most important transport components provided on a social rather than a financial basis (although roads have also been provided as revenue-producing projects by both private and public enterprise).

OBJECTIVES OF A UTTF

The three most important objectives for a UTTF might be

1. In the short run, to encourage transport users to pay the costs incurred in the use and provision of existing transport facilities;
2. In the long run, to encourage renewal or expansion of transport facilities for which users are prepared to pay and the contraction of facilities for which users are not prepared to pay; and
3. At all times, to provide a financial mechanism to enable the providers of transport facilities to supply the services that transport users require and are prepared to pay for.

The provision of services for which users are not prepared to pay is not included as a UTTF objective because such provision usually implies the transfer of resources from some classes of people to others. For example, proponents of rail transport would like to see rail transit systems supported by taxes on highway use, regardless of the wishes of the highway users. But an important characteristic of the 800 or so U.S. governmental trust funds is that the money paid in is eventually expended in the interests of the contributors. Decisions that involve the transfer of resources from one group to another are essentially political and should, therefore, be dealt with by the appropriate political processes.

Basis of Pricing and Investment Rules

The conventional way of dealing with transportation pricing is to set a definition of costs and base prices on them. Typically, analysts look at different kinds of costs, such as maintenance, traffic control, signaling, and capital expenditure, and assess prices to different classes of user by a cost allocation that seems reasonable. This approach suffers from the disadvantage that there is often no unique way of allocating common costs (such as the entrance hall to an apartment block) between different users. This paper will therefore attempt an alternative approach used in competitive markets: Prices are determined not by costs, but by demand—by what the market will bear. If, at market-determined prices, an asset earns a surplus of revenue over expenditure, this is taken by the decision maker as a signal that the asset should be renewed, expanded, or duplicated. (For example, if revenues from competitively determined rail fares are just sufficient to cover crew and fuel costs, the service would be run until the rolling stock wears out; if the revenues are sufficient to cover replacement of rolling stock, the service would be continued until the track wears out; if the revenues are large enough to finance the renewal of track, the track is renewed; if a profit remains after all expenditure over and above the minimum required to attract capital into the industry, expansion of the whole system would be indicated.) If an asset makes a loss under prices determined by competitive markets, this is taken as a signal that the asset should not be renewed unless a case is made to do so for reasons not connected with transport.

The Pricing Rule

The objective of efficient transport pricing is to ensure that every user of transport facilities meets his or her share of the costs associated with use, no more and no less. Only in this way can one assure that the extra cost entailed in the production of a little more travel is balanced by the extra satisfaction obtained from it. Two separate elements comprise the costs associated with the use of transport facilities:

1. Direct costs—costs imposed as a result of resources directly consumed in making the transport facility available (e.g., wages, fuel, wear and tear, and atmospheric pollution).
2. Congestion costs—costs imposed by users on one another, when the demand for a service at the price charged exceeds the available supply. These costs arise out of scarcity, which, in principle, enables additional charges to be levied for the use of the scarce facility. In this sense, congestion costs arise because of underpricing (i.e., because sufficient rents are not charged for the use of scarce resources).

A system of efficiency prices, under which users are charged the costs that arise out of their travel, including a rent to ensure that the demand for scarce facilities is tailored to the available supply, can be called user-cost pricing. The appropriate price can be called the user-cost price (UCP).

Where there is no congestion, the UCP will consist only of direct costs (i.e., of the value of the resources directly consumed as a result of the provision of the good or service in question). Under a rational economic system, no service would be provided unless users pay at least the direct costs; for if direct costs are not explicitly met, each additional unit of service provided is more likely to reduce society's assets than to increase them. If, when direct costs are charged, the demand for the facility exceeds capacity so that potential users have to queue up, the UCP includes an additional element to balance supply and demand. The UCP is therefore not equivalent to cost in the popular sense of the word. For example, the appropriate charge for the use of a parking meter may exceed the direct cost of supervision and cash collection. But, failure to collect the scarcity rent element of the UCP for street parking would result in inefficient pricing in the sense that the pricing system would do little to allocate the limited number of parking spaces to the most urgent uses nor would it encourage the provision of off-street parking at economic prices.

Where congestion is heavy and persistent, as in city centers, the UCP congestion or rent element could be substantial. Calculations made by Mohring (4) suggest that, under the conditions prevailing on traffic arteries in Minneapolis and St. Paul, the UCP could exceed 40 cents/vehicle-km (66 cents/vehicle mile) in peak periods, and 20 cents/vehicle-km (33 cents/vehicle mile) in the off-peak. Imposition of such charges would yield substantial financial surpluses, which, according to Mohring, "suggest that road expansion might well yield substantial benefits."

The Investment Rule

It is therefore apparent that, if transport charges were based on the UCP, in the short run the total revenues earned by providers of transport facilities would not necessarily equal the total costs. Some facilities would make profits while others have losses. In the private sector of the economy, investment resources tend to flow to those industries and uses that make profits and to avoid those that have losses. Investment in profitable industries tends to increase capacity and reduce the rates of profit; however, disinvestment in loss-making industries tends to cut out their least-profitable elements and thus increase profits. In theory, this process continues until all industries yield a normal return on investment. If these investment rules were applied to transport facilities, given constant returns to scale, they too would be expanded or contracted until each element yielded normal profits. Furthermore, Mohring (4) and others have demonstrated that, given constant returns to scale, each component of a cost-minimizing transportation system would stand on its own feet, and such a system would require neither subsidies from society at large nor from one mode to another. Thus, imposition of user cost pricing would not only induce travelers to select travel modes that minimize total travel costs but would also generate funds required for capital investment.

From the national point of view, an initial assumption of constant return to scale is unlikely to lead to serious error. Where scale economies or diseconomies are shown to be important, the investment rule can, if appropriate, be adjusted so as to encourage investment where there are systemwide economies of scale and to discourage it where increased size leads to external diseconomies.

Economic efficiency does not require only that a complete transport enterprise should be profitable. In theory, each and every segment should, at equilibrium, earn normal profits (5). For the efficient or equitable allocation of resources, there is no special virtue in users of one bus line subsidizing the users of another or in off-peak passengers subsidizing peak users. But there are practical limits to the extent to which it is possible to vary prices, in the private as well as in the public sector.

APPLICATION TO UNITED STATES TRANSPORT SECTOR

Before these ideas are applied to selected elements of the United States transport sector, some preliminary general points should be made:

1. The requirement that direct costs be paid by users implies that the UTTF would be primarily concerned with financing infrastructure. The payment of direct costs of transport services by users should, by definition, provide adequate revenues to finance the operating and maintenance costs for all modes. The prime function of the UTTF should be the collection of charges to finance the roads, railroads, and other indivisibles of the transport system. It could do so by combining the criterion of financial profitability with the imposition of prices for the use of infrastructure as close as practicable to the UCPs.
2. To avoid waste and misallocation of resources, all significant elements have to be debited to the project under consideration and valued at the highest prices obtainable in alternative uses. For example, this point applies to land and to the use of government personnel, such as members of the Army Corps of Engineers.
3. The application of the UCP is of special interest in cities, where it would involve the imposition of additional charges for the use of congested roads. The economic, technical, and political problems of imposing such charges have been discussed extensively elsewhere (6); therefore, this paper will confine itself to the interurban elements of the United States transport sector. However, the pricing and investment framework described here for application outside United States cities would be entirely consistent with the application of UCPs within them.

Railroads

On economic and financial grounds, a railroad can charge what the market will bear and base its investment program on the replacement or duplication of assets that earn an acceptable profit under a system of market-determined prices. Joy, when chief economist to the British Railways Board, asserted that British railways followed just such a market-based pricing policy (7):

> In future, investment will be made only in assets which convey existing traffics at a long-run marginal cost which is covered by their respective revenues, or in assets for new traffics which meet the same criterion. The use of market-based prices will provide a clear indication of the opportunities for profitable investment in replacement or capacity-increasing assets.

Such a strategy could only be considered by a rail system, such as British Railways, that is free of economic regulation and able to decide what to carry, how to carry it, and at what price. Freedom from economic regulation for U.S. railroads would require major changes in the powers and activities of the Interstate Commerce Commission (ICC).

The requirement that all direct costs be covered is likely to endanger the future of intercity rail passenger service, which is now provided almost exclusively by the National Railroad Passenger Corporation (Amtrak). Amtrak's ridership has risen by an average of 8 percent/year since its establishment in May 1971, but the taxpayer subsidy has risen from $40 million to $500 million/year and is estimated to reach $1 billion/year by 1984 (8). None of Amtrak's 41 routes covers its operating costs, but its services weaken those provided (on an economic basis) by the intercity bus industry. The technical characteristics of a train make it slower than an airplane and costlier than a bus; therefore, it is difficult to see an economic future for passenger train services in the United States.

On the other hand, some rail freight services, for example in the Southern, Union Pacific, and Santa Fe Railroads, generate revenues that are reported to exceed total costs. Application of the proposed criteria, coupled with freedom from ICC regulation, would enable such companies to expand profitable services and phase out the unprofitable ones.

Air Transport

Commercial air services in the United States have some flexibility in setting their rates with the objectives of filling their seats and covering their costs. They also have the option of varying the frequency of their services. The industry recovers 99 percent of its costs from fares (9); the balance is accounted for by a portion of the Federal Aviation Administration's (FAA's) expenditure and a 5 percent subsidy to local-services air carriers (8). The FAA finances air traffic control facilities in the United States, for which it gets reimbursed by an 8 percent tax on air tickets paid into the Airport and Airways Trust Fund. Neither the FAA nor individual airports levy increased landing charges at

Table 1. Disbursement for U.S. highways in 1976.

Expense	Disbursements ($000 000 000s) Federal	State	Total Federal and State	County and Town-ship	Munici-pality	Total County, Township, and Municipality	Total
Capital outlays	0.375	10.580	10.955	1.470	1.870	3.340	14.295
Interest on debt	-	0.896	0.896	0.093	0.235	0.328	1.224
Debt retirement	-	0.920	0.920	0.170	0.440	0.610	1.530
Total	0.375	12.396	12.771	1.773	2.545	4.278	17.049
Operating expenses							
Maintenance	0.098	3.165	3.263	2.520	2.100	4.620	7.883
Administration	0.260	1.237	1.497	0.350	0.370	0.720	2.217
Law enforcement	-	1.424	1.424	0.205	1.000	1.205	2.629
Total	0.358	5.826	6.184	3.075	3.470	6.545	12.729
Total	0.733	18.222	18.955	4.808	6.015	10.823	29.778

hours of peak congestion to ration demand and finance expansion. For this reason, and also because of the undercharging of general aviation (e.g., noncommercial flying), American airports could benefit from applying user-cost pricing to their operations. This would have the effect of diverting traffic from congested airports to uncongested ones.

Water Transport

The inland waterborne transportation industry has been characterized by heavy governmental expenditure and the absence of waterway user charges, in accordance with the principle that "navigable water . . . shall be . . . forever free . . . without any tax, impost, or duty" (Northwest Ordinance, Art. 4, 1787). In the fiscal years 1965-1974, more than $3.2 billion (current dollars) was spent by the U.S. Army Corps of Engineers on the construction of shallow-draft navigation projects. State and local governments provided lands, rights-of-way, and river-port facilities, such as docks, warehouses, and elevators. A DOT study calculated that, if present trends continue, the U.S. taxpayer would pay $1 billion in 1990 to enable the domestic marine industry to increase its revenues by $21 million and to reduce the revenues of other transport modes by $60 million (10, p. 119).

The effect of subsidized waterway rates on other transport modes illustrates the need to apply uniform pricing and regulatory principles to all transport modes (11).

Railroads are frequently prevented from raising rates on "captive" traffic to a point where they can earn a reasonable return on total investment on the grounds that such rates would be unreasonably high and therefore unlawful. . . . Water carriers raise or lower rates in response to market conditions; railroads are compelled by law to provide service only at published prices. This amounts to forcing a railroad to give traffic to its competitor, who responds to market conditions in a way rail carriers cannot respond and receives subsidies rail carriers do not receive. . . . User charges and increased rail pricing freedom would lower transportation costs to society and encourage a more equitable distribution of resources within the transportation system.

Intercity Buses

According to DOT, "The bus industry is unique because transportation is provided by private companies which receive neither direct subsidies nor tax exemptions" (10, p. 151). Intercity buses pay federal, state, and local taxes that exceed by more than 25 percent their fair share for the use of public roads (10, p. 166). They serve 15 000 places, provide their own terminals, carry more passengers (340 million in 1976, compared to 220 million carried by air and 18 million by Amtrak) than any other mode at the lowest cost [3 cents/passenger-km (5 cents/passenger mile) compared to 5 cents/passenger-km (8 cents/passenger mile) by air and 9 cents/passenger-km (15 cents/passenger mile) by Amtrak], and have the lowest fuel consumption [(30.6 passenger-km/L (116 passenger miles/gal) in 1976 compared to Amtrak's 11.6 passenger-km/L (44 passenger miles/gal)] (12).

Despite its many advantages, the intercity bus is losing ground to more costly and speedier modes, such as air and private automobile transport. It also has to compete with the heavily subsidized Amtrak rail services. The typical Amtrak trip of 364 km (226 miles) costs Amtrak $44. Of this, the passenger pays $16, and the taxpayers pay $28. But, the same journey by bus would cost the passenger $17 and the taxpayer nothing (8). Taxpayers are thus made to pay Amtrak a subsidy that enables it to undercut a more efficient competitor in the same way that the provision of free waterways enables the water carriers to undercut some railroad freight operations.

The introduction of user-cost pricing for all transport modes would have little direct effect on intercity bus services, as they already base their pricing and investment policies on commercial principles consistent with user-cost pricing. The indirect results could be considerable: user-cost pricing could (a) increase the costs of operating private automobiles and (b) reduce the services of Amtrak. Intercity buses would stand to benefit from both effects.

Roads

Figures published by the Federal Highway Administration (13) form the basis of Table 1. The table shows that, in 1976, federal and state expenditure on roads (i.e., all expenditure other than by counties, townships, and municipalities) amounted to about $19 billion, of which about one-third was spent on maintenance, police, and administration of the current system to make it available for public use; two-thirds were expended on constructing and expanding the network. The table below breaks down the expenses by highway length (1 vehicle-km = 0.62 vehicle mile).

U.S. Highway	Federal and State	County, Township, and Municipality	Total
Vehicle kilometers (000 000 000s)	1015.23	1252.58	2267.81
Expenditure per vehicle kilometer ($)			
Capital	1.26	0.34	0.75
Operating	0.61	0.52	0.56
Total	1.87	0.86	1.31

These total outlays, when divided by the 1015 billion vehicle-km (631 billion vehicle miles) of estimated rural travel, amount to 1.9 cents/vehicle-km (3 cents/vehicle mile); 0.6 cent/vehicle-km (1 cent/vehicle mile) to operate the existing system and 1.3 cents/vehicle-km (2 cents/vehicle mile) to construct and expand it. If all roads (urban and rural) and all travel in the United States were considered, the expenditure per unit travel would average 0.6 cent/vehicle-km (1 cent/vehicle mile) on operation of the existing system and 0.7 cent/vehicle-km (1.2 cents/vehicle mile) on construction and expansion of it.

Assuming an average fuel consumption of 6.38 km/L (15 miles/gal), these figures suggest that a tax of 4 cents/L (15 cents/gal) [equivalent to 0.6 cent/vehicle-km (1 cent/vehicle mile)] would be roughly equivalent to the element of the UCP payable to administer and operate the existing highway system. This tax would be somewhat higher than existing fuel taxes, which consist of the 1.1 cents/L (4 cents/gal) federal tax and state taxes ranging from 1.3 to 2.9 cents/L (5 to 11 cents/gal), but an increase in fuel tax would be consistent with the administration's energy policy. The tax on diesel fuel would have to be substantially higher than that on gasoline if it is to be used as a method of charging for roads, because of the low fuel consumption of diesel-engined vehicles.

The costs of constructing and expanding the road system, assuming for the moment that 1976 expenditure was optimal, would have required at least a further 5.3 cents/L (20 cents/gal) tax on rural road use, if the fuel tax were considered an appropriate source of funds for capital investment. However, because a considerable proportion of highway construction costs is due to the effects of heavily loaded axles, a tax on commercial vehicles would seem a more appropriate source of funds, particularly if it could be varied in proportion to pavement damage caused, which is reckoned to be proportional to the fourth power of the axle load (e.g., a 4-Mg axle load damages a road 16 times as much as does a 2-Mg axle load). If all the $17 049 billion of capital expenditure were charged to the nation's 1976 population of 28 197 900 commercial vehicles (13), the annual tax on each truck and bus would average $605; if the capital expenditure were charged entirely as an annual tax payable by the nation's total 1976 population of 143 538 500 vehicles (13), the average annual tax per vehicle would be $119.

The purpose of these arithmetic exercises is not to recommend a particular combination of taxes but to indicate that there appears to be no insuperable difficulty in devising a tax structure that would enable all road users to be charged the total cost of U.S. roads. However, two problems remain:

1. Any charging system that relies on taxes on fuel and vehicle parts, supplemented by annual registration fees, would involve a considerable degree of averaging and would not meet the test of being a market-determined price, based on what the market would bear.
2. The use of such taxes would tell us nothing about the optimal size of road networks, nor whether they should be expanded or contracted, as the profits from any road system could be arbitrarily increased by the taxes.

Tolls can, and are, being used to charge for many roads (particularly for costly sections such as bridges), and the toll-road industry is, in fact, developing new pricing methods to enable charges to be assessed against moving vehicles (6, pp. 15-20). However, a considerable amount of averaging has to be accepted as a fact of life. The diseconomies of averaging are particularly evident in the absence of charges for the use of congested urban streets, but the absence of road pricing need not prevent improvements in the financing of other elements of the transport system. Tax rates on road use already vary from state to state.

Extremes of underinvestment in road networks can be avoided by allowing private suppliers to build new road sections and to be reimbursed from the fuel and other taxes earned on their roads. The appropriate amounts could be determined by traffic counts and could replace or supplement tolls. Overinvestment would be more difficult to deal with, particularly if associated with high taxes and poor planning, but a vigilant and educated electorate would tend to exert its influence to cut taxes and improve planning.

OTHER SOURCES OF FINANCE

The UTTF need not, of course, be the sole source of finance for transport infrastructure. There would always be room for grants from public or other agencies to finance unprofitable services. But such grants should be deliberately and specifically voted by appropriate political levels. There would be no advantages (and many disadvantages) in giving the UTTF powers to switch funds from profit-making to loss-making concerns. There is no reason in equity why users of profitable transport services should be made to subsidize unprofitable ones.

Nor is there any reason for the UTTF to monopolize the financing of profitable transport projects. Other public or private agencies could be allowed (even encouraged) to finance and operate roads, railways, and ports, but it would be desirable that similar pricing and investment rules be used throughout the U.S. transport sector.

CONCLUSION

Given the sources of funds described and clear pricing and investment rules, it is possible to envisage a UTTF that would collect revenues that broadly reflect economic user charges in the transport sector and use the revenues to provide any element of transport infrastructure that is likely to be profitable in the financial sense. However, if the recommended pricing and investment rules were followed, there would be no obvious advantage of a UTTF over the existing funding methods. On the other hand, a UTTF not bound by strict investment and pricing criteria would have considerable potential for the misallocation of resources that are scarce. On balance, there seems to be a stronger case for the adoption of investment standards or criteria for publicly financed projects than for the establishment of a UTTF.

ACKNOWLEDGMENT

Thanks are due to Anthony Churchill, Ron Cooke, Charles Dale, Damian Kulash, Herbert Mohring, Ali Sevin, David Starkie, Floyd Thiel, and Alan Walters for their comments and suggestions, and to Betty Easter for editing the draft, but neither they, nor the World Bank, necessarily endorse the assumptions, data, or conclusions presented in this paper.

REFERENCES

1. Transportation Finance: Choices in a Period of Change. Congressional Budget Office, March 1978.
2. B. Adams. Transportation Policy for a Changing

America. U.S. Department of Transportation, Feb. 6, 1978.
3. Congressional Record, Feb. 24, 1977, p. S2965.
4. H. Mohring. The Benefits of Reserved Bus Lanes, Mass Transit Subsidies, and Marginal Cost Pricing in Alleviating Traffic Congestion. In Issues in Urban Economics (P. Mieszkowski and M. Straszheim, eds.), Johns Hopkins Univ. Press, Baltimore, 1979.
5. R. H. Strotz. Principles of Urban Transportation Pricing. HRB, Highway Research Record 47, 1964, pp. 113-121.
6. Problems in Implementing Roadway Pricing. TRB, Transportation Research Record 494, 1974, 35 pp.
7. S. Joy. Pricing and Investment in Railway Freight Services. Journal of Transport Economics and Policy, Vol. 5, No. 3, Sept. 1971, pp. 231-246.
8. J. C. Miller. What's To Be Done About Amtrak? Wall Street Journal, Eastern Ed., Aug. 25, 1978, p. 8, column 4.
9. Financing Waterway Development: The User Charge Debate. Congressional Budget Office, July 1977.
10. Trends and Choices. U.S. Department of Transportation, Jan. 1977. NTIS: PB 282 230.
11. E. B. Hymson. Analysis of Rail-Water Price Competition. TRB, Transportation Research Record 635, 1977, pp. 6-11.
12. N. Bade. Intercity Bus Transportation: A Changing Market. Paper presented at 57th Annual Meeting, TRB, 1978.
13. Selected Highway Statistics, 1976. Federal Highway Administration, 1978.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Institutional Factors in the Implementation of Automobile-Restrictive Measures

Part 1: Implementation Experience with Transportation Air Quality Measures in the Denver, Colorado, Urban Area

Jack Kinstlinger, Colorado Department of Highways, Denver

In recent years, Denver's high altitude, topography, rapid growth, and heavy reliance on the automobile have combined to cause a severe air pollution problem. According to the Colorado Air Pollution Control Commission, the principal cause of the pollution is the use of motor vehicles. The Denver region developed an air quality plan that was submitted to the U.S. Environmental Protection Agency as part of the state implementation plan for air quality. The Denver element of the plan relies on strategies that reduce emissions at the tailpipe rather than strategies to restrict automobile use. Several institutional and attitudinal factors played a role in determining that automobile-restriction measures were not acceptable: (a) the no-problem syndrome, (b) the no-solution syndrome, (c) lack of public acceptance, (d) possibility of unequal burdens, (e) changing economic impacts, (f) agency priorities, and (g) difficulty in resolving conflicts. As the Denver region moves from planning to implementation of air quality strategies, it will be important for the state to transcend parochial political interests and take the difficult stands necessary. The state must also be careful not to make decisions in a vacuum. Ascertaining the public's opinion on air quality strategies will be critical to their successful planning and implementation.

Denver is known for its attractive environment and healthy climate. In recent years, however, Denver's high altitude, topography, rapid growth, and heavy reliance on the automobile have combined to give the "Mile-High City" a severe air pollution problem. The Denver region is currently experiencing frequent violations of the National Ambient Air Quality Standards (NAAQS) for carbon monoxide, ozone, nitrogen dioxide, and suspended particulate matter. For example, in 1977 the second highest recorded 8-h average carbon monoxide concentration was 120 percent greater than the 8-h carbon monoxide standard [22.8 mg/m^3 (19.8 ppm) versus 10.4 mg/m^3 (9 ppm)].

According to the Colorado Air Pollution Control Commission (APCC) the principal cause of the pollutants is the use of motor vehicles. The commission estimates that vehicular sources account for 93 percent of the carbon monoxide emissions, 85 percent of hydrocarbon emissions (which are a primary precursor to ozone), 75 percent of particulate emissions, and 37 percent of nitrogen oxides. For three of the four standards, automobile use is the primary cause of the violations (1, pp. 11-20).

States in which there are areas that do not now meet the NAAQS must prepare revised state implementation plans (SIPs) that will ensure compliance with the air quality standards by December 31, 1982. Under certain circumstances, attainment of the standards for carbon monoxide and ozone may be extended to December 31, 1987. These revised SIPs must be submitted to the U.S. Environmental Protection Agency (EPA) by January 1, 1979. If the plans do not demonstrate the required compliance to EPA's satisfaction, severe sanctions on federally funded highway, sewer, and other construction can be imposed on the states and local governments.

DEVELOPMENT OF THE COLORADO SIP

Final responsibility for development of the Colorado plan rested with the APCC, an independent body appointed by the governor with the consent of the senate.

In order to prepare the statewide plan, the APCC considered the plans submitted by Denver and the four other nonattainment areas.

The Denver Regional Council of Governments (DRCOG) was designated by the governor as the principal participating agency for preparation of a proposed Denver regional element of the SIP. The DRCOG established two key working committees to prepare a draft plan. These were the Clean Air Task Force, which consisted of representatives of key interest groups in the Denver regional community and whose membership was jointly appointed by the DRCOG chairman and the governor and the Air Quality Policy Committee, which was composed of an equal number of voting representatives from the DRCOG and the Colorado APCC. The policy committee also had one nonvoting member from the Colorado Highway Commission and the Board of the Regional Transportation District, the local transit operating agency. The Clean Air Task Force and the Air Quality Policy Committee met regularly from July to November to prepare a draft proposed air quality plan for submittal to the DRCOG (2, p. 34). On November 17 and 18 the APCC held public hearings on the SIP, which included the Denver element. After adoption by the state APCC, the governor submitted the SIP to the EPA on January 2, 1979.

RECOMMENDATIONS OF THE DENVER PLAN

The Denver plan relies almost entirely on strategies to reduce automobile emissions at the tailpipe and strategies that encourage the voluntary use of modes of transportation other than the single-occupant automobile. For example, the plan asks the state legislature to expand the automotive inspection and maintenance program to include 1968 and later vehicles instead of 1977 and newer automobiles. Other strategies include a doubling of transit ridership by 1982, a vanpool demonstration program, a regional bicycle plan, expansion of the carpool matching service, studies of a regional traffic signal control system and high-occupancy vehicle (HOV) treatments, and the preparation of a regional parking management plan "designed to complement other transportation control strategies. . . ." (2, p. 34).

All in all it is a modest, but perhaps realistic, plan. According to the preliminary assessment conducted by the APCC, the proposed measures will not permit attainment of the standards for carbon monoxide and ozone by 1982 nor for ozone by 1987, unless ozone standards are liberalized. In other words, Denver has come up with a plan that will not make compliance with the national standards possible, either by 1982 or 1987.

Some members of the Clean Air Task Force and the Air Quality Policy Committee issued a minority report that makes more stringent recommendations. The minority report supports the majority's inspection and maintenance proposal but otherwise characterizes the adopted plan as being composed chiefly of preexisting highway and transit plans, of calling for innumerable studies, and of relying on voluntary action. The minority plan, among other things, recommends three strategies to limit automobile use:

1. Use of private automobiles restricted one day a week, through a mandatory program to begin January 1, 1982, if voluntary efforts are inadequate (this was formally included in the SIP just prior to adoption and submittal by the governor);
2. Parking management plans to be developed by all local governments that must consider numerous strategies, including a moratorium on all new parking facilities, a surtax on all parking in the region, and preferential parking for HOVs; and
3. Development of automobile-restricted zones in addition to the one planned for the central business district.

Many factors, of course, played a part in the decision not to include automobile restrictions in the majority plan. Included are the institutional and attitudinal factors listed below.

The No-Problem Syndrome

Because of the difficulty in determining air pollution's causative role in diseases or disorders, it is asserted by some interests that Denver's air quality situation is an aesthetic, but not a public health, problem. It is then argued, quite logically, that if there is no problem, no remedial steps are necessary.

The No-Solution Syndrome

Although they concede that there is a problem, some opponents to automobile restrictions believe constraints on automobile use offer no solution. For example, this group theorizes that proposals to limit downtown parking would force shoppers to shop in the suburbs and encourage businesses to relocate to the suburbs. Such changes, it is said, would increase vehicle travel distances and, consequently, air pollution because of the greater distances to be driven and the decreased access of public transit in the suburbs. Also, they insist that restrictions should not be considered unless there is absolute proof that they will solve the problem. Such proof clearly is not feasible.

To complicate things even further, a recent study by DRCOG on the impact of air quality from changes in land-use and transportation patterns showed some interesting results. Changes in development patterns, densities, and modal splits between automobiles and transit showed insignificant changes in carbon monoxide levels. The inability of the predictive models to relate development densities with travel behavior may account for these unexpected conclusions. For ozone, a more pernicious pollutant in Denver, only an increase in the highway level of service showed an improvement leading to near attainment of the proposed 195 $\mu g/m^3$ (0.1 ppm) standard by the year 2000. This is due to the fact that ozone pollution, which results from mixing hydrocarbons and oxides of nitrogen in sunlight, increases with congestion and resulting speed reductions, remains essentially constant with decreases in vehicle kilometers of travel because reduction in vehicle kilometers of travel results in a constant balance of hydrocarbons and nitrogen oxides, and is reduced only with increases in traffic speed.

Lack of Public Acceptance

Numerous surveys of residents of the Denver metropolitan area indicate a high level of concern about the region's air pollution problem. A recent survey commissioned by the APCC revealed that air pollution was believed to be the area's most serious problem. On the other hand, this and other surveys also reveal an unwillingness to alter automobile driving habits. This uncertainty about the public's reaction to automobile-restriction measures has been a significant factor in the reluctance of elected officials to enact stringent air pollution measures. If a clear mandate to clean the air and to change life-styles were evident to the local and state legislators, they would pass the appropriate legislation.

Possibility of Unequal Burdens

Referring to the concept of a no-drive day, one day a week, the draft Denver element of the SIP depicts one of the disadvantages (2, p. 34): "...(it) may place serious burdens on those for whom there is no practicable alternative to the automobile, especially one-car households."

Inspection-maintenance programs and increased parking rates are other strategies that may affect most acutely those who are least able to afford the sanctions.

Changing Economic Impacts

One of the major factors in resisting automobile constraints is the fear of economic dislocations. Many communities, especially those that represent the core city areas, are concerned that transportation controls may exacerbate the problems of inner-city stagnation; they are concerned that the attractiveness of the city vis a vis the suburbs for working and living will be tipped in favor of the suburbs. There also exists a fear that, if the Denver region implements automobile-use restrictions, it may be at a disadvantage when competing for new businesses with cities that have less severe air pollution problems and hence less stringent automobile restraints. Continued air pollution could also place the Denver area at such a disadvantage. A study recently completed by Cambridge Systematics reported (3, pp. 1-36):

> In addition to their intended results, air quality transportation measures may change the competitive position of one area relative to other locations in the metropolitan region. Although they are unplanned, these secondary economic impacts are not unimportant. If they are adverse, they can impose economic hardship, producing a loss of retail sales, a contraction of job opportunities, vacancies in buildings, and declines in the value of property as economic activity shifts to other parts of the region. Fear of such losses lies behind the common opposition of the downtown business community to transportation controls and associated policies.

Agency Priorities

Each agency involved with transportation in Denver has its own agenda consistent with its particular mission. Even though most of these agencies are pursuing programs that are compatible with clean air goals, these goals are rarely a high priority within each agency. In order for air quality programs to receive a higher priority, each agency will have to reallocate its financial and staffing resources. Most entities are unwilling to do this when such programs are not synonymous with their highest priorities. No agency, with the possible exception of the APCC, has moved automobile-restriction measures to a position of high priority. Hence, there is little institutional push for these measures.

The priorities and activities of several agencies illustrate this problem. For example, the 1978-1982 transportation improvement program of DRCOG describes potential highway projects as being evaluated by a volume-to-capacity ratio, a hazard index, and a sufficiency rating. Air pollution mitigation factors are added on as one of the criteria, but not the primary one (4, p. 27).

The Denver Regional Transportation District feels that all of its actions improve air quality and, therefore, tends to emphasize objectives that indirectly affect air quality, such as fleet modernization and expansion, efficient use of road space, and park-and-ride facilities.

Indicative of the problem of priority that faces local traffic engineering divisions is the situation of the Denver Department of Public Works. Long known for its innovative approaches to improving vehicular flow and access to the central business district, this division is understandably reluctant to implement strategies that are contrary to these objectives.

The Colorado Department of Highways is not immune from the conflict of priorities. Since the air quality problem is so closely tied to transportation, many agencies, interest groups, and individuals look to the department to implement and influence the implementation of many SIP strategies. Often the expectations are beyond financial or legislative limits placed on the department. The department's primary mission is to build, operate, and regulate an effective and safe highway transportation system. Efficiency, economy, and safety, not cleaner air, are the forces that motivate the development and programming of highway actions. On the other hand, each major highway project is subject to an environmental assessment in which air quality impacts are calculated and compared to federal ambient standards and requirements within the SIP for air quality. The assessment is also reviewed by the Colorado Department of Health and appropriate federal agencies. On this basis, air quality does not drive the highway program but rather is a check on the development of a particular project.

In recent years, however, the highway department has become much more sensitive to the issue of air quality and energy conservation and has programmed and expanded several million dollars of Interstate highway funds for park-and-ride lots operated by Denver's Regional Transportation District and has made state and federal highway funds available for carpooling and vanpooling activities as well as bikeway construction.

Some clean air activists have charged that these steps are not sufficient and that the highway department as well as local government should formulate transportation improvement projects, highway and transit, with the express purpose of improving air quality with other transportation objectives that receive less emphasis and should avoid adding to the highway capacity in the Denver area. This is an unreasonable challenge to transportation planners since it is contrary to the mission and legal charge of transportation agencies. In many instances, such an objective could lead to the closing of major streets to vehicular traffic or the indiscriminate designation of existing lanes for HOVs. The closure of state highways that are major arterials for leading traffic to and through communities would be a disservice to the traveling public. The closure of local city streets, however, is more likely to be accepted and is being considered by the city of Denver and the Regional Transportation District. Dedication of new lanes for HOVs is being considered but with much greater caution. The experience on the Santa Monica Freeway in Los Angeles has put all transportation professionals on their guard. Preliminary studies have indicated that the total breakdown of traffic flow in the non-HOV lanes would clearly be unacceptable to the public in terms of accidents and congestion and would also be damaging to efforts to clean the air. The department supports careful consideration of other automobile disincentives, including no-drive days on a voluntary basis, judicious use of parking charges to discourage excessive automobile use, and more compact mixed land-use developments where it can be clearly shown that single-occupant automobile use will be reduced without disproportionate degradation of social, economic, or environmental conditions.

Denver is experiencing a 7 percent annual growth in vehicle kilometers of travel, and its buses are filled to capacity during peak periods. On this basis a deliberate policy of no expansion of highway capacity, as has been proposed by many air quality activists, would be unacceptable. One exception is along corridors that lead to the center city of Denver, where bus line-haul capacity can more readily be expanded and where downtown congestion and excessive dedication of scarce land for parking purposes make an increase of automobile traffic particularly undesirable. In fact, the DRCOG's Transportation Committee, at the highway department's urging, has adopted a policy of giving transit expansion preference over highway expansion along radial corridors that lead to the center city of Denver.

Difficulty in Resolving Conflicts

As the agency responsible for developing the Denver element of the SIP, the DRCOG is in a sensitive position. DRCOG receives funding support from local units of government—both suburban and urban. The problem of changing economic impacts is essentially a problem of urban versus suburban fortunes. Therefore, in order not to jeopardize support from either faction, the DRCOG tends to avoid those issues of economic impacts. Another element in DRCOG's fragility is the fact that its governing board consists of local elected officials. This dilemma is characterized in the Cambridge Systematics study as follows (3, pp. 1-36):

> The Denver Regional Council of Governments, recently designated as Denver's metropolitan planning organization, has been reluctant to become actively involved in either the planning or implementation of air quality measures, in large part because of their controversial nature. Because of its voluntary membership and lack of independent funding sources, DRCOG must be very sensitive to the views of its diversified member jurisdictions. As a result, DRCOG has difficulty in making policy choices on controversial issues, and the policies which it adopts are moderate enough to satisfy the majority of its members.

Shifting economic balances is not the only controversy that can complicate DRCOG's role. Only two weeks after the Denver element was officially adopted, the city of Boulder, one of the members of DRCOG, considered a move to withdraw from the council of governments. Boulder's stated reason for such a consideration is the fact that the region could lose some of its federal funds because of the Denver element's inadequacy in meeting federal criteria. Boulder's air is cleaner than Denver's, and it might be penalized even though it meets the air quality standards.

CONCLUSION

Of the factors that I have cursorily examined, the two most critical are the fears of changing economic impacts and the difficulty of regional councils of government in resolving conflicts. In fact, these problems probably underlie many of the other factors. It may also be true that some of the other concerns are put forth to divert our attention from the real issues.

As we move from planning to implementation of transportation control strategies, I think the lessons provided in the development of the SIP will be valuable. It is imperative to devise strategies that will not adversely affect the economy of one area to the benefit of another.

The state must play its part in addressing the issue. The state should have the expertise, authority, and broad perspective to transcend parochial political interests to take the difficult stands necessary. In this role, however, the state must be careful to avoid making decisions in a vacuum. Previously, the APCC developed regulations that did not sufficiently incorporate the other actors into the process and the regulations were not well received or enforced. The state needs the advice and support of local governments, the state legislature, and the public at large in order to select strategies that will be workable.

Finally, and perhaps most importantly, it is essential to ascertain the opinion of the public. In the discussions that occurred during the preparation of the Denver element, a recurring argument was that a given strategy would not be acceptable to the public. When the public hearings were held, however, many observers were surprised at the support voiced for stronger measures and the willingness of citizens to change their life-styles. Such demonstrations of public will are necessary to influence the APCC in its drafting of the final SIP, to encourage the state legislature to adopt strict laws, and to convince public agencies to rearrange their priorities. But are those who show up at meetings truly representatives of the public? Perhaps not, but regardless, they are the ones who showed concern and interest and, therefore, their views should guide officials.

Overall, the Denver process is workable. A broad range of interests and perspectives was brought together under difficult time constraints and a modest and realistic plan was produced. It will be interesting to follow the machinations of the institutions once the implementation of the programs begins.

ACKNOWLEDGMENT

Special thanks to Greg Henk and Mike Huffaker of the Transportation Planning Division of the Colorado Department of Highways for their assistance in preparing this paper. The contents reflect my views and do not necessarily represent those of the Colorado Highway Commission.

REFERENCES

1. Summary: Colorado Revised State Implementation Plan for Air Quality. Colorado Air Pollution Control Commission, Dec. 1978.
2. Draft Plan for the Denver Regional Element of the State Air Quality Implementation Plan. Denver Regional Council of Governments and Colorado Air Pollution Control Commission, Sept. 28, 1978.
3. Cambridge Systematics, Inc. Implementation and Administration of Air Quality Transportation Controls: An Analysis of the Denver, Colorado, Area—Executive Summary. U.S. Department of Transportation, Rept. DOT-P-78-00-2, April 1978.
4. Transportation Improvement Program 1978(79)-1982(83). Denver Regional Council of Governments, Jan. 1978.

Part 2: Traffic Restraints in the San Francisco Bay Area

Elizabeth Deakin, Cambridge Systematics, Inc., Berkeley, California

Implementation experience with five different transportation system management measures in the San Francisco Bay Area is examined: neighborhood traffic restrictions, road and congestion pricing, toll increases, parking management, and ramp metering. The conclusions are that proposals to restrain automobile use have had little success when proposed by planning agencies, but when implemented at the initiative of local government, they have gained substantial (if not unanimous) acceptance. Although automobile-restraint measures may have adverse differential community impacts, it should not be assumed that automobile restraints are always unacceptable to the public and that it will not be possible to develop the community support necessary to succeed.

Automobile-restrictive measures have received increased attention at both the local and the federal levels of government. From the local perspective, restraints on the use of automobiles have most frequently been suggested when congestion, parking shortages, noise, and neighborhood disruption are problems. Federal interest has grown out of two separate but complementary initiatives: (a) the joint Urban Mass Transportation Administration-Federal Highway Administration (UMTA-FHWA) requirement for transportation system management (TSM), which aims to bring about a more balanced and efficient use of the private automobile as well as more effective utilization of existing transportation facilities and services (1), and (b) the requirement of the Clean Air Act Amendment of 1977 for the implementation of transportation measures that improve air quality, as part of a state implementation plan (SIP) for areas that cannot otherwise meet the national ambient air quality standards.

Although TSM, transportation-air-quality measures, and local traffic restraints can be compatible, the three have not often converged in joint planning and implementation efforts. In the Bay Area, TSM activities are performed by transit operators, the California Department of Transportation (Caltrans), and the county transportation organizations, in cooperation with the Metropolitan Transportation Commission, the metropolitan planning organization (MPO). Local proposals for street and highway operational improvements are integrated into the TSM or transportation improvement program (TIP) process. Less traditional initiatives, such as automobile restrictions, are only occasionally proposed in that forum, at least partly because the only regular source of funding for city-level innovations is the state Transportation Development Act (California, 1978), which supports community transit and paratransit activities. So far, transportation-air-quality planning has been conducted in a separate process from the more conventional transportation activities, and, although the cities have been represented in the process and citizen participation forums have been provided, discussions have not yet focused on the implementation of specific local projects. Meanwhile, proposals for automobile restraints emerge periodically at the city or neighborhood level and are planned and evaluated in relative isolation from the regional transportation activities.

The separation of TSM, transportation-air-quality, and local automobile-restriction planning activities is partly a reflection of the separate motivations that underlie the three efforts. However, it also points out some ways in which transportation organizations have not fully adapted to the implementation process required for small-scale, local-impact projects. This paper reviews several recent proposals to restrain automobile use in the Bay Area and concludes that local support is the main key to success.

BACKGROUND

The San Francisco Bay Area is a nine-county region that has a population of 4.8 million and a land area about the size of Massachusetts. Three major cities—San Francisco, Oakland, and San Jose—are located around the Bay's rim.

Responsibilities for transportation are shared among a variety of regional, local, and state agencies. Caltrans is responsible for state freeways and highways and also provides funding for certain transit and paratransit proposals. Public transportation is provided by several bus operators and by the Bay Area Rapid Transit (BART). There is no regional transit authority. The Metropolitan Transportation Commission was created by the California legislature in 1970 to provide comprehensive regional transportation planning consistent with the other regional planning activities conducted by the Association of Bay Area Governments (ABAG) and several special districts and authorities. County agencies play an active role in both highways and transit, and the cities frequently get involved in community-level transportation planning and services.

The Bay Area enjoys a high-quality transportation system, excellent freeways, and plentiful transit service. Transportation always has been of interest to Bay Area residents, and occasionally it has been the source of controversy. Two of the better-known freeway controversies occurred in San Francisco (the Embarcadero Freeway) and in Oakland (the Grove-Shafter), and public support for transit alternatives has been strong. Continued heavy reliance on the automobile, however, has contributed to periodic violations of the national ambient air quality standards and to local problems of congestion and neighborhood disturbance. For these reasons, measures to restrain automobile use have been proposed.

Neighborhood Traffic Restrictions

Berkeley, California, which has one of the most extensive traffic management programs in the United States, provides a good example of implementation experience. Berkeley is a city of 120 000, situated across the Bay from the Golden Gate Bridge and north of Oakland. The land slopes gently upward from the Bay inland to the central business district (CBD), where steep hills climb to some 335 m (1100 ft) at the eastern city boundary.

Most of the city is on a grid system of roads except for the hill area in the northeast, where the topography and winding streets discourage through traffic. The University of California at Berkeley is located in the central section of the city, just east of the CBD. The university, the downtown, and an industrial area along the Bay are the major trip attractors.

Berkeley's 1968 master plan adopted as policy the reduction of automobile traffic through residential neighborhoods and classified streets as major, collector, and local. However, through-street designation and coordinated traffic engineering improvements were insufficient to reduce traffic shortcuts through the flatlands neighborhoods, which lie on the paths between the suburbs and university and the other trip attractors. Thus, in 1972 the city undertook a neighborhood traffic study, which resulted in the traffic management plan (TMP). The plan built on the success of traffic diversion devices installed in the 1960s in a low-income neighborhood in southwest Berkeley. Through the use of traffic diverters, chokers, and stop signs, the plan is designed to transform the flatlands neighborhood's grid street system into cul-de-sacs, offset intersections, and narrowed and winding streets, with through traffic transferred to the arterials and collectors.

The TMP was developed by a consortium of neighborhood associations, working closely with the city's planning commission. Active lobbying by residents of the neighborhoods convinced the city council to hire a consultant to develop detailed proposals. After an intensive, citywide citizen participation campaign, more than 70 traffic control devices were approved for installation in 1975. Costs have been estimated at $0.5 million, not including city staff time.

The TMP was evaluated after six months of experience (2). In general, traffic volumes changed as expected, with significant decreases on protected streets. Traffic increased on several of the arterials and some collector streets, but traffic operations improvements absorbed much of the increase, and capacity utilizations at major intersections actually improved. The traffic devices have had marginal effects on (a) the number of automobile accidents (down slightly), (b) transit operations, and (c) the provision of city services. Police and fire departments occasionally have expressed concern that response time has increased in the protected neighborhoods, but no incidents or statistics have been cited. Overall travel times have not changed significantly, although the elimination of shortcuts has caused increases for some of the trips that formerly were made through the neighborhoods.

Despite strong support from most residents of the neighborhoods that benefit from the traffic restraints, the TMP has been controversial. Residents of other Berkeley areas have complained about the cost of the program and have questioned the equity of restrictions of movement on city streets. Some residents of protected neighborhoods also are critical of the TMP; for example, they dislike the circuitous paths they must follow to and from their homes. Those who live along the streets where traffic has worsened understandably object to the diverters. Members of the business community have expressed some concern that the traffic restraints foster an exclusionary image of Berkeley, although they have not taken a formal stand regarding the plan.

Twice, opponents have placed a proposal on the ballot to remove the traffic control devices; both times it was soundly defeated. Opposition to the plan has not stopped at the city boundaries. In 1977, an Oakland resident whose trip to work was lengthened by traffic diverters took the city to court. Several Berkeley residents, including some who resided on the streets to which traffic had been diverted, joined in the suit. The trial court found the devices illegal on the ground that they are not state-approved traffic control devices; the city is appealing that decision.

Perhaps the greatest disappointment with the TMP has come from the city's inability to improve transit as an alternative to automobile use. Transit improvements are considered as important as the traffic restraints; it was hoped that more commuters could be induced to use buses or BART to get to work, so that an overall reduction in traffic would result. So far, however, it has not been possible to obtain new services from the conventional transit operators, and Proposition 13 has further dimmed hopes in this regard. The city has obtained a state grant to develop paratransit. Studies will begin in mid-1979 and will include an exploration of new funding sources for community services as well as alternatives for commuters.

Road Pricing and Congestion Pricing

Road pricing was suggested both for Berkeley and for congested Bay Area freeways. Neither proposal survived past the preliminary stages. The Berkeley road pricing proposal arose in the context of UMTA's exploration of potential sites for pricing experiments. The city council approved the initial exploration of pricing schemes in hopes that they could alleviate congestion and encourage transit use. Possibilities were areawide pricing and pricing during peak hours in the Tunnel Road corridor, a major commuting route into Berkeley for residents of the Contra Costa County suburbs.

The study had just begun when an adverse article in a major newspaper aroused a raft of protests. Council members and city staff were inundated with calls from residents who interpreted the proposal to mean that they would have to pay a toll to get to work and to the grocery store. Merchants complained that road pricing would destroy their ability to compete with suburban shopping areas. Officials in other cities objected to the burden that would be placed on their residents who work in Berkeley and hinted that countermeasures might be taken. Opponents of the TMP were furious that yet another proposal to limit use of public thoroughfares was being considered.

In the wake of the protest, the council voted to end the study. Several members stated off the record that they felt the publicity had been seriously misleading; however, they also believed that the resulting political environment would not support a rational exploration of options. In retrospect, both city officials and the consultant group that conducted the study believe that a more extensive public information and involvement program might have saved the study.

Congestion pricing for Bay Area freeways was proposed in a series of research reports from the university economics department and in studies for the state transportation board. The main recommendation was to establish tolls on the bridges (and perhaps at other points) that would reflect the full cost of automobile travel during congested periods. It was suggested that such a pricing strategy would result in shifts to transit and carpools, thereby alleviating congestion as well as improving air quality and reducing energy consumption. The governor, however, was persuaded that such pricing would discriminate against the poor, and his opposition effectively stopped the proposals.

Toll Increases

Tolls have been increased on both the Bay Bridge (which is state owned) and the Golden Gate Bridge. The increases were controversial and generated strong opposition at the local level and in the state legislature. The Bay Bridge experience illustrates the difficulties that arise with proposals that would restrain automobile use through pricing mechanisms. There, an initial proposal was to raise the toll for occasional users and to eliminate discount tolls for regular commuters. A modest increase ($0.25) in the toll for occasional users was adopted, but the proposed elimination of the commuter discount was dropped after vigorous protest from commuter organizations and officials from cities that have a large commuting population.

The resulting tolls have an insignificant impact on automobile volumes. In real dollars, they are lower than the previous tolls were when they were first established. Although there appears to be some public acceptance for periodic toll increases to cover costs, many believe that when a bridge is fully paid for, tolls should be reduced to cover only upkeep. However, when tolls are discussed in terms of automobile restrictions or disincentives rather than revenues only, they generate substantial controversy. Equity for lower-income groups and those without high-quality transit alternatives is a concern, but perhaps more fundamental is the widely held belief that pricing has no place on a facility that is supported by user taxes and public funds.

San Francisco Parking Programs

Two parking programs in San Francisco illustrate the directions automobile restrictions may take. One program restricts long-term on-street parking to residents, and the other taxes parking fees in some areas of the city.

The resident parking-permit program is in operation in several San Francisco neighborhoods. For a small fee, residents may purchase bumper stickers that allow all-day parking—an exception from the parking regulations in the areas. Only 2-h parking is permitted during daytime hours without a sticker.

The resident parking program was established after much hard work by neighborhood groups, who developed the initial proposals and enlisted widespread community support, then worked with city agencies and officials to get the program approved. Many of the merchants in the affected neighborhoods supported the program because they thought that commuters were tying up parking spaces that otherwise could be used by shoppers.

Preliminary results indicate a 25 percent reduction in the number of automobiles parked at midday in most of the restricted areas. In some areas, however, parking still is scarce; it appears that residents there simply may own more automobiles than there are on-street parking spaces. An informal survey of vehicles that crossed the Golden Gate Bridge pointed to another possibility: Since a number of San Francisco parking permits were spotted, nonresidents may be obtaining permits through a black market or from friends who live in restricted areas.

For several years, San Francisco has had a tax on parking fees in certain areas of the city. A 25 percent surcharge, designed to raise revenues, went into effect in October 1970 and generated about $5.5 million a year. The tax was favored by environmentalists, who expected to see shifts from automobile to transit use, and by property owners, who hoped that the funds raised would help keep real estate taxes down. Retailers, parking facility operators, and automobile commuters mounted considerable opposition, however, and in July 1972 the tax was reduced to 10 percent.

A study of the San Francisco parking tax, conducted by the Urban Institute, found that long-term parking was affected more than was short-term parking, as commuters more than shoppers adopted travel patterns that avoided the parking tax. A 2-6 percent reduction in the number of vehicles parked and a 30 percent decline in total demand for parking services (a combination of the number of automobiles and the duration of occupancy) occurred. Traffic on city streets was reduced not more than 2 percent. Although the available data could not support a detailed analysis, the surcharge appeared to have negligible impact on downtown retail activity (3).

In the past year, several of the largest parking operators in the Nob Hill and Russian Hill areas have eliminated maximum daily parking fees. Instead, they charge a flat half-hourly rate that would make all-day parking cost $8-$10. The operators have found that there is sufficient short-term demand for parking that they make more money under the new scheme. Several businesses were in favor of this development: They said that San Francisco workers have plenty of ways to get to the city without driving and added that parking for shoppers is vital if the downtown is to compete successfully with suburban malls.

Ramp Metering Proposals for the Eastshore Freeway

The Eastshore Freeway regularly experiences congestion in the morning peak at several points from Richmond to the Bay Bridge. However, proposals to use ramp metering to alleviate these problems and make optimum use of existing capacity have not yet been pursued beyond preliminary studies. A common objection to this automobile-restrictive measure is that drivers from the outermost suburbs would be given a clear shot to the bridge, whereas those from closer-in communities, who pay for their locational advantage in higher housing prices and the like, would experience delays. Other concerns are that the metering would divert traffic to city arterials that parallel the freeway and would undermine vanpooling and carpooling programs for outlying areas by making private automobile travel easier and faster. Officials from several communities have made it known that they would oppose any ramp metering proposal for these reasons, regardless of any overall benefits that might result.

CONCLUSIONS

In the Bay Area, proposals to restrain automobile use or to impose disincentives to automobile use have had little success when proposed by planning agencies. Local initiatives, on the other hand, have been implemented and have gained substantial, if not unanimous, acceptance.

What lessons can be learned from this experience?

1. We should not assume automatically that automobile restraints are unacceptable to the public. Automobile restraints generate serious concerns about social equity, differential impacts on different communities, and government's right to control travel choices. Nevertheless, people have shown a willingness to move ahead with measures that provide substantial benefits, particularly when the travelers affected by the measures have reasonable alternatives.
2. Any proposal for automobile restraint must have substantial community support if it is to succeed. This need not mean that all such proposals must be locally initiated; rather, the first step of the planning process should be to involve the affected interests in an initial exploration for workable strategies. It does not appear likely that any automobile-restriction proposal developed by technical staff and then put forth for public comment will have enough support to be implemented.
3. We should explore mechanisms for providing assistance to local efforts that have the potential to reduce automobile use. In several instances, greater benefits might have been obtained if a locally initiated effort had been tied to a broader TSM or transportation-air-quality proposal, but the staff and funds to look at such factors simply were not available. Perhaps cities could apply for earmarked funds or "lend-a-planner" assistance through a competitive process in which proposals that

have potential air quality, energy conservation, or other TSM benefits would be considered.

REFERENCES

1. Transportation Improvement Program. 40 Federal Register 181, Sept. 17, 1975, pp. 12976-12981.
2. Six Months Experience: Berkeley Traffic Management Plan. De Leuw, Cather and Company; city of Berkeley, 1976.
3. D. J. Kulash. Parking Taxes as Roadway Prices: A Case Study of the San Francisco Experience. Urban Institute, Washington, DC, Working Paper 1212-9, Dec. 1973.

Part 3: Experience of the Service and Methods Demonstration Program with Automobile-Restrictive Measures

Carla Heaton, Transportation Systems Center, U.S. Department of Transportation, Cambridge, Massachusetts

Findings are presented from the perspective of the Urban Mass Transportation Administration's Service and Methods Demonstration Program on the implementation of various physical or operational strategies designed to either alter the supply of road space available to vehicular traffic or to reallocate the supply of road space among different classes of vehicles. These findings include the need for objective technical information on the impacts of similar strategies in other locales and the early and continuous involvement of potentially affected groups, the importance of a quick response to any early construction or operational problems, and the importance of the relationship of enforcement to the political feasibility of a project.

The Urban Mass Transportation Administration's (UMTA's) Service and Methods Demonstration (SMD) program sponsors the development, demonstration, and evaluation of innovative transit operating techniques and services that use existing technology. The program has been involved in two broad categories of automobile-restrictive concepts through the conduct of demonstrations, concept feasibility studies, and case study evaluations. One group consists of physical or operational strategies designed to either alter the supply of road space available to vehicular traffic or to reallocate the supply of roadspace among different classes of vehicles. Within this category are automobile-restricted zones (ARZs), transit malls, residential neighborhood traffic and parking restraint schemes, and various priority treatments for high-occupancy vehicles. The second category of automobile-restrictive measures includes pricing incentives and disincentives aimed at particular groups of travelers; examples would be parking surcharges to discourage general or commuter automobile travel to the downtown and various forms of road pricing. Most of the SMD program activities related to automobile-restrictive strategies have been of the first type; efforts in the pricing area have only recently gotten under way (1).

The SMD program implicitly recognizes that inadequate knowledge and experience regarding innovative service concepts can serve as an institutional barrier to implementation. Urban planners, decision makers, and the general public appear to want objective technical information on the impacts of these strategies in other locales as well as a reasonable prognostication of the impacts and likely barriers to implementation in their own locale. Initial support for these concepts is enhanced by credible examples of success in other places; for example, merchant endorsement of the Boston ARZ plan was spurred by the success of the nearby Quincy Market and an encouraging talk from a major retailer located on Philadelphia's transit mall. Progress in implementating such innovations can be impeded by inadequate knowledge. For example, the difficulties that the SMD program has experienced in finding sites willing to implement road pricing measures can be attributed in part to the general uncertainty about the likely nature and magnitude of impacts and public reaction.

Aside from its recognition of the knowledge base as a significant institutional factor, the SMD program has operated on the assumption that a carrot-stick approach may be the only politically feasible means of implementing automobile-restrictive measures. Thus, projects that entail physical, operational, or pricing restraints on automobile use also include a complementary package of incentives or improvements. The ARZ projects being sponsored in Boston, Memphis, New York City, and Providence illustrate this principle in that they all include a number of visible improvements (for example, improved transit service and pedestrian amenities) that are intended to maintain access to and within the area, minimize adverse impacts on peripheral areas, and provide a more pleasant environment for people within the area. Similarly the two SMD-sponsored demonstrations that involve reserved freeway lanes for high-occupancy vehicles (on Los Angeles' Santa Monica Freeway and Miami's I-95) have encompassed transit service improvements, park-and-ride lots, and carpool matching programs that provide alternatives for single-occupant automobiles denied access to the lanes. A soon-to-be-implemented demonstration in Madison, Wisconsin, will provide transit improvement and incentives along with peak-period parking surcharges.

LESSONS LEARNED FROM THE SMD EXPERIENCE WITH AUTOMOBILE-RESTRICTIVE MEASURES

It is critical that potentially affected groups be involved as early as possible in the planning and design process. This is especially important in the case of merchants and other business owners, who are justifiably concerned about the economic impacts of automobile restrictions, both during construction and afterwards.

It is essential that an effective public information program be launched well in advance of project implementation. This lesson was well demonstrated in the Santa Monica Diamond Lane project, which, like Boston's

Southeast Expressway project, met with an early demise. Throughout its five-month period of operation, the Santa Monica project faced intense public opposition, which was fueled in part by the lack of advance publicity about the project.

There is need for a flexible and effective mechanism to respond to problems and complaints that arise during the construction phase or early in the operation phase. Even when detailed feasibility and design work has been performed ahead of time, some unforeseen difficulties are bound to surface. These problems must be resolved quickly before public annoyance and opposition mount. In the Boston ARZ project, which opened in the summer of 1978, staff members from the city Department of Traffic and Parking had to spend considerable time during the first few weeks responding to merchant complaints about delivery restrictions. Eventually restrictions were modified to accommodate these business concerns.

The degree and effectiveness of enforcement can be critical to the operational and political feasibility of the project. In Boston's Southeast Expressway project, the stepping up of enforcement was one of the major factors leading to public opposition and project termination. In Miami, the minimum requirement for carpool size was eventually reduced from three to two persons because police were unable to enforce the more stringent criterion.

The Boston ARZ experience presents a somewhat different perspective on enforcement. To date, the project has pursued a vigorous program of ticketing and towing parking violators within and around the zone. This enforcement program has cost considerably more than anticipated (since the city pays more for towing service than it receives in fines), but it has been viewed as an essential component of the project, both to promote public awareness of the concept and to provide capacity on peripheral streets for the diverted traffic.

CONCLUSION

Efforts under UMTA's SMD program plus local initiative on the part of several pioneering areas have begun to broaden the U.S. base of experience with automobile-restrictive strategies. It is hoped that other locales will be inspired by these examples to implement similar measures; however, the process of diffusion should not move too hastily or with too little attention to important site-specific details, particularly the local institutional environment. Individuals responsible for planning and implementation of the next generation of projects should place emphasis on early and careful feasibility, design, and planning work, liaison with potentially affected groups, and well-designed public information campaigns.

REFERENCE

1. D. Kendall and others. Service and Methods Demonstration Program Annual Report. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Rept. UMTA-MA-06-0048-78-6, July 1978.

Discussion

John H. Suhrbier, Cambridge Systematics, Inc., Cambridge, Massachusetts

An examination of recent transportation programs indicates (at least) two major changes in orientation:

1. A shift in emphasis from the construction of high-capital highway and transit facilities to the improved management of existing transport facilities. Objectives concerning air quality and energy consumption, in many cases, have become as important as the traditional concern with mobility.
2. Key issues associated with the implementation of a project can often be characterized as being institutional in character. Cost, funding source, and design considerations are still important; however, increased attention is being devoted to questions of public acceptability; political support; choice of lead agency; regulation; the consistency of agency priorities; and the appropriate roles of state, regional, and local agencies in the planning and implementation process.

Specific states and urban areas are involved in debate concerning the appropriate role for transportation agencies in managing the use of the private automobile. Unfortunately, the result sometimes has been interagency conflict and stalemate rather than effective, implementable decisions.

The preceding papers examined the implementation experience of representative transportation system management actions in a number of U.S. cities and identified a variety of institutional issues that have either aided or served as a barrier to success. Specific topics addressed include the role of public involvement, interagency coordination, the problems of enforcement, the role of a metropolitan planning organization, the relative costs and effectiveness of different measures, and design considerations.

The papers by Kinstlinger and Deakin provide case studies of implementation experiences in the Denver urban area and the San Francisco Bay area, respectively. Heaton provides a perspective of the SMD program conducted by UMTA. Although Heaton's comments reflect the national orientation of the SMD program, particular attention is devoted to recent activities within the Boston metropolitan area.

In any discussion of transportation measures directed toward the improvement of air quality or energy conservation, an important point is whether such actions may be characterized as (a) disincentives or restrictions on the use of the automobile or (b) incentives to use modes other than the single-occupant automobile or whether there is indeed any difference between an incentive and a disincentive. Are such measures designed and implemented independently and in relative isolation of one another? Or is it possible to consider a coordinated package of interrelated measures that affect a range of available transportation modes and provide changes in a number of cost, travel time, and promotional variables? In an examination of potential institutional issues, the choice of attitude is perhaps most important of all. Are we viewed by the public as designing positive incentives, which implies that benefits exceed costs, or negative discentives, with the associated implication that costs to the public exceed benefits? This question provides an important framework for these papers.

Sandra Rosenbloom, University of Texas, Austin

A number of major common themes appear in the papers presented; these same themes appear and reappear in the literature on transportation innovations as well. The most striking feature in each case is that the transportation improvements being considered are so different from traditional transportation efforts. These differences are subtle as well as obvious. The obvious difference between traffic and environmental management schemes and traditional approaches is that the former attempt to regulate and control the demand for transportation facilities, but traditional transportation improvements have concentrated on significantly expanding the supply of facilities and services. Other comparisons are also obvious; these new approaches are short range, generally noncapital, and often (but certainly not necessarily) cheaper than traditional transportation responses.

This discussion highlights the new and different role for transportation planners that the federal government mandates may be creating. Such mandates require the planners to assume an elitist role in determining (or at least accepting mandated) societal problems and then in fashioning a solution for people, whether they are interested or not. I am not sure that many planners really want to be in the position of telling people what is good for them: Those who feel that they know what is good for individuals or society in general certainly are having an uphill fight, as the papers presented here appear to chronicle.

These newer approaches are very different in the way they are perceived and responded to by the public; measures such as automobile-restrictive techniques rarely have a constituency. They lack groups in society who see direct benefit from their initiation and actively support them. In fact, such measures are frequently opposed by large segments of the public as well as by special affected interests. For example, experience in Seattle and Boston seems to suggest that such projects can only be successful when they in fact create particular benefits for specific individuals or businesses, who will then be willing to actively fight for their implementation. There is also an ironic twist to this situation; a planner cannot be assured of the implementation of a promising automobile-restrictive measure simply because there is no controversy when the measure is first discussed. Public meetings and formal hearings can be held without any opposition voiced, but once, for example, the ground is broken for the park-and-ride lot or streets are made one-way, citizen complaints may suddenly create significant obstacles to implementation. Schemes that lack any strong support or constituency can fail even when exposed to only minor conflict.

Many of the transportation system management and planning strategies proposed are also different from traditional transportation approaches because the benefits to either individuals or society are not obvious and there is some conflict about how well the strategies respond to goals the public has really articulated. Such strategies require transportation planners to expend a great deal of time in calculating the benefits individuals will accrue, whether they perceive them or not. Moreover, not only do planners have to sell the public on how much good a particular measure will do them, the public must also be convinced that the risks of the measure are small or nonexistent. Unfortunately, a number of such measures hold significant risks for affected parties; for example, downtown business people have real fears about losing business to suburban malls if parking bans are implemented.

Also, it is exceedingly difficult to use aggregate statistics on the overall benefits of proposed schemes to convince affected individuals of the overall good to society. Individual benefits from such strategies are often extremely small, perhaps a trip-time savings of 2 or 3 min on a 25-min trip; such numbers sound very good when aggregated but may be meaningless to any one individual.

Kinstlinger suggests that the problems these measures are designed to address may not be perceived as very significant by any one individual either. In essence, many of the strategies considered in these papers and in the literature may create disruptions in people's travel habits in response to a problem people do not perceive, in order to obtain a solution that will not do them much good.

Another significant difference between traditional transportation approaches and those discussed in this set of papers is the extraordinary amount of organization and interagency cooperation that must be achieved. Many funding sources and different organizations may be involved in the newer approaches. This requires the expenditure of significant amounts of time in the structuring of interagency coordination. Problems with the funding sources were not addressed in the papers presented, but my own work clearly reveals that different federal and state sources often have different rules and regulations. Attempts to conform to at least perceived discrepancies among them often create problems in implementation.

Other interorganizational problems are at least briefly mentioned in the papers. One of the most common problems facing endeavors of this kind is the need to coordinate the activities of a number of public and private agencies. In Seattle, for example, a tremendous amount of staff time was required to meet with all affected parties in a carpool program. The Boston ARZ had similar experiences and also illustrates the amount of time and resources lead agencies must expend to get certain measures implemented. It is clear that a tremendous amount of coordination is required, whether or not crucial agencies are supportive of the measure. Even more significant institutional problems arise when key agencies are not supportive.

One particular institutional problem that is only briefly mentioned in the papers is the problem of securing the cooperation of essential organizations whose whole orientation is in opposition to the thrust of the automobile-restrictive or transit-enhancement measure. The most conspicuous example is the need for law enforcement officials to enforce parking restrictions and priority lanes. Many law enforcement agencies identified in our study were quite unwilling to provide such enforcement, not because of the cost but rather because they did not like traffic enforcement activities and gave them a low priority. In Boston, the police would not even change the working hours of the meter maids, so that there is no enforcement at all during the first 1.5 h of the parking bans implemented there. In Miami, police departments in three communities along the transit priority treatment simply refused to continue enforcement activities even though they were paid to do so; in Los Angeles the highway patrol gave back the money they were paid to enforce the Los Angeles Diamond Lane.

The literature calls such problems institutional barriers, but in fact such problems can be predicted, given an understanding of how organizations work. Just as the benefits of certain automobile-restrictive measures seem small or immeasurable to individuals, such benefits may seem insignificant to a number of public organizations as well. These agencies are asked to incur significant expenses (as in Seattle where more transit peak-hour service was promised) or to face significant risks (as in downtown parking bans in San Francisco) or simply to become involved in activities that are not highly regarded (such as enforcing priority lanes on freeways). Given the large number of reasons not to become involved, it is not surprising that most of the examples we have of automobile-restrictive measures were conceived in response to federal mandates and sanctions. They were rarely developed out of local initiative.

The papers presented here are important because they give planners a clear idea of the dangers involved in assuming easy implementation of rational low-cost transportation improvements. There are clearly a number of lessons to be learned from an in-depth analysis of self-conscious case studies such as these. The most glaring feature common to all the papers is that transportation improvements, or any measures that significantly affect people's behavior, require tremendous foresight and detailed planning and ultimately are tested in a very real political arena.

Publication of these papers sponsored by Committee on Social, Economic, and Environmental Factors and Committee on Citizen Participation in Transportation Planning.