Optimal Pricing of Urban Trips with Budget Restrictions and Distributional Concerns

By David M. Nowlan*

1. Introduction

Urban travel raises remarkably similar policy issues from one large city to the next. These arise from perceptions of excessive congestion, especially in the central areas, overuse of the car, land-use sprawl related to under-priced transport facilities, disproportionate commuter travel at peak times, public transport deficits, and concerns about the effect of transport prices on low-income residents. To address these problems, pricing policies are widely advocated. Current attention has focused in particular on the potential use of pricing to optimise traffic congestion at peak times, a focus encouraged by the emerging technological possibility of establishing finely-tuned user charges for urban road use. Recent work, such as that reported on by Newbery (1990), has helped confirm the general view that congestion costs are high and that the private costs of commuting by car are well below the public costs.

Underpricing of urban roads is often used as a reason for keeping public transport prices low. This in turn leads to transport deficits which financially straitened cities have difficulty supporting. Attempts to raise transport prices to cope with the financial problem frequently result in protests by and on behalf of the urban poor who claim a disproportionately high use of public facilities. Although these issues are all issues of urban travel, policy towards them is typically undertaken piecemeal. Policies for improved car pricing focus on the costs of traffic congestion, including pollution from cars. Public transport policy is dominated by the financial concerns of transport authorities; and the political reaction to price proposals is often conditioned by considerations of distributional equity among the population.

* Department of Economics, University of Toronto. An earlier version of this paper was presented at the Sixth World Conference on Transport Research in June 1992. Financial support for this research was provided in part by the Humanities and Social Sciences Committee of the University of Toronto under a programme of the Social Sciences and Humanities Research Council of Canada.
This piecemeal approach ignores the interdependencies among costs, finance and distributional concerns. Policies that appear optimal from the perspective of any one may be distinctly inferior when viewed from the perspective of another. This has been increasingly recognised by those who have proposed an "integrated approach" (May and Gardner, 1990) or a "package approach" (Jones, 1991) to urban transport issues. The idea of an integrated approach has been encouraged in part by the perception that motorists will give greater support to road charges if the revenue obtained is used within the transport sector. With its focus on a better understanding of the interdependencies among the different issues, this paper is very much in the spirit of the integrated approach.

A single public sector, the transport sector, is assumed to provide infrastructure for private-vehicle travel and infrastructure and operating services for public vehicles. Efficient pricing requires, as in standard models of urban transport, that travellers face the appropriate marginal cost, including congestion or other external costs, of their use of a service. Over the whole transport sector, however, a budget constraint is imposed. There is an upper limit placed on the amount of general government revenue available to meet a deficit in the transport sector; if that upper limit is zero, the sector as a whole is required to break even. In addition, the sector may be required to take into account the effect of its pricing policies on residents at different income levels. Given these restrictions, the objective of the public transport sector is to set prices on its different services and to invest in infrastructure in such a way that the aggregate well-being among the urban community is maximised.

Some of the results in this paper will be familiar, since over the years the transport economics literature has addressed all three of the issues of marginal-cost pricing, budget restrictions and distributional concerns, although the last has been dealt with less formally than the others. Emphasis will be placed on those results and interpretations that are new or at least less familiar. These include the importance for price balancing, in the face of a budget constraint, of cross-revenue elasticities among transport services and the role of both price adjustments and investment adjustments in meeting distributional goals. The most general result that emerges from the analysis is that the optimal price for a service may be decomposed into three multiplicative terms: a standard term representing marginal costs including external costs; a term representing the budget constraint; and another term representing the distributional characteristics of the users of a particular service.

This result is shown as equation (18) in Section 5 (or as equation (19) under slightly different assumptions). It collapses nicely to standard marginal cost pricing when the budget constraint is not binding and when the distributional characteristics of the service users are just the community average.

The framework of the paper is described in the next section and the first-order optimality conditions are set out in Section 3. Section 4 analyses the role played by the transport sector's budget constraint and Section 5 discusses the use of transport prices and infrastructural investment to address distributional issues. A concluding section provides a brief summary of the results and comments on the implications for practical policy of the first-order efficiency conditions that are derived below.
2. The Basic Model

The urban setting for the model consists of a given and finite number of individual residents. A transport sector provides an assortment of services to this population, services that are differentiated by mode, by time of travel and by location of travel. Differentiation among services may be coarse-grained or fine but, by assumption, the agency responsible for the transport sector is able to set separate prices for each service, prices that are uniform to all consumers. This assumption of uniform prices means that any distributional concerns or financial constraints to which prices respond will appear in price differences among services and not through different prices for different individuals for a given service.

Transport modes include publicly provided road rights-of-way for private vehicles along with various public or publicly controlled services such as buses, commuter rail and subway. The time of travel may be divided simply into peak and off-peak travel or it could be more finely divided. Service location could be any segment of a road or public transport service that can be priced separately. The question of optimum partitioning of services for pricing purposes, whether by mode, time of day or location, is not addressed in this model; the model is based on a given partition. It would be an interesting and reasonably straightforward extension to study optimum service partitions, with the improved efficiency of finer partitioning traded off against the additional costs of a more complicated pricing structure.

An individual resident’s use of one of these services once is a “trip.” The subscript \( i \) will be used to index the services, where \( i \) ranges over all modes, all time periods and all locations. Individual residents will be differentiated by a superscript \( h \). Thus, an individual’s consumption per time period (annually, say) of type-\( i \) trips is given by \( x^i_h \), and the vector of all types of trip by the individual is \( X^h \), the sum of all type-\( i \) trips is \( x_i \), and the vector\(^1 \) of all trip types by all people is \( X_i \).

Each individual has a utility function, the value of which is positively related to his or her consumption of trips and consumption of a numéraire good \( z^i \). The numéraire good, which may be thought of as “income”, has a normalised price of 1. Because of possible congestion or other externalities associated with the use of the transport facilities, an individual’s utility function also includes the total use of each service, which is the vector \( X_p \) and the infrastructural or investment resources (in numéraire terms) that have been devoted to the provision of services by the transport sector. Infrastructural costs that can be associated with a specific service \( i \) will be denoted \( f_i \), with the vector of all such investments given by \( F_i \). Joint or overhead costs that cannot be assigned to any particular service are shown by the variable \( J \). Thus the utility of any individual \( h \), \( u^h \), is given by

\[
    u^h = u^h(z^h, X^h_p; J, F_i)
\]

with

\(^1\) Vector variables are shown in bold, upper case italics with a subscript indicating the counter over which the vector ranges (\( X_i \) for example); scalar variables are in italics, not bold type, and either lower case, if they are one of a larger set (\( x_i \) for example), or upper case if they are singletons (\( J \) for example).
\[
\frac{\partial u^h}{\partial t^h} \geq 0; \quad \frac{\partial u^h}{\partial x^h_i} \geq 0; \quad \frac{\partial u^h}{\partial x^h_i} \leq 0; \quad \frac{\partial u^h}{\partial J} \geq 0; \quad \text{and } \frac{\partial u^h}{\partial F_i} \geq 0.
\]

By using in the utility functions a vector of total trips undifferentiated by consumer, the usual "anonymity" assumption has been adopted, that congestion (or other forms of externality such as pollution) is affected by the number of trips of a given type but not by who the trip-makers are.

A social welfare function, \( W \), determined outside the transport sector, provides the necessary relationship among the utilities of all individuals in the area. This is assumed to have the general, separable form:

\[
W = W(u^1, u^2, \ldots, u^h, \ldots).
\]

The total cost of providing transport services is given by a cost function, \( C \), that shows in numeraire terms the cost of infrastructure, \( J \) and \( F_i \), and the operating cost of services \( X_i \). Thus,

\[
C = C(X_i, J, F_i).
\]

(Anonymity enters again in this formulation since costs are affected by total trips of type \( i \) but not by who makes the trips.)

There is assumed to be an overall resource constraint in this urban area which restricts to a given amount \( R \) the total use of resources for transport services and all other goods \( z \), where \( z \) is the sum of all \( z^h \). \( R \) could be made a function of the resources devoted to transport services to reflect the possibility that improved transport facilities might enhance production, but the simpler formulation has been chosen in order to focus on other issues.

A revenue constraint on the transport sector is now added. Total revenue available to the sector is the sum of the revenue from the priced transport services plus any transfer that might exist from general public funds. The price of each service, a policy variable under the control of the sector, is given by \( p_i \); the vector of all prices is \( P \). Thus, total service revenue is given by

\[
\sum_i p_i x_i.
\]

With the transfer to the sector denoted by \( T \), the revenue constraint may be written as

\[
\sum_i p_i x_i + T \geq C(X_i, J, F_i).
\]

The transfer amount \( T \) is not under the control of the transport sector, and could be zero.

The services consumed by each traveller, \( x_i^j \), are assumed to be utility-maximising amounts for the individual, given the prices set by the transport sector, the infrastructural resources devoted to the sector and the aggregate levels of demand for each service. This leads to aggregate demand functions of the sort \( x_i = x_i(P_i, J, F_i) \). It is more convenient to work with the inverse demand functions \( p_i = p_i(X_i, J, F_i) \) which exist, given the direct demand functions, provided that the Jacobian determinant of the matrix

\[
\begin{bmatrix}
\frac{\partial x_i}{\partial p_i}
\end{bmatrix}
\]

is not zero in the neighbourhood of a demand equilibrium.
Optimal Pricing of Urban Trips

D. M. Nowlan

The inverse demand functions reflect the fact that the price for any given service is uniform across users. Given this, it is each $x_i$, the aggregate use of a service, and not each $x^h_i$, the use by an individual, that the transport sector can affect by varying price $p_r$. $x_i$ is clearly the appropriate quantity variable in the model, but because the utility functions have been defined in terms of individual consumption levels, $x^h_i$, a relationship between the $x_i$ and the $x^h_i$ must be established. This is done by regarding each individual's consumption of service $i$ as a share of total service consumption. An incremental share value, $s^h_i$, is then defined such that

$$s^h_i = \frac{dx^h_i}{dx_i}$$

at the optimum points of the model, with

$$\sum_h s^h_i = 1.$$ 

This share variable turns out also to be a useful way of addressing the issue of distributional concerns with uniform prices. If fully individualised prices across all services were permitted, then not only would the share variable be unnecessary but the welfare costs of adjusting to binding financial or distributional constraints would in general be lower.

In summary, the basic model consists of a social welfare function, $W$, to be maximised subject to two constraints, an overall resource constraint for the community and a revenue constraint for the transport sector:

**maximise** \( W = W(u^1, u^2, \ldots, u^h, \ldots) \) with respect to \( z^h \in Z^h \), \( x_i \in X_i \), \( J \) and \( f_j \in F_j \)

**subject to** \( R - \sum_h s^h_i = C(X_i, J, F_i) \geq 0 \)

and \( \sum_i p_i x_i + T - C(X_i, J, F_i) \geq 0. \)

---

3. First-Order Conditions for Optimal Prices

With shadow prices introduced for each of the constraints, $\lambda_1$ and $\lambda_2$ respectively, the Langrangian expression to be maximised may be formed:

$$L = W(u^1, u^2, \ldots, u^h, \ldots) + \lambda_1[R - \sum_h z^h - C(X_i, J, F_i)] + \lambda_2[\sum_i p_i x_i + T - C(X_i, J, F_i)]$$

First-order necessary conditions for a level value of this function may be written for four sets of variables: (1) the $z^h$ for all $h$; (2) the $x_i$ for all $i$; (3) the infrastructural cost variables $f_j$ for all $i$, and $J$; and (4) the shadow prices $\lambda_1$ and $\lambda_2$.

$$\frac{\partial L}{\partial z^h} = \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial z^h} - \lambda_1 = 0, \forall h \quad (1)$$

$$\frac{\partial L}{\partial x_i} = \sum_h \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial x_i} s^h_i + \sum_h \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial x_i} \frac{\partial C}{\partial x_i} + \lambda_1 \left( p_i + x_i \frac{\partial p_i}{\partial x_i} + \sum_j x_j \frac{\partial p_j}{\partial x_i} - \frac{\partial C}{\partial x_i} \right) = 0, \forall i \quad (2)$$

$$\frac{\partial L}{\partial f_j} = \sum_h \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial f_j} - \lambda_1 \frac{\partial C}{\partial f_j} + \lambda_2 \sum_j \frac{\partial p_j}{\partial f_j} = 0, \forall i \quad (3)$$
\[
\frac{\partial L}{\partial J} = \sum_h \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial J} - \lambda_1 \frac{\partial C}{\partial J} + \lambda_2 \sum_i \frac{\partial p_i}{\partial J} - \lambda_2 \frac{\partial C}{\partial J} = 0 \quad (3a)
\]

\[
\frac{\partial L}{\partial \lambda_1} = R - \sum_h z^h - C = 0 \quad (4)
\]

\[
\frac{\partial L}{\partial \lambda_2} = \sum_i p_i x_i + T - C = 0 \quad (4a)
\]

These first-order conditions are sufficient to represent a welfare optimum provided that the global conditions for a maximum and the local second-order conditions are satisfied. For decreasing-cost services, these second-order conditions require among other things that the relevant demand curves have a sufficiently large absolute slope relative to the cost curves, a condition that is assumed to be met.

Equation (1) describes the way in which distributional issues are taken care of independently of transport pricing policies. These first-order equalities require the numeraire good, \( z^h \) or "income," to be distributed in such a way that the social welfare value of an incremental unit of the good is equalised across every member of the community. That this equalisation exists is an important assumption made, implicitly if not explicitly, in most optimal-pricing models. Its relaxation is dealt with in Section 5.

The equalized \( W \) value of a unit of numeraire is seen from equation (1) to be \( \lambda_1 \), the shadow price on the first constraint. Without loss of generality, \( \lambda_1 \) will be set equal to 1 (this simply adjusts the measurement scale of the \( W \) function). If equations (2), (3) and (3a) are divided through by equation (1) they can be cleared of the partial derivatives of \( W \) and expressed in terms of numeraire-good partial utilities. Individual utilities in numeraire units will be written as \( v^h \), so that

\[
\frac{\partial v^h}{\partial x^i} = \frac{\partial u^h}{\partial x^i} \frac{\partial u^h}{\partial x^h}.
\]

Applying this definition to equation (2) and using equation (1) yields the following:

\[
\sum_h \frac{\partial v^h}{\partial x^i} s^h_i = -\sum_h \frac{\partial v^h}{\partial x^i} \frac{\partial C}{\partial x^i} + \lambda_2 \left( R + \sum_i x_i \frac{\partial p_i}{\partial x^i} + \sum_j x_j \frac{\partial p_j}{\partial x^j} - \frac{\partial C}{\partial x^j} \right) \forall i \quad (5)
\]

The left-hand side of equation (5) consists of a sum of terms involving the marginal utility in numeraire units of each individual's consumption of service \( i \). It is this marginal utility, \( \frac{\partial v^h}{\partial x^i} \), that each person consuming service \( i \) will equate to the price \( p_i \) of the service (for any community member not consuming service \( i \), \( p_i > (\frac{\partial v^h}{\partial x^i}) \) and \( s^h_i = 0 \). Since this price is the same for everyone, the left-hand side may be written as

\[
\sum_h p_i s^h_i.
\]

With

\[
\sum_h s^h_i = 1.
\]

\[\text{For convenience, these conditions are written as if the community resource constraint and the sectoral budget restriction are equalities and not inequalities. With inequalities, the Kuhn-Tucker conditions are}
\]

\[
\lambda_1 \left( R - \sum_h z^h - C \right) = 0 \quad \text{and} \quad \lambda_2 \left( \sum_i p_i x_i + T - C \right) = 0
\]

\[\text{The case where} \lambda_2 = 0 \quad \text{and the sectoral budget just balanced or was in surplus will be discussed below.}\]
this reduces simply to \( p_i \), the price of the service. Thus, (5) may be rewritten as

\[
p_i = -\sum_h \frac{\partial v_h}{\partial x_i} + \frac{\partial C}{\partial x_i} - \lambda_2 \left( p_i + x_{i1} \frac{\partial p}{\partial x_i} + \sum_{m} x_{ij} \frac{\partial p}{\partial x_i} - \frac{\partial C}{\partial x_i} \right), \quad \forall i
\]

(6)

Except for the \( \lambda_2 \) term, equation (6) is the familiar marginal cost condition for optimal transport pricing. For the moment, assume that the financial constraint on the transport sector is not binding so that \( \lambda_2 = 0 \). Equation (6) then says that the optimal price for transport service \( i \) is a price that equals the marginal congestion cost to the urban community of an extra trip, the term

\[-\sum_h \frac{\partial v_h}{\partial x_i},\]

plus the marginal operating cost of a trip, \( \partial C/\partial x_i \). Notice that the congestion-cost term is a sum of costs to all individuals whose welfare is being taken into account, including those, if any, who do not consume service \( i \). "Congestion" thus includes the costs of externalities, including pollution costs, over the whole jurisdiction and not just the cost of time delays experienced by travellers.

Equation (6) raises an interesting issue that has been more of an annoyance than a stumbling block for the theory of optimal congestion prices. The first term on the right-hand side is a sum of congestion effects across all consumers. If person \( h \) takes into account some sort of “self-congestion” represented by the \( h \)th component of this sum, then the utility adjustment to the price of the service must be not simply the marginal utility of the service, \( \partial v_h/\partial x_i \), but this marginal utility minus share \( s_i \) of the \( h \)th term of the congestion sum. This in turn means that in general there could not be a single, uniform optimal price. Optimal prices would have to be individualised. Strotz (1964) came to this conclusion in an early contribution to the literature on optimal urban transport pricing, and resolved the difficulty by assuming that for each person congestion was a parameter the size of which was not affected by the individual’s travel decisions. If all members of the community are identical, then uniform prices will be optimal and the problem does not arise. But if heterogeneity is to be retained among the members of the community, then the Strotz assumption or some equivalent must be introduced. Some writers simply assume that the self-congestion effect is so small that the costs of incremental travel are more or less the same for everyone. More recent literature, especially in club theory, uses in similar circumstances a measure-theoretic formulation that assumes a dense continuum of consumers with a congestion integral that remains constant no matter how many individuals are removed from under the integration sign.

In the present model, the marginal congestion cost plus the marginal operating cost of service \( i \), the first two terms on the right-hand side of equation (6), will be labelled \( \Gamma_i \). Thus, the first result of the model says that, with no binding financial constraint on the transport sector and with redistribution taken care of outside the transport sector, the optimal price for a trip is given by

\[
p_i = \Gamma_i, \quad \forall i
\]

(7)

From equations (3) and (3a) the optimal marginal investment rules for service-specific and for overhead infrastructural expenditures, \( f_i \) and \( J \) respectively, may be derived. Let
both the cost $C$ and the infrastructural expenditures be measured in numeraire units. The terms $\partial C/\partial f_i$ and $\partial C/\partial J$ will then be identically equal to 1 and the optimality conditions become

$$1 = \sum_{h} \frac{\partial v^h}{\partial f_i} - \lambda_2 \left( 1 - \sum_{j} x_j \frac{\partial v^h}{\partial f_j} \right) \forall i$$  \hspace{1cm} (8)

and

$$1 = \sum_{h} \frac{\partial v^h}{\partial J} - \lambda_2 \left( 1 - \sum_{j} x_j \frac{\partial v^h}{\partial J} \right) \hspace{1cm} (9)$$

These conditions say that a numeraire unit of spending on either overhead or service-specific infrastructure should at the margin equal one numeraire unit of improved utility across the community (because of reduced congestion or other positive externalities), with a correction if the budget constraint is binding, that is, if $\lambda_2 \neq 0$. This correction says that if a numeraire unit of spending generates exactly a numeraire's worth of additional revenue through higher prices with constant trip volumes, then the budget constraint is irrelevant and the basic optimality conditions remain, but if less than a unit's worth of revenue is generated, the aggregate utility improvement must be greater than one numeraire unit. If more than a unit's worth of revenue is generated per unit additional infrastructural spending, then the community's aggregate utility improvement should be less than one unit at the margin. These investment optimality conditions will be used in the next section which examines the budget constraint in more detail.

4. The Budget Constraint

If transport services were priced according to equation (7) and if the revenue generated by these prices plus any subsidy to the transport sector were insufficient to cover the costs of the sector, then the budget constraint would be binding, $\lambda_2 > 0$ and revenue-constrained optimal prices would have to take into account the final term on the right-hand side of equation (6). This equation is easier to interpret if it is rewritten with the partial derivatives of the inverse demand functions replaced by elasticities and cross-elasticities of those functions. Thus,

$$x_i \frac{\partial p_i}{\partial x_i} = \frac{p_i}{\bar{p}_i} x_i \frac{\partial p_i}{\partial x_i} \quad \text{and} \quad x_j \frac{\partial p_j}{\partial x_i} = \frac{p_j}{\bar{p}_j} x_j \frac{\partial p_j}{\partial x_i}$$

may be written as

$$\bar{p}_i \bar{p}_j \xi_{ii} \quad \text{and} \quad \frac{p_j}{\bar{p}_j} x_i \frac{\partial p_j}{\partial x_i} \xi_{ji},$$

where $\xi_{ii}$ and $\xi_{ji}$ are respectively the elasticity of $p_i$ with respect to $x_i$ and the cross-elasticity of $p_j$ with respect to $x_i$. Also, $\xi_{C_i}$ is defined as the elasticity of the cost function with respect to $x_i$. With these changes, equation (6) becomes

$$p_i \left( 1 + \lambda_2 + \lambda_2 \xi_{ii} + \lambda_2 \sum_{j \neq 1} \frac{x_j p_j}{x_i \bar{p}_i} \xi_{ji} - \lambda_2 \frac{C_i}{x_i \bar{p}_i} \xi_{C_i} \right) = -\sum_{h} \frac{\partial v^h}{\partial x_i} + \frac{\partial C}{\partial x_i}, \forall i$$  \hspace{1cm} (10)
Optimal Pricing of Urban Trips

D. M. Nowlan

The left-hand side of (10) may be more readily interpreted if it is rewritten to incorporate within square brackets an expression that represents the overall revenue elasticity of changes in service levels $x_i$:

$$
\rho_i \left( 1 + \lambda_2 \left[ 1 + \xi_{ii} + \sum_{j \neq i} \frac{x_{ij}}{x_i} \xi_{ij} - \frac{C}{x_i} \xi_C \right] \right).
$$

To see that the square bracketed terms measure the revenue elasticity of changes in $x_i$, consider a reduction in the use of service $i$ induced by an increase in its price. The elasticity of the inverse demand function $p_i = p_i(X, J, F)$ shows the percentage increase in price that will accompany a unit percentage decrease in service consumption; this is $\xi_{ii}$. Similarly, the inverse demand functions $p_j = p_j(X, J, F)$ will provide the cross-elasticities $\xi_{ij}$ showing the percentage change in the price of service $j$ that will keep $x_j$ constant in the face of an increase in $p_i$. If the services $j$ and $i$ are substitutes, the cross-elasticities will be negative; if complements, they will be positive.

Suppose, in an effort to raise revenue, $x_i$ is reduced by 1 per cent through an increase in $p_i$. The first term in the square bracket, "1", says that one effect will be a revenue decrease by the same percentage as the reduction in $x_i$. The second term shows an offsetting revenue increase by the percentage increase in price $p_i (\xi_{ii} < 0)$. Taken together the two terms show the percentage revenue effect in the $i$ service of a 1 per cent reduction in $x_i$.

The third term in the square brackets shows through the cross-elasticities $\xi_{ij}$ the percentage effect on prices in the $j$ services as $x_i$ is cut back by 1 per cent (with substitute services, $\xi_{ij} < 0$). Each of these price effects is weighted by the relative revenue from the $j$ service to the revenue from the $i$ service. Taken together, the components of the third term in the square bracket above show the revenue changes from other services of an $x_i$-increase in the $i$ service, as proportions of $i$-service revenues. The final term within the square brackets shows the decrease in operating costs associated with a decrease in $x_i$, again as a percentage of $i$-service revenue. The sum of the terms in the square bracket therefore shows the total revenue change, as a proportion of $i$-service initial revenue, that will accompany the price changes necessary to reduce $i$-service consumption by 1 per cent while keeping all other service use constant. The square bracket thus shows the total revenue elasticity associated with changes in service use $i$ which are induced by changes in the price of $i$. Using $\Xi_i$ to indicate this total revenue elasticity associated with service $i$, the left-hand side of equation (10) may be written as $p_i (1 + \lambda_2 \Xi_i)$ and the whole equation reduced to

$$
\rho_i = \sum_{j \neq i} \Gamma_{ij} \frac{\partial x_j}{\partial x_i} + \frac{\partial C}{\partial x_i} = \frac{\Gamma_{ii}}{\Phi_i}, \forall i
$$

(11)

where $\Phi_i = 1 + \lambda_2 \Xi_i$. If the revenue constraint is not binding, $\lambda_2 = 0$ and $\Phi_i = 1$. Equation (11) then collapses to (7), with optimal prices equal to marginal costs.

The total revenue elasticities, $\Xi_i$, must be zero or negative if the second-order conditions of the optimising problem are to be met. In these circumstances, the prices $p_i$ will be at or above the marginal costs. (If $\Xi_i$ were positive, equation (11) would entail a
p_i below marginal cost which could never be optimal.) Given $\lambda_2$, the ratio of $p_i$ to marginal cost $\Gamma_i$ is directly related to the absolute size of the revenue elasticity $\Xi_i$. If the (negative) revenue elasticity is larger absolutely, the denominator of (11), $\Phi_i$, is smaller and $p_i$ larger relative to marginal cost. In a financially constrained transport sector, optimal prices should exceed marginal costs in proportions that depend on the revenue elasticities of price-induced service reductions in the different services. This is an intuitively plausible rule and bears a clear resemblance to the Ramsey inverse-elasticity mark-up rule (Ramsey, 1927, and Boiteux, 1956) that is well established in the taxation and regulatory economics fields.

In fact, equation (10) or (11) reduces exactly to the standard form of the Ramsey rule if all cross-elasticities are zero and if marginal costs are regarded simply as marginal operating costs with no congestion costs. In these circumstances, equation (10) becomes

$$\frac{\Delta C}{\Delta x_i} = \frac{\lambda_2}{1 + \lambda_2} \varepsilon_{ii}.$$  

With zero cross-elasticities, $\xi_{ij}$, the elasticity of the inverse demand function is just the inverse of the elasticity of the original demand function, $\xi_{ij} = 1/\varepsilon_{ij}$, where $\varepsilon_{ij}$ is the standard elasticity of demand for service $i$ with respect to its own price. Equation (10) then becomes

$$\frac{\Delta C}{\Delta x_i} = \frac{\lambda_2}{1 + \lambda_2} \frac{1}{\varepsilon_{ii}}$$

for all $i$, the standard Ramsey form with no cross-elasticities. Walters (1968, pp. 115-17) has an early discussion of the effect on optimal road prices of a revenue-constrained transport authority. Forsyth (1977) also recognises the potential importance of revenue constraints on transport pricing and applies optimal taxation theory to the problem, using a model based on Walters’ earlier work. Neither Walters nor Forsyth incorporate congestion into their revenue-constrained models, and the discussion in both papers is marred by the implicit assumption that the elasticities of the inverse demand functions are simply the inverse of the elasticities of the normal demand functions. Implicit function theory (see for example Friedman, 1971, pp. 236-42) may be used to show that the relationship between the two elasticities is more complex. Using the notation in this paper and assuming only two services, $i$ and $j$, then

$$\xi_{ii} = \frac{1}{\varepsilon_{ij}\varepsilon_{ij}} \text{ and } \xi_{jj} = \frac{-e_{jj}}{\varepsilon_{ii} \varepsilon_{jj} - \varepsilon_{ji} \varepsilon_{ij}}.$$  

A formulation for optimal prices in a multi-service regulated firm similar to equations (10) and (11) above (but without congestion terms) may be found in Brown and Sibley (1986, pp. 194-97).

The cross-elasticity terms in equations (10) and (11), which are bound to be important in a multi-service transport sector, serve to draw the sector’s attention to the need to consider the prices of other services as the price of any given service, $i$, is being adjusted for revenue purposes. Suppose, for example, that public transport is divided simply into
peak and off-peak services, and that the transport authority decides to raise revenue by
increasing the peak-service price per trip. If this were the only price change, many riders
would simply shift their trip to an off-peak time by moving their trip time just across the
peak/off-peak boundary. The revenue-constrained transport authority would be alerted
by the cross-elasticities of revenue in the optimal-pricing equations to raise off-peak
prices at the same time. If this in turn would cause consumers to shift to cars, then price
increases for road services should be part of the package. With public transport more
finely divided into peak, shoulder and off-peak services, a budget-inspired price increase
in peak-period trips should be accompanied by companion increases in the prices of
shoulder and off-peak trips if cross-service revenue and demand elasticities are high. If,
on the other hand, the cross-service revenue elasticity of some service such as town centre
parking was very small while its own service revenue elasticity was high (that is, its
demand elasticity was low) then the transport sector would find that parking prices in the
town centre could be raised substantially to generate revenue with minimal welfare costs.
In some cases, a revenue-enhancing price increase might be accompanied by price
reductions elsewhere in the system. Suppose the transport sector seeks to raise revenue
by increasing the price of park-and-ride services, then it should also be contemplating
some reduction in the price of complementary services such as the transport services used
by those who park and ride.

Consider as a further example two services, a bus route and a subway line, serving part
of a city, and suppose that the transport authority, faced with a more restrictive budget,
proposes to raise the price of the bus service. If the demand elasticity for bus travel is very
high, this might seem to suggest that the revenue elasticity would be low and that a price
increase for that service would be unwise. But the high elasticity might simply be
signalling that users can easily switch to the subway, a substitute service. The optimal
response of the authority would then be to raise the prices of both services. If all
passengers in the corridor served by the two services were captive, but if many of the users
viewed the services as substitutes, then the new optimal prices should be set above
marginal costs by whatever proportions were necessary to keep constant the number of
riders on each service. This is a general result. If all users of the services under the control
of the transport sector were captive, in the sense that differential price changes for services
might alter the demand for each service but not aggregate trip demand, then price increases
above previously optimal levels induced by a tighter budget constraint should apply to all
services in proportions such that the relative service demands remain unchanged. A
totally captive user group is highly unlikely, of course. In the face of price increases,
passengers would probably slip away to services not under control of the integrated sector,
or they would travel less. While precise balance among prices is defined by equations (10)
and (11) only when prices are optimal, the structure of those equations can usefully direct
the transport sector's attention to the issues that should be considered when price changes
are being contemplated for the purpose of balancing the sector's budget.

The question still to be addressed is, when will the budget constraint be binding; when
will $\lambda_2 > 0$? For any given subsidy level, $T$, whether or not the budget constraint is binding
depends principally on the responses of the utility levels of the individuals in the
community to changes in the congestion-producing \((X_i)\) and the congestion-easing \((F_i\) and \(J)\) terms. Suppose the transport sector is required to cover its costs, so \(T = 0\). In this case, if the utility functions of all individuals are homogeneous of degree zero or less in the arguments \(X_i, F_i\) and \(J\) then marginal cost prices will provide at least enough revenue to cover the costs of the sector, and these prices will be optimal.\(^3\) If, however, the utility functions have greater than zero-degree homogeneity then \(\lambda_2 > 0\) and the set of optimal prices will depart from marginal costs in ways discussed above.

These propositions may be demonstrated by noting that the degree of homogeneity of any \(v^h\) (utility in numeraire terms) with respect to \(X_i, F_i\) and \(J\) depends on the sign of the following expression:

\[
X_i \frac{\partial v^h}{\partial X_i} + F_i \frac{\partial v^h}{\partial F_i} + J \frac{\partial v^h}{\partial J}.
\]

If this is zero or negative then the utility costs of congestion across the whole transport sector may be said to be constant or increasing; in this case, equiproportionate increases in all \(X_i\), all \(F_i\) and \(J\) would lead to constant or lower utility levels. If, however, this expression is positive, then a form of decreasing costs exists; equiproportionate increases in all \(X_i\), all \(F_i\) and \(J\) would lead to higher utility levels.

Suppose utility congestion costs are constant, so that the expression in the above paragraph is equal to zero. From this, each

\[
\frac{\partial v^h}{\partial x_i} = -\lambda_2 \left( \sum_i f_i \frac{\partial v^h}{\partial f_i} + J \frac{\partial v^h}{\partial J} + \sum_i x_i \frac{\partial v^h}{\partial x_i} \right).
\]

This in turn may be substituted into (10) or (6) to yield

\[
\sum_i x_i \frac{\partial v^h}{\partial x_i} = \sum_i f_i \frac{\partial v^h}{\partial f_i} + J \frac{\partial v^h}{\partial J} + \sum_i x_i \frac{\partial C}{\partial x_i}
\]

(again letting \(\Phi\) stand for the bracketed price adjustment term in equation (10)). The investment optimality conditions, equations (8) and (9), may now be used to replace the terms

\[
\sum_i \frac{\partial v^h}{\partial f_i} \text{ and } \sum_i \frac{\partial v^h}{\partial J}
\]

in the above equation, yielding

\[
\sum_i x_i \frac{\partial v^h}{\partial x_i} = \sum_i f_i \left[ 1 + \lambda_2 \left( 1 - \sum_i x_i \frac{\partial p}{\partial f_i} \right) \right] + J \left[ 1 + \lambda_2 \left( 1 - \sum_i x_i \frac{\partial p}{\partial J} \right) \right] + \sum_i x_i \frac{\partial C}{\partial x_i}
\]

(12)

Equation (12) was derived assuming constant utility returns to scale in service levels and infrastructural expenditures. If \(\lambda_2 = 0\), \(\Phi_i = 1\) and the left-hand side of (12) becomes total sectoral revenue. Still with \(\lambda_2 = 0\), the right-hand side becomes the sum of service-specific infrastructural costs plus overhead costs plus operating costs. This is strictly true only if marginal operating costs equal average operating costs at the equilibrium (that is, local constant returns to scale) so that the final term on the right-hand side of (12) equals total operating costs. Constant marginal operating costs — a common assumption —

\(^3\) This is the multi-service analogue of the zero-homogeneity condition on a single-good congestion function introduced by Oakland (1972).
would give this result. If marginal operating costs are above or below average operating costs there will be inframarginal surpluses or losses respectively associated with the final term of (12). If infrastructure, $F_i$, and overhead, $J$, were measured in natural units rather than numeraire units, the condition for budget balance with optimum prices and constant utility costs of congestion would be that costs must equal

$$\sum_i f_i \frac{\partial C}{\partial f_i} + j \frac{\partial C}{\partial J} + \sum_i x_i \frac{\partial C}{\partial x_i}$$

which is the condition that the overall cost function, $C(X, f, J)$, and not just operating costs, exhibits constant returns to scale in the neighborhood of the optimum. In the present case, $\lambda_2 = 0$ is a balanced-budget solution to equation (12). With constant congestion costs in the sense defined above, marginal cost pricing yields total revenue that exactly covers all costs. The budget constraint with $T = 0$ is not binding.

If the utility functions have a degree of homogeneity less than zero in service levels and infrastructural expenditures, aggregate revenue shown on the left-hand side of equation (12) will be greater than costs shown on the right. Marginal-cost pricing in such circumstances will still be optimal but revenue will exceed costs over the whole transport sector, an excess that should be transferred back to the community in some lump sum fashion. Alternatively, if the degree of homogeneity of the utility functions is greater than zero, the left-hand side of (12) will be less than the right-hand side and revenue at marginal cost prices will be inadequate to cover sectoral costs. In this case, $\lambda_2 > 0$ and $\Phi \leq 1$ and prices must be adjusted as described earlier in this section.

In summary, if the utility costs of aggregate trip volumes remain constant or rise in the face of equiproportionate increases in trips and in infrastructure expenditures, then unadjusted marginal cost pricing is optimal and will produce aggregate revenue sufficient to cover sectoral costs, even with no transfer or subsidy to the sector. A shortfall of revenue below assignable costs, $f_i$, in some services will be covered, or more than covered, by revenue surpluses in other services, and aggregate revenue surpluses will be at least sufficient to cover overhead costs $J$. However, if utility costs fall in the face of equiproportionate increases in service levels and transport-sector spending on infrastructure—a form of increasing returns to scale—then pricing each service at its marginal cost will not yield enough revenue to cover costs in the transport sector. Such prices are efficient only if the sector receives a subsidy sufficient to make up the revenue shortfall. Otherwise, the price of each service has to rise above its marginal cost by a proportion that reflects the overall revenue elasticity associated with price and service-level changes in the sector, as shown by equation (11).

If prices for a service could be differentiated by individual user, then the response to a binding budget constraint could be much more finely tuned. Relatively higher prices

---

4 This result is similar to that of Berglas (1984) in which cross-service subsidies are optimal among local public services. He assumes that tastes with respect to congestion (service "quality" in his model) are the same across all individuals, so the congestion terms in the utility functions are suppressed. He allows numbers and types of individuals among communities to be variables. The mobility that this implies, along with his assumptions about the aggregate cost functions for local services, yield a cross-service subsidisation optimum that occurs in conjunction with constant returns in the cost function in the neighborhood of the equilibrium.
would be assigned to those users with low demand elasticities (high revenue elasticities); in essence, the captive users would bear the burden of the price increases. The outcome would be better in social welfare terms than the outcome with uniform prices, an apparent paradox that is accounted for by the assumption, reflected in equation (1), that the numeraire is distributed so that its marginal welfare value is the same for everyone no matter what the structure of efficiency prices. Because of this, the distributional burden of individual pricing is irrelevant, just as the distributional burden of differential service pricing (relative to marginal costs) is irrelevant with uniform prices. That this is unlikely to be the case in practice leads to the issue discussed in the next section on distributional considerations.

5. Distributional Considerations

As long as equation (1) is satisfied transport pricing should not have to respond to distributional concerns, at least not to concerns that are based on an acceptance of the social welfare function W. To put this another way, the efficiency conditions with and without budget constraints derived in the above two sections (frequently called Pareto-efficient prices) describe socially optimal prices only if the numeraire good ("income") is distributed in the way required by equation (1). This interdependence between optimal lump sum redistribution and socially efficient prices has long been recognised in the theoretical welfare economics literature, especially by Graaff (1957) and Samuelson (1954 and 1969), and has been emphasised in a well-known paper on public utility pricing by Feldstein (1972). Graaff (1957, p. 171) ends his classic book on welfare economics with the penultimate paragraph: "Similarly, the economist cannot say at what level the National Coal Board should set the price of coal. He can merely make clear (or attempt to make clear) the probable consequences of setting it at various levels. If he succeeds in this task it will almost certainly become more widely appreciated that tinkering with the price mechanism is one of the more feasible and generally satisfactory ways of securing whatever distribution of wealth is desired." Feldstein (1972, p. 32) begins his paper with the observation that "In practice, optimal lump sum redistribution is impossible and the distributional aspect of public pricing is an important policy consideration." Strotz's (1964) paper on transport pricing recognises that there will in general be many sets of efficiency prices satisfying the Pareto first-order conditions, and that distributional considerations must be used to choose among these sets. Similarly, Mohring's (1970) paper on peak-load pricing emphasises the need to assume costless, lump sum income transfers, although he also points out the contradiction of supposing both that such lump sum transfers can take place and that public utilities (including transport services) of the sort he analyses are faced with budget constraints (which imply an inability to transfer to them lump sum subsidies) that may force them to set second-best, revenue-constrained prices. In his wide-ranging discussion of urban transport pricing, Forsyth (1977, p. 28 and pp. 41-42) argues that "equity and efficiency, in economies such as we operate, cannot be separated either analytically or in practice" and he goes on to suggest that "if prices are
to reflect distributional considerations, the urban transport market should prove a fairly
easy one to put this into effect."

It is in fact commonly acknowledged that public sector prices have distributional
implications, and this acknowledgement surfaces frequently when urban road or transport
prices are under discussion. How might the implementation of transport prices take such
distributional considerations into account?

Suppose the assumption is retained that there exists a given welfare function $W$ and
that from this, given some distribution among individuals of income or the numeraire good
$z$, a set of individual welfare weights, $\beta^h$, can be defined:

$$\beta^h = \frac{\partial W}{\partial u^h} \frac{\partial u^h}{\partial x^h}.$$ 

These weights measure the relative social value, according to $W$, of a unit of numeraire
going to different people.\(^5\) Equation (1) says that for a welfare maximum these weights
should all be equal, but if we disbelieve that this equality exists in the real world of policy
then equation (1) must be replaced by an assumption of differentiated weights. Most
people's social welfare functions would probably assign higher weights to lower income
people.

Recall that by combining equations (1) and (2), the basic efficient-pricing conditions
of equations (5) and (6) were obtained. If, instead of equation (1), the welfare weights $\beta^h$
are substituted into equation (2), and the definition

$$\frac{\partial v^h}{\partial x^h_i} = \frac{\partial u^h}{\partial x^h_i} + \frac{\partial u^h}{\partial x^h}$$

is retained, the efficiency condition for trip-pricing becomes

$$\sum_h \beta^h \frac{\partial v^h}{\partial x^h_i} s^h = -\sum_h \beta^h \frac{\partial v^h}{\partial x^h_i} + \lambda_1 \frac{\partial C}{\partial x_i} - \lambda_2 \left( p_i + x_i \frac{\partial p_i}{\partial x_i} + \sum_{j \neq i} x_j \frac{\partial p_j}{\partial x_j} - \frac{\partial C}{\partial x_i} \right) = 0, \forall i \quad (13)$$

Equation (13) collapses to equation (5) if the $\beta^h$ are all equal, and equal to the common
social welfare value $\lambda_1$, which was set equal to 1. In that case the optimal uniform price
for service $i$ is given by the right-hand side of the equation, and in the absence of a budget
constraint ($\lambda_2 = 0$), a uniform price for all individuals is itself optimal even if differentiated
prices were possible. However, even with $\lambda_2 = 0$, equation (13) cannot in general be
satisfied with a single, uniform price for service $i$ if the $\beta^h$ are not all equal. Suppose $p_i$
is set equal to the marginal cost of the service, the first two terms of the right-hand side
of (13). Individuals, as before, will adjust their marginal valuations, $\partial v^h/\partial x^h_i$ to the price
$p_i$, but with differentiated welfare weights this adjustment to $p_i$ will not in general bring
the left-hand side into equality with the right-hand side. Recalling that

$$\sum_h s^h \equiv 1$$

one can see that equation (13) will be satisfied if each $\beta^h(\partial v^h/\partial x^h_i)$ is equal to the right-hand
side, but because of the variations in $\beta^h$ this will require offsetting variations in the
$\partial v^h/\partial x^h_i$. One way to satisfy equation (13), therefore, would be to differentiate by
individual using the price of each service $i$ so as to bring about the necessary offsetting

\(^5\) For a discussion of this procedure see Starrett (1988, pp. 11-15) and Harberger (1978).
variation in trip marginal valuations. Those who had been assigned a higher $\beta^h$ (perhaps those with lower incomes) would pay a lower price $p^h$, and vice versa.

If the assumption is retained that service prices must be uniform to all users, then individually differentiated prices cannot be used to achieve equation (13). However, an optimal uniform price can be found, one that will depend on the distributions of service use and marginal costs among $\beta^h$ types. These distributionally constrained optimal prices will in general be different from the uniform prices that satisfy the Pareto-efficiency conditions of Sections 3 and 4.

To find the new optimal prices, let the average value of $\beta^h$ across all individuals in the community be $\bar{\beta}$. Without loss of generality, this average value can be set equal to 1. Continue to assume for the time being that the budget constraint on the transport sector is not binding so that the right-hand side of (13) consists only of the overall marginal trip cost, which is the sum of marginal congestion cost (a shorthand way of describing the negative community externality associated with an incremental trip) plus marginal operating cost. With uniform prices $p_i$ for each service, each user will adjust $\partial x^h / \partial x^t_i$ to the service price, so the left-hand side of (13) becomes

$$p_i \sum_h \beta^h s^h_i.$$

The expression under the summation sign is a weighted average of the welfare weights of the users of service $i$. If this particular transport facility caters for users with relatively low welfare weights (perhaps a commuter rail line used by high-income members of the community), then the summation will have a value less than the community average $\bar{\beta}$; if the service is used more by high welfare-weight users (perhaps off-peak transport service), then the summation will have a value above the community average. Let the value of

$$\sum_h \beta^h s^h_i$$

be represented by $B_i$, and call this variable the distributional characteristic of service $i$, using a term introduced by Feldstein (1972). The left-hand side of (13) may then be written as $p_i B_i$.

Turning now to the cost side of (13), consider the term involving congestion externalities of an incremental trip,

$$-\sum_h \beta^h \frac{\partial x^h}{\partial x_i}.$$

Define $g^h$ as the share of the total externality that is borne by individual $h$:

$$g^h = \frac{\partial x^h}{\partial x_i} + \sum_h \frac{\partial x^h}{\partial x_i}.$$

The externality term can now be written as

$$-\sum_h \beta^h g^h \left( \sum_h \frac{\partial x^h}{\partial x_i} \right).$$

---

6 The model used in this paper is different from that used by Feldstein, but his "distributional characteristic" is similar in definition to $B_i$ and identical to it in concept and use.
Define

\[ \sum \beta^h g^h \]

as the cost-distributional characteristic and let it be represented by \( B^*_i \). The congestion term then becomes

\[ -B^*_i \sum \frac{\partial v^h}{\partial x_i}. \]

Equation (13) may then be written as

\[ p_i B_i = -B^*_i \sum \frac{\partial v^h}{\partial x_i} + \lambda_i \frac{\partial C}{\partial x_i}, \forall i \]

(14)

From this point forward, it will be assumed that \( \lambda_i = 1 \). This implies that the social welfare cost of using an incremental unit of numeraire in the transport sector is the community average welfare weight \( \bar{\beta} \). Alternative assumptions about the cost-distributional characteristic, \( B^*_i \), may be considered. Suppose that the congestion (including pollution) costs and the operating costs are spread evenly across the whole urban community. This could be interpreted as \( B^*_i = \bar{\beta} = 1 \). Equation (14) then becomes (15),

where \( \Gamma \) is used as in equations (7) and (11) to represent marginal congestion plus marginal operating costs:

\[ p_i = \frac{\Gamma}{B_i}, \forall i \]

(15)

Equation (15) provides a straightforward price-adjustment relationship if service prices must bear a redistributional burden. Price should exceed marginal cost if the distributional characteristic of the service is lower than the community mean welfare weight \( \bar{\beta} = 1 \) and vice versa. If the welfare weights are inversely related to individual incomes, then services attracting relatively high-income users should have distribution-constrained optimal prices above marginal costs, while those with relatively low-income users should have prices below marginal cost. Although this is an intuitively plausible result, the simplicity in (15) of the optimal price formulation was achieved only with a strong assumption about the community-wide effect of marginal trip costs.

An alternative route to equation (15) would have been to define the mean of the \( \beta^h \) distribution of users of service \( i \) as \( \bar{\beta}_i \) and to assume that this distribution is uncorrelated with the marginal share of service use \( s^h_i \). With these assumptions, the expected value of the left-hand side of (13),

\[ \sum \beta^h \frac{\partial v^h}{\partial x_i} s_i^h \]

becomes

\[ \sum \bar{\beta}_i p_i s_i^h \]

which reduces to \( \bar{\beta}_i p_i \). Equation (15) would then be

\[ p_i = \frac{\Gamma}{\bar{\beta}_i} \]

This may be an easier formulation for transport planners to work with since it uses as the distributional characteristic simply the average \( \beta^h \) value of the service users rather than the
more general $B_i$ which incorporates any correlation that exists among $\beta^h$ and $\phi^i$ of the users. In the rest of this section, $\beta_i$ can be substituted for $B_i$ whenever the latter appears, but its use requires the more restrictive assumption of no correlation between the $\beta^h$ and $\phi^i$ of the users.

We may also use an alternative assumption about the set of individuals who bear the cost of a marginal trip. Rather than the assumption that this is the whole community, suppose that only those who use service $i$ are affected, and suppose that this cost effect is distributed in proportion to individual service use. In this case, the cost-distribution characteristic, $B_i^*$, is equal to the basic service-distribution characteristic, $B_i$. Equation (14) then reduces to

$$
\beta_i = \Gamma_i + \left( \frac{\lambda_i - B_i}{B_i} \right) \frac{\partial C}{\partial x_i} = \Gamma_i^a
$$

(16)

With $B_i$ not too far from the community average $\bar{\beta} (= \lambda_i)$, and with marginal operating costs small in comparison with congestion costs, $\Gamma_i^a$ will be close to $\Gamma_i$, and equation (16) will be approximately the same condition as equation (7) which was derived on the assumption that distributional concerns were handled outside the transport sector. The intuition behind this result is that the new assumptions assign the marginal congestion cost of service $i$ as well as all of its marginal benefit to the service users. It will always be optimal therefore to price trips at (approximately) marginal cost, without an adjustment for the distributional characteristic of the service.

These results are not the end of the story, however. If the optimal service-specific investment rule, equation (8), is revisited, it will be seen that distributional considerations can affect this optimality condition in ways that depend on the assumption made about the distributional characteristic associated with such spending.

Continuing to ignore the $\lambda_2$ term for the time being, and recalling that $\lambda_1 = 1$, equation (8) becomes

$$
1 = \sum_{h} \beta^h \frac{\partial \nu^h}{\partial f^h}
$$

when distributional considerations exist. Proceeding as before, define $f^h$ as individual $h$'s share of total community utility relief from incremental infrastructural spending,

$$
f^h = \frac{\partial \nu^h}{\partial f} + \sum_{h} \frac{\partial \nu^h}{\partial f^h},
$$

and define $B_i^f$ as

$$
\sum_{h} \beta^h f^h,
$$

the infrastructure-distribution characteristic. Equation (8) may then be written as

$$
\frac{1}{B_i^f} = \sum_{h} \frac{\partial \nu^h}{\partial f^h}, \forall i
$$

(17)

If the distribution characteristic of the benefits associated with infrastructural spending are spread evenly among the members of the community, then $B_i^f = \bar{\beta} = 1$ and (17) is reduced to (8). This distributional assumption is consistent with the one that led to equation (15); in that case, congestion externalities were assumed to be spread uniformly
among the whole community. Thus, if prices are adjusted according to equation (15), the optimal investment rule, equation (8) remains the correct rule.

Suppose, however, that the benefits from improved infrastructure are confined to the users of the improved service. This is very similar to the assumption that congestion affects only users that led to the optimal price rule of equation (16). In this case $B'_i$ will reflect the distributional characteristic of the service, $B_a$, and may even be equal to it. What has happened in this case is that the distributional role of the transport sector has been transferred from trip-pricing policies to investment policies. The pricing rule, equation (16), is just the Pareto-efficiency rule but the investment equation, (17), says that for those services with users having a distributional characteristic higher than average, that is, with $B_i$ (or $B'_i$) > 1, service-specific investment should be expanded beyond the point given by equation (8) where a numeraire unit of investment yields a numeraire unit's worth of aggregate utility improvement. For services with lower-than-average distributional characteristics, infrastructural investment should be less than the optimum given by (8). Only the service-specific investment rule, equation (8), will be affected. Since overhead investment, $J$, by definition is not assignable to individual services, it makes sense to continue to assume that its beneficial effects,

$$\sum_k \frac{\partial y^h}{\partial J}$$

in equation (9), will be spread over the whole community with an expected welfare weight of $\overline{\beta} (=1)$, the community average.

With distributional assumptions between the alternative assumptions considered above, equation (14) will hold without collapsing to either of the extremes, (15) or (16), and it, together with (17), show that if optimal redistribution is not undertaken outside the transport sector, and if in consequence the sector has to respond to the distributional goals of the community, then either prices or investment or both will have to adjust away from the more usual efficiency prices and investment conditions that were derived in previous sections.

The budget constraint may now be reintroduced. If the assumptions about the distribution of welfare weights that led to equation (15) are retained, then equation (13), with $\lambda_2 > 0$ and with $\Xi_j$ standing for overall revenue elasticity associated with $x_j$ as in equation (11), may be rewritten as

$$p_i \ B_i \left(1 + \frac{\lambda_2}{B_i} \overline{\Xi}ight) = \Gamma_i.$$ 

Letting $\Phi^*_i$, stand for the term in brackets, this becomes

$$p_i = \frac{\Gamma_i}{B_i \ \Phi^*_i}, \forall i$$

(18)

Equation (18) is the most general form of the revenue and distribution-constrained optimal service price condition when prices and not investment bear the whole of the distributional effect. The term $\Phi^*_i$ in the denominator differs from $\Phi_i$ in equation (11) by the $1/B_i$ value that multiplies $\lambda_2$. The effect of this, when the budget restriction is binding and when the distributional characteristic of service $i$ differs from the community average.
welfare weight, is to alter the budget-driven price adjustment. With a higher-than-average distributional characteristic, \( \Phi^* > \Phi \), and the price \( p \), will be less than otherwise above the marginal trip cost \( \Gamma \). This effect is separate from but reinforces the direct effect of \( B \) in the denominator of (18). Notice that if \( B = \tilde{\beta} \), \( \Phi^* \) collapses to \( \Phi \) and (18) becomes identical to (11). If the budget restriction is not binding, \( \Phi^* = 1 \) and (18) is reduced to equation (15). With no binding budget restriction and with a service having a distributional characteristic equal to the community average welfare weight, (18) becomes (7), the simplest form of optimal, marginal-cost pricing. If assumptions about the cost-distribution characteristic do not allow (14) to be reduced to (15), then the general optimal service price condition becomes

\[
p_i = \frac{\Gamma_i^s}{B_i^s \Phi^*}, \quad \forall i
\]

(19)

where \( B_i^s = B_i / B_i^* \) and \( \Phi^* \) is now equal to

\[
\left( 1 + \frac{\lambda}{B_i^s} \Xi \right).
\]

This pricing condition is likely to be accompanied by the new infrastructural investment condition (17).

Under some circumstances, the conditions under which the budget constraint will be binding will remain those derived in Section 4 above. Suppose that both the cost-distribution characteristic and the congestion-distribution characteristic are equal to the community average welfare weight: that is, \( B^* = B = \tilde{\beta} = 1 \). This means that the pricing equation (18) and the investment equation (8) give the relevant optimality conditions. The left-hand side of equation (12) would then be

\[
\sum_i x_i B_i^s \Phi^*
\]

and the right-hand side would be unchanged. If the distributional characteristics \( B_i \) have a mean of \( \tilde{\beta} \) and are distributed independently of both the trip volumes \( x_i \) and the revenue elasticities \( \Xi_i \), then the expected value of the left-hand side becomes

\[
\sum_i x_i B_i \Phi_i
\]

which is just the left-hand side of (12). In these circumstances, the condition for a binding or a slack budget constraint is therefore unchanged. More generally, equation (12) will become

\[
\sum_i x_i B_i^s \Phi^* = \sum_i f \left[ 1 + \lambda^2 \left( 1 - \sum_j x_j \frac{\partial p_j}{\partial f} \right) + \lambda \left( 1 + \lambda^2 \left( 1 - \sum_i \frac{\partial p_i}{\partial f} \right) \right) + \sum_j x_j \frac{\partial C}{\partial x_j} \right].
\]

Either a positive correlation of the adjusted distributional characteristic \( B_i^s \) with trip volumes \( x_i \), or a positive correlation of the congestion-distributional characteristic \( B_i \) with service-specific investment \( f_i \), or both, would make the revenue constraint more difficult to satisfy. Zero homogeneity of the utility functions in service levels and infrastructural expenditures, the "constant cost" condition, would in these circumstances yield service revenue at marginal-cost prices that was insufficient to meet total costs. Without a subsidy to the sector, optimal prices would have to be set above marginal costs.
6. Conclusions

Under some circumstances, an integrated transport sector can set each service price at the service's marginal cost, where this cost includes marginal congestion cost, and at least break even financially across the whole sector. Prices would be optimal and deficits associated with some services would be offset or more than offset by surpluses from others. This result will occur if the utility of individuals is unaffected or worsened by equiproportionate changes in the use of all services and in the service-specific and overhead investment spending by the sector. This might be said to entail constant or increasing congestion costs. If this condition does not hold, the sectoral budget constraint will be binding and this will require the optimal service prices to be adjusted in direct proportion to each service's revenue elasticity. The revenue elasticity is defined as the percentage change in revenue when the price of a service is raised by the amount that reduces service use by one per cent. This result says quite plausibly that if the transport sector needs to raise prices in order to balance its budget, the burden of price increases should be borne by services with high revenue elasticities. The cross-service revenue elasticities that contribute to the value of a service's overall revenue elasticity draw attention to the need to consider not only the demand elasticity for any service that is a candidate for a price increase but also the effect of such a price increase on the demand for other services. A number of examples are provided in Section 4 of the role of revenue elasticities in determining optimal budget-constrained prices.

If lump-sum income transfers cannot be made among members of the urban community, then service prices must respond to distributional goals and an additional adjustment factor is introduced, as shown in equations (18) or (19). This new factor raises prices for services used relatively heavily by people with lower-than-average welfare weights (if welfare weights were inversely related to incomes, these would be people with higher-than-average incomes), and lowers prices for services used by people with relatively high welfare weights. In some circumstances, the optimal response may also be to expand investment in services used by those with higher welfare weights, and to reduce investment in infrastructure for services with low-welfare-weight users. The budget constraint and the distributional concern could obviously produce adjustment factors that worked against each other. For example, if the demand elasticity for a particular service was very low, its revenue elasticity would probably be high, and with a binding budget constraint it would be a candidate for prices well above marginal costs; but if this service had mainly high welfare-weight (low-income) passengers, the distributional adjustment factor would assign lower prices to it.

These results are all based on first-order conditions that maximise a community's social welfare function. Although policy conclusions in the welfare-theoretic literature are commonly based on efficiency results derived in exactly this way, there is quite properly a mounting scepticism about their applicability to real-world policy issues. Some of the strongest criticism is based on the view, long represented in the literature, that lump-sum transfers — defined as income or numeraire transfers with no resource or allocative implications — cannot exist. Every feasible tax or subsidy has resource
implications. Results based entirely on Pareto-efficiency conditions, such as those in Sections 3 and 4 of this paper, are therefore unachievable and of no use for policy purposes. Hammond (1990) includes this among several other arguments in his broadside attack on the usefulness of Pareto efficiency as a guide to policy. Many critics would regard the abandonment of optimal lump-sum transfers and the introduction of distributive considerations, as in Section 5, to be an improvement since policy-makers are then provided with results that reflect more closely their day-to-day policy concerns.

However, the abandonment of lump-sum transfers along with the introduction of distributional goals by no means puts to rest the question of policy usefulness of the first-order results. These results consist basically of a set of simultaneous equations having as parameters such things as service marginal costs, the elasticities and cross-elasticities of the inverse demand functions and the individual welfare weights derived from some social welfare function. The numerical values of these parameters are the values they would have if service prices and infrastructural investment were all at optimal levels. If in some urban area the actual prices and investment levels are not optimal, how are the parameters necessary to calculate optimal levels to be estimated? Without additional and perhaps heroic assumptions, they cannot be based on the existing situation, even if the relevant existing parameters could be measured. The best that might be hoped for is some iterative movement towards an optimal outcome, with prices and investment adjusted initially on the basis of the current and known parameters, then readjusted as the parameters of a new configuration were calculated, and so on. There is no guarantee that such an iterative approach would converge on an optimal outcome, but the possibility of such convergence might be sufficient to encourage some interest in the optimality conditions.

Even this idea of converging on the optimal configuration runs into difficulties, however. If the assumptions of the model are accepted, then the first-order conditions define an optimal configuration of prices and infrastructure. If a community that was not at this optimal configuration could move without cost from where it was to the optimum, then doing so would unambiguously be good policy. But any move from what exists, any price adjustment or investment decision, is bound to have costs as well as benefits, and the typical welfare-theoretic model, including the one of this paper, provides no way of evaluating the net community benefit associated with some move away from the current status. Even if it were possible to know beforehand what an optimal configuration would look like, and possible to get to it through one or more feasible paths of price and investment changes, it may be that the net community welfare evaluation of the stages required to go from the present to the optimal configuration, including the evaluation in the final and subsequent stages of the optimal state itself, is negative, meaning that the move should not take place. The first-order conditions cannot tell us whether a move from what exists to the optimum is desirable.

Faced with this, Hammond (1990) and others have urged economists and policy-makers to turn their attention away from standard first-order effects and towards the evaluation of any proposed policy that would take the economy from the status quo configuration to some new configuration. Welfare weights derived from a welfare function would still be needed (or a sensitivity to welfare weights calculated), and
assumptions would have to be made about the stability of parameters such as demand elasticities (or about the stability of the underlying functions). These assumptions would obviously be more risky as the proposed situation departed further and further from the pre-existing situation, but Hammond is surely right to argue for this refocusing of policy-oriented welfare economics.

Where does this view leave first-order results, even those that do not assume lump-sum transfers such as the results of Section 5? The answer to this lies in understanding what the Hammond-type approach leaves out, which is the origin of the policy changes to be evaluated. The merit, the usefulness, of first-order conditions derived from a well defined model lies in the inspiration they provide for productive policy proposals. It is not feasible to evaluate all possible policy changes, which are infinite, and certainly inefficient to evaluate randomly generated proposals. As a matter of reality, the proposed policy agenda of a sector or agency is usually relatively narrow in scope with the items having been inspired, for better or worse, by some mechanism, an idea perhaps, an irritant more probably. It can and should be agreed that first-order conditions do not provide a set of equations from which an optimal configuration can be directly calculated, and it should further be agreed that they contain no evaluation of the merit in real policy time of making a move from the status quo; but even with these important understandings of the limitations of optimality conditions, their role as inspiration for proposed policy remains intact. This may lead back to the iterative approach described above, but that would be only one among many possible policy responses.

A model that is formulated to embrace generally agreed policy concerns should yield optimality conditions that help suggest productive directions for policy changes. An integrated transport sector that has concerns about pricing its services to reflect the cost of congestion, about balancing its budget and about the effects of its prices on individuals at different income levels, will find the first-order optimality equations derived in this paper useful, at least as a source of relationships to consider when formulating policy.

References


*Date of receipt of final manuscript: June 1993*