10 URBAN TRANSPORT PRICING AND FINANCE

Prices allocate resources and raise revenue. Urban transport pricing is complicated by the multiplicity of objectives pursued and by the institutional separation of road infrastructure from operations, of infrastructure pricing from charging, and of roads from other modes of transport. In the interests of both urban transport integration and sustainability, developing countries should move toward prices reflecting full social costs for all modes; to a targeted approach to subsidization reflecting strategic objectives; and to an integration of urban transport funding.

THE ROLE OF PRICES IN URBAN TRANSPORT

In most markets, prices have two main functions: to ration and allocate the use of resources between the production of different products, and to finance production. In transport in general, but in urban transport in particular, the performance of these functions is subject to three major complications:

a. The separation of responsibility for infrastructure from that of service provision

b. The pursuit of multiple objectives, particularly in public transport policy

c. The separation of infrastructure financing from charging.

In this chapter we address each of these complications and suggest ways in which municipal governments may improve their effectiveness in the provision of transport services by developing better pricing and funding arrangements.

CHARGING FOR THE USE OF INFRASTRUCTURE

Roads are "congested" when traffic volume reaches a level at which the flow of traffic is significantly impaired. In such circumstances the extra cost to society of the marginal trip is not only the cost that the extra vehicle itself incurs but also the sum of the marginal delays that it imposes on all existing road users. Unless the cost charged to, and perceived by, each user reflects that extra cost imposed on society, there will be an incentive to make road trips for which the benefit to the trip maker is lower than the extra total cost to society.[1]

In practice, in most countries—industrialized and developing—urban roads are provided to their users without any direct charge. The only payments from the private user to the public supplier that vary with the amount of road use come indirectly in the form of taxes (primarily on fuel). Underpricing is endemic. Even in developing countries with relatively higher taxes on fuel, taxes do not reflect the costs of urban congestion.

The fact that these taxes do not cover costs has several adverse effects. First, it distorts the choice of mode in favor of road transport, particularly private cars. Second, it encourages excessive use of the infrastructure (which may cause "excess" congestion). Third, because there is no direct revenue, it is not logically possible to use conventional commercial investment criteria in deciding how much capacity should be provided. Fourth, because the revenues do not accrue to the responsible local authority, there may be inadequate money for proper maintenance of the existing infrastructure. For all these reasons, it is desirable to ensure that the price charged to users at the margin covers the full social cost of their trips.

CONGESTION PRICING

The concept of congestion pricing is that road users should pay a price that reflects the short-run marginal social cost of road use, and that hence varies according to the prevailing level of congestion.[2] Congestion pricing has long been advocated by economists on the grounds that, unlike the various administrative controls which are often used to manage traffic, it gives correct incentives over the whole range of dimensions involved in travel decisions, including choice of destination, time of travel, mode of transport, route, and so on. Moreover, if congestion pricing were applied in the context of a flexible land and property market, the city would evolve toward a more compact form, with more mixed land use, fewer resources devoted to the spread of the road network into surrounding areas, and more funds available for upgrading infrastructure in the already urbanized area.[3] Above all, unlike administrative restraint alternatives, congestion pricing yields revenue rather than being a cost to the public purse (box 10.1).

For these reasons the World Bank strongly advocated congestion pricing in the 1975 urban transport policy paper (World Bank 1975), and has subsequently examined its applicability in industrialized countries.[4] Nevertheless, attempts to introduce it in Kuala Lumpur and Bangkok in association with Bank projects foundered, as did the initiative in Hong Kong, China, in the early 1980s. To date, the most sophisticated application remains that in Singapore, where an area licensing scheme (ALS) first introduced in 1973 has now been developed into a much more sophisticated electronic pricing system. Instead, various surrogate measures, such as the use of parking restraints and pricing policies, have been adopted in many countries. Even in industrialized countries, there is growing concern that these surrogates have not worked well. Direct charges for urban roads have been introduced to generate revenue in some Norwegian cities, and congestion pricing is now under serious consideration in the Netherlands and the United Kingdom.

Many of the objections raised against congestion pricing have already been overcome. Initial concerns about the cost and reliability of the technology have been superseded by developments in electronics. Fears about intrusion on privacy—which contributed to undermining the Hong Kong experiment—can be overcome through the choice of a technology not relying on any centralized record of vehicle movements. Fears about the effects of congestion charging on lower income groups, which have been a significant obstacle to its political acceptability in highly motorized societies,[5] are much less justified in developing-country cities, where most people remain dependent on public transport or nonmotorized modes.[6] Even some of the rapidly motorizing major cities of Eastern Europe still offer much more extensive and frequent public transport services as an alternative to the private car than are available in many of the Western countries. Specific inequities that do arise in the shift to a more direct and efficient system of charging for road use can also be compensated by provision of case-specific subsidies, such as free or reduced-price smart cards or stickers during a transition period. Congestion pricing is increasingly being viewed as a concept whose time has come.

Congestion prices can be charged with different degrees of precision through a variety of tech-

BOX 10.1 THE NAMSAN TUNNELS IN SEOUL: SIMPLE ROAD PRICING REDUCES CONGESTION AND FINANCES TRAFFIC MANAGEMENT

Traffic congestion in Seoul increased dramatically during the 1980s and early 1990s despite extensive construction of new urban freeway and subway lines. In 1996 the Seoul metropolitan government commenced charging 2,000 won (US$2.20) for the Namsan #1 and #3 tunnels, two corridors with high private vehicle use linking downtown Seoul to the southern part of the city. Charges were set for one- and two-occupant private vehicles (including driver) and collected in both directions per entry or exit from 7:00 a.m. to 9:00 p.m. during weekdays and from 7:00 a.m. to 3:00 p.m. on Saturdays. Private cars with three or more passengers, taxis, and all kinds of buses, vans, and trucks were exempted from charges, as was all traffic on Sundays and national holidays.

In the two years following commencement of the congestion pricing scheme for the Namsan #1 and #3 tunnels, there was a 34 percent reduction in peak-period passenger vehicle volumes, the average travel speed increased by 50 percent, from 20 to 30 km/h, and the number of toll-free vehicles increased substantially in both corridors. On the alternative routes, traffic volumes increased by up to 15 percent, but average speeds also increased as a result of improved flows at signalized intersections linked to the Namsan corridors and increased enforcement of illegal on-street parking on the alternative routes.

The whole of the annual revenue from the two tunnels (equivalent to about US$15 million) goes into a special account used exclusively for transport projects, including transport systems management and transport demand management measures throughout the city.

Source: Hwang, Son, and Eom 1999.

niques of different degrees of technological sophistication and cost. While, theoretically, different prices can be set for each link in the network for each time period, in practice a rather rougher approximation may be used both for reasons of practicability of application and for predictability of response from drivers. Three principal forms of congestion pricing have been developed to date:

Cordon pricing, or area licenses, can be implemented with simple technology, to charge for the right to access or circulate within limited geographical areas, with some degree of time differentiation. The principal application has been the area licensing scheme for downtown Singapore in operation between 1975 and 1998. The scheme being prepared for central London is of this type.

Time-dependent tolling of individual roads or road lanes can charge for congestion on major highways and improve traffic flows on the affected facility (though not necessarily beyond it). In developing countries it has been used on a number of tunnels in Hong Kong, the Namsan tunnels in Seoul, as well as on the expressways in Singapore.

Electronic road pricing (ERP) enables more precise differentiation of charges by road, time of use, and type of vehicle for whatever area is covered. Recent developments in intelligent transport system technologies make this much more attractive, and several large test applications have been undertaken. In 1998 Singapore replaced its ALS by an ERP system that applies to the city center and some major access roads, and is gradually being extended as needed (box 10.2).[7]

In principle, once roads are charged at a price equal to the marginal social cost, the decision by the road authority as to whether and when to invest in expansion of the road should depend on the relationship between congestion-pricing revenues and the costs of the expansion.[8] In practice, where the pricing system is only roughly applied, it may still be necessary to undertake a cost-benefit analysis of proposed investments. Nevertheless, the application of the pricing scheme will create the proper base on which to undertake the appraisal and will have generated a revenue pool that can be used to finance extra capacity.[9]

The fact that efficient congestion prices, or parking charges used as a proxy for them (see discussion below), yield revenues in excess of the total direct costs of congestion is sometimes used as an argument against them. This should not be seen as a problem but as an opportunity. High user prices are quite apposite to the shortage of street space relative to built area in many developing-country cities. The distributional implications are very progressive because those who do have cars are predominantly high-income earners likely to attach the highest value to reduced travel time and journey predictability. They may also be democratically viable, since only a minority of the wealthier citizens would be called upon to pay the increased charges due to low car ownership. For all of these reasons, congestion pricing, accompanied by use of the revenues to improve public transport and other city amenities to the benefit of the poor, can be recommended for the typical cash-strapped municipality in a developing or transition economy.

BOX 10.2 ELECTRONICS IMPROVES ROAD-PRICING EFFICIENCY IN SINGAPORE

In September 1998 the government of Singapore replaced the manually enforced ALS by ERP. It generally covers the same area as that for the ALS, but has been extended to work on principles similar to ALS, but with extension to approach and bypass roads. All vehicles are required to have an electronic in-vehicle unit (IU) that accepts credit in the form of a smartcard. Tolls are automatically paid when the vehicle passes under a gantry and a liquid crystal display indicates the current credit balance. Tolls do not fluctuate in relation to actual traffic volumes but are adjusted quarterly to ensure optimum traffic speeds. The system cost $200 million in Singapore dollars (US$125 million) to implement, one-half of which was for the free fitting of IUs.

The ERP system was not implemented to increase government revenue. ERP charges are generally less than the corresponding ALS fees, although the ERP system's per-pass charging principle means that those motorists who use the priced roads must now pay more. Overall, revenue is about 40 percent less than that previously collected from the ALS, but electronic charging gives greater flexibility to set charges that are just sufficient to keep the roads free of congestion.

Once new traffic patterns stabilized, weekday traffic volume entering the restricted zone dropped by 20 to 24 percent, from 271,000 vehicles per day to between 206,000 and 216,000. With those lower traffic volumes, average traffic speeds in the zone increased from 30–35 km/h to 40–45 km/h. Improvements are less clear on the three expressways in the ERP scheme, and the Land Transport Authority is reviewing ERP charges to further optimize traffic flow.

Source: PADECO 2000.

FUEL TAX AS A SURROGATE USER PRICE

For the moment, however, the tax element of the pump price of fuel is the main charge directly associated with road use. Fuel price is thus very important in urban transport. For the private motorist, it is usually the only incurred cost that is perceived in making a marginal trip. It may thus directly affect the amount of travel undertaken, choice of transport mode, and choice of vehicle technology, and may less directly affect the trade-off between location and transport expenditure.

In an ideal world, the price of fuel should cover its resource cost (the border price); maintenance of roads and congestion costs should be charged for directly through highly differentiated tolls; environmental costs should be charged for through emissions charges; and any redistribution objectives should be pursued through nondistorting lump sum taxes.

In reality, that prescription is not achieved currently in either industrialized or developing countries. Most roads are open access (particularly in urban areas), and cannot be subject to highly differentiated direct charges. The technology for location- and pollutant-specific emissions charging does not exist. Lump sum transfers or naïve direct distributional taxes are not possible. In the absence of direct charging mechanisms, fuel taxation is often looked to as a proxy for a number of other, theoretically preferable, taxes or charges relating to road maintenance, congestion, and environmental and distributional objectives. It is thus important to examine the potential of fuel taxation to perform these various functions.

As a charge for ***road maintenance,*** fuel tax does not differentiate well by vehicle type, and needs to be supplemented by axle-weight charging. Similarly, as a ***congestion charge,*** fuel tax does not reflect the structure of congestion well because fuel consumption is not sufficiently sensitive to vehicle speed variation. As an ***environmental charge,*** it is likely to be a very good proxy for carbon emissions taxation, but many other emissions are not proportional to fuel consumption and vary by fuel type, vehicle type, and extent of emissions technology employed. As a ***distributional*** instrument, it has the right combination of low price elasticity and high income elasticity, and is therefore a very promising redistributor in developing countries.

Fuel tax does have one great advantage compared with taxation on vehicles (vehicle excise duties, registration taxes, insurance charges, and so on), which does not vary with vehicle use. That concerns the relationship with public transport fares, which vary directly with use (period passes are relatively unpopular in industrialized countries because of the element of advance payment involved). If taxation on private cars is not related to use, the consequence is that the cost of car use for the marginal trip may appear very inexpensive in comparison with that of public transport. One of the great advantages of increasingly including payment of what are presently fixed charges within fuel taxation, or other distance-based charges,[10] is that it puts private and public transport pricing structures on a more-comparable, and hence less-distorting, basis.

Reliance on fuel taxation as a surrogate for direct charging faces some severe administrative difficulties. There are multiple transport fuels, used in multiple sectors, to achieve multiple objectives. Diesel, which may merit heavy taxation for local air pollution reasons, is also the primary fuel for road freight transport, for which reason, as an intermediate input rather than a final consumption good, it would justify a relatively low tax burden. Kerosene typically carries low taxation, or is even subsidized, because of its use for domestic heating in many countries. Unfortunately, if these fuels carry low tax, there is a high likelihood of substitution of diesel for gasoline vehicles, or the adulteration of gasoline with kerosene or diesel, either of which could be environmentally very damaging. Particularly where there is spatial differentiation of fuel taxation for congestion charging reasons, fuel carrying could also be a problem in everything except the very largest cities.

Above all, however, fuel taxation is usually a national prerogative, with revenues accruing to the national treasury rather than to the municipality. For this reason it is likely to be difficult to fit it into, or coordinate it with, a city-level strategy, except on the basis of a countrywide agreement on allocation of all transport-related tax revenues and expenditure responsibilities.

OTHER LESS-DIRECT TAXES AND CHARGES FOR ROAD USE

Other vehicle taxes may also perform a function of allocating road costs equitably and efficiently among vehicle categories. In particular, annual charges based on standard axles are widely used to correct for the fact that road wear is much more a function of axle weights than of fuel consumption. Even though these taxes may be fixed centrally, the revenues yielded sometimes accrue directly to a local jurisdiction.

Some taxes have a very clear allocation function. In Denmark and Hong Kong, acquisition taxes that approximately triple the cost of cars have been major factors limiting motorization (presently some 330 and 60 cars per 1,000 inhabitants, respectively).[11] The most extreme form of this is in Singapore, which in 1990 introduced a system in which a rationed number of entitlements to own a motor vehicle (the so-called Certificates of Entitlement) are auctioned each month. In combination with preexisting taxes, this has resulted in new cars retailing at some four or five times their world market price. Other relevant experiences are those of Tokyo, where prospective car purchasers are required to provide proof of having an off-street parking space; and of various Chinese cities, where entitlements to register a motorcycle have been sold and local circulation of motorcycles registered elsewhere banned.

Parking charges are often used in place of direct road-use charges. While they cannot deter people from driving through an area or reflect the different distances or routes drivers take, they can be varied by time and place in such a way as to capture a significant part of the congestion externality and encourage better spatial or temporal distribution of demand for movement on the roads.[12] They are already manipulated in quite sophisticated ways in many European cities, and both the United Kingdom and France are debating taxation of downtown parking spaces provided by employers for their employees.[13] Transferable permits for a given number of available parking spaces would be a way of auctioning the right for commuters to drive to work.

In summary, an ideal charging regime for the use of infrastructure should allocate all of the associated costs (congestion, environmental, wear and tear on the infrastructure, and so on) to the users incurring them, as far as possible directly in proportion to the costs imposed. That is not completely achievable even in principle because there are some costs (such as lighting, street cleaning, and so on) that have the characteristic of public goods and hence must be allocated more arbitrarily. Because the imposition of the different types of cost varies between vehicle types according to a range of characteristics of vehicles and their use, only the simultaneous application of a number of different charging devices could meet this ideal.

The design of a charging system thus needs to be constructed from a carefully designed combination of different charging components. For simplicity's sake, it may be desirable to limit the number of these components, but a number of general elements can be recommended.

- Fuel should never cost less than border price in any circumstances.
- Where direct congestion charges are possible, these should be varied by time of day and structured according to the congestion-creating equivalents of different vehicle types.
- Wear-and-tear costs should preferably be recouped on a basis variable both by wearing effect and distance traveled (standard axle kilometers). Where this is not possible, the

wear-and-tear cost may be incorporated partly in a fuel tax surcharge and partly in vehicle category–related charges (preferably also based on actual annual use of each vehicle).
- Marginal environmental impact costs and marginal public costs of accidents might be recouped through local fuel tax and insurance surcharges.
- Any sumptuary, or "luxury" taxation, levied on road users for general taxation (non-transport) reasons should be levied on passenger transport rather than freight transport in order to minimize economic distortion.
- Where diesel fuel is taxed at a low rate, some compensating tax should be imposed on the procurement and use of diesel-powered light passenger vehicles to minimize adverse environmental consequences.

PUBLIC TRANSPORT PRICING AND FINANCE

Setting public transport prices and raising the necessary finance, raises problems because of the multiple objectives faced by decisionmakers. The primary objective of public transport pricing is to generate revenue that can ensure an efficient and adequate supply of public transport service. Public transport pricing may also be expected to contribute to the reduction of congestion and environmental impact of road traffic, efficient coordination between public transport modes, and the reduction of poverty. It is commonly argued that if urban public transport is to satisfy these latter objectives, it cannot be expected to cover its full costs. Urban public transport is consequently subsidized in many major cities in industrialized countries.

Similar policies have traditionally been applied in the transition economies and some postcolonial developing countries, but many of these countries are no longer in a position to fund such policies, and their public transport sectors are facing decline because of cash starvation. This section considers how to establish a basis for setting policies on pricing and cost recovery in urban public transport that best reconciles the multiple objectives.

EFFICIENT SUPPLY

There are two important aspects of supply efficiency. First, it is necessary to provide the most beneficial range of services with the resources available—"doing the right thing." Second, it is necessary to supply the required services at the least-possible cost—"doing the thing right." Neither is simple and neither is well achieved in most developing countries.

The difficulty in doing the right thing is that there may be legitimate reasons, both on equity and efficiency grounds, why a pure market outcome would not be optimal. For these reasons, municipal governments may intervene to define both the services to be provided and the fares to be charged. In the absence of the capability to make any more sophisticated calculation, this typically takes the form of setting flat fares for a defined network of basic services. Public sector monopolies are often protected because of the difficulty of enforcing a flat-fare policy with multiple operators.

Such monolithic simplicity is rarely optimal. Transport users have different preferences (partly as a result of differences in income). It may therefore be possible to increase consumer welfare while increasing profitability by supplying different products to different segments of the market. In urban bus transport, this means the provision of "premium" services—such as express services or air-conditioned services—at premium prices. Price discrimination, in this sense, may benefit all users. Operators should therefore be encouraged to examine such revenue-enhancing strategies as an alternative to reliance on subsidy. This strategy has already been widely adopted as a means of maintaining urban bus service in Seoul, Buenos Aires, Bangkok, and many other developing-country cities.

Doing the thing right is what commercial competition usually ensures, since the threat of bank-

ruptcy is a powerful stimulant to internal efficiency. The possibility of subsidy weakens that incentive. Particularly in parastatals, subsidies strengthen the likelihood that organized labor will appropriate part of it to support better wages and working conditions than a competitive market would be able to support. In the United Kingdom it was estimated that over one-half of subsidies aimed at lowering fares or improving service quality actually "leaked" into benefits for management or workers or were lost through reduced efficiency of operation.[14] Exploring measures to reduce subsidy requirements through improved efficiency and reduced costs is thus the first step to take in formulating a public transport subsidy strategy.

Much of the argument for transport deregulation and privatization (see chapter 7) derives from a need to reduce subsidies. Attempts to increase efficiency of operation through the introduction of competitive pressures within the sector may enable lower prices to be charged without recourse to subsidy. The implication is that, for any level of cost recovery lower than 100 percent, subsidies should be specifically targeted at the objectives sought and should be embodied in competitively tendered service contracts.

The macroeconomic effects of urban transport subsidies also need to be considered. In the short run, they may reduce the general price index if transport has a disproportionately heavy weight in the index. Particularly where urban transport is a significant item in the consumption pattern of strongly unionized labor groups, there may be some associated relief from wage pressure. In the longer run, however, it is inevitable that the inflationary pressure will spill over to other items that will have to be taxed to finance the subsidy, or through increased direct income taxation–inducing compensatory wage claims. Deficit budgeting has similar inflationary consequences. The general World Bank position is that subsidy is the wrong tool to deal with inflation; there is nothing special about the transport sector to vitiate this view.

THE IMPACTS ON ROAD CONGESTION

In industrialized countries it is commonly argued that public transport should be subsidized in order to entice trip makers out of private cars and hence reduce road congestion. Where road congestion is concentrated in particular locations or at particular times of the day, this implies a highly differentiated structure of subsidy. Where road congestion is systemwide and pervasive throughout the day, general public transport subsidy may be called for on this argument.

Despite the superficial attractiveness of this line of reasoning, there are several considerations that counsel caution in the use of public transport subsidy as a countercongestion instrument. First, there is the supply efficiency impact discussed earlier. Only if the benefits resulting from a more efficient modal split were greater than the disbenefit of any reduction of supply efficiency associated with subsidy would there be a net benefit overall.[15]

Second, there is a problem of targeting. Where congestion is limited in time or spatial extent, it becomes increasingly difficult to target public transport subsidies as a response to the problem. Highly peaked road congestion calls for peak-hour public transport subsidies, which may have the perverse side effect of shifting public transport demand from off-peak to peak periods. For spatially variable congestion, differential subsidy policy by route may be extremely difficult to structure even if competitively tendered service franchising allows each route to be treated differently.

Third, there is a problem of fiscal cost. Where the basic purpose of the subsidy is to divert traffic from the private auto to public transport, the ideal situation would require very high cross-elasticity of demand for auto use with respect to public transport pricing. Empirical evidence (albeit mostly from industrialized countries) suggests that in fact this cross-elasticity is very low (perhaps 0.1) in the short run. While it may be somewhat higher in the long run due to effects on car

ownership, it would still appear to be an instrument of relatively weak leverage.

Finally, there is a problem of perverse land-use effects. Public transport subsidy to countervail systematic undercharging for private transport will mean that all transport is subsidized. This will tend to generate excessive travel and inappropriately sprawling land use. Measures to redress inadequacies in the system of charges for private transport should always be a condition attached to the financing of "second-best" subsidies.

PUBLIC TRANSPORT COORDINATION

Allocation of traffic between competing modes within an urban transport system would be efficient if prices were everywhere equal to short-run marginal social costs. The general prescription for covering fixed costs where these will not be recouped by short-run marginal cost pricing is to minimize distortion by loading the fixed costs most heavily on those parts of the market for which the demand is least likely to be affected by a price increase. This is known as the "Ramsey pricing rule" (box 10.3). In multimodal urban transport systems, which are, as a whole, in financial equilibrium, this may imply cross-subsidy between modes, particularly when the modes have different cost structures. The application of this approach to urban public transport pricing requires consideration, not only of the implications of the relative congestion externalities of alternative modes but also of their different cost structures.

Where two modes are operated within a single organization, and entry is regulated, it is possible to reconcile different levels of cost recovery for the two with any overall cost-recovery requirement (including breakeven overall). It is quite common for this to occur in metropolitan public transport undertakings (though it should be remarked, in

BOX 10.3 RECOVERING FIXED COSTS: THE RAMSEY PRICING RULES

The Ramsey inverse elasticity mark-up rule concerns the allocation of fixed costs between alternative products within a production agency. If a multimodal transport system is viewed as a single supply agency, the Ramsey pricing rule can be interpreted as a rule about the allocation of system-fixed costs between modes. For example, in typical bus operations, over 90 percent of costs vary with respect to either the number of vehicles employed or the number of bus kilometers run. Short-run marginal cost pricing would nearly cover full cost. The same is not the case for rail systems, where typically only 50 or 60 percent of costs are directly related to the service provided. In such conditions, the most efficient outcome may involve different levels of total cost coverage by the modes and transfers between them.

If cross-elasticities between alternative transport forms are zero, the objective of efficiently covering total costs will be achieved by adding mark-ups to the marginal social cost in proportion to the reciprocal of the price elasticity of demand. The rule then becomes rather more complex. Consider, for example, the ratio of peak and off-peak prices. If the peak demand were completely inelastic and the off-peak demand infinitely elastic, the whole of the burden would fall on the peak price. More realistically, some users may shift from peak to off-peak as relative prices change. With nonzero cross-elasticities, the ratio of the mark-ups of different service should take into account the cross-elasticity as well as the own-price elasticities.

Source: Authors.

passing, that where it does occur, it is often regarded as a form of illegitimate cross-subsidy).

However, when the modes are operating in competition with each other, attempting to apply the inverse elasticity rule creates a conflict between the interests of pricing for allocative efficiency and those of pricing for commercial viability. Competitive operation of such mixed-mode systems on a purely commercial basis will result in a suboptimal level of patronage for those modes with the greatest proportion of fixed cost (empty trains and overfull buses). It will also affect the structure of services provided. For example, commercial operation of buses and metros within an urban area may lead to a smaller amount of restructuring of bus services to take advantage of the complementarity of the modes and a larger amount of trunk service by buses than would be optimal. Recent experience in Rio de Janeiro suggests that concessions of urban rail operations to the private sector may make the achievement of fare integration more difficult.

Not all modal coordination considerations militate in favor of public transport subsidy. Because congestion must, by definition, be greater at peak than at off-peak hours, the marginal social cost of the peak period private car user must always exceed the average cost. The marginal social cost of peak-hour public transport may also exceed its average cost, because peak service requires extra vehicle capacity costs and may impose high marginal labor costs (resulting from labor agreements requiring payment for a minimum number of hours per day exceeding the number of peak hours, and premiums for any splitting of working shifts). The marginal social cost of peak public transport provision thus depends not only on the level of congestion but also on the disparity between peak and off-peak demand levels and the industrial relations context in which the service is being supplied. For cities with relatively low levels of road congestion but high public transport peak-to-off-peak demand ratios, optimal peak public transport subsidies may be small (or even, conceivably, negative).

EQUITY

In many cities public transport fares are controlled in an attempt to maintain an affordable service to the poor who have no alternative form of travel. While in principle lump sum transfers would be a preferable way to redistribute income, in practice they are rarely feasible politically. Increases in public transport fares are therefore politically sensitive. (Five people died in riots following a fare increase in Guatemala in April 2000.) The question is thus not whether the objective of maintaining affordable public transport service is important and desirable, but whether it can be practically and cost-effectively achieved with the policies typically adopted for the purpose. The main characteristics to review in this context are the general fare level, fare structures, fare concessions, and fare discrimination.

Fare levels. In most low- and middle-income developing countries, the average income of public transport users is below that of motorized private transport users, and below the overall average income. If fares are set below cost in such circumstances and operating deficits covered either from a progressive income tax or a tax on private motoring, the distributional (poverty alleviation) impact will be positive. If, however, the control on fares is not directly funded, the long-term effect will be to reduce first the quality and eventually the quantity of public transport available. Whether the poor actually benefit from price controls then depends on the tradeoff between less expensive fares and poorer service. There is considerable evidence, even in relatively poor countries, such as the Kyrgyz Republic, that the poor are willing to pay more for a better service than that provided at existing controlled fares. In the extreme, the poor get no benefit at all from setting very low fares if that causes supply to disappear altogether. Decisions on the control of fares should thus be taken in the context of an assessment of the effects of the control on the quality of service and in the context that the continuation of an appropriate level of service can continue to be financed.

Fare structure. Flat fares are frequently adopted throughout a municipal or conurbation area in the belief that this is equitable. Where longer work journeys have been forced on the poor, either through racial discrimination in housing (as in South Africa) or through market-unaware housing planning (as in many former command economies), flat fares may indeed compensate for other forms of discrimination against the poor. The danger, however, is that in larger cities, a flat fare may need to be set at such a high level that it encourages shorter-distance travelers to seek alternative modes, and hence to undermine any "within-mode" cross-subsidy. It is therefore recommended that such spatial distribution issues be handled through specifically designed services and fares, rather than through a general flat-fare system.

Designing a service and fare structure that is both equitable and efficient is particularly difficult in multimodal systems. There may be an efficiency case for structuring bus services to feed high-capacity trunk rail systems and for allowing the buses to cross-subsidize the rail system. In many cases, however, the average bus user is poorer than the average rail user, so what may look efficient from a coordination viewpoint involves the poor subsidizing the rich. The lack of multimodal ticketing accentuates this problem The moral is that wherever explicit attempts are made to secure modal coordination within urban public transport systems, it should be in the context of an integrated fares and charges strategy taking particular cognizance of the effects on the poor.

Fare reductions or exemptions. The distribution of poverty-oriented subsidies should ideally be explicitly related to the income levels of service users. There may be some vulnerable categories of passenger (schoolchildren and senior citizens not in receipt of a full income) that can be easily distinguished and for which charging a lower price is possible. The advantage of discrimination on this basis is that it can be applied to all services and can even be nuanced (for example, to exclude retired people with incomes above some minimum level). In many countries, and particularly in the transitional economies, there is a pervasive tradition of providing free or reduced-fare transport for a wide range of public servants. This raises two major problems. First, the categories in receipt of free fares are often not the most needy, so that the redistribution effects of supporting the concessions by internal cross-subsidy are perverse. Second, where there are a large proportion of non-fare-paying passengers, it becomes more difficult to enforce payment by those who are supposed to pay. Explicit consideration of the extent and justification of fare exemptions and the extent and means of combating fare evasion should therefore be a sine qua non of any fare-determination process. Moreover, where reduced fares are mandated for the support of other sector policies (health, social security, policing, and so on), the costs of the subsidies should be charged directly to the other sector budgets.

Fare discrimination. Where residential locations are highly segregated by income group, specific routes may also be identified for subsidy on income-distribution grounds. For example, at one stage a fare differential existed between Lines 1 and 2 of the metro in Santiago, Chile, with the line serving the lower-income population charging lower fares per kilometer. Less well targeted approaches may have perverse effects. For example, the provision of minimum levels of network density and frequency throughout an urban area at the standard fare level will tend to provide higher levels of service or lower price to areas with low residential density than would be commercially sustainable. In middle-income countries, these areas are areas of high incomes and high car ownership. If such a policy were pursued within a system covering its costs in total, the effect would be a perverse cross-subsidy from the poor to the rich. Maintaining specified network density or frequency by external subsidy has other disadvantages. In particular, there will be a systematic tendency to overprovision, because both those who receive the subsidized service and those who supply it develop a vested interest in high levels of provision.

THE DYNAMICS OF POLICY REFORM: MAKING SUSTAINABILITY PARAMOUNT

There is a close relationship between charges for infrastructure and charges for public transport service. If road use were efficiently priced, there would be no economic second-best case for public transport subsidy. Logically, this suggests the need not only for action to secure optimal pricing for both infrastructure and services but also for careful attention to the dynamics of change. Eliminating public transport subsidy in the absence of efficient infrastructure charges might simply further distort the choice of transport mode. The counsel of perfection would thus be to link any move to improved cost recovery for public transport to the progress made in introducing efficient road pricing.

Within the context of political reality, however, that may be a difficult prescription to observe, particularly where, because of fiscal collapse, the short-term alternative to fare increases is not continued subsidy but loss of service. Generating an adequate cash flow is the most serious problem for many public transport operators in developing countries. Attempts to control fares at uncommercial levels in the absence of a secure fiscal foundation threatens sustainability of service, not only for parastatals (as in many cities in West Africa) but also for franchised private operators (as in Jamaica). Without a secure financial basis, it is impossible to achieve the other objectives discussed. For that reason it is suggested that, while increased cost recovery through the farebox should not be the primary concern of public transport policy per se, the maintenance of a sound financial basis should be at the heart of the pricing policy process.

This is not always well understood. In transitional economies, where parastatal operators are not accustomed to commercial accounting principles, it may be particularly difficult to ensure that proper provision is made for the financing of depreciation. Even where some of the capital investment is sunk in long-lived assets that are fixed or cannot be profitably resold or devoted to any other use, system sustainability still requires enough revenue to cover costs of operation, maintenance, and replacement of rolling stock.

URBAN TRANSPORT FINANCING

The financing of municipal transport is complex and difficult both because of the separation of road infrastructure from operations and because of the multiple objectives that public authorities are pursuing in urban transport policy. In this section we first consider the problems of financing public sector infrastructure, then examine the possibilities of private sector participation in infrastructure finance, and finally consider the institutional possibilities of mobilizing the pricing devices that have been discussed above to solve the financing problem.

PUBLIC FINANCING OF INFRASTRUCTURE

The sources of funding for public sector investments may include transfers from central government, local borrowing, local taxation, and service charging. Public expenditures on urban transport in capital cities may be fully financed by the central government. More generally, and almost exclusively in noncapital cities, the main responsibility for finance will rest with the regional or municipal government with some degree of counterpart funding from the central government. That sharing of financial responsibility can induce overinvestment in infrastructure if it takes the form of automatic counterpart funding of locally generated schemes. Projects should therefore be subject to some formal economic appraisal of investments as a condition for cost sharing.

Local borrowing may be secured against general municipal revenues or occasionally against toll revenue. Some municipalities are at least as creditworthy as their national governments and may be able to issue bonds of their own. That is not usually the case for smaller cities, however, and borrowing may have to be by government on their behalf. In this context it is essential that all claims

for capital resources be based on, and assessed according to, a common investment appraisal criterion and that the municipality should meet the servicing costs of its borrowing.

Local taxing capability is also very limited in many countries, often being restricted to property taxation and various minor licensing revenues. In many developing countries, sales tax and income tax are central government prerogatives. Other local taxes, or taxlike impositions, are sometimes used to supplement user charges for transport services. The French "versement transport" is an earmarked payroll tax. The "vale-transporte" in Brazil is an obligation imposed on employers to finance part of the commuting costs of their employees.

On the principle that "he who benefits should pay," it may seem desirable to capture in tax revenues part of the benefit due to infrastructure investment accruing to local residents or businesses, and to use these funds to finance investments. A "betterment" tax, which appropriates part of the gain in the value of properties resulting from infrastructure investments, does this in principle but is difficult to assess and apply in practice. It is also collected after the investment rather than in advance. Development charges and infrastructure contributions, when assessed systematically, can provide for the extension of infrastructure as cities expand, but can rely on the existence of very strong development controls, which are often lacking in developing countries. The national government of Japan has issued administrative guidance on integrated rail and new town development mandating contributions of the land developer to the rail enterprises with several provisions to be included in their agreement.[16]

Direct charges for service from users of a particular piece of locally funded infrastructure is normally treated as a trading revenue that accrues automatically to the supplying authority—but these direct charges for service are very rare. Indirect charges, such as fuel taxation, usually accrue to the central government. A critical question is whether congestion charges are to be treated as a charge for service or a tax. One of the reasons that road pricing has made so little progress in industrialized countries is that it is often legally classified as a tax, accruing to the central treasury. The British Parliament has recently passed a law to allow municipal authorities to keep road-charging revenues, and hence gives municipalities an incentive to introduce congestion charges.

MOBILIZING PRIVATE FINANCE

Where public transport service is franchised to the private sector, the financing of vehicles and their supporting infrastructure normally becomes the responsibility of the franchisee. The main difficulty in this context is that unless the contracts are well defined, and of reasonable duration, it may be difficult for a private operator to finance the necessary vehicles. For example, one of the major difficulties confronting the introduction of competitive tendering of franchises in Uzbekistan is that in the absence of any history of contracting, neither borrowing for nor leasing of vehicles is possible.

Recent experience in some Latin American and East Asian countries has shown that private finance can also be mobilized for urban transport infrastructure through concession arrangements. Urban expressways have been totally privately financed in Argentina, Malaysia, and Thailand, and the first urban metro in Bangkok was privately financed. In other cases, such as the concessioning of the existing suburban railways in Argentina, where prices were fairly tightly constrained, public contribution was necessary to make private financing of new investment viable.

The need for public sector contribution is not a reason for forgoing private participation, but it does emphasize the need to establish both the principles and procedures through which public contributions should be appraised. Essentially, if the private concessionaire is able to exploit consumer surplus of users of the new infrastructure,

the public sector should require that the value of external and nonuser benefits be sufficient to justify the required contribution on its normal evaluation conventions. The danger is that the government will become committed without any clear understanding of the costs and benefits.

One of the major impediments to private sector initiative in urban transport infrastructure is the problem of acquiring many fragmented parcels of land and assembling them into a linear right-of-way. For "greenfield" developments, transport infrastructure may so improve the value of adjoining land that it should be in the interests of landowners to facilitate infrastructure construction. This potential gain has been effectively exploited in a number of urban rail systems in Japan through a process of land assembly and reparceling (box 2.3 in chapter 2).[17] The codification of a basis for such private sector initiatives is worth exploring as a means of facilitating private sector investment in large-city transport infrastructure. In areas of existing development, however, that goal is more difficult to achieve. For this reason the public sector usually uses its powers of eminent domain to assemble and provide the right-of-way, as in the cases of the private investments in new metro systems in both Bangkok and Manila.

URBAN TRANSPORT FUNDING

It is widely agreed that urban transport planning and operation should be an integral part of urban strategy. It is even more obvious that the activities of the different modes and functions within the transport sector should be well integrated. By implication that requires consistency of financial arrangements within an overall strategic framework. In most countries, however, no more than lip service is paid to that prescription. We therefore need to explore how financing arrangements might be structured to secure more effective integration both within the sector and between sectors.

The usual institutional context for this has two complicating characteristics. First, responsibility for public transport is increasingly being decentralized to the cities in the absence of an adequate local fiscal base. This means either that the sector has to be financially self-supporting or that it must be dependent on intergovernmental transfers. Second, the supply of transport facilities and services is typically very fragmented, both between private and public sectors and functionally within the public sector itself. This means that the provision of service usually depends on separate and independent financing arrangements for the modes.

Intergovernmental transfers

With respect to intergovernmental transfers, a mechanism is needed that channels finance in a way that neither distorts the allocation of resources between alternative modes or instruments nor weakens the incentives to efficient operation of the individual modes. Unfortunately, the two most commonly used transfer mechanisms—counterpart funding of infrastructure investments and direct subsidy of bus operations—fail these tests.

Counterpart funding by national governments of urban transport infrastructure investment, particularly in capital cities, is common, while current account support to the municipalities for transport purposes is rare. The effect of this is to bias the cities in the allocation of resources at their disposal in whatever way attracts the most generous counterpart funding. The partiality of cities for road investment in many countries, and even for heavy rail investments in some countries, often derives from the high degree of counterpart funding that such investments attract. It may also undermine collaboration, because authorities see more benefit in competing for the "cheap" counterpart funds than in collaborating to develop effective programs.

Three devices can contribute to limit or eliminate this rent-seeking behavior. First, allocation of funds from central to local government should be based on the quality of the project. Second, all such investments in conurbations may be

required to be in conformity with an agreed-on conurbation development plan. Third, support can be channeled through block grant arrangements that leave the cities free and responsible to decide how central government contributions are best used.

Subsidization of public transport operations by the central government is typically channeled through deficit funding of publicly owned transport companies. This discourages the search by municipalities for more efficient supply mechanisms, and has ultimately failed to provide a sustainable basis for public transport operations in most of the developing and transitional economies where it has been applied. To avoid this it is necessary to replace the funding of publicly owned supply agencies by the provision of untied funds with which the municipalities can procure services, in whatever way they find it most effective. In some cases where municipal government has been captured by the supply agencies or their employees, it may be necessary to link the provision of such funds with requirements for the introduction of a competitive procurement process.

Intrasectoral coordination

The pricing principles set out earlier clearly do not result in balanced budgets for each and every urban transport mode. In the case of congested urban roads, short-run marginal cost pricing would yield a surplus over road maintenance and operation costs. While in many sectors this would justify increasing capacity so that short-run costs fell, that may not be possible in the case of urban roads because of the high land and environmental costs of increasing the capacity. In contrast, setting public transport fares below cost in order to redistribute income, to act as a countervailing distortion to the underpricing of road space, or to affect modal split for environmental reasons would all leave a deficit in public transport financing. This might be implemented in a competitive regime through negative price franchising of transport services. Even ***within*** the public transport sector, an optimum pricing structure might require different cost-recovery rates for road- and rail-based systems because of differences in cost structures between modes and indivisibilities in supply. In all these cases, a systemwide approach to urban transport finance makes more sense than a strict financial balance for individual modes or suppliers.

Political attitudes to cross-subsidy are somewhat paradoxical. Within the bus sector, cross-subsidy of unprofitable routes or schedules by profitable ones is often viewed as a good thing and the loss of that capability to be the major disadvantage of competitive regimes. Yet, at the same time, there is great resistance to the transfer of surpluses from one mode to another. Road pricing is frequently opposed because the benefits of the surplus are not likely to be returned to those who pay the charges. The future of congestion charging will depend largely on the political acceptability of the proposed uses for the large revenues that will be generated.

Public opinion on this appears to be changing in the industrialized countries. The emerging professional consensus that "net revenues from congestion charging (after some reduction of other road-user taxes in cases where they are already high) should be devoted largely to transport improvements" certainly reflects the results of public opinion surveys. European discussions usually give particular emphasis to use of a significant portion of the revenues for support to public transport, partly on grounds that drivers in congested areas benefit directly through reduced private travel time and cost from the decisions of others to use public rather than private transport. Various formulas have been proposed to combine transport service objectives with social and fiscal desiderata so that the benefits of the scheme would be distributed as fairly as possible and would be perceived as being fair.[18] The emergence of such a view could be very significant in influencing policies in developing countries.

Integrated urban transport finances

These considerations suggest the need for a mechanism to integrate urban transport

finances, whether secured from users, from local taxes, or from intergovernmental transfers, and to ensure their efficient allocation among uses. For smaller, unitary cities, an existing all-purpose administration might be quite adequate without the creation of any special new institution. Many Chinese cities appear to operate very satisfactorily this way. In these circumstances there may already be sufficient fungibility of funds both within the transport sector and between sectors to serve the area well. But the larger the city size and the more complex its jurisdictional and functional organization, the more likely it is that a "ring fenced" fund and authority will be beneficial.

One straightforward solution for larger conurbations is the creation of an urban transport fund into which all local transport trading profits, transport-related intergovernmental transfers, or local tax allocations should be paid, and from which all local public sector transport expenditures should be financed. The creation of such a fund does not depend on any specific tax source being earmarked for transport, although it would be an essential part of such a scheme that local congestion charges be treated as municipal trading revenues and not as general tax revenues.[19] The essential features to be sought in an arrangement would thus be the pooling of all available transport revenues and their allocation between uses on the basis of the contribution made to the overall municipal development objective.

Neither does the creation of such a fund require any unique local political structure. While the funds would need to be administered by a professional urban transport executive, there are many possible structures for the strategic control of the executive. Where there is a strong local democratic process with competent administration, it might be directed by the local political authority (as is being considered in Buenos Aires); in multiple-jurisdiction authorities, it might be responsible to a joint committee of contiguous authorities (as is being developed in many large, multijurisdictional Brazilian conurbations). It might be administered by a "special purpose district," as is common in the United States. The controlling authority might even include direct user representation, as is commonly advocated for national road funds. These alternatives are considered in more detail in chapter 11.

CONCLUSIONS: A STRATEGY FOR URBAN TRANSPORT PRICING AND FINANCING

The essence of the urban transport pricing and financing issue is that in a system where the modes are highly interactive and policy objectives are complex, the separation of responsibilities for road infrastructure from operations, for infrastructure provision from infrastructure charging, and for roads from that of other modes creates significant policy distortions. The irony is that with such a heavy demand for road space, and such palpable undercharging for its use, cities are short of financial resources to support the investments and the modes of transport that can best contribute to the relief of urban transport problems. In the interests of urban transport integration and sustainability, developing countries should move toward prices reflecting full social costs for all modes, to a targeted approach to subsidization reflecting strategic objectives, and to an integration of urban transport funding. At the same time, however, it is desirable to retain institutions and objectives for the individual modes or components that give high incentive to operational efficiency and cost-effectiveness. A monolithic public sector monopoly is thus not the solution.

The components of a strategy to achieve this logical integration include:

On charging for road infrastructure

- Vehicular users of congested urban road space should be charged a price at least equal to the short-run marginal cost of use.
- Cordon pricing and tolling of specific roads may make an interim contribution, but the

long-term solution must lie in a more thorough application of congestion charges.
- Fuel tax, although a poor surrogate for congestion or road maintenance impact pricing, should in the absence of direct charging be structured along with vehicle license duties to give the best available proxy.
- Taxes on different fuels should be structured to reflect their relative contributions to urban air pollution, again in conjunction with the structuring of vehicle license duties.
- Parking charges, although a poor proxy for congestion charges, should always cover the full opportunity cost of land used for parking.
- Where parking policy is the only available proxy for efficient pricing, controls need to cover all forms of parking space (including that provided privately by employers for employees), and should be designed to secure a level of restraint equivalent to that of efficient prices.
- All road congestion charges, or fuel tax surcharges operating as a proxy for them, should accrue to the municipal or metropolitan authority, and not to the central government treasury.

On public transport pricing and finance

- Pricing principles for public transport modes should be determined within an integrated urban strategy and should reflect the extent to which road infrastructure is adequately charged.
- Given the high level of interaction between modes, and the prevalent undercharging of road use, no absolute value should be ascribed to covering all costs from fares, either for public transport as a whole or for individual modes.
- Transfers between roads and public transport services, and between modes of public transport, are potentially consistent with optimal pricing strategies.
- In the interests of efficient service supply, transport operators should operate competitively with purely commercial objectives, with financial transfers achieved through contracts between municipal authorities and operators for the supply of services.
- Any noncommercial objectives imposed on operators should be compensated directly and transparently, where appropriate by nontransport line agencies in whose interests they are imposed.
- In the absence of appropriate contracting or other support mechanisms, the sustainability of public transport service should be paramount and generally have precedence over traditional price regulation arrangements.

On urban transport financing

- Given the degree of interaction between modes, urban transport financial resources should be pooled within an urban transport fund administered by the strategic transport authority at the municipal or metropolitan level.
- Intergovernmental transfers should normally be made to the fund and should be structured in such a way as to avoid distorting the efficient allocation of resources within the transport sector at the local level.
- Private sector financing for transport infrastructure should be raised through competitive tendering of concessions that may be supported by public contributions as long as these have been subject to proper cost-benefit analysis.
- When allocating funds to urban transport, the relationship between transport policy and other sector policies, in particular housing, should be borne in mind.

NOTES

1. That does not mean that the optimum would be a situation without any congestion at all. In urban areas, where the cost of extra road capacity is usually very high, some people may be willing to pay a surcharge over their own direct costs for the benefits that they receive from making a

trip, but the benefits of eliminating congestion are not sufficient to justify the high costs of the extra capacity.

2. Mohring 1999.
3. Wheaton 1996.
4. Hau 1992b.
5. See, for example, Richardson and Bae 1998.
6. Equity concerns have not been a major concern in Singapore's repeated raising of taxes and charges on motor vehicle ownership and use, and family budget surveys make it clear that the vast majority of these taxes and charges are paid by the 30 percent of the households with the highest incomes.
7. See Willoughby 2000b.
8. Hau 1992a.
9. Roth 1996.
10. For a discussion of distance-based charges, see Litman 1999.
11. Other countries, too, have sometimes justified new taxes on car acquisition as measures to ration access to road space, but the relatively low rates of the tax make it harder to identify specific impact. For the case of the United Kingdom, see Newbery 1990.
12. Verhoef, Nijkamp, and Rietveld 1995.
13. Conseil National des Transports 1999; and Department of Environment, Transport, and the Regions 1998.
14. TRL 1985.
15. Given widespread evidence that demand for public transport is more elastic with respect to level of service—particularly network density and frequency—than to price, policies aimed at securing high levels of service quality may well be more effective than those directed at prices.
16. PADECO 2000.
17. PADECO 2000.
18. Small 1992.
19. Without such a provision, municipalities are inevitably resistant to the introduction of direct charges on local road users. The recent resurgence of interest in congestion charging in the United Kingdom has resulted partly from the passing of legislation to allow local authorities to keep any road congestion charge revenues.