OPTIMAL PRICING OF URBAN TRIPS: A "WHOLE-SECTOR" APPROACH

David M. NOWLAN
Professor of Economics
University of Toronto
Toronto - Canada

INTRODUCTION

Urban travel raises remarkably similar policy issues from one large city to the next. These include excessive congestion, especially in the central areas, overuse of the automobile, land-use sprawl related to under-priced transportation facilities, disproportionate commuter travel at peak times, public-transit deficits, and concerns about the effect of transportation prices on low-income residents. Because every one of these issues is related to the price of travel, there is widespread interest in using price policies to address them. Perhaps the most studied of these is the potential use of pricing to optimize travel congestion, especially automobile congestion in peak times. Recent work, such as that reported on by David Newbery in Britain (Newbery 1990), have helped confirm the general view that congestion costs are high and that the private costs of automobile commuting are well below the public costs.

Underpricing of urban road use is often used as a reason for keeping public-transit prices low, the argument being that aggregate welfare will be reduced if higher transit prices cause more travellers to switch to automobile travel which is priced inefficiently low. This in turn leads to transit deficits which financially straitened cities have difficulty supporting. Attempts to raise transit 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. Meanwhile, both low fares for public transit and the absence of efficient congestion charges for automobiles have been among the factors that have encouraged more travel, more commuting travel in particular, and that correspondingly have reduced the compactness of our cities and metropolitan areas, resulting in district-by-district imbalances between jobs and residences, especially in cities with strong central business districts.

Although the issues I have described are all issues of urban travel, policy towards them is typically undertaken piecemeal. Policies for improved automobile pricing focus on the costs of automobile congestion, including automobile pollution. Public-transit policy is dominated by the financial concerns of transit authorities; and the political reaction to price proposals is often conditioned by considerations of distributional equity among the population.

This piecemeal approach ignores, however, the interdependencies among the three sorts of issue, the issues of efficiency, of finance and of distributional concerns. Policies that appear optimal from the perspective of any one of these issues may be distinctly inferior when viewed from the perspective of another. This paper proposes an analytical framework that may be used to understand better the nature of the interdependencies among the different issues. Given these interdependencies, a pricing structure that addresses optimally all three types of issue is derived. In this framework, the public authorities that provide urban infrastructure and operating services for both private-vehicle and public-vehicle travel are modelled as a single public sector, the transportation sector.

This whole sector provides a variety of services to urban travellers over the course of a daily cycle of travel. Some of these services are for automobiles, some for public transit; some are congested and some not. Efficient pricing requires, as in standard models of urban transportation, that travellers face the appropriate marginal cost, including congestion cost, of their use of a service. But, over the whole sector, a budget constraint must also be met. There is an upper limit placed on the amount of general government revenue available to meet a deficit in the transportation sector; that upper limit may be zero, which means that 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 the well-being of residents at different income levels; it may, for example, be told that its pricing policies should not worsen the existing distribution. Given these restrictions, the objective of the public transportation sector is to set prices on its different services in such a way that the aggregate well-being among the citizens it serves is maximized.

My approach to modelling this set of issues is to draw on and merge two streams of literature on optimal pricing. One of these literatures, well known to transportation planners and economists, analyzes optimal transportation pricing in the face of congestion; the other, better known to those specializing in industrial and regulatory policy, involves the optimal pricing of the output of regulated private industries which necessarily are constrained to have aggregate revenues at least equal to aggregate costs. The transportation-pricing model was early explored by economists such as Herbert Mohring (Mohring 1976) and it continues to form the basis for much of the current discussion of optimal pricing. The principle of optimal, regulated prices was set out in such well known papers as Boiteux 1956 and Baumol and Bradford 1970; this literature is closely related to that on optimal taxation associated with early work by F. Ramsey (Ramsey 1927).

A pricing model that draws on both these approaches is described in the next section and the first-order optimality conditions are set out in section 2. Section 3 analyzes the role played by the transportation sector's budget constraint, section 4 discusses the use of transportation prices to address distributional issues, and the final section discusses some of the implications of this whole-sector approach to optimal pricing.

1. THE BASIC MODEL

The transportation sector provides an assortment of travel services to the urban public, services that may be differentiated by mode, time of travel and location of travel. Modes may include private automobile, buses, commuter rail, and subway. The time of travel may be divided coarsely into peak and off-peak travel, but these travel times may be divided much more finely into smaller time intervals. Travel locations would ideally be defined in such a way that adequately differential pricing of differential congestion could occur, but location might most simply be thought of as road segments lying between cordon lines surrounding a principal commuting destination, the central business district perhaps. Whatever service distinctions are introduced, the agency is able by assumption to set separate prices for each.

An individual's use of one of these services once is a "trip." Although trips may be differentiated in several dimensions and indexed by subscripts, perhaps "i" for mode, "j" for time of day and "k" for location, for simplicity of notation, with no loss in generality, I will use just the subscript "i" to index these services, where i ranges over all modes, all time periods and all locations. Individual consumers are differentiated by a superscript "h". Thus, an individual's (annual) consumption of type-i trips is given by x_i^h and the vector of all types of trip by the individual is X_i^h ; the sum of all type-i trips is x_i , and the vector of all trip types by all people is X_i.

Each individual has a utility function defined over his or her consumption of trips and an index of other goods, z^h; the vector of z^h over all individuals is given by Z^h. z will be our numeraire good with a price of 1. The individual's utility function also includes a congestion function, γ^h, which is a function of all trips X_i , and of the fixed-cost resources that have been devoted to the provision of travel services by the agency. By using in the congestion functions X_i , which is a vector of trip types undifferentiated by consumer, I have adopted the usual "anonymity" assumption that congestion is affected by total trips of a given type but not by who the trip-takers are. The fixed costs of transportation infrastructure in the congestion functions include an amount J that represents joint resources that cannot be assigned to any particular service, and amounts c_i specific to each service i. The vector of all fixed costs identifiable with particular services is given by C_i . Thus $\gamma^h = \gamma^h(X_i , J, C_i)$, and the utility of individual h, u^h, is given by $u^h = u^h(z^h, X_i^h , \gamma^h(X_i , J, C_i))$. A social welfare function, W, determined outside the transportation sector, provides the necessary relationship among the utilities of all individuals in the area. I assume that this can be written in the separable form: $W = W(u^1, u^2, \dots u^h, \dots)$.

The total cost of providing transportation services is given by a cost function, Γ, that shows in numeraire terms the cost of infrastructure, J and C_i , and the operating cost of services X_i. Thus, $\Gamma = \Gamma(X_i , J, C_i)$. (Again I have incorporated the anonymity assumption that costs are affected by total trips of type i but not by who makes the trips.) There is an overall resource constraint in this urban area which restricts to a given amount R the total use of resources for transportation services plus all other goods z, where z is the sum of all z^h.

The components of the model so far are similar to those of many optimal congestion-pricing models. Now what I want to add is the revenue constraint for the transportation sector. Total revenue is the sum of the revenue from the priced transportation services plus whatever transfer exists from general public funds. The price of each service, a policy variable under the control of the sector, is given by p_i, and the vector of all prices by $\mathbf{P}_i$. Thus total service revenue is the inner product $\mathbf{P}_i \cdot X_i$. With the transfer denoted by T, the revenue constraint may be written as $\mathbf{P}_i \cdot X_i + T = \Gamma(X_i, J, C_i)$. The transfer amount T is not under the control of the transportation sector, and could be zero.

The services consumed by each traveller, x_i^h, are assumed to be a utility-maximizing amounts for the individual, given the prices set by the transportation sector and the amount of congestion. From this we can define an inverse demand function for each service as $p_i = p_i(X_i, J, C_i)$. This formulation precludes the transportation sector's setting differentiated prices for each traveller, a limitation that has a welfare cost.

In summary, the basic model consists of a social welfare function, W, to be maximized subject to two constraints, an overall resource constraint for the community and a revenue constraint for the transportation sector:
MAXIMIZE $W = W(u^1,..., u^h,...)$ with respect to $z^h \in Z^h$ and $x_i^h \in X_i^h$, $\forall h$; J and $c_i \in C_i$

s.t. $$R - \sum_h z^h - \Gamma(X_i, J, C_i) \geq 0$$

and $$\mathbf{P}_i \cdot X_i + T - \Gamma(X_i, J, C_i) \geq 0,$$

$u^h = u^h(z^h, X_i^h, \gamma^h(X_i, J, C_i))$ with $\frac{\partial \gamma^h}{\partial X_i} \geq 0$; $\frac{\partial \gamma^h}{\partial J} \leq 0$; $\frac{\partial \gamma^h}{\partial C_i} \leq 0$; and $\frac{\partial u^h}{\partial \gamma^h} \leq 0$. Each price in the vector $\mathbf{P}_i$ is a function of X_i, J, and C_i.

2. FIRST-ORDER CONDITIONS FOR OPTIMAL PRICES

Shadow prices or opportunity costs for each of the constraints are introduced, λ_1 and λ_2. These multiply the constraints in the Lagrangian expression which is to be maximized:

$$\mathcal{L} = W(u^1, u^2, ... u^h, ...) + \lambda_1[R - \sum_h z^h - \Gamma(X_i, J, C_i)] + \lambda_2[\mathbf{P}_i \cdot X_i + T - \Gamma(X_i, J, C_i)]$$

.

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^h for all h and all i; 3) the infrastructure-cost variables c_i for all i, and J; and 4) the shadow prices λ_1 and λ_2. In writing these I have made use of the fact that $\frac{\partial x_i}{\partial x_i^h} = 1$ (recall that

$x_i = \sum_h x_i^h$) to show the partial derivatives of γ^h, Γ and P_i with respect to x_i^h as $\frac{\partial \gamma^h}{\partial x_i}$, $\frac{\partial \Gamma}{\partial x_i}$, and $\frac{\partial p_i}{\partial x_i}$ respectively.

$$\frac{\partial \mathcal{L}}{\partial z^h} = \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial z^h} - \lambda_1 = 0;\ \forall h \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial x_i^h} = \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial x_i^h} + \sum_h \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial x_i} - \lambda_1\frac{\partial \Gamma}{\partial x_i} + \lambda_2[p_i + x_i\frac{\partial p_i}{\partial x_i} + \sum_{j \neq i} x_j\frac{\partial p_j}{\partial x_i} - \frac{\partial \Gamma}{\partial x_i}] = 0, \forall h, \forall i \tag{2}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial c_i} &= \sum_h \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial c_i} - \lambda_1\frac{\partial \Gamma}{\partial c_i} + \lambda_2\sum_j x_j\frac{\partial p_j}{\partial c_i} - \lambda_2\frac{\partial \Gamma}{\partial c_i} = 0\ ,\quad \forall i \\ \frac{\partial \mathcal{L}}{\partial J} &= \sum_h \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial J} - \lambda_1\frac{\partial \Gamma}{\partial J} + \lambda_2\sum_j x_j\frac{\partial p_j}{\partial J} - \lambda_2\frac{\partial \Gamma}{\partial J} = 0 \end{aligned} \tag{3}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_1} &= R - \sum_h z^h - \Gamma = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= P_i \cdot X_i + T - \Gamma = 0 \end{aligned} \tag{4}$$

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

Equations (1) describe the way in which distributional issues are taken care of independently of transportation pricing policies. These first-order equalities essentially mean that lump-sum transfers of the numeraire good, "income," are being made so as to equalize the social welfare value of a unit of numeraire to every member of the community. That this can be done is an important assumption made, implicitly if not explicitly, in most optimal-pricing models. Its relaxation is dealt with in section 4.

The W value of a unit of numeraire is seen in equations (1) to be λ_1 , the shadow price on the first constraint. Dividing equations (2) and (3) by λ_1 allow us to clear these equation sets of the partial derivatives of W and also to express the partial utilities of the consumers in terms of the numeraire good z. Individual utilities in numeraire units I write as v^h, so that $\frac{\partial v^h}{\partial x_i^h} \equiv \frac{\partial u^h}{\partial x_i^h} \div \frac{\partial u^h}{\partial z^h}$. Dividing equations (2) through by λ_1 in this way gives us:

$$\frac{\partial v^h}{\partial x_i^h} = -\sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial x_i} + \frac{\partial \Gamma}{\partial x_i} - \frac{\lambda_2}{\lambda_1}[p_i + x_i\frac{\partial p_i}{\partial x_i} + \sum_{j \neq i} x_j \frac{\partial p_j}{\partial x_i} - \frac{\partial \Gamma}{\partial x_i}] \text{ ,} \forall h,\ \forall i \qquad (5)$$

The left-hand side of equation (5) is the marginal utility to person h of the use of service i, in numeraire terms. It is this marginal utility that the utility-maximizing individual adjusts to the price p_i set by the transportation sector. All other private costs, such as the length of time to take trip i, are already accounted for; the only remaining private adjustment is to adjust to the price of the trip. Thus, p_i may be substituted for the marginal utility on the left-hand side of the equation.[1]

It is also helpful for subsequent interpretation if the partial derivatives of price on the right-hand side of (5) are converted to inverse elasticities and cross elasticities. $x_i\frac{\partial p_i}{\partial x_i}$ and $x_j\frac{\partial p_j}{\partial x_i}$ may be written as $p_i\xi_{ii}$ and $\frac{p_j x_j}{x_i}\xi_{ji}$ respectively, where ξ_{ii} is the own-service inverse elasticity of x_i and ξ_{ji} the cross-service inverse elasticity.[2] As long as x_i and x_j are not complementary services, $\xi_{ji} \leq 0$; the greater the substitutability between services i and j, the larger the absolute value of ξ_{ji}. Making these changes to (5) and rearranging, we get

$$p_i\left(1+\frac{\lambda_2}{\lambda_1}+\frac{\lambda_2}{\lambda_1}\xi_{ii}+\frac{\lambda_2}{\lambda_1}\sum_{j\neq i}\frac{p_j x_j}{p_i x_i}\xi_{ji}\right) = -\sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial x_i} + \frac{\partial \Gamma}{\partial x_i} + \frac{\lambda_2}{\lambda_1}\frac{\partial \Gamma}{\partial x_i},\ \forall i \qquad (6)$$

The right-hand side of (6) shows the cost of a trip x_i in terms that are familiar from standard congestion models of transportation pricing. These are the congestion cost itself[3] plus the marginal cost, if any, of servicing an extra trip. In addition, there is a third term representing the worsening of the budget problem when operating costs rise. If there were no budget constraint assigned to this sector, or if the budget constraint was not binding, then λ_2 would be zero. The left-hand side of (6) would collapse simply to p_i, and the right-hand side would contain the usual two marginal cost terms. This is the normal transportation pricing result. However, with a binding budget constraint, $\lambda_2 > 0$ and the term in brackets will lead to optimal departures from normal marginal-cost prices. As long as $|e_{ii}| < 1$, then $\xi_{ii} < -1$ and the bracketed term is necessarily less than 1. Thus, the budget constraint drives all optimal prices above the sum of marginal congestion plus service costs by proportions that depend differentially on the different elasticities and cross elasticities of each type of trip. Other things equal, the larger the absolute value of ξ_{ii} (i.e., the lower is $|e_{ii}|$), the lower the value of the bracketed term and the higher the optimal price for that service. Within the context of this model, that result is the equivalent of the well known Ramsey inverse-elasticity pricing rule for optimal prices of a revenue-constrained regulated monopoly.

Clearly the magnitude of the shadow price or opportunity cost of the sector's budget constraint, λ_2 is an important determinant of optimal prices. In the next section I establish the conditions under which this shadow price will be positive.

3. THE BUDGET CONSTRAINT

If transportation prices are set optimally, the significance of a budget constraint imposed on the whole sector depends on the nature of the congestion terms, γ^h, in each person's utility function. These are homogeneous functions: $\gamma^h(0) = 0$. If, for everyone, their homogeneity is of degree zero or more in their arguments X_i, C_i, and J, then standard marginal cost prices will provide sufficient revenue to cover the costs of the sector, and these prices will be optimal. λ_2 will be zero in this case. Surpluses generated from some services must be used to cover deficits that may exist in other services, or to cover deficits attributable to unassigned costs, J. If the γ^h functions have less than zero homogenity then $\lambda_2 > 0$ and the set of optimal prices will depart from marginal costs in ways described by equation (6).

These propositions may be demonstrated by noting that the degree of homogeneity of any γ^h depends on the sign of the following expression: $X_i \cdot \frac{\partial \gamma^h}{\partial X_i} + J\frac{\partial \gamma^h}{\partial J} + C_i \cdot \frac{\partial \gamma^h}{\partial C_i}$. If this is zero or positive then congestion costs over the whole transportation sector are constant or increasing; if it's negative, then congestion costs are decreasing. From this expression, we can substitute for $\frac{\partial \gamma^h}{\partial x_i}$ in equation (6). Using Λ_i to represent the terms within the brackets on the left-hand side of (6), and assuming that the γ^h functions have zeroth degree homogeneity, this yields

$$p_i\,(\Lambda_i) - \frac{\lambda_1+\lambda_2}{\lambda_1}\frac{\partial \Gamma}{\partial x_i} = -\sum_h \frac{\partial v^h}{\partial \gamma^h}\left[\frac{C_i}{x_i}\cdot\frac{\partial \gamma^h}{C_i} + \frac{J}{x_i}\frac{\partial \gamma^h}{\partial J} + \sum_{j\neq i}\frac{x_j}{x_i}\frac{\partial \gamma^h}{\partial x_j}\right]$$

which in turn may be written as

$$x_i\, p_i\,(\Lambda_i) - x_i\frac{\lambda_1+\lambda_2}{\lambda_1}\frac{\partial \Gamma}{\partial x_i} = \sum_i c_i \sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{C_i} + J\sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial J} + \sum_{j\neq i} x_j \sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial x_j}$$

Substituting again from (6) gives us

$$\sum_i x_i p_i\,(\Lambda_i) - \frac{\lambda_1+\lambda_2}{\lambda_1}\sum_i x_i\frac{\partial \Gamma}{\partial x_i} = \sum_i c_i \sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{c_i} + J\sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial J} \qquad (7)$$

Equation (7) may be interpreted using equation set (3), which defines the optimal amounts of resources to use in various forms of infrastructure. I define the c_i and the J to be resources in numeraire units, so the marginal cost of altering the size of the various infrastructure dimensions is for all of them identically 1. Using this, and dividing (3) by λ_1 and rearranging gives the optimal investment rules shown in equation (8) on the next page. Substituting from (8) into (7) gives equation (9).

The left- and the right-hand sides of equation (9) are joined with an "equals" sign because of the initial assumption that the γ^h congestion terms had zero

$$\sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial c_i} + \frac{\lambda_2}{\lambda_1}\sum_j x_j \frac{\partial p_j}{\partial c_i} = 1 + \frac{\lambda_2}{\lambda_1}\ , \forall i$$

$$\sum_h \frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial J} + \frac{\lambda_2}{\lambda_1}\sum_j x_j \frac{\partial p_j}{\partial j} = 1 + \frac{\lambda_2}{\lambda_1} \tag{8}$$

$$\sum_i x_i\, p_i(\Lambda_i) - \frac{\lambda_1+\lambda_2}{\lambda_1}\sum_i x_i \frac{\partial \Gamma}{\partial x_i} = \sum_i c_i\left(1+\frac{\lambda_2}{\lambda_1}-\frac{\lambda_2}{\lambda_1}\sum_j x_j\frac{\partial p_j}{\partial c_i}\right)+J\left(1+\frac{\lambda_2}{\lambda_1}-\frac{\lambda_2}{\lambda_1}\sum_j x_j\frac{\partial p_j}{\partial J}\right) \tag{9}$$

homogeneity. If, under these circumstances, the shadow price of the budget constraint, λ_2 , is set to zero, each Λ_i becomes 1, so the left-hand side is just the total revenue from all services less an amount to pay for operating costs[4]. The right-hand side, with $\lambda_2 = 0$, becomes simply the sum of the infrastructure costs. So, net revenue has exactly matched fixed costs with all prices set to marginal service costs. The budget constraint is not binding.

If the γ^h functions have a degree of homogeneity greater than zero, then the left-hand side of equation (9) will be greater than the right-hand side, and marginal cost pricing, which will still be optimal, will yield a surplus of revenue over costs for the whole sector. However, if the γ^h functions have a degree of homogeneity less than zero, then the left-hand side of (9) will be less than the right-hand side, and pricing at marginal cost with a $\lambda_2 = 0$ would not yield enough revenue to cover costs. Since this violates the budget constraint, λ_2 must be positive and prices must be above marginal costs in the manner shown by equation (6).

In summary, if congestion costs remain constant or rise in the face of equi-proportionate increases in trips and in infrastructure costs, then unadjusted marginal cost pricing is optimal, but revenue shortfalls in any service must be met by surpluses generated in other services. (With constant or increasing-cost congestion functions, all services cannot be in deficit.) If congestion costs fall in the face of equi-proportionate increases in services and cost outlays, then marginal cost pricing is not optimal; λ_2 will be positive and optimal prices are all above marginal costs in inverse proportion to the value of the bracketed terms in equation (6). Let me turn now to consider the possibility that transportation prices will be required to satisfy distributive goals.

4. DISTRIBUTIVE CONSIDERATIONS IN PRICING

As long as equation (1) is satisfied, then transportation 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. However, if z^h is not distributed in such a way that (1) is satisfied, then we cannot use the results from (1) in generating equation (5), from which the basic pricing equation is derived. Instead, we might proceed in the following way.

Take the first and second terms on the right-hand side of equation (2), multiply and divide each by $\frac{\partial u^h}{\partial z^h}$ and define $\beta^h \equiv \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial z^h}$, so that

$$\frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial x_i^h}\frac{\frac{\partial u^h}{\partial z^h}}{\frac{\partial u^h}{\partial z^h}}+\sum_h \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial x_i}\frac{\frac{\partial u^h}{\partial z^h}}{\frac{\partial u^h}{\partial z^h}}=\beta^h\frac{\partial v^h}{\partial x_i^h}+\sum_h \beta^h\frac{\partial v^h}{\partial \gamma_h}\frac{\partial \gamma^h}{\partial x_i},\ \forall h,\forall i$$

The β^h are sometimes called "welfare weights;" they measure the relative social value, according to W, of a unit of income going to different people.[5] Most social welfare functions would likely assign higher weights to lower income people. Using the above expression, we may now re-write equation (5) to get

$$\beta^h\frac{\partial v^h}{\partial x_i^h}=-\sum_h \beta^h\frac{\partial v^h}{\partial \gamma^h}\frac{\partial \gamma^h}{\partial x_i}+\lambda_1\frac{\partial \Gamma}{\partial x_i}-\lambda_2[p_i+x_i\frac{\partial p_i}{\partial x_i}+\sum_{j\neq i} x_j\frac{\partial p_j}{\partial x_i}-\frac{\partial \Gamma}{\partial x_i}],\forall h,\forall i \qquad (10)$$

For any trip type i, the right-hand side of (10) is the same for all h, just as it was in equation (5). But the left-hand side has a significant difference. It has a marginal utility in numeraire units, which the maximizing consumer equates to price, but this is multiplied by a welfare weight which in general will be different for each person. Since the product of the two terms on the left-hand side is the same for everyone, the implication of (10) is that optimal prices should be differentiated by person, with people having a higher β^h paying a lower price, and vice versa. Our assumption so far, however, has been that the transportation sector cannot price differentially to different people (if it could, it would also be efficient to differentiate according to the different price elasticities of different people — see note 2). Retaining this assumption, the constant price of type-i trips is p_i , and we can re-write the left-hand side of (10) as $\beta^h\left(p_i-\left(p_i-\frac{\partial v^{h*}}{\partial x_i^h}\right)\right)$, where $\frac{\partial v^{h*}}{\partial x_i^h}$ is the optimal price of i-type trips for person h. Using Ω_i to represent the right-hand side of equation (10), this gives us the uniform price of i-type trips:

$$p_i = \frac{\Omega_i}{\beta^h} + \left(p_i-\frac{\partial v^{h*}}{\partial x_i^h}\right) = \frac{\Omega_i}{\beta^h} + l_i^h\ ,\forall h,\forall i \qquad (11)$$

The l^h term represents the loss at the margin for person h that results from the difference between the optimal, differentiated price and the uniform price p_i . If the term is negative, then this is a gain for that person.

Write the mean β for the whole population as $\bar{\beta}^h$ and suppose that β types using any given service i are distributed normally about a mean $\bar{\beta}_i^h$. If optimal redistribution had occurred outside the transportation sector, then all β^h would be equal, and equal to λ_1; if this is substituted for the β^h in equation (11), then all losses are zero. For all i, the equation collapses to equation (5). However, with nonoptimal initial distribution, the best the transport sector can do is to adjust each service's price so as to reduce l_i^h to zero for the mean user (assuming symmetrically distributed user types around each service mean, $\bar{\beta}_i^h$). This entails, as equation (11) indicates, lowering p_i below the base-model optimum if $\bar{\beta}_i^h > \bar{\beta}^h$ and raising it if $\bar{\beta}_i^h < \bar{\beta}^h$. If p_i^* represents the optimal price with optimal redistribution, then the optimal prices without such redistribution will be given by

$$p_i = \frac{\bar{\beta}^h}{\bar{\beta}_i^h} p_i^* \tag{12}$$

With this pricing rule, the conditions that lead to a binding or a slack revenue constraint for the whole sector will still be those discussed in section 3, provided only that the distribution of service use, x_i, over types of service is independent of the distribution of the mean welfare weights, $\bar{\beta}_i^h$, over types of service.

5. IMPLICATIONS OF WHOLE-SECTOR OPTIMAL PRICING

The distinguishing feature of whole-sector pricing is the interrelated effect of the several goals — efficiency, budget balance, and distributional equity — on the prices of all services. This interrelationship can be most conveniently summarized in the following equation, which combines equations (6) and (12). The optimal, whole-sector prices for all services i is given by

$$p_i = \frac{\bar{\beta}^h}{\bar{\beta}_i^h} \frac{\Phi_i}{\Lambda_i} \tag{13}$$

where $\Phi_i = -\sum_h \frac{\partial v^h}{\partial \gamma^h} \frac{\partial \gamma^h}{\partial x_i} + \frac{\partial \Gamma}{\partial x_i} + \frac{\lambda_2}{\lambda_1} \frac{\partial \Gamma}{\partial x_i}$, the right-hand side of (6), and, as in equation (9), $\Lambda_i = 1 + \frac{\lambda_2}{\lambda_1} + \frac{\lambda_2}{\lambda_1} \xi_{ii} + \frac{\lambda_2}{\lambda_1} \sum_{j \neq i} \frac{p_j x_j}{p_i x_i} \xi_{ji}$, the bracketed term on the left-hand side of (6). If optimal redistribution takes place outside the transportation sector, then all the welfare weights, β^h, are identically equal to λ_1 and (13) is simply the pricing rule given by (6). If the budget constraint is not binding, then $\Lambda_i = 1$ for all i, and the pricing rule becomes the usual congestion-model, marginal cost pricing rule, modified by the distribution effect, if any.

This model provides a formalization of what has recently been called an "integrated approach" (May and Gardner 1990) and a "package approach" (Jones 1991). A few simple examples may help draw out its implications. Suppose that congestion is subject to constant or increasing costs across the whole sector, so that the budget constraint is not binding. Optimal prices should then be set at marginal congestion plus service costs for each service, modified up or down in inverse proportion to the average welfare weights of the service users. Suppose welfare weights are inversely related to income, and suppose that the average income of automobile commuters was just the population average, the average income of commuter rail commuters was higher than the population average, and that of public transit commuters was less than the community average. Then optimal prices for private-vehicle road use in commuting periods would equal marginal costs, prices for commuter rail travel would be above marginal cost and prices for public transit would be below marginal cost. The full costs of an individual service is not well defined if joint costs, J, exist across the whole sector. If there are no joint costs, then whether service revenues cover, exceed or fall short of service costs depends on the effect on the congestion terms of each service separately. Surpluses in one sector offset deficits in another, at optimal prices.

Suppose that service and infrastructure expansion lowers congestion costs, which is the condition under which the budget constraint is binding. $\Lambda_i < 1$ for all services and, neglecting distributional effects, optimal prices will be above marginal costs by a proportion that depends mainly on the price elasticity of demand for the service. If, for example, the price elasticity of off-peak public transit use was less than that of peak use, then, other things equal, the Λ_i of off-peak use would be smaller and the price of off-peak use would be proportionately higher. Or suppose that service i was commuter parking in the central area. A very inelastic response to higher parking prices is sometimes taken to be an indication that such a policy is neither optimal nor effective. But, within the context of the whole-sector model, a low price elasticity (a high $|\xi_{ii}|$) is a signal for a higher optimal price, relative to marginal cost.

NOTES

1. Equation (5) 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, including person h whose trip consumption is being varied. 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 on the left-hand side, but this marginal utility minus the h^{th} term of the congestion sum. This in turn means that there would have to be a different optimum price for every individual, because with this change the right-hand side would not be the same for everybody. This problem has typically been handled 1) by accepting the mathematical problem as theoretically insurmountable unless everyone has identical congestion terms; 2) by acknowledging the problem but ignoring it by assuming that there are so many consumers that the congestion effect will be approximately the same no matter whose self-congestion is removed from the sum; and 3) by adopting a measure-theoretic formulation that assumes a dense continuum of consumers with a congestion integral that remains constant no matter how many individual consumers are removed from under the integration sign. Since I want to retain heterogeneity among the consumers, I have accepted implicitly either the second or the third alternative.

2.The inverse elasticity ξ_{ii} is the proportionate change in price that accompanies a unit proportionate change in consumption. The cross-service inverse elasticity, ξ_{ji} , is the proportionate change in the price p_j that is necessary to keep demand for x_j constant in the face of a unit proportionate increase in the consumption of x_i (which entails a fall in p_i). Inverse elasticities and normal elasticities are related this way: $\xi_{ji}=e_{ji}/(e_{jj}e_{ii})$, $j\neq i$; and $\xi_{ii}=1/e_{ii}$. If we allowed the transportation sector to charge differentiated prices to different consumers, then the price effects in equation (5) would be specific to each consumer h. Optimal prices would then involve, with $\lambda_2 \neq 0$, charging higher prices to those consumers with high own-service inverse elasticities (low normal price elasticities).

3. Notice that the congestion-cost term is a sum of costs to all persons whose welfare is being taken into account. There is no reason why this set of people would include only those who travel. In other words, "congestion" includes the costs of externalities, including pollution costs, over the whole jurisidiction and not just the cost of time delays experienced by travellers.

4. It is usually assumed that marginal operating costs are constant, so that the second term on the left-hand side of equation (9) exactly covers total operating costs. If, however, marginal operating costs are rising or falling, then there may be some inframarginal surpluses or losses associated with this term.

5. David Starrett discusses this procedure in Starrett 1988, pp.11-15.

REFERENCES

Baumol, William J. and David F. Bradford. 1970. "Optimal Departures from Marginal Cost Pricing," *The American Economic Review*. Vol. 60. Pp. 265-283.

Boiteux, M. 1956. "Sur la gestion des monopoles publics astreint á l'équilibre budgetaire," *Econometrica*. Vol. 24. Pp. 22-40.

Jones, Peter. 1991. "Gaining Public Support for Road Pricing through a Package Approach," *Traffic Engineering + Control*. April, 1991. Pp.194-196.

May, A.D. and K.E. Gardner. 1990. "Transportation Policy for London in 2001: The Case for an Integrated Approach," *Transportation*. Vol. 16. Pp.257-277.

Mohring, Herbert. 1976. *Transportation Economics*. Cambridge, Mass: Ballinger.

Newbery, David M. 1990. "Pricing and Congestion: Economic Principles Relevant to Pricing Roads," *Oxford Review of Economic Policy*. Vol. 6, No. 2. Pp. 22-38.

Ramsey, F.P. 1927. "A Contribution to the Theory of Taxation," *Economic Journal*. Vol. 37. Pp. 47-61.

Starrett, David A. 1988. *Foundations of Public Economics*. Cambridge: Cambridge University Press.