

OPTIMAL TRANSIT PRICES UNDER INCREASING RETURNS TO SCALE AND A LOSS CONSTRAINT

By Kenneth Train*

The typical urban public transport system comprises several companies or agencies, each of which provides public transport on a certain mode or in certain geographical areas. Each transport unit is required (either by law or to prevent bankruptcy) to have its revenues meet some proportion of costs. If there are increasing returns to scale, this constraint might compel each unit to price above marginal cost.

In some urban transport systems it might be possible to relax the constraint that each unit must cover a proportion of its costs, and instead require that total revenues from all the units must cover a proportion of total costs of the units. That would allow one unit (which might face decreasing returns to scale) to subsidise another unit (perhaps with increasing returns to scale). An agency could be established which would administer prices and cross-subsidisation. The advantage of this arrangement is that the welfare loss incurred if not all prices were set equal to marginal cost would be less than if each unit faced a separate constraint.

The question of most interest is what price each unit should charge for its service so as to maximise welfare, given the constraint that all units' revenues cover a proportion of all units' costs. This paper presents a method for empirically determining optimal prices. The method is applied to two transport units in the East Bay Area of the San Francisco Bay Area. The method is quite general, however, and can be used to determine optimal prices whenever different prices can be charged for transport on different modes, by different companies, or in different areas.

The two transport units which are the subject of the present paper are Alameda-Contra Costa (A.C.) Transit, which provides bus service, and Bay Area Rapid Transit (BART), which provides rail service. The paper consists of five sections. (1) The first section presents the first order conditions for constrained welfare maximisation which Boiteux [1] developed for the French National Railway. It is shown that the optimal price for each service is a function of the quantities demanded of each service, the elasticities of demand for each of the services with respect to the price of each service, and marginal and average costs of each service. (2) A method is presented for determining demand quantities and elasticities from a

*Urban Travel Demand Forecasting Project, Institute of Transportation Studies, University of California, Berkeley. This research was supported in part by National Science Foundation grants to the University.

behavioural model of consumer mode choice. This method of demand analysis has been used extensively in recent transport research [5, 7, 9, 10, 11, 12]. Demand quantities and elasticities for BART and A.C. Transit service are estimated by this method. (3) Estimates of marginal and average costs are given. (4) The cost and demand estimates are used to estimate the prices which should be charged by A.C. Transit and BART. (5) Limitations of the preceding analysis are discussed.

BOITEUX'S THEORETICAL SOLUTION

Boiteux's analysis is quite general. Each of a number of consumers maximises utility subject to a budget constraint. Each of a number of firms in the private sector maximises profits subject to a production function and given the prices for all goods and services. A public producer (which can be one unit or a group of units) produces several goods and is required to have revenues minus costs equal to some number, perhaps zero. Boiteux derived the following first order conditions for constrained welfare maximisation:

$$\sum_{i=1}^{n-1} (p_i - \pi_i)(ij) = \beta y_j \quad j = 1, \dots, n-1 \quad (1)$$

where:

- n = the number of goods in the economy,
- p_i = the price of good i ,
- π_i = the marginal cost of good i ,
- y_j = the net output of good j by the public producer (perhaps zero),
- (ij) = the partial derivative with respect to the j -th price of the compensated demand function for the public firm's output of the i -th good,
- β = a constant, chosen so as to bring about compatibility among the conditions (1) and the constraint on revenues minus costs of the public producer.

These $n - 1$ conditions were derived under the normalisation $p_n = \pi_n$. Thus there are $n + 1$ unknowns (i.e. β and p_i , $i = 1, \dots, n$) and $n + 1$ equations (i.e. conditions (1), the normalisation, and the constraint on revenues minus costs of the public unit).

An intuitive interpretation of conditions (1) is not obvious. Boiteux and Drèze [2] wrote at considerable length concerning interpretation. Suffice it here to state that conditions (1) are the same as the first order conditions for profit maximisation for a monopolist, with two changes: (a) for the monopolist, β is set equal to one; and (b) for the monopolist, the partial derivatives of the uncompensated demand curve are relevant, rather than those of the compensated demand curve.

Since $(ij) = (ji)$, conditions (1) can be rewritten as:

$$\sum_{i=1}^{n-1} \frac{p_i - \pi_i}{p_i} \eta_{ji} = \beta \quad j = 1, \dots, n-1 \quad (2)$$

where η_{ji} = the elasticity of compensated demand for good j with respect to the price of good i . For empirical analysis the income effect on demand is assumed to be zero or negligible, and so the η 's are simply elasticities of (uncompensated)

demand. Thus the optimal prices are a function of elasticities of demand, marginal costs, and the public producer's constraint (as reflected in β).

In applying this analysis to pricing of A.C. Transit bus and BART rail services, the two services are considered to be outputs of one public producer. Let bus service be denoted by 1, rail service by 2, and a composite of all other goods by 3. Thus conditions (2) become:

$$\frac{p_1 - \pi_1}{p_1} \eta_{11} + \frac{p_2 - \pi_2}{p_2} \eta_{12} = \beta$$

$$\frac{p_1 - \pi_1}{p_2} \eta_{21} + \frac{p_2 - \pi_2}{p_2} \eta_{22} = \beta$$

Equating the left-hand terms and rearranging produces

$$\left(\frac{p_1 - \pi_1}{p_1} \right) (\eta_{11} - \eta_{21}) + \left(\frac{p_2 - \pi_2}{p_2} \right) (\eta_{12} - \eta_{22}) = 0 \quad (3)$$

BART is required by law to cover operating costs (but not capital costs) with revenue. Let A.C. Transit be required to cover all costs.¹ Therefore, the constraint on the public producer BART/A.C. Transit is

$$(p_1 - ATC_1) y_1 + (p_2 - AOC_2) y_2 = 0 \quad (4)$$

where

AOC = Average operating cost

ATC = average total cost

Equations (3) and (4) can be solved for the two unknowns, p_1 and p_2 .

DEMAND QUANTITIES AND ELASTICITIES FROM A LOGIT MODEL OF CONSUMER MODE CHOICE

The logit model is a model of individual choice behaviour which estimates the probability that a person with given alternatives to choose from, with measured attributes for each alternative, will choose a particular alternative.

Consider an individual with personal attributes denoted by a vector s who is faced with a choice among J alternatives. Each alternative exhibits attributes denoted by a vector x_j . The individual has a utility function U which depends on attributes of the alternative x (which can assume values x_j , $j = 1, \dots, J$), and personal attributes s . Without loss in generality, this utility function can be separated into two parts: a part, V , which is "representative" and gives the utility which the average or representative individual with personal attributes s and consuming an alternative with attributes x would derive, and a part, ε , which is peculiar to the particular individual. This is represented as:

¹ Since subsidisation of A.C. Transit will be continued, A.C. Transit is required only to cover some proportion of costs. This proportion is not fixed by law, however.

$$U(x, s) = V(x, s) + \epsilon(x, s) \quad (5)$$

The individual will choose alternative i , given that he maximises utility, if and only if

$$U(x_i, s) > U(x_j, s) \quad \forall j = 1, \dots, J, j \neq i.$$

If we know V , which is representative utility, but not ϵ , then, using (5), we can state that the probability, P_i , that a person drawn at random from a population, if he has personal attributes s and faces alternatives with attributes $x_j, j = 1, \dots, J$, will choose alternative i is:

$$P_i = \text{Prob}(V(x_i, s) + \epsilon(x_i, s) > V(x_j, s) + \epsilon(x_j, s)) \quad \forall j = 1, \dots, J; j \neq i \quad (6)$$

McFadden [6] shows that using the Weibull distribution for ϵ gives the following specification to equation (6):

$$P_i = \frac{e^{V(x_i, s)}}{\sum_{j=1}^J e^{V(x_j, s)}} \quad (7)$$

To make the problem empirically tractable, it is assumed that representative utility is linear in parameters. Thereby (7) becomes

$$P_i = \frac{e^{\alpha'z(x_i, s)}}{\sum_{j=1}^J e^{\alpha'z(x_j, s)}} \quad (8)$$

where z is a vector valued function of x and s , and α is a vector of parameters.

It is possible to estimate α by a maximum likelihood method with a sample of individuals facing alternatives $1, \dots, J$. This method is described in [6].

For the purpose of determining optimal pricing for BART and A.C. Transit, we need to know the quantity, measured in passenger miles, demanded of each service, and elasticities of demand at various price combinations. The logit model expressed in equation (8) is particularly amenable to this type of analysis.

Assume that in the short run the number of individuals needing to travel at peak hours² is fixed. Each individual can choose between travelling by bus, rail, and auto. Denote bus by 1, rail by 2, and auto by 3. Thus, by (8), the probability that person k ($k = 1, \dots, K$) will choose mode $i, i = 1, 2, 3$, is:

$$P_i^k = \frac{e^{\alpha'z(x_i^k, s^k)}}{\sum_{j=1}^3 e^{\alpha'z(x_j^k, s^k)}} \quad (9)$$

where x_j^k is the vector of attributes (cost, time, etc.) of travel mode j faced by person k and s^k is the vector of attributes of person k . Since the cost of an alternative is a function of its price, P_i^k will depend on the price of each alternative.

²The optimum prices derived herein are optimal peak period prices.

An estimate of total demand in passenger miles for bus and rail at given prices for all modes is:

$$N_i = \sum_{k=1}^K m_{ki} p_i^k \quad i = 1, 2 \quad (10)$$

where m_{ki} is the number of miles which person k would travel on mode i if he chose that mode.³

Elasticities of demand for bus and rail with respect to changes in bus prices can be determined by recalculating P_i^k for a small change in bus price and aggregating over k for a new N_i . The percent change in N_i , $i = 1, 2$, for a percent change in bus price is the elasticity of demand for mode i , $i = 1, 2$, with respect to bus price. In a similar manner, elasticities with respect to rail prices can be determined. Note that the demand elasticities estimated in this way are not necessarily constant, but can vary over bus and rail prices.

The vector α used in the present study is that estimated as model (1) in [7]. The attributes of alternatives which are included in this model are: cost; on-vehicle time multiplied by after-tax wage; walk time multiplied by wage; wait time at first bus or rail stop multiplied by wage; transfer time multiplied by wage; and a pure auto mode preference variable which assumes a value of one for the auto alternative and zero otherwise.⁴ In addition, a family income variable is included, which assumes a value equal to family income with a ceiling of \$10,000 for the auto mode and zero for bus and rail modes. This income variable can be seen as an auto preference variable which varies with family income.

In calculating N_i , $i = 1, 2$, several concessions were made to the data. First, the sample used for calculating N_i was not drawn at random from the population. Rather, the sample was restricted to workers who lived near bus and rail lines and found it possible to take each of the three modes (bus, rail and auto). Furthermore, some individuals who met these criteria had a higher probability of being chosen than others simply because of the sampling method used. To correct these anomalies, each individual in the sample was weighted by some number which reflected his relative probability of having been chosen before the sample was taken. The sample and weights are described more fully in [5]. The sample and weights used in that paper are precisely those used in the present study.

Second, data were unavailable at the time of this study on the number of miles each person in the sample would travel on each mode if he took that mode. Because of this, the number of miles the person would travel by auto was taken as a proxy

³This specification is not exactly correct. A traveller will often use several modes during his trip. For example, he may take a bus to the rail station and take rail for the remainder of his trip. Thus, the accurate specification of N_i is:

$$N_i = \sum_{k=1}^K m_{k1}^i P_1^k + m_{k2}^i P_2^k + m_{k3}^i P_3^k$$

where m_{kj}^i = the number of miles the individual would travel on mode i if his primary mode was j . There are two reasons for not choosing this more appropriate specification. First, as explained below, the data necessary was unavailable for the sample used in this analysis. Second, it would be necessary to build and estimate a model of the choice of access mode and principal mode; that is beyond the scope of this paper.

⁴Walk, wait, and transfer times assume a value of zero for the auto alternative.

for the number of miles he would travel on bus or rail. There are obviously errors involved in this approach, but in the absence of the correct data it was necessary to use it. In any case, auto miles is probably very highly correlated with bus miles and rail miles, and there seems to be no reason to expect systematic errors.

Estimates of the number of passenger miles demanded and of demand elasticities were calculated for various price combinations. The estimates behaved as expected over the range of prices considered. Bus (rail) demand fell with its own price and rose with the price of rail (bus). Own price elasticity of bus (rail) demand rose in magnitude with own price and fell with rail (bus) price. Cross-price elasticities of bus and rail rose with prices of both services.

AVERAGE AND MARGINAL COSTS

Average costs

Merewitz and Pozdena [8] conducted a thorough study of the costs of BART rail service. Short-run cost functions were estimated, using time series data for BART. With these functions, they estimated (p. 4) that, at twenty-five million vehicle-miles per year, the average operating cost of BART rail service is 2.18c per seat-mile. Assuming that there are usually 25 per cent more passengers than seats in each vehicle during peak periods, the average operating cost per passenger-mile is 1.74c. This is the average cost figure used for BART rail service in the present study.

Merewitz and Pozdena also estimated short-run cost functions, using a time series of cross-sectional data from eleven rail services in North America. The estimated costs for BART obtained from these models were similar to those obtained with the models based on a time series of BART data. Long-run cost functions were derived from the short-run functions using a method developed by Keeler [3], but the long-run functions are not relevant for the purposes at hand.

Average cost is a function of quantity, so using only one average cost figure is equivalent to assuming that average costs are constant within the relevant range. This assumption, while inaccurate, greatly facilitates computation of the optimal fares and does not seem to be significantly incorrect for BART. According to Merewitz and Pozdena, a doubling of the quantity supplied would change average cost by about 20 per cent.⁵ For the ranges of quantities with which this study is concerned, average cost can be considered constant with little loss of accuracy.

An estimate of average total cost for A.C. Transit was presented by Lee [4]. This estimate is 2.5c per seat mile. Assuming 25 per cent more passengers than seats, the estimate of average total costs per passenger-mile is 2.0c. Lee's cost estimate was derived in a manner analogous to that used by Merewitz and Pozdena for BART rail service, so the two average cost figures are comparable (though one is average operating cost and the other is average total cost). There seems to be a consensus

⁵ At 25 million vehicle miles, 46 per cent of the operating costs are fixed and the remaining 54 per cent are variable. The fixed operating costs include all maintenance and other costs associated with operating the rail service without any traffic, and do not include any capital interest or depreciation costs. Doubling quantity decreases average fixed operating cost by 50 per cent and increases average variable operating costs by about 2.5 per cent (according to the first equation on page 28). This translates into a decrease of 21.65 per cent in average operating costs.

among researchers that constant returns to scale exist for A.C. Transit bus service [13], which means that average costs are constant.

Marginal costs

Merewitz and Pozdena [8, p. 2] estimated the marginal cost of BART service to be 1.1 cents per seat-mile. Assuming 25 per cent more passengers than seats, marginal cost per passenger-mile is 0.88c.⁶

It is questionable whether this method is proper for estimating marginal cost per passenger mile. Merewitz and Pozdena calculated the incremental cost of running a vehicle an extra mile, and divided by the number of seats in the vehicle, to obtain their estimate of marginal cost per seat mile. Dividing this figure by 1.25 to obtain marginal cost per passenger-mile is not valid unless exactly 25 per cent more passengers ride in a vehicle than there are seats. If one more passenger could ride in the vehicle with negligible effect on costs, the marginal cost per passenger mile must be approximately zero. In determining the optimal prices in the next section, both estimates of marginal costs (zero and 0.88c) are considered.

Since A.C. Transit faces constant returns to scale in terms of vehicle miles, its marginal cost is equal to its average cost: 2.0c. As with BART service, it seems that the correct estimate for marginal cost per passenger mile is perhaps zero.

THE OPTIMAL PRICES

For welfare maximisation, equations (3) and (4) must hold. Figure 1 charts the price combinations for which they do so. Curve A connects the price combinations for which equation (4) holds. Curve B connects the price combinations for which equation (3) holds if marginal costs are considered to be equal to zero for both bus and rail. Curve C connects the price combinations for which equation (3) holds if marginal costs are considered to be equal to 2.0c and 0.88c for bus and rail, respectively.

Curves A and B intersect at the point at which bus price is approximately 1.95c per mile and rail price is about 1.78c per mile. These are the estimated optimal prices when marginal costs are considered to be zero. With average costs for bus and rail at 2.0c and 1.74c respectively, there need be little or no cross-subsidisation. Any subsidisation that need occur would be from rail to bus. This analysis indicates, however, that pricing at average cost by each service would come close to attaining the pricing requisite for constrained welfare maximisation. There would be no need for an agency to administer prices and cross-subsidisation.

Curves A and C intersect with bus price approximately equal to 2.42c per mile and rail price about 1.28c per mile. These are the estimated prices when marginal costs for bus and rail are considered to be 2.0c and 0.88c respectively. In this case, the bus service would subsidise to a significant degree the operation of rail service. An agency to administer prices and cross-subsidisation would be necessary.

⁶ Marginal cost is assumed constant over the relevant range for the same reason as for constant average cost. See the third equation, p. 28, in [8].

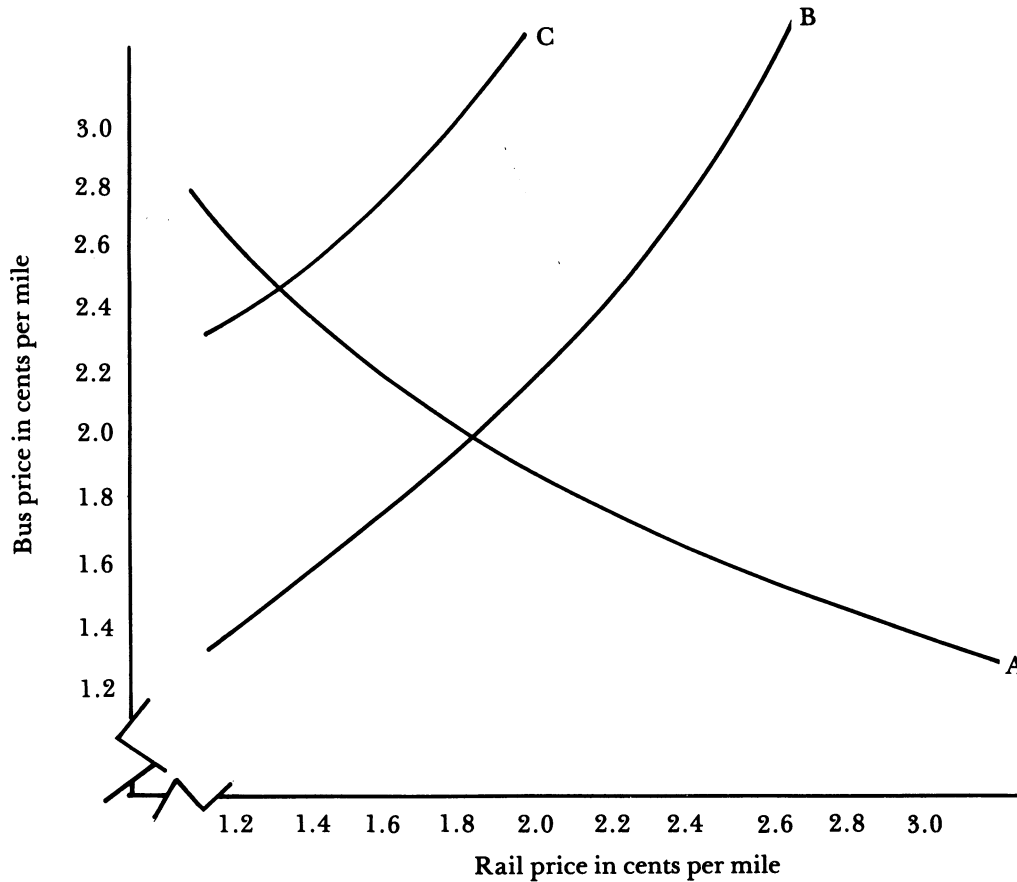


FIGURE 1

CAVEAT

Several problems arising from the preceding analysis are important and warrant discussion. First, the marginal cost figures upon which the analysis is based do not include all social costs. Two marginal cost figures were considered in the analysis for both bus and rail. Under the assumption that an extra person could ride transit without increasing the number of carriers, marginal cost was considered to be zero. Under the alternative assumption that increasing volume necessitated increasing the number of carriers, a positive marginal cost figure was calculated. Yet both these figures are too low under their respective assumptions. If extra carriers are not provided when volume increases marginally, then during peak hours the extra volume results in crowding within the carriers and increases the chances that persons waiting for carriers will be passed by and forced to wait longer to board a carrier. These social costs indicate that the true marginal cost exceeds zero. Under

the alternative assumption that extra carriers are added when volume increases, marginal cost varies between zero and the cost of an extra carrier. Perhaps the best approximation to marginal cost would be some average or expected marginal cost. This expected marginal cost would be equal to the cost of the extra carrier divided by the number of people expected to use the carrier. If the carrier were expected to be completely full, the marginal cost would equal the values used in the analysis above. If, however, the extra carrier were not expected to be full, the marginal cost would exceed the values used above.

Related to the inaccuracies in calculated marginal cost for bus and rail is the assumption in Boiteux's model that price equals marginal social cost for all private goods. In particular, the analysis assumes that the price of auto use equals its marginal social cost. Because of pollution and highway congestion during peak hours, the marginal social cost of auto use exceeds the private cost. A more appropriate model would incorporate this deviation from perfect competition and analyse the effects of various transit fare policies on auto use.

Another limitation of Boiteux's model is that, like most welfare maximisation models, it assumes that the existing distribution of income is equitable. If this assumption cannot be considered accurate, then the welfare implications of Boiteux's model are not as straightforward as the above analysis indicates. This issue is particularly important for the above analysis because, in the East Bay Area of the San Francisco Bay Area, buses serve lower income persons, generally, than do trains. Hence, subsidisation of rail by bus revenues would be a form of regressive taxation. Any determination of truly optimal fares would take into consideration the income distribution effects of various fare policies.

Lastly, the analysis is based on the assumption of a constant number of people travelling on all the modes. This assumption is necessary until a behavioural model of individual trip generation is developed. Such a model could be coupled with the logit model of mode choice to determine the number of passenger-miles demanded on each mode without assuming that the sum of demand over all modes is constant. When an adequate model of trip generation is developed, a study similar to the present work should be undertaken, and its results should then be taken more seriously.

REFERENCES

- [1] M. Boiteux: "Sur la gestion des monopoles publics astreints à l'équilibre budgétaire." *Econometrica*, 1956. Reviewed in English in [2], pp. 27-35.
- [2] Jacques H. Drèze: "Some Postwar Contributions of French Economists to Theory and Public Policy, with Special Emphasis on Problems of Resource Allocation." *American Economic Review*, Supplement, June 1964.
- [3] Theodore E. Keeler: "Railroad Costs, Returns to Scale and Excess Capability." *The Review of Economics and Statistics*, May 1974, pp. 201-208.
- [4] Douglas B. Lee, Jr.: *Cost Components for Selected Public Transportation Modes in the San Francisco Bay Area*. Institute of Urban and Regional Development, University of California, Berkeley, January 1974.
- [5] Daniel McFadden: "BART Patronage and Revenue Forecasts for Flat Fares." Working Paper No. 7407, Travel Demand Forecasting Project, Institute of Transportation and Traffic Engineering, University of California, Berkeley, December 1974.

- [6] Daniel McFadden: "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, editor: *Frontiers of Econometrics*, Academic Press, 1973.
- [7] Daniel McFadden: "The Measurement of Urban Travel Demand." *Journal of Public Economics*, 3, 1975.
- [8] Leonard Merewitz and Randall Pozdena: "A Long-Run Cost Function for Rail Rapid Transit Properties." Working Paper No. 240, Institute of Urban and Regional Development, University of California, Berkeley, September 1974.
- [9] Antti Talvitie: "Models for Transportation Systems Analysis: An Analytical Approach." Unpublished paper, University of Oklahoma. 1975.
- [10] S. L. Warner: "Stochastic Choice of Mode in Urban Travel: A Study in Binary Choice." Ph.D. dissertation, Department of Economics, Northwestern University, 1967.
- [11] Peter L. Watson: "Predictions of Intercity Modal Choice from Disaggregate, Behavioral, Stochastic Models." Highway Research Board, *Highway Research Record*, No. 446, pp. 28-35, 1973.
- [12] Richard B. Westin: "Predictions from Binary Choice Models." *Journal of Econometrics*, April 1974.
- [13] Personal communication with T. Keeler and K. Small.