Pricing Options for Urban Transportation Modes

ATHANASSIOS K. BLADIKAS and WILLIAM H. CROWELL

ABSTRACT

Urban transit is in many ways in serious financial trouble, and its primary source of revenue--fares--is politically difficult to establish and change. In addition to raising funds, fares must often redistribute income, spur local development, and help reduce automobile use. This multipurpose situation is similar to the one that Boiteux handled in his examination of the publicly owned French electric utility industry. Bailey and Willig did similar analyses of long-distance telephone rates. The goal was to have prices that efficiently related to services' marginal costs and elasticities of demand, generated an acceptable level of profits (or deficits), and maximized the public's welfare from these services. The pricing analyses presented follow a similar approach; the Metropolitan Transportation Authority in New York City is used as the case study. Net benefit changes (revenues plus consumer surplus) were shown to be greater from efficiently set peak and off-peak fare differentials than for flat fares. These results were found to be fairly insensitive to changes in the fare elasticities of demand. In fact, the differential-fare approach looked best when the gap between peak and off-peak elasticities and marginal cost values was greatest. This was not an attempt to find the optimal or welfare-maximizing set of fares, but rather to show which transportation pricing options maximized public welfare within given budgetary constraints.

Considerable attention has been given in recent years to the manner in which transportation modes are priced in urban areas. Growing transit deficits, overcrowded highway networks, the high social cost of capacity expansion, and the growing realization that current pricing methods are often inefficient have all combined to make the revision of such methods a regular part of public policy discussions. In addition, the notion of viewing transportation services in a systems context rather than merely as separate modes has received widespread acceptance. The field of transportation, and public utilities in general, is greatly complicated by the multiple and frequently changing policy goals that pricing, financing, and subsidization decisions often face. Fares may be raised one year to decrease the deficits for fiscal reasons and then lowered in another year in order to reduce environmental pollution, relieve congestion, or stimulate economic activity; both fare change decisions affect the demand not only for transit but also for private automobile use. Attempts to provide transportation services efficiently in such a policy setting are often difficult and frustrating.

In the classical economic sense, a particular mode is being used inefficiently when the demand or supply or both of its services are greater or less than they would be if the true cost and demand functions were known and reflected in the price structure. In terms of pricing theory, a system of interacting modes would operate efficiently if the prices charged in each market were set equal to marginal costs (the cost increment required to provide an additional unit of service). This concept is particularly important for transport modes because of the typical peaking of demand during rush hours that requires a physical plant and staffing level well above that required in off-peak periods. The concept of marginal cost pricing is therefore often referred to as "peak-load" and "congestion-cost" pricing and has received extensive treatment in the transportation economics literature (1-3). However, marginal cost pricing for transportation services may lead to operating deficits that are beyond available subsidy levels, forcing planners to deal with what economists call "second-best" pricing. On the basis of the original insights of Dupuit, Hotelling (4), and Steiner (5), methods were sought to solve the paradox of industries in which costs are decreasing. These analyses were furthered by Baumol and Bradford (6), who showed in probably the most widely quoted analysis of the problem what Ramsey had demonstrated in 1927 (7), namely, that any government entity (or private firm) that wanted to use efficient (but not strictly marginal cost) pricing schemes could use both demand elasticities and marginal cost information to determine how the fare in each market should differ from the marginal cost.

Most of the efforts using this type of Ramsey pricing approach have dealt with the pricing policies of public utilities, with a very important contribution coming from Marcel Boiteux of France, who demonstrated how the electric utility industries could determine prices in their various service areas (8). Willig and Bailey (9) developed a multiservice pricing model for various interrelated telephone services that maximized the consumer benefits of an existing system within specified budget constraints. It is this type of analytical approach that is pursued in this paper.

CONSUMER SURPLUS AND THE PRICE ELASTICITY OF DEMAND

The operating position of a transit agency could generally be shown as in Figure 1 by fare and output levels F_2 and Q_2 somewhere between the breakeven (Q_1) and optimum (Q_3) levels. As output is expanded by lowering the fare, the gap between average cost (AC) and average revenues increases; for example, BD increases and DF decreases as a greater deficit is incurred for the sake of a more optimal pricing scheme. The demand curve (D'D") indicates what consumers are willing to pay for the provided services. According

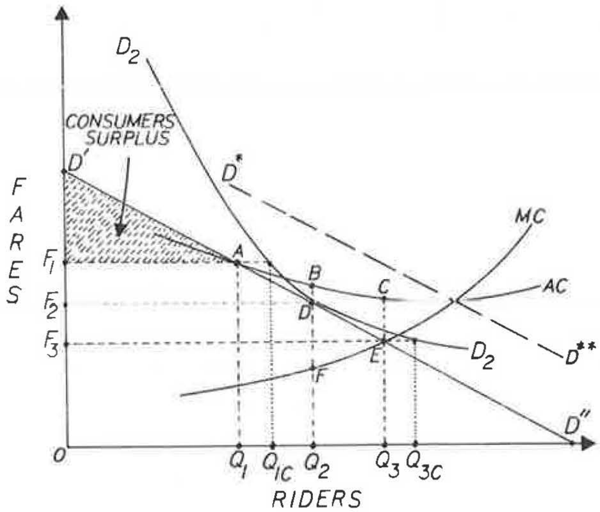

FIGURE 1 Transit supply and demand functions.

to Hicks' concept of "equivalent variation," the consumer welfare gained from a reduction in price would be equal to the maximum amount that a consumer would pay to bring about the reduction (10). The surplus value at any fare level F_i would equal 1/2 ($OD' - OF_i$) • Q_i, which clearly increases as the fare is lowered.

An operator or public official wanting to maximize this type of consumer benefit (which is closely associated with ridership) faces a number of constraints. In terms of allocational efficiency, by lowering the fare below F_3, riders would not be covering marginal costs (MC). This constraint places an upper limit on welfare at fare level F_3 = MC. The second limitation is that caused by the existence of increasing returns (decreasing average costs), which produce deficits equal to CE when prices are equal to marginal cost without any price differentiation (i.e., the same price at all times). Therefore, if one wishes to eliminate deficits, fares have to increase to F_1. This places a lower limit on welfare, at least for publicly owned systems that may not be profit seekers. This does not mean that increasing returns and optimal pricing cannot be compatible with a profitable operation. One can envision a hypothetical shift in the demand function of transit, possibly caused by a major increase in retail gasoline prices, for example, to D*D** in Figure 1, where a fare set equal to marginal cost would also be equal to average cost. At such an expanded output level, the operator in this hypothetical situation would be functioning with the lowest possible average costs.

Measurement of the consumer surplus level in a market is a direct fallout of the estimation of the demand function for that good or service, and the demand function can be determined from estimates of the price elasticity of demand. The important factor here is that there are significant differences in transit demand elasticities for different modes and in various transit "markets," especially as delineated by such factors as time of day, day of the week, and direction. Even with transit elasticities given, assumptions need to be made about the overall shape of the demand functions. As demonstrated in Figure 1, one could assume that the relation between price and demand is a linear one (D'D"), with a constant slope (dQ/dF) but with elasticity increasing with price [e.g., with $E = -(dQ/dF) \cdot (F_i/Q_i)$, as $F_i \rightarrow D'$ and $Q_i \rightarrow 0$, F_i/Q_i increases, and with dQ/dF constant, this results in a steadily increasing fare elasticity as fares increase]. A second assumption could be that fare and demand adjustments will occur in the same proportion over the full range of fares, for example, that elasticity will be constant (curve D_2D_2 in Figure 1). This type of hyperbolic function would result in an elasticity greater than that of the linear function for fare decreases and less than that for fare increases. In Figure 1, after a fare increase from F_2 to F_1, the new demand level Q_{1C} under the constant elasticity assumption is greater than Q_1 under linear assumptions, whereas the reverse holds for Q_{3C} and Q_3 for a fare decrease to F_3. Consumer surplus is also greater under the constant elasticity assumptions, because the area under D_2D_2 is not only greater than that under D'D" at all fare levels but is actually infinite (the demand curve D_2D_2, as a hyperbolic function, will never intersect the fare axis). A linear demand function will be of the following form:

$$Q = a + bF \tag{1}$$

and the constant elasticity demand functions will be of the following form:

$$Q = KF^E \tag{2}$$

where

Q = demand level (ridership),
F = fare, and
E = fare elasticity.

OUTLINE OF RAMSEY PRICING METHODOLOGY

To set the stage generally for this analysis, assume an operation with a set of n markets with price and output sets ($p_1, p_2, \ldots, p_n$) and ($x_1, x_2, \ldots, x_n$). If the goal were to maximize some measure of consumer surplus as a function of market prices such as $W = f(p_1, p_2, \ldots, p_n)$, subject to the profit constraint $B = g(p_1, p_2, \ldots, p_n) = M$, then, by using the LaGrangian multiplier method, it would be desirable to maximize the following function:

$$h = f(p_1, p_2, \ldots, p_n) + \lambda[g(p_1, p_2, \ldots, p_n) - M] \tag{3}$$

Therefore, consumer surplus can only be maximized if

$$\partial h/\partial p_i = 0$$

or

$$\partial W/\partial p_i = \lambda \partial B/\partial p_i \quad \text{for all } i \tag{4}$$

According to the Hicksian concept of consumer surplus, the welfare of an individual consuming quantity x_1 at price p_1 would be increased at a rate equal to x_1 for a one-dollar price reduction and he would therefore be willing to pay up to that amount to bring about the price change (10). Accordingly, for a change in price of p_1, the rate of change in welfare is

$$\partial W/\partial p_i = -x_i \tag{5}$$

From Equations 4 and 5,

$$-x_i = \lambda \partial B/\partial p_i \tag{6}$$

If MC_i, MR_i, and E_i represent marginal cost, marginal revenue, and price elasticity of demand, respectively, then

$$MR_i = p_i + x_i \cdot (\partial p_i/\partial x_i) \tag{7}$$

This assumes that $\partial p_i/\partial x_j = 0$; that is, that the cross-elasticity of demand between markets is zero.

Marginal profit can now be defined as follows:

$$\partial B/\partial p_i = (MR_i - MC_i) \cdot (dx_i/dp_i) \quad (8)$$

Combining Equations 6, 7, and 8 and using the definition of the elasticity of demand [Bladikas and Crowell give a complete derivation (11)], the following expression may be obtained:

$$[(p_i - MC_i)/p_i] \, E_i = K \quad (9)$$

where $K = (1 + \lambda)/\lambda$.

The relationship shown in Equation 9 states that for every good or service i, the amount by which price p_i varies from marginal cost will depend on the elasticity of demand for that good. If all E_i's are equal, prices in all markets will differ from marginal cost by the same proportions. However, when demand elasticities differ, these proportions will differ; E_i thereby determines, along with marginal cost, the proper level of prices (or fares in this context); for example, as E_i increases, $(p_i - MC_i) \div p_i$ decreases, and

$$E_i[(p_i - MC_i)/p_i] = E_j[(p_j - MC_j)/p_j] = K \quad (10)$$

where K is the Ramsey pricing constant.

THE BASIC PRICING MODEL

On the basis of what has been presented so far, a set of computer-based models was developed to recalculate the fares of a given set of urban transportation "markets" so that Equation 10 is satisfied for all of them. The following data must be provided for each transit market to be analyzed:

- Marginal cost,
- Current fare levels,
- Ridership levels, and
- Price elasticity of demand.

Alternative values for marginal costs and elasticities are provided where no exact measurements are available to allow the model to perform a sensitivity analysis of different combinations of these values. After the foregoing information has been read, the model applies the Ramsey pricing methodology for each combination of the given marginal cost and elasticity alternatives. The procedure begins by the calculation of total profit (or loss) from all modes and time periods at the current fare and ridership levels, as follows:

$$SUMP = \sum_{i=1}^{I} \sum_{j=1}^{J} Q_{ij} \, (F_{ij} - C_{ij}) \quad (11)$$

where

SUMP = total profit (or loss),
Q = ridership,
F = fare,
C = marginal cost,
I = number of modes, and
J = number of time periods.

After the total profit (or loss) has been obtained, the demand functions for every mode and time period are determined from Equation 1 or 2 depending on the user's choice. The iterative process begins by setting the demand for the first mode and time period equal to a very high value and obtaining the associated Ramsey constant (K). By using the Ramsey constant, the elasticities, marginal costs, new prices (fares), and quantities (ridership) are computed for all modes and time periods in a way that will satisfy Equation 10. The new prices and demand levels determine new profit levels, which are added and compared with the initial total profit (SUMP). If they are not at least as large as SUMP, a new iteration begins by decreasing slightly the very high demand of the first mode and time period and repeating the process. The iterations continue until the current profit becomes at least equal to the original one. When this happens, control is passed to a subroutine that calculates and prints the following output variables for every mode and time period:

- Original and final fares (PRICE and PNEW),
- Original and final ridership (QUANT and QNEW),
- Original and final elasticities (EL and ELNEW),
- Marginal cost (COST), and
- Percent changes in ridership, revenue, fare, and welfare (%DQ, %DREV, %DP, %DW).

In addition to the foregoing information the output contains the following cumulative results for all modes and time periods:

- Total change in revenues (CHREV),
- Total change in cost (CHCOST),
- Total change in profits (CHPROF),
- Total change in welfare (CHWELF),
- Total net change (CHNET, the sum of CHPROF and CHWELF), and
- Effectiveness index (EINDEX, the ratio CHNET/CHPROF).

The process is repeated until all combinations of alternative elasticity and marginal cost values have been analyzed.

Minor modifications of the algorithm produce a variety of pricing models that can be grouped into four major categories according to the following:

1. The programming language and implementation hardware (FORTRAN for mainframes and BASIC for microcomputers),
2. The shape of the demand function used (linear or hyperbolic),
3. The iteration procedure (constant increments or increments that are proportional to demand and inversely proportional to elasticities), and
4. The presence or absence of special constraints on fares or ridership for any mode or time period.

Any model type can be picked from each of the four categories. It is possible to use, for example, a mainframe-based, constant-elasticity, constant-iteration-increment, unconstrained model. Therefore, there is a total number of 16 (2^4) possible model types. The variable-iteration-increment models are simply faster but less accurate than the constant-iteration-increment models provided that the constant increment is small (about 1/100 for the ridership for linear demand and 1/100 of 1 cent for hyperbolic demand functions). Full program listings may be found elsewhere (11).

A MULTIMODAL CASE STUDY

The New York City area was chosen for an application of the model in a large, multimodal urban setting.

Every conceivable mode of public transportation exists in New York City, but not all existing modes were included because some are insignificant when compared with others in terms of their annual ridership. The Metropolitan Transportation Authority (MTA) is the major provider of public transportation services in the New York City area. The multimodal nature of the MTA allows the results produced by the pricing models to validate one of their major aims--to efficiently adjust prices while maintaining the total system's profit (or deficit). In an area where different organizations own and operate the commuter railroad, the tolled road facilities, the buses, and the rapid transit system, it would be institutionally more difficult to implement a Ramsey pricing approach, because some modes may benefit considerably more than others. In New York City, however, a resident of Forest Hills working in Manhattan's financial district may go to work by (a) driving a car and paying a toll collected by the Triboro Bridge and Tunnel Authority (TBTA), (b) taking the Long Island Rail Road (LIRR) commuter rail line, or (c) using the city's transit system run by the New York City Transit Authority (NYCTA). But TBTA, LIRR, and NYCTA are all MTA subsidiaries, and a change in the price structure will simply reallocate revenues in the various MTA "pockets." Shifting of funds among MTA entities has been in effect since the 1960s through the use of TBTA surpluses to finance deficits of the NYCTA and commuter rail operations.

Input Data

The data used for this multimodal case study are presented in Tables 1-3. Prices do not reflect the latest increases that became effective in January 1984. Only two time periods are used (peak weekday and off peak for all other times) and only the four major modes within the MTA are considered. NYCTA's bus and rail operations are treated as two separate modes, whereas LIRR was the only commuter rail operation. "Tunnel and bridge" represents vehicular traffic entering Manhattan from Queens, Brooklyn, and the Bronx via TBTA tolled facilities.

Ridership estimates were obtained from UMTA Section 15 reports and annual reports of TBTA and LIRR. Three elasticity alternatives are used. Basically, there are two ways of determining elasticities: use values from ridership changes due to fare changes in either the system under consideration or similar systems in other areas. For New York's transit system, the Regional Plan Association (RPA) (12) estimated demand functions and related price elasticities of demand that were -0.16 for the subway and -0.31 for the bus services. However, these are overall elasticities across all time periods, and other studies (13-16) show that no demand formula is applicable in all situations. Such factors as the transit mode in question, trip purpose, time of day or day of the week, and similar specifics effectively create an array of demand situations or "markets."

The elasticity values for the subway and bus modes presented in Table 2 vary about the average figures suggested by the studies just mentioned. The base elasticity figure for the TBTA tunnels and bridges was based on these studies and on a brief analysis of the changes in demand levels after toll increases. Unfortunately, no data were available on elasticities by time of day, so it was assumed that the ratio of peak to off-peak elasticity was the same as that for subway services. The same ratio was used for the commuter rail elasticities, a basic figure that was extrapolated from estimates made by the New York State Department of Transportation (17).

The marginal cost values calculated for each of the transport services are rough estimates. No rigorous empirical attempt was made to determine these cost functions, because data were not sufficient for the systems in question. For this reason, a range of cost values was tested and it is assumed that marginal costs for all modes and time periods remain constant over the range of demand levels being considered.

TABLE 1 Multimodal Case Input Data: Price and Ridership

Mode	Price ($)	Ridership (000,000s) Peak	Off Peak
Subway	0.75	623	415
Bus	0.75	334	281
Commuter rail	2.00	43	40
Tunnel and bridge	1.25	50	186

TABLE 2 Multimodal Case Input Data: Cost and Elasticity Alternatives

Period	Mode	Cost ($) 1	2	3	Elasticity (absolute value) 1	2	3
Peak	Subway	0.69	0.92	1.15	0.08	0.10	0.13
	Bus	0.71	0.94	1.18	0.15	0.20	0.25
	Commuter rail	3.00	4.00	5.00	0.15	0.20	0.25
	Tunnel and bridge	0.70	0.80	1.10	0.08	0.10	0.13
Off peak	Subway	0.35	0.46	0.58	0.15	0.20	0.25
	Bus	0.35	0.47	0.59	0.30	0.40	0.50
	Commuter rail	1.50	2.00	2.50	0.30	0.40	0.50
	Tunnel and bridge	0.35	0.40	0.55	0.15	0.20	0.25

TABLE 3 Multimodal Case Input Data: Likely Demand Functions

Mode	Linear	Constant Elasticity
Subway	$Q = 1.261*10^9 - 2.24*10^8(F)$	$\ln Q = 20.734 - 0.16(\ln F)$
Bus	$Q = 8.05*10^8 - 2.54*10^8(F)$	$\ln Q = 20.191 - 0.31(\ln F)$
Commuter rail	$Q = 1.09*10^8 - 1.27*10^7(F)$	$\ln Q = 18.019 - 0.31(\ln F)$
Tunnel and bridge	$Q = 2.74*10^7 - 3.02*10^7(F)$	$\ln Q = 19.244 - 0.16(\ln F)$

Given the current prices and quantities and the basic overall elasticities, linear (Equation 2) and constant elasticity (Equation 3) demand functions can be produced for all modes, as shown in Table 3. The linear demand functions imply that the fare at which ridership will be eliminated (Fe) is $5.43 for subways, $3.17 for buses, $8.45 for commuter rail, and $9.06 for tunnels and bridges. These values imply further that current consumer surplus [1/2 x (Fe - F) x Q] is $2.45 billion for subways, $0.74 billion for buses, $0.27 billion for commuter rail, and $0.92 billion for tunnels and bridges. The total consumer surplus is therefore $4.4 billion. These figures are given to provide a base for comparing the relative magnitudes of the output variables produced by the models.

Results from the Constant-Iteration-Increment Models

Only results from the constant-iteration-increment models are presented, because the variable-iteration-increment models are simply faster and cruder tools. Because the prices suggested by the models are not very different for the various cost and elasticity alternative combinations, the results of the first cost and elasticity alternative only are shown in Table 4.

It is obvious from the results that all low-elasticity peak periods are charged more than the high-elasticity off-peak periods. Some of the price changes are rather significant. For example, peak-period commuter rail passengers are charged 72 percent more and off-peak bus passengers have their fares reduced by 39 percent. However, in spite of the large individual price variations, the model produces minor overall changes for the average transportation services user in the area. If ridership and revenues for all modes and time periods are added together, under the original pricing scheme 1.972 billion trips are made, which generates $1.701 billion of revenues. Thus the price of the average trip is 86.24 cents. If the prices suggested by the model are implemented, total ridership increases by 1.6 percent to 2.004 billion, the average trip becomes 2.2 percent less expensive (84.35 cents), total revenues are reduced by 0.5 percent or $8.9 million, and consumer surplus is increased by $18 million, or 0.4 percent. Therefore, if a systems approach is taken, the overall changes are for all practical purposes negligible. This is a perfect illustration of the impacts of Ramsey prices. They simply redistribute revenues among the various "markets" by charging differential prices, but without producing any significant overall changes.

MODELS WITH CONSTRAINTS

An area's transportation system not only satisfied the basic mobility needs of its population, but also can be used to serve or promote other purposes and causes. Therefore, although the prices suggested by a Ramsey pricing methodology may make economic sense, there might be other overriding concerns that dictate a different price structure. Peak-period vehicular traffic may have to be limited because of environmental reasons, whereas it might be desirable to increase off-peak travel for economic development reasons. Because of political realities, transit fares may not be able to increase above a certain level and may not fall below another because of system capacity constraints. The constrained versions of the models provide the user with a tool that can model such special local situations. Only a ridership constraint example is presented here. A variety of constrained model results may be found elsewhere (11). Tables 5 and 6 give the results of setting a lower limit on the ridership of the subway peak period, which was limited to 580 million passengers (round-off errors produce a QNEW of 579.9 in Table 5). The linear demand model was used and the detailed results from only the first cost and elasticity alternative combinations are shown. Cumulative changes produced by all nine cost and elasticity alternatives from the constrained and

TABLE 4 Sample of Results from Linear Demand Function Unconstrained Model

Mode	Period	Changes in Ridership (%DQ)	Revenue (%DREV)	Fare (%DP)	Welfare (%DW)	Final Ridership (QNEW) (000,000s)	Final Fare (PNEW) ($)
Subway	Peak	-2.96	32.99	37.06	-5.84	604.5	1.03
Bus	Peak	-2.81	15.40	18.73	-5.54	324.6	0.89
Commuter rail	Peak	-10.81	53.48	72.09	-20.46	38.4	3.44
Tunnel and bridge	Peak	-0.19	2.15	2.35	-0.38	49.9	1.28
Subway	Off peak	4.13	-24.55	-27.55	8.43	432.1	0.54
Bus	Off peak	11.85	-32.32	-39.48	25.09	314.3	0.45
Commuter rail	Off peak	3.65	-8.96	-12.17	7.43	41.5	1.76
Tunnel and bridge	Off peak	6.83	-41.82	-45.54	14.13	198.7	0.68

Note: Cumulative results (in millions of dollars, except EINDEX): CHREV = -8.9, CHCOST = -9.1, CHPROF = 0.26, CHWELF = 18.1, CHNET = 18.4, EINDEX = 0.0069.

TABLE 5 Results from Constrained Model

Mode	Period	Changes in Ridership (%DQ)	Revenue (%DREV)	Fare (%DP)	Welfare (%DW)	Final Ridership (QNEW) (000,000s)	Final Fare (PNEW) ($)
Subway	Peak	-6.92	73.58	86.48	-13.36	579.9	1.40
Bus	Peak	-6.77	35.31	45.13	-13.08	311.4	1.09
Commuter rail	Peak	-14.45	67.95	96.31	-26.81	36.8	3.93
Tunnel and bridge	Peak	-4.25	46.66	53.18	-8.33	47.9	1.91
Subway	Off peak	-0.11	0.63	0.74	-0.22	414.5	0.76
Bus	Off peak	7.29	-18.78	-24.29	15.11	301.5	0.57
Commuter rail	Off peak	-0.57	1.33	1.91	-1.14	39.8	2.04
Tunnel and bridge	Off peak	2.48	-14.46	-16.53	5.02	190.6	1.04

TABLE 6 Comparison of Results from Constrained and Unconstrained Models

Alternative		Variable ($000,000s)					
Cost	Elasticity	CHREV	CHCOST	CHPROF	CHWELF	CHNET	EINDEX
Constrained							
1	1	449.7	-57.6	507.3	-520.5	-13.2	-0.0261
1	2	263.2	-49.0	312.2	-309.0	3.1	0.0101
1	3	130.9	-40.8	171.7	-154.3	17.4	0.1011
2	1	459.0	-94.8	553.8	-547.5	6.3	0.0114
2	2	271.1	-91.5	362.6	-341.6	20.9	0.0577
2	3	123.9	-85.7	209.7	-175.3	34.3	0.1637
3	1	492.4	-14.5	637.4	-607.5	29.9	0.0469
3	2	300.2	-15.2	452.2	-407.5	44.7	0.0988
3	3	134.9	-15.2	287.3	-229.4	57.9	0.2016
Unconstrained							
1	1	-8.9	-9.1	0.265	18.1	18.4	0.0069
1	2	-9.9	-10.5	0.597	24.3	24.8	0.0042
1	3	-13.7	-13.8	0.077	30.8	30.9	0.0004
2	1	-27.3	-27.8	0.562	23.8	24.3	0.0043
2	2	-34.3	-35.7	1.373	31.0	32.3	0.0023
2	3	-44.8	-45.3	0.414	40.1	40.5	0.0098
3	1	-50.6	-51.1	0.523	31.3	31.8	0.0061
3	2	-67.1	-68.1	1.046	41.1	42.1	0.0040
3	3	-85.7	-86.7	0.967	52.0	52.9	0.0055

unconstrained versions of the model are also presented for comparison purposes.

The unconstrained linear demand model produces a peak subway ridership that ranges from 593.4 to 609.4 million for the various cost and elasticity alternatives. Therefore, by restricting peak subway ridership to 580 million, that ridership can effectively be reduced between 2.3 and 4.8 percent. With the exception of the first cost and elasticity alternative, which produces a negative net change and therefore a negative effectiveness index, all other cost and elasticity combinations continue to produce positive overall net changes after the constraint has been placed. However, the effectiveness indices of the constrained model are one to three orders of magnitude smaller than they were in the unconstrained model. This deterioration of the constrained model's effectiveness is not so much a result of reductions in net change (for the last two cost and elasticity combinations the constrained model actually produces slightly higher net changes) as it is mainly from the rather dramatic profit increases of the constrained model (two to three orders of magnitude higher than those of the unconstrained model). The high profit changes are produced from the higher price that the constrained model has to charge the peak-period subway users in order to turn away some of them and reach the specified limit. The higher peak-period subway fare produces in turn higher prices for all other modes and time periods through the Ramsey constant (Equation 10). The higher prices produced another interesting difference. The positive net changes produced by the unconstrained version were a result of increased benefits to both the transportation system (more profits) and its users (higher welfare). In addition, the user benefits were from one to three orders of magnitude higher than the system benefits. This situation was reversed dramatically when the constraint was placed. Although the net changes remained roughly the same (except for the first cost and elasticity alternative combination), only the transportation system received benefits, whereas the users had to suffer a considerable reduction in their welfare.

Naturally, there are limits to what one can do with prices alone. Unrealistically high or low constraints either will produce no results at all, because the model will not be able to meet the nonreduction of the current profit criterion, or will produce unreasonable results. For example, if an off-peak subway ridership goal of 480 million is forced on the system, the model indicates that the new fare for this mode and time period should be -$0.03. This indicates that if it is desired to attract 480 million passengers to the subways during the off-peak period, they have to be paid 3 cents a trip. The correct interpretation is that the constraint was too high, and it is impossible to attract 480 million passengers through pricing incentives alone. Generally, it is impossible to know in advance whether a constraint is reasonable. Its feasibility has to be investigated in most cases by trial and error.

MODEL LIMITATIONS

The model results appear to depend on the assumptions that those deterred from peak-period use of all modes will shift to off-peak trip making. Because the model treats the various modes as independent rather than substitute services, with no measure of cross-elasticity of demand among the modes, it is difficult to determine what portion of the 30,000 former daily users of rush-hour commuter rail service would (a) travel at another time period but still use the railroad, (b) make the same trip but by a different mode, or (c) not make the trip at all. The answer to this question is crucial to the understanding of the impact of such fare and toll adjustments on the economy of the city. It must be remembered, however, that (a) one can still drive into Manhattan without using a tolled facility and (b) there are other transit and paratransit options available (e.g., express bus, taxi, and various semilegal van services), most of which are not under MTA control. The very presence of untolled bridges directly adjacent to tolled river crossings shows that there are other economic or political considerations that prevent the efficient control of the city's transportation network and that tend to move feasible pricing policies away from Ramsey-type solutions.

Environmental policies might conflict with any program for higher transit fares in any time period. The models call for considerably higher subway and commuter rail fares, whereas an increase in vehicular trips across the MTA's toll facilities is projected. According to the results in Table 4, for example, annual transit use is expected to increase by 1.3 percent, and bridge and tunnel crossings would increase by 5.3 percent. Therefore, these types of changes in transportation pricing could possibly conflict with the environmental goals of reduced vehicular travel.

The major limitation of the Ramsey pricing models was seen during the discussion of ridership constraints. As is the case with all goods and services, travelers do not choose transportation services only on the basis of their fares. Comfort and other amenities are quite often more important factors than the fare, but they are not included in the model. In addition, the models do not assume any interaction between the transportation system and the economy in which it operates and serves. Demand levels are determined without a check to see whether the local economic activity is capable of producing these levels or, conversely, what impacts the new demand levels will produce on the local economy.

DISTRIBUTIONAL ASPECTS OF RAMSEY PRICING

The changes that the models produce for the overall average fares are very minor. However, the charges paid by individual travelers in different time periods vary considerably, which implies a redistribution of the system's cost burdens. The results in Table 4, for example, imply a 33 percent, or $155 million, increase in peak-period subway user charges, whereas revenue expectations for other periods are lowered by $78 million, or 25 percent. At this point, the differences between equity and efficiency must be stressed. The Ramsey pricing model and the welfare-revenue trade-off considerations discussed previously both dealt with the defining of efficient pricing policies, that is, those that increase total welfare without any consideration of its distribution before or after the price change. Concerns over the equity of pricing or taxation policies, however, can also be made a part of fiscal decisions. These two very distinct economic characteristics are inexorably linked in the realities of the political process and in economic thinking as well. Although either one may be treated separately for analytical ease, the links between them must always be kept in mind.

A consumer surplus methodology can be envisioned that would bring the public's concern over the welfare of certain groups into the fare policy decision process by including not only marginal costs and revenues but also some form of distributional weights as an additional constraint on fare selection. With three rough income classifications (low, medium, high) and the eight time-period and modal combinations of the case study, the following distribution could be defined:

l_{ij}, m_{ij}, h_{ij} = portion of service X_{ij} that is purchased by the low (medium, high) income group

where X_{ij} is, for example, off-peak bus trips.

Let us assume that some set of weights has somehow been specified that expresses quantitatively the desire of the policymakers to influence transit pricing through the addition of certain income characteristics of the riders in the various transit submarkets. When combined with the l, m, h (for all ij) measures of ridership income distribution, a composite equity index of the following form could be constructed:

$$I_{ij} = w_l l_{ij} + w_m m_{ij} + w_h h_{ij}$$

where w_l, w_m, and w_h are the weights that the agency places on a unit of benefit that each of the three income groups receives. If the additional equity-based distributional variable (I_{ij}) is to be included, the basic Ramsey optimizing formula for the most efficient pricing policy (Equation 10) should be changed to the following:

$$(P_{ij} - MC_{ij})E_{ij}/P_{ij} = K_{ij}/I_{ij} \qquad (12)$$

With the higher values of I_{ij} given to favored groups, this would mean that the higher the values of I_{ij}, the lower the price for that service.

The net impact of any public program or policy on the general welfare or that of specific groups depends on the benefits received and the costs borne by the groups in question. Just as the incidence of the fare must be traced, so also must the distributional characteristics of the other taxes used to support transit operations be estimated. Typically, the assumed impact of the fare receives the bulk of the public consideration, whereas the other fiscal tools that actually raise the majority of the transit funds are left relatively undiscussed. Crowell has shown that the mechanisms employed by local tax systems are relatively regressive (18). Policy arguments, therefore, in favor of providing, say, low-fare benefits to riders principally on the basis of their income redistributive effect have two basic weaknesses:

1. The profile of the city's transit ridership shows that the median income of riders is actually above that of the general population, especially in congested peak periods when the system's costs are highest and an even higher-income population of work-trip travelers predominates; and
2. The tax revenues used to support these services are mainly from fairly regressive local tax sources that would effectively cancel any distribution of the services' benefits supposedly for the poor. In addition, the benefits from improved service or a low-fare policy may be shifted by market forces to the landowners whose property value is increased through improved transit access; and with a property tax system that drastically underassesses the land portion of real estate, it is very difficult to "capture" the benefits accrued by these high-income individuals (19).

MODEL SENSITIVITY TO INPUTS

The models are for all practical purposes insensitive to the elasticity assumptions. This insensitivity to elasticity values is not very obvious from the summary results that have been presented because of the huge total annual riderships of the systems, which produce relatively large overall changes for a difference of even 1 cent in price. But, given a marginal cost alternative, changing the elasticities produces differences in the suggested fares that are at most 4 cents (or 1 percent), whereas in most cases they remain identical. On the other hand, given an elasticity alternative, changing the marginal costs produces differences in the suggested prices that typically range between 5 and 10 percent and go as high as 40 percent for the commuter rail mode. Typically, therefore, the models are about 10 times more sensitive to changes in marginal cost than to changes in demand elasticity. Unfortunately, marginal cost is the only variable for which "hard" estimates are not generally available. Higher off-peak and lower peak-period marginal cost values would mean a lower potential payoff from a Ramsey-type pricing approach.

SUGGESTIONS FOR PUBLIC POLICY AND FURTHER RESEARCH

The results of the pricing model applications and the transit taxation options point out two principal areas for public attention:

1. The current flat-fare approach that is in vogue among transportation agencies generates a lower level of benefits to its users than would be possible under a pricing method that took into consideration the demand and cost characteristics of the systems under their control. This same kind of pricing methodology could also be used to account for such social costs as pollution, inefficient use of scarce resources, or traffic congestion. Differential fares have been used in other areas around the world and in parts of the United States, and of course are very common in such private industries as the telephone companies, airlines, and other market situations with large fixed investments and considerable peaking of demand.

2. Municipalities should begin to look less at nontransportation taxes as sources of earmarked support for transportation operations and more at the use of rational pricing of transportation services as a way of raising revenues, avoiding unnecessary expansion of highway or transit systems, and increasing the benefits received by the travelers in their region. The pricing model showed in part how a more sensible pricing approach could be shaped. In addition, the ability of a local government to carry out income distribution through the transit system is severely limited because of the use of the system by many travelers who are not poor and the considerable economic disruption that levying a fairly heavy progressive taxation on a local level might cause in terms of the flight of firms and upper-income citizens trying to avoid the tax.

The results of the pricing modeling effort showed that Ramsey pricing could be applied to a multimarket transportation agency, but the model's predictive ability is rather sensitive to changes in the cost assumptions. Unfortunately, no accurate cost information exists. Section 15 of the Urban Mass Transportation Act is a step in the right direction, but transit is still very much behind other industries as far as data on unit costs are concerned. In addition, more up-to-date information on the users' income distribution could help to resolve some of the conflicts surrounding the claims of the regressivity (or progressivity) of various transportation prices. Accurate cost and user characteristics data, together with the inclusion of modal and time-period cross-elasticities, would make future versions of the models very precise planning and policy tools that can answer questions on the equity as well as efficiency of differential pricing.

ACKNOWLEDGMENT

Research for this paper was supported by UMTA, U.S. Department of Transportation.

REFERENCES

1. H. Mohring. Optimization and Scale Economics of Urban Bus Transportation. American Economic Review, Vol. 62c, Sept. 1972.
2. W. Vickrey. Pricing, Metering, and Efficiently Using Urban Transportation Facilities. In Highway Research Record 476, HRB, National Research Council, Washington, D.C., 1973, pp. 36-48.
3. J.F. Kain. How to Improve Urban Transportation at Practically No Cost. Department of Economics, Harvard University, Cambridge, Mass., May 1970.
4. H. Hotelling. The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates. Econometrica, Vol. 6, 1938.
5. P.O. Steiner. Peak Loads and Efficiency Pricing. Quarterly Journal of Economics, Vol. 71, Nov. 1957.
6. W. Baumol and D. Bradford. Optimal Departures from Marginal Cost Pricing. American Economic Review, Vol. 60, June 1970.
7. F. Ramsey. A Contribution to the Theory of Taxation. Economic Journal, Vol. 37, March 1927.
8. M. Boiteux. On the Management of Public Monopolies Subject to Budgetary Constraints. Journal of Economic Theory, Vol. 3, Sept. 1971.
9. R. Willig, E. Bailey, and I. Ramsey. Optimal Pricing of Long Distance Telephone Service. In Pricing in Regulated Industries: Theory and Applications, University of Arizona Press, Tucson, 1977.
10. J.R. Hicks. The Four Consumer Surpluses. Review of Economic Studies, Vol. 11, 1943-1944, pp. 31-41.
11. A.K. Bladikas and W.H. Crowell. Pricing Options for Urban Transportation Modes. Final Report. UMTA, U.S. Department of Transportation, Aug. 1984.
12. Financing Public Transportation. Paper 98. Regional Plan Association, New York, 1976.
13. T.A. Domencich, G. Kraft, and J.-P. Valette. Estimation of Urban Passenger Travel Behavior: An Economic Demand Model. In Highway Research Record 238, HRB, National Research Council, Washington, D.C., 1968, pp. 64-78.
14. Transit Development Program--Bi-State Transit System. Bi-State Development Agency, St. Louis, Mo., 1974.
15. The Effect of Fare on Transit Ridership in the Atlanta Region: Summary Report. Metropolitan Atlanta Rapid Transit Authority, Atlanta, Ga., 1973.
16. J. Cavala and R. Roess. The Effect of Fare Reductions on Public Transit Ridership. Report UMTA-74-6-1. UMTA, U.S. Department of Transportation, 1974.
17. D.T. Hartgen and S.M. Howe. Transit Deficits; A Projection for New York State. Presented at 55th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1976.
18. W. Crowell. An Analysis of Transit Finance and Pricing Options. Ph.D. dissertation. Department of Economics, New York University, June 1977.
19. D. Netzer. The Case Against Low Subway Fares. New York Affairs, Vol. 1, Winter, 1974, pp. 14-25.

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