



**Economic and Social
Commission for Asia
and the Pacific**



**Asian Institute of
Transport
Development**

SUSTAINABLE TRANSPORT PRICING AND CHARGES

Principles and Issues



United Nations



**Economic and Social Commission
for Asia and the Pacific
TCTID Division
United Nations Building
Rajadamnern Lok Avenue
Bangkok 10200, Thailand
Tel: +66-2-2881371, 2881555, 2881501
Fax: +66-2-2806042, 2883050, 2881067
E-mail : tctidd.unescap@un.org**



**Asian Institute of
Transport Development
E-5, Qutab Hotel
Shaheed Jeet Singh Marg
New Delhi 110 016, India
Tel: +91-11-6856117, 6856113
Telefax: +91-11-6856113
E-mail: aitd@vsnl.com**

ST/ESCAP/2139

This publication has undergone a process whereby drafts were examined by a group of experts and were also subjected to peer review. The views expressed in the publication are reflections of this process and do not necessarily reflect the views of the United Nations Secretariat. The publication has been issued without formal editing.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the United Nations Secretariat concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Mention of firm names and commercial products does not imply the endorsement of the United Nations.

Contents

<i>Foreword</i>	i
<i>Preface</i>	iii
<i>Glossary</i>	vii
<i>Abbreviations</i>	x
<i>Overview and Recommendations</i>	xi
1. Transport and Sustainable Development	1
Transport and Economic Development	1
Concept of Sustainable Development	3
Sustainable Transport	11
Transport Pricing and Sustainable Development	20
Ecological Economics	24
2. Pricing and Charges for Transport Services and Infrastructure	28
Optimal Pricing Policies	28
Marginal Cost Pricing in Practice: Fixed Capacity, Peak-Load Pricing, Uniform Pricing, Indivisibilities and Joint Costs	36
Marginal Cost Pricing in Practice: Monopoly, Externalities, Strategic Interdependence and Uncorrected Externalities	51
Profitability, Taxation and Income Distribution	58
Financing and Private Sector Participation	62
3. Pricing and Charges for Road and Urban Transport	69
Pricing for Road Infrastructure	69
Public Transport Pricing	87

4.	Pricing and Charges for Railways	99
	Railway Pricing: The Context	99
	Railway Infrastructure	102
	Passenger and Freight Services	108
5.	Pricing and Charges for Maritime and Air Transport Sectors	114
	Ports and Inland Waterways	114
	Maritime Transport	126
	Airports	130
	Air Transport	134

TABLES

1.1	Creating Competition	14
1.2	Selected Environmental Effects of Principal Transport Modes	15
1.3	Environmental Sustainability	16
1.4	Social Sustainability	17
2.1	Pricing Policy Objectives and Conflicts	32
3.1	Classification of Road Transport Costs	70
4.1	Railway Problems and Causes	100
5.1	Port Charges on Ships	122
5.2	Port Charges on Cargo	123

DIAGRAMS

1.1	The Environment - A Schematic Representation of Sustainable Development	5
1.2	Sustainability: Synergies and Trade-offs	18
1.3	Materials Circulation and Energy Flow in the Environment	24

FIGURES

1.1	Costs and Benefits of Environmental Improvements	12
2.1	Welfare Maximization	33
2.2	Long-Run Marginal Cost (year 1)	35
2.3	Short-Run Marginal Cost (year 0)	35
2.4	Costs and Economies of Scale	37
2.5(a)	Economies of Scale and Excess Capacity	38
2.5(b)	Full Capacity Utilization	39
2.5(c)	Excess Demand	40
2.6	Daily Demand Cycle	42
2.7	Peak and Off-Peak Demand	43
2.8	Peak and Off-Peak Pricing	44
2.9	Peak Loads and Optimal Capacity	46
2.10	Optimal Pricing and Flat Load Curves	47
2.11	Indivisibilities	49
2.12	Indivisibilities and Investment	49
2.13	Price Determination under Monopoly vs Perfect Competition	53
2.14	The Optimum Pollution Charge	54
2.15	Optimal Traffic Flows and Congestion Pricing	56
2.16	Pricing with Uncorrected Externalities	57
2.17	Average Cost Pricing	59
2.18	Price Discrimination	60

Foreword


The development of effective and efficient transport services is essential as we move further into the era of globalization. Unfortunately, not all ESCAP member countries are in a position to immediately meet this challenge. Financial resources for infrastructure are limited, transport users are not fully aware of the environmental impacts of their actions and access and mobility are impaired. Part of the core of these problems is the issue of sustainable transport pricing and charges.

Transport pricing is a complex issue because of the multiplicity of sustainability objectives, the institutional separation of infrastructure from operations, pricing from tax components of charges and transport modes from each other. As a result, many governments in the region have yet to introduce transport pricing that covers all direct costs, let alone prices that adequately reflect social and environmental costs.

This publication suggests the adoption of efficient transport prices that can promote a better balancing of economic, social and environmental concerns. It is also intended to form a basis for a deeper applied analysis of pricing in specific transport subsectors. I hope that readers will find the analysis and recommendations helpful.

I am happy to place on record my appreciation for the financial assistance extended by the Royal Netherlands Government and the Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) to the preparation of this publication.

I also acknowledge the contribution of the Asian Institute of Transport Development (AITD), in hosting a regional seminar with ESCAP on Transport Pricing and Charges held at New Delhi in November 2000, assisting in the peer review and publishing this document. I am impressed by the high level of objectivity of the Institute, its regional outreach and its capability to draw on a wide range of academic and professional resources. I am particularly encouraged by the deepening of cooperation between ESCAP and AITD in the spirit of the Memorandum of Understanding between the two organizations.


Kim Hak-Su
Executive Secretary
ESCAP

Preface

The role of transport in the development process is manifold and well recognized. As a result, policy makers usually face a multiplicity of objectives to regulate a complex set of interrelated activities, which affect diverse groups of society. Provision of transport services involves costs that need to be reflected in the prices charged for these services. This is a thorny problem of political economy, especially for the developing countries which have constantly to grapple with the problem: whether these costs should be fully reflected in prices at all levels of service; and if not, how to fund the gaps.

An equally important component of the problem is the definition of sustainable development, which concerns social measures designed to take note of the effects of economic behaviour and economic policies on the environment. This publication clearly identifies the sources of externalities that can be incorporated into the theories of resource allocation and prices. Essentially, it pinpoints two sets of externalities, one related to health and quality of life and the other to the adverse impact of the depletion of exhaustible resources on the future growth prospects of the economy.

The document combines the principles of microeconomic theories and welfare economics with the technological and organizational characteristics of transport industries and provides a comprehensive analysis of the problems of transport pricing. It strongly recommends that in the interest of promoting sustainable development, the users of transport services should ordinarily be required to pay the social costs of providing these services.

There is often a tendency to overlook the important fact that the problem of transport pricing (and, more generally, of transport economics) is an integral part of the overall problem of resource use in the geographical space of the economy. This volume highlights the fact that the essential job of transportation activities is to link different locations for the movement of goods and people. The existing pattern of land-use in an economy determines the structure of demand for transport services. On the other hand, the provision of transport services and their pricing determines the location of economic activities in a growing economy.

Another important contribution of this volume is the suggestion that information technology can be harnessed to overcome the administrative difficulties associated with

differential pricing mechanisms. For example, only a few decades ago flat rates were the norm in some transport sectors, because the cost of monitoring any differentiated fare structure was considered prohibitively expensive. However, the phenomenal progress in information technology in recent times has made the changeover to marginal cost pricing economically viable.

The treatise also emphasizes the need for analyzing the pricing problems along with the related issues of institutional arrangements regarding ownership, management and control of enterprises responsible for providing transport services. There is plenty of evidence from all over the world to show that state-owned enterprises operating under soft budgetary constraints have tended to ignore problems of incentives and information, and are in need of meaningful institutional reforms. Hence, the emphasis on deregulation, privatization and franchising.

Given that some transport sub-sectors can significantly raise their efficiency frontiers and that presently prices are not reflecting costs, there is need for a transition period from current pricing regimes to those that may support sustainable transport development. This transition needs to be very carefully managed. The role of regulatory bodies that *inter alia* regulate prices becomes critical in this regard. Market mechanisms function effectively only when appropriate market and non-market institutions are in place to support the working of the markets. It is, however, important to remember that these institutions do not come up automatically in response to the demands made by the opening up of market opportunities. They need to be consciously created with care and foresight.

I hope this volume will take the debate forward in a substantial manner by aiding policy makers in improving the efficiency with which transport infrastructure and services are utilized and financed. It will also contribute significantly to the creation of an environment that is more conducive to attracting private sector participation in the provision and operation of transport services and infrastructure facilities by using prices as one of the instruments for internalizing the externalities generated by the sector.

The Asian Institute of Transport Development (AITD) fosters regional cooperation by sharing expertise and facilities with its member countries. It also disseminates findings of important research studies on issues having a bearing on balanced and sustainable

development. This publication, which has been brought out jointly with UN-ESCAP, is an effort in that direction.

K. L. Thapar
Director
AITD

About the Publication

The publication is divided into two parts. The first part comprises overview and recommendations. The second part consists of five chapters. The first two chapters are concerned with the generic policy and theoretical issues to be addressed. The remaining three chapters deal with charging policies to promote sustainable development on a sector-by-sector basis. The details of the structure are given below:

Overview and Recommendations

Chapter 1: clarifies the main issues involved in devising pricing systems to promote sustainable development.

Chapter 2: defines optimal pricing policies and addresses the range of complexities that arise due to the nature of the cost structure and market distortions found in the industry.

Chapter 3: examines urban transport policy and the role of pricing and financing in promoting sustainable development and an appropriate allocation of resources for public transport.

Chapter 4: examines the competitive context of railways and the scope for private sector participation in the provision of both infrastructure and rail services. It also includes an analysis of the problems of allocating track costs and calculating access charges to train operators; and identifies methods of charging passengers and freight users for rail services.

Chapter 5: examines pricing and charging methods for infrastructure and services in other sectors, including ports, inland waterways and maritime transport, airports and air transport.

Glossary

The following is a glossary of the main technical terms used in this document:

Average Costs and Average Cost Pricing : Average costs are the total costs of providing a transport service, including infrastructure, divided by some measure of output, such as vehicle-km. They are relevant to cost recovery, since prices that are set equal to average costs will ensure that total costs are recovered. Total costs include both fixed and variable costs.

Capital Costs : Capital costs comprise the consumption of fixed capital and interest payments, and usually represent a high proportion of infrastructure costs. They differ from annual capital expenditure that may or may not cover all the capital costs. If annual expenditure is less, then the quality of the transport assets will deteriorate.

Congestion : Congestion arises when traffic exceeds infrastructure capacity and the speed of traffic declines. It can be defined as a situation where traffic is slower than it would be if traffic flows were at a low level.

Congestion Costs : Congestion costs comprise direct costs, including time costs and opportunity costs of time lost to third parties due to delays, and environmental costs.

Cost Recovery : This is an approach to infrastructure charging whereby fixed and variable costs are recovered in full or in part.

Cost-relatedness : This means that charges cover at least marginal cost and relate to these in a non-distortion generating way.

Cross-subsidization : Cross-subsidization involves supplying transport services to one group of consumers (users) at a loss, which is made up by profits on services provided to other consumers (users). It can be viewed as a particular way of allocating rents associated with the transport activity.

Depreciation : Depreciation is an accounting charge for the decline in value of an asset spread over its life.

External Costs or Externalities : External costs are those which the user of a

service does not pay for; they include pollution, noise, health, accidents costs. Failure to acknowledge such costs may result in excessive use of a transport service.

Fixed Costs : Fixed costs are those which are independent of traffic flow or usage.

Flat-rate Charge : A charge, normally applied to transport infrastructure, which does not vary with usage.

Internal Costs : Internal costs are those which the user pays for.

Interest : Interest charges reflect the opportunity cost of capital. In the public sector, the interest rate is usually comparable to the refinancing cost of government loans.

Investment Expenditure : This reflects the annual expenditure on fixed assets with lives greater than one year, such as infrastructure and vehicles. Such expenditure is normally 'capitalized' with a depreciation rate and an interest rate reflecting the opportunity of capital invested.

Maintenance Costs : Maintenance costs represent the costs necessary to maintain the existing infrastructure. A distinction can be made between maintenance which is time-related and that which is use-related.

Marginal Costs : Marginal costs are specific variable costs related to the provision of a service or the use of infrastructure. Short-run marginal costs are the additional operating and maintenance costs associated with a marginal increase in output without any increase in physical capacity. If external costs are also included, this is referred to as marginal social cost. Long-run marginal costs include also the capital costs of increasing capacity to accommodate an increase in output; they are difficult to measure. Linking charges to long-run marginal costs would lead to significant inefficiencies where excess transport capacity exists.

Operating Costs : These are running costs associated with operation of transport services.

Price Discrimination : This is where users are charged according to their willingness or ability to pay. Users valuing a service highly will make a greater contribution to fixed costs than those who can afford to pay less.

Ramsey Pricing : This involves setting charges according to the elasticity of demand

of each user or group of users.

Running Costs : The costs necessary to keep a particular asset or service in operation. They do not enhance the value of the asset.

Structural Maintenance : Maintenance of a capital nature such as road resurfacing. The benefits of this type of maintenance are reaped over a number of years.

Sunk Costs : The cost of assets with zero resale value or which have exceeded their economic life.

Total Costs : The sum of fixed and variable costs, or of capital and running costs.

Two-Part Tariffs : Two-part tariffs comprise a fixed charge plus a variable charge. In principle, the latter would be related to marginal costs and the former would be set to contribute to fixed costs.

Variable Costs : Those costs that vary with traffic levels. Examples include wear and tear to infrastructure and congestion costs.

Abbreviations

AC	Average Cost
ADB	Asian Development Bank
AITD	Asian Institute of Transport Development
BOT	Build-Operate-Transfer
CO	Carbon Monoxide
CO ₂	Carbon Dioxide
CRS	Computer Reservation System
ERP	Electronic Road Pricing
EU	European Union
FFP	Frequent Flyer Programme
GC	Generalized Cost
GDP	Gross Domestic Product
GRT	Gross Registered Tonnage
HC	Hydrocarbon
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
LRMC	Long-run Marginal Cost
MB	Marginal Benefits
MC	Marginal Cost
MEC	Marginal Environmental Cost
MPC	Marginal Private Cost
MR	Marginal Revenue
MSC	Marginal Social Cost
NNP	Net National Product
NRT	Net Registered Tonnage
NO ₂	Nitrogen Dioxide
NO _x	Oxides of Nitrogen
OECD	Organization for Economic Cooperation and Development
PM ₁₄	Particulate Matter of size 14 microns
SO ₂	Sulphur Dioxide
SRMC	Short-run Marginal Cost
UK	United Kingdom
UNCTAD	United Nations Conference on Trade and Development
UN-ESCAP	United Nations Economic and Social Commission for Asia and the Pacific
UNIDO	United Nations Industrial Development Organization
USA	United States of America
USO	Universal Service Obligations
VOC	Volatile Organic Carbon
WCED	World Commission on Environment and Development

Overview and Recommendations

Transport improvements promote economic growth and social development by increasing mobility and improving access to resources and markets. However, in recent years, there has been growing recognition of the need to promote sustainability, sustainable development, and sustainable transport in order to bring about improvements in transport systems and policies. In the context of this document, sustainability is interpreted in its widest sense and incorporates its economic, financial, environmental and social dimensions. This overview provides a summary of the issues addressed in the publication as well as a range of policy recommendations for transport pricing and charges aimed at promoting sustainable development.

Sustainable Development

1. Sustainable development has been defined as development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs.
2. A policy of sustainable development, when applied to transport, should seek to secure a balance between equity, efficiency and inter-temporal concerns. This is predicated on:
 - (i) the maintenance of high and stable levels of economic growth and employment to generate the necessary resources to achieve sustainable development;
 - (ii) protection of the environment through the prudent and efficient use of natural resources and the development of renewable resources, wherever possible; and
 - (iii) social progress, which addresses the needs of all sections of society through reducing unemployment, poverty and pollution and their impact on health and quality of life.

Sustainable Transport Policy

3. An environmentally sustainable transport system could be defined as one where transportation does not endanger public health or ecosystems and meets the needs for access and mobility consistent with (i) use of renewable

resources below their rates of regeneration, and (ii) use of non-renewable resources below the rates of development of renewable substitutes.

4. Sustainable development should be promoted to the extent possible through transport prices that are equated with marginal social cost. Distorted pricing conveys wrong or misleading information about resource scarcity and thereby provides inadequate incentives for the efficient use of resources and capital assets.
5. A free market mechanism promotes efficiency and for such a mechanism to be in place, competition should be encouraged. It, however, needs to be ensured that there are no significant adverse externalities and distributional consequences.
6. Transport policy can pursue multiple objectives while promoting sustainable development. These may relate, among others, to energy efficiency, environment-friendly technologies, and responsive markets.
7. A sustainable transport policy will require intervention in the market system to ensure that:
 - (i) the direct or indirect use of natural resources is such that they can at least be replaced by (a) their natural regeneration (e.g. hydroelectric energy for electric traction), or (b) discovery of new deposits of the currently used exhaustible resource (e.g. oil or natural gas reserves), or (c) the use of a new renewable resource (e.g. wind or solar power), or (d) conserving the use of resources per unit of transport output, or (e) a combination of these; and
 - (ii) the damage to the environment is within such limits that the productivity of other economic activities and the quality of life, in terms of health and security against accidents, do not deteriorate over time.
8. The global scarcity value of the natural resources used in the provision of transport infrastructure and services and the external costs due to pollution and degradation of the environment (that is, the social cost of transport), should be built into the price of providing or using transport facilities and services.

Role of the Government

9. The role of the government should change from that of a supplier and quantitative regulator to that of a facilitator of competition, private sector participation and custodian of environmental and social interests. In cases where markets do not function efficiently, the government intervention would be necessary to set efficient charges involving users in decision-making.

Pricing Objectives

10. Pricing should be used as a method of resource allocation. Optimal pricing must balance economic efficiency, equity and transaction costs.
11. Economic efficiency should take precedence over other objectives, such as: income distribution, relationship with macroeconomic policy, etc.

Marginal Cost Pricing

12. In a 'first-best world' characterized by perfect competition, the market will set prices equal to marginal social cost and thereby maximize social welfare. However, conditions necessary for a 'first-best world' rarely exist. Hence, the markets will not set transport prices equal to marginal social cost. This underscores the need for regulation and targeted policies.
13. Transport infrastructure requires heavy capital investment, is long lasting and has few alternative uses. As a result, infrastructure facilities – roads, railbeds, bridges, piers, runways, locks, harbours, tunnels, etc. – have potential for significant economies of scale.
14. When the invested cost of infrastructure remains fixed over a period of time, the marginal cost in the long run also remains constant. The financial results would vary according to the utilization of capacity. Where the utilization is less than the projections (i.e., demand was overestimated) and the prices are fixed on the basis of long-run marginal cost, losses will result. Profits will, however, accrue if the utilization exceeds projections (i.e., demand was underestimated). In general, therefore, the policy should be that the prices are set in relation to short-run marginal cost, which may lie above, below, or equal to long-run marginal cost.

Peak-Load Pricing

15. Peak-load pricing enables an even rate of capacity utilization. Beyond an optimal level, however, the increase in price would lead to welfare losses.
16. The appropriate scale of operation needs to be determined by establishing the optimal timing of transport investments under conditions of growing demand.

Monopoly

17. Monopolies and oligopolies distort markets which, if left unregulated, fail to lead to optimal transport prices thereby adversely impacting on both social welfare and sustainable development. In such circumstances, government regulation and intervention is necessary to align transport prices with marginal costs.

Externalities

18. The internalization of externalities is a fundamental requirement in devising transport pricing policies to promote sustainable development. Transport generates many negative externalities or external costs, including noise, accidents, pollution and congestion. If the externality costs are not borne by those who generate them, then the market mechanism fails to allocate resources efficiently. The 'polluter pays' principle suggests that users should be made aware of the external costs they generate by imposing on them pollution tax equal to the marginal environmental cost. This would also reduce the volume of transport activity to the socially optimal level.

Modal Split

19. The transport industry abounds in situations where particular modes charging prices at marginal social cost compete with other modes that do not charge prices on this basis. Since there are strong interdependencies between alternative modes of transport, it may be appropriate to reduce prices on certain modes in order to effect an appropriate modal split. This is especially the case where uncorrected externalities or subsidies exist in the case of one of the alternative modes.

Financial Sustainability

20. Sustainable development is inextricably linked to financial viability. The transport policy must create a framework that ensures economic optimality

in the allocation of resources and financial viability, consistent with promoting sustainable development.

Profit Targets

21. Profit targets need to be used to stimulate innovation, reduce costs and improve efficiency. Prices and outputs should be chosen in a way that meets the profit target with minimum loss of efficiency.

Price Discrimination and ‘Ramsey Pricing’

22. Price discrimination is the most efficient method of meeting a profit target. It allows marginal costs to be equated to price in a dynamic sense inasmuch as there can be times when they are less than the price and also times when they are more than the price, but, over a finite time horizon, they tend to equalize. Broadly captured by the principles enunciated by Frank Ramsey, price discrimination works best over networks.

Equity and Cross-Subsidization

23. The equity issues are important and governments are legitimately concerned about them. It is, therefore, necessary not only to identify pricing policies that meet profit targets with minimal losses in economic efficiency, but also to evaluate the effects of such policies on social equity.
24. When an operator is required by the government in public interest to provide transport services at fares that are persistently below the relevant marginal cost, a mechanism needs to be evolved to fund the gap. If owing to a severe fiscal constraint, the government is unable to finance the gap with subsidies, service providers are forced to resort to cross-subsidization between different products. It results in charging some users above the marginal cost to offset losses made on traffics or services where prices are fixed by the government at levels that do not cover the relevant marginal cost.
25. Cross-subsidization, especially when it is dictated by the government policy in consideration of equity of income distribution, obscures the economic efficiency or profitability of a service. In such cases, it would be desirable or more transparent to bear the element of subsidy on government account. It is, therefore, important to identify the equity implications of a particular pricing policy and develop specific strategies to strike a balance between the need for stimulating economic efficiency and the need for more equitable distribution among different income groups of users.

Private Sector Participation

26. In several areas of transport, competition is not a natural phenomenon. Such situations would need to be corrected through careful design of sector reforms. The government has two options in this regard:
- (i) creating direct competition in the supply of services, often referred to as ‘competition in the market’; and
 - (ii) creating competition for the right to supply transport services through concessions or other contracts, often referred to as ‘competition for the market’.

Role of a Regulator

27. The main task of the regulator should be to ensure that there is healthy competition and that the prices are aligned to marginal cost.
28. The regulator would also have to ensure that operators in the private sector have sufficient incentive to invest in services which, though not profitable, are socially desirable. One method could be to clearly specify universal service obligations (USO) in the scope of the service providers’ obligations.

Pricing and Charges**Road and Urban Transport**

29. Urban transport pricing should fulfil three functions: ration and allocate the use of competing resources; signal the need for investment; and help in generating funds for the development of the related sectors.
30. The price charged to the users should cover the full social cost of their trips. In most countries, both developed and developing, urban roads are provided to the users without imposing on them any direct user charge. The only payment from the private user to the public supplier comes indirectly in the form of taxes, primarily on fuel. These taxes, however, do not reflect the costs of congestion with the result that the cost perceived by the marginal road user – his own private cost – does not take into account the extent to which he slows down the movement of other road users. This has several adverse effects. First, as rail and some other public transport infrastructure is paid for directly through fares, there is a distortion in the choice of mode. 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, since the revenues do not accrue to the local authority, there may be inadequate money available for proper maintenance of the existing infrastructure.

31. The charging regime should allocate, as far as possible, all associated costs (congestion, environmental, wear and tear, etc.) to the users directly in proportion to the costs imposed by them. As the imposition of different types of costs varies between different types of vehicles, only the simultaneous application of a number of different charging devices could meet this ideal.
32. Price discrimination would be the best means of recovering fixed costs. Where imposition of direct congestion charges is possible, these should be varied according to the congestion creating equivalents of different vehicle types. Behavioural adjustment to reduce externalities should be rewarded in the form of lower charges.
33. Urban transport infrastructure and public transport pricing have strong interdependencies and, therefore, any 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.
34. For reasons of 'second best', there may be need for financial transfers between the exchequer and public transport services, or between roads and public transport services, or between different modes of public transport. These transfers should be achieved through contracts between municipal authorities and operators for the supply of services.
35. If public transport cannot be subsidized to compensate for the inadequate road pricing policies, then financial sustainability of the public transport service should take precedence over price or fare regulation.

Pricing and Charges

Railways

36. Pricing policies based on the principle of marginal social cost pricing moderated by price discrimination should be pursued in order to ensure the financial sustainability of rail industry.
37. The recent past has witnessed restructuring of the railway systems in many countries resulting in the dismantling of their vertically integrated activities.

This has led to separation of ground infrastructure like tracks, yards, stations, etc. from train operations. In such cases, the train operators are required to buy access rights to use the infrastructure. Where a single entity provides both infrastructure and train services, the task is one of cost allocation. On the other hand, where separate entities are involved, it is a matter of determining the charges to be levied on train operating companies by the entity responsible for infrastructure.

38. Access or infrastructure charges should consist only of those costs that are relevant to the specific pricing, investment, or operating decisions under consideration. If the sum of route-based access charges fails to cover the total revenue requirements of the infrastructure operator, then the principle of price discrimination could be adopted.
39. Price signals for the efficient production and allocation of railway infrastructural resources should be based on avoidable (marginal) costs of changes in the use of the existing network and changes in the network itself.

Pricing and Charges

Ports and Inland Waterways

40. The pricing in the port sector should be based on the principle of long-term marginal social cost, which also accounts for externalities relating to environment, congestion and accidents. This would encourage efficient use of existing facilities besides providing guidance on investment or disinvestment in port facilities and services.
41. The joint and common costs, commonly known as indivisible costs should be allocated according to 'what the traffic can bear'. This price discrimination policy should equally be applied to recover inescapable costs, such as those for quays and breakwaters. The difference in the two cases is that whilst joint and common costs are escapable, inescapable costs are fixed even in the long term.
42. The principle of price discrimination should be applied if the ports are required to be self-financing or if a specific financial target is to be met.
43. Calculating port charges would involve determination of appropriate cost centres and collection of information on the utilization of assets corresponding to a given cost centre.

44. It is recognized that cost accounting techniques that split cost between shipping and cargo cost are not precise. Also, any formula-embedded rationale has to reckon with the respective bargaining powers of the usually highly organized and powerful shipowners on the one hand, and shippers who are typically scattered and often much less able to effectively negotiate with the port authorities, on the other.
45. The issues involved in the pricing of inland waterways infrastructure facilities are similar to those for ports.

Pricing and Charges

Maritime Transport

46. In charter shipping, the prices are equal to short-run marginal cost (SRMC) and may lie above or below long-run marginal cost (LRMC), depending on the forces of supply and demand. The charter markets are characterized by near-perfect competition; hence are highly efficient. Attempts to regulate the market through protectionism and cargo reservation would lead to reduction in social welfare.
47. In liner shipping, the freight rates are based on the principle of 'what the traffic will bear'. Since there are widely differing values to the cargoes shipped, marginal costs associated with individual cargoes can vary and, in some cases, may even be below the average marginal cost. The liner operators, therefore, practise price discrimination, whereby the same service may be sold to different shippers at different rates, the aim being to recover total cost.
48. Recent years have witnessed decline of conferences and emergence of carrier alliances on a global basis. Besides, mergers and acquisitions have led to a greater degree of consolidation. Another notable feature witnessed in the recent past has been the vertical integration of related activities. All these developments have the potential to adversely impact competitive impulses. This underscores the need for bringing the liner shipping services within the purview of regulatory and anti-trust legislation.
49. Marine pollution risks, congestion and other externalities should be quantified, internalized and taxed wherever practicable. This should be undertaken by the countries where the ships are registered.

Pricing and Charges**Airports**

50. Airport pricing should be based on the principle of marginal social cost pricing. The issue of social costs is, however, problematic. The theoretical solution is to calculate airport charges on the basis of social accounting prices, with due allowance for indirect costs, such as congestion and noise.
51. The pricing structure should ensure that the charges are always enough to discourage excess demand; are never below the short-run marginal cost, including social costs; and conform to 'what the market will bear'. This would enable the airports to raise additional revenues in order to be self-financing.
52. Airport pricing systems should encourage the efficient use of existing facilities and also provide guidance on investment or disinvestment in port facilities and services. The methods to achieve these objectives are the same as enunciated for seaport pricing.
53. Airports possess a degree of locational monopoly which may give rise to monopoly pricing, a situation that needs to be safeguarded.

Pricing and Charges**Air Transport**

54. Pricing policies, based on the principle of marginal cost pricing moderated by price discrimination, should be pursued in order to ensure the financial sustainability of airline industry.
55. There is considerable scope in the industry for differentiating price and service quality. Computer-based yield management techniques offer the scope to maximize capacity utilization as also tap consumers' willingness to pay for the desired service.
56. Deregulation of air transport industry will enhance competition, thus maximizing consumer welfare, including reduction in fares.
57. The air transport industry in most countries is still owned and operated by governments or state-owned undertakings and is heavily regulated. The future is, however, likely to see increasing liberalization in the grant of traffic rights and tariff formation. The industry is also witnessing formation of alliances

on a global basis besides a degree of mergers and acquisitions. These developments would have both positive and negative effects. Large carriers would benefit, while smaller and weaker airlines would suffer.

Charging Structures

58. Any transport charging framework should have the following key characteristics:
- (i) comprehensibility – the structure should be clearly and easily understood by users whose behaviour it is meant to influence and should not impose undue transaction costs to identify the appropriate information;
 - (ii) transparency – the structure should provide clear information to users on the make-up of charges, and hence not confer undue advantage on particular industry participants, e.g. through information asymmetry;
 - (iii) stability – charges should not fluctuate or alter in an arbitrary or unpredictable manner, except where significant short-term cost changes are being signalled. If congestion (scarcity) pricing is introduced, short-run prices could be unstable. However, future average levels could be projected in some cases by establishing a long-run avoidable cost around which short-run prices might be expected to fluctuate;
 - (iv) measurability, cost effectiveness and objectivity – the data required to derive charges should be objectively measurable, cost-effective to collect, and unambiguous to apply; and
 - (v) cost-reflectivity – the charges should be, to the extent possible, cost-reflective in order to meet the objective of economic efficiency.

Concluding Remarks

Sustainable development, economic and social, is inextricably linked with sustainable transport systems. These systems, like any other, are resource bound in terms of material, energy and human skills, and also in terms of efficient management of resources. In the context of sustainability, therefore, the processes of consumption, conservation, replenishment and augmentation of resources are indeed significant. The object of the present study is to examine the extent to which prices and markets can help these processes. There are, however, some special features of transport systems, which must be taken into account.

Transport is one of the fastest growing energy demand sectors. The present known reserves of primary energy sources, especially hydrocarbons, are non-

renewable and, therefore, inadequate to sustain the rate of consumption and demand, particularly of the energy-importing developing countries. Investments in search for new reserves of primary energy sources and for alternative fuels hold out hopes for the future, whilst technological improvements have continually brought down specific consumption.

Whilst the logic of conservation demands less movement of men and materials, the prime drivers of development, namely, globalization and urbanization, call for increase in movement. This twin phenomenon has also accentuated the divide between income groups within countries and between countries, raising equity issues, which make market and price mechanisms more complex.

Viewed at the global level, energy consumption and the energy resource base are unevenly distributed. Consequently, cartels and oligopolies have often shaped the production and distribution patterns of hydrocarbons. Furthermore, no consensus has so far been reached on how to conserve resources efficiently and mitigate environmental pollution.

Development policies and strategies are the primary tools that address these major issues and thus determine the extent of sustainability; transport is only a secondary derivative.

It is within these constraints that the present study attempts to examine the extent to which pricing of transport infrastructure and services can be utilized as an instrument for the efficient use of energy and materials as well as the allocation of resources consistent with the goals of sustainable development. It is a complex task and many conditions need to be attached to the recommendations. There is, however, a central message that runs through the principles considered in this publication, namely:

- (i) It is necessary to determine the full cost of transport services, economic and social, internalizing the environmental costs, and also to determine the cost of improvement at the margin and let market determine the norms of consumption, investment and resource allocation, as far as possible.
- (ii) It is desirable to recover all or most of such costs from the user. This is not necessarily inconsistent with economic, financial, environmental/ecological and social/poverty alleviation objectives.

- (iii) However, reality dictates that intervention will be necessary, which means a change in the role of the government:
 - (a) to become a facilitator of competition and private investment in this sector, rather than be a provider itself;
 - (b) to be a custodian of environmental and social interests;
 - (c) to set standards of safety, pollution etc. and to establish adequate regulatory institutions;
 - (d) to legislate, as necessary, for taxes and incentives; and
 - (e) to provide, where essential, an identified time-bound subsidy.

1

Transport and Sustainable Development

Introduction

Transport improvements promote economic growth by increasing mobility and improving physical access to resources and markets. In recent years, however, there has been growing recognition of the need to promote sustainability, sustainable development, and sustainable transport in planning improvements in transport systems and policies.

Concern about sustainability originates from the growing awareness that there are several dimensions to the impact of transport and the costs that it imposes – economic, social and ecological. Promoting sustainable development is, thus, concerned with seeking an optimal balance between economic, social and ecological objectives, whereas promoting sustainable transport requires that full account is taken of these factors in the development of transport systems.

This chapter aims to clarify the issues involved in devising pricing systems to promote sustainable development. It comprises four sections: section I examines the role of transport in economic development and the need to promote sustainable development; section II defines sustainable development and its main dimensions; section III deals with the concepts of sustainable transport and sustainability; section IV identifies the implications of promoting sustainable development for transport pricing.

Section I

Transport and Economic Development

Transport is pivotal to economic development. On the one hand, the achievement of economic growth and poverty reduction requires good physical access to resources and markets, whilst on the other, quality of life is generally dependent on the quality of physical access to employment, health services, homes, education and other amenities. Fromm¹ identified that transport has the following four broad functions in assisting economic development:

¹ Fromm, G., 1965. Introduction: an approach to investment decisions. *Transport Investment and Economic Development*. Washington. Brookings Institute.

- (i) as an input into the production process permitting goods and people to be transferred between and within production and consumption centres;
- (ii) transport improvements can shift production possibility functions by altering factor costs and reducing levels of inventory tied-up in the production process;
- (iii) increasing factor mobility and permitting factors of production, especially labour, to be transferred to places where they may be most productively employed; and
- (iv) increasing the welfare of individuals by extending accessibility to a range of facilities and providing superior public goods, such as improved social cohesion and security.

Macroeconomic studies have shown that investment in transport promotes growth by increasing the social return to private investment without ‘crowding out’ other productive investment. Further, the economic rate of return on transport projects has been estimated by the World Bank to be 50% higher than that secured in other sectors.² Many economists have emphasized the linkage between transport provision and economic development by distinguishing direct from indirect or multiplier effects. The former stem from the cost and time savings resulting from transport improvements, whilst the latter flow from the substantial input of resources needed to construct modern transport infrastructure. Indeed, some argue that efficient transport services are a necessary prerequisite for national economic development³, while others argue that economic development is a complex process with transport permitting the exploitation of the natural resources and talents of a country. Transport is thus seen as a necessary, but not sufficient, condition for development.⁴

In broad terms, *rural transport improvements* increase market access and thereby lower agricultural production costs and also facilitate the development of the non-agricultural rural economy. *Urban transport improvements* increase labour market efficiency and access to amenities. *Inter-urban transport improvements* facilitate domestic and international trade by speeding up the movement of freight and people. Conversely, in many developing countries, the inadequacy of transport infrastructure and

² Gwilliam, K. M., and Shalizi, Z., 1996. *Sustainable Transport: Sector Review and Lessons of Experience*. TWU 22 10/96 Washington. World Bank.

³ Rostow, W. W., 1960. *The Stages of Economic Growth*. Cambridge. Cambridge University Press.

⁴ Ahmed, Y., O’Sullivan, P., Sujono and Wilson, D., 1976. *Road Investment Programming for Developing Countries: An Indonesian Example*. Evanston. Northwestern University.

the inefficiency of transport services are recognized as being amongst the main bottlenecks to socio-economic development and social integration⁵.

Although transport improvements have a major role to play in fostering economic development, there is growing interest in sustainability, sustainable development, and sustainable transport. The nature and scope of these concepts, and their implications for transport planning and policy, are only beginning to be explored. Several factors have contributed to the growing interest in these concepts. These include the increasing awareness that transport can have an adverse environmental impact that can impose significant economic, social and ecological costs. This has highlighted the need to review the impact of transport development from a broader perspective.

Section II

Concept of Sustainable Development

Sustainable development is wider in scope than economic development and strives for an optimal balance between economic, social and ecological objectives. Interest in sustainability originally reflected concerns about long-term risks of current resource consumption, keeping in view the goals of ‘intergenerational equity’ (i.e., being fair to future generations). The concept of ‘sustainable development’ has emerged in recent years and has contributed to the debate on development and the environment. In particular, it has served to strike a balance between the demand for economic growth and the need for conservation and environmental protection.

In 1987, the World Commission on Environment and Development (the Brundtland Commission) proposed that sustainable development may be defined as ‘development that meets the needs of the present without compromising the ability of future generations to meet their own needs’.⁶

This brief statement has fostered many more elaborate definitions and has enabled the identification of four main dimensions with regard to sustainable development, namely, the environmental, the economic, the financial and the social.

Environmental and Ecological Dimensions

Ecological economics (a discipline concerned with valuing ecological resources)

⁵ Thoopal, R. K., 2000. *Railway Pricing and Charges*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

⁶ WCED (World Commission on Environment and Development), 1987. Brundtland Report. *Our Common Future*. Oxford. Oxford University Press.

defines sustainability in terms of *natural capital*, the value of natural systems to provide goods and services, including clean air, water and climatic stability. Preserving these services is equivalent to a business maintaining the value of its productive assets. Ecological economists argue that consumption should not deplete natural capital at a rate faster than it can be replaced by viable and durable human capital. This suggests, for example, that non-renewable resources, such as petroleum, should not be depleted without sufficient development of substitutes, such as renewable energy sources (for a more technical discussion of the above, see Annex 1 at the end of this chapter).

Ecological economics attempts to account for non-market costs of economic activities, which tend to be ignored in traditional economics or even considered positive economic events by indicators, such as gross domestic product. This requires determining the economic value of non-market goods and services, such as the benefits that a wetland site for a proposed new airport provides in terms of improving water quality and supporting fishing industries.

Sustainable economics maintains a distinction between *growth* (increased quantity) and *development* (increased quality). It focuses on social welfare outcomes rather than simply measuring material wealth, and questions common economic indicators, such as gross domestic product, which measure the quantity but not the quality of market activities. Unlike neoclassical economics, sustainable economics does not strive for ever-increasing consumption, but rather for *sufficiency*.

Sustainability tends to reflect a conservation ethic that minimizes resource consumption and waste. This requires changing current economic policies that encourage production and consumption. For example, many countries minimize energy prices in order to keep transport utilities affordable. That reflects a *consumption* ethic. A *conservation* ethic might increase energy prices (perhaps through a carbon tax) while implementing programmes to increase vehicle fuel efficiency, improve alternative modes of transport, and increase industrial efficiency, so that manufacturers and consumers can meet their needs with less resource consumption.

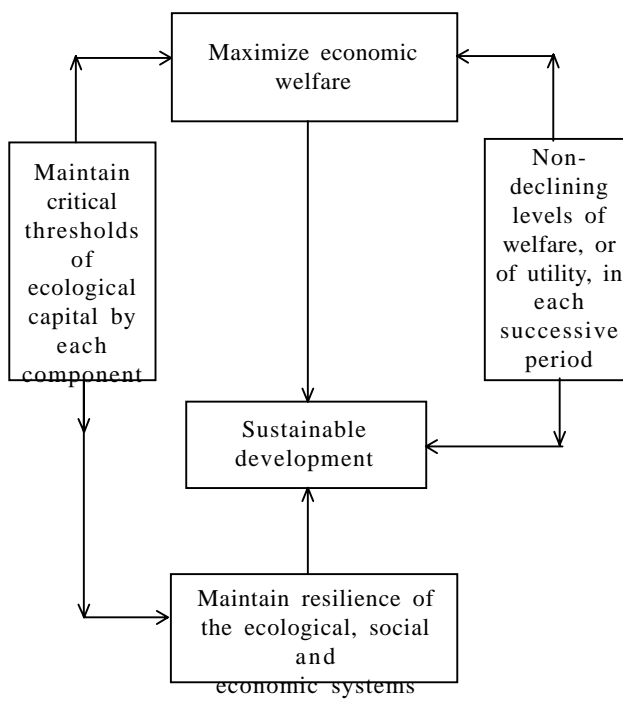
Interpreting resource consumption in these terms has led recently to ecological definitions of sustainability, which have implications for economic development. Rao⁷ suggests that society can draw upon economic and ecological resources to such an extent that the productive capacity to produce material well-being remains intact. He further elaborates that sustainable development is the process of socio-economic development whereby the worth of capital stocks (i.e., resources), valued at the

appropriate shadow prices, remains constant, or undiminished, at each time interval, for ever. This concept embraces the assumption that the ‘rental’ or ‘return’ on the productive capacity of an economy, measured by the adjusted or corrected net national product (NNP), remains constant, or does not decline, under conditions of sustainable development. Diagram 1.1 provides a schematic representation of sustainable development. It attempts to describe the factors which define when a system is sustainable and how this relates to the maximization of social welfare.

The term ‘resilience’ used in the diagram refers to the ability of an ecosystem to absorb internal and external disturbances to its environment and restore balance (or equilibrium). The term ‘threshold of ecological capital’ refers to the level below which the loss of ecological capital cannot be compensated by other types of resources, and may result in a threat to the resilient power of the ecological and economic system.

In essence, sustainable development is the achievement of continued economic development without detriment to the environment and natural resources – a definition which inextricably links the environment, ecosystems and economic development. The responsibility for achieving this has to be shared globally in a fair, equitable and proportionate measure.

Diagram 1.1 : The Environment – A Schematic Representation of Sustainable Development



Economic Dimension

Well-functioning markets are normally efficient mechanisms for allocating resources, both between alternative uses and over time. However, the following fundamental

7 Rao, P. K., 2000. *Sustainable Development: Economics and Policy*. Blackwell. London.

conditions must be met, if a market is to function efficiently:

- (i) property rights over all resources must be clear and secure;
- (ii) all scarce resources must be subject to prices determined on the basis of supply and demand;
- (iii) there should be no significant externalities;
- (iv) competition should prevail;
- (v) there should be no public goods; and
- (vi) issues of myopia, uncertainty and irreversibility should not arise.

If these conditions are not met, the free market will fail to allocate resources efficiently, both today and over time. Further, environmental degradation will occur and sustainable development will not be achieved. In other words, too many resources will be used up today and too few resources will be left for the future.

Much of the mismanagement and inefficient use of resources and the environment can be traced to malfunctioning, distorted or totally absent markets. In such situations, prices do not reflect the true social and economic costs and benefits from resource use. Such prices convey wrong or misleading information about resource scarcity and provide inadequate incentives for the management to efficiently use resources and capital assets, or to facilitate sustainable development.

Panayotou⁸ summarized some major forms of market failure, which are relevant to the transport industry and which inhibit sustainable development, including:

- (i) externalities, spillover effects and non-priced intersectoral linkages;
- (ii) public goods; (iii) uncompetitive markets;
- (iv) myopic planning horizons and 'too high' discount rates;
- (v) risk and uncertainty;
- (vi) irreversibility; and
- (vii) policy failures.

In addition, ill-defined or totally absent property rights and the existence of unpriced resources with non-existent or thin markets will also lead to inefficient economic development and misallocation of resources.

Externalities

⁸ Panayotou, T., 1992. *Economics of Environmental Degradation*. The Earthscan Reader in Environmental Economics. Markandya, A. and Richardson, J. Earthscan. London.

Externalities exist when the activities of one group of individuals/countries (either consumers or producers) affects the welfare of another group without any payment or compensation. There are many examples of external costs or negative externalities associated with transport: air travellers impose noise costs on those living below aircraft flight paths; heavy trucks inflict dirt and vibration on those living adjacent to roads; cars impede pedestrian movement in towns and cities; shipping pollutes beaches through oil discharges. At the global level, excessive energy consumption by one group of countries can also impose negative externalities on another group. Externalities may also be positive in nature and give rise to certain external benefits.

A distinction should be made between *pecuniary and technological externalities*⁹. The former occur when, say, a firm's costs are affected by price changes induced by other firms' actions in buying and selling factors of production. By way of example, a new road may destroy the view for particular residents thereby causing a reduction in their welfare – this is a technological externality. If the new road also takes business and income away from a local garage and transfers it to a service station on the new road, then the reduced income suffered by the garage proprietor is a pecuniary externality, since the effect is indirect.

The distinction is a fine one since, in practice, both forms of externality occur simultaneously, but it is an important one. Technological externalities are real resource costs that need to be accounted for if optimal efficiency is to be ensured. Pecuniary externalities do not involve resource costs, in an aggregate sense, but they do have implications for the distribution of income. In the above example, the new road leads to additional income and profit for the service station owners, at the expense of the garage proprietor. Pecuniary externalities do not affect total economic benefits, but rather affect the distribution of costs and benefits amongst society.

If technical externalities are not internalized through measures, such as taxes or specific charges, then there will be inefficiency, and the conditions necessary for sustainable development will not exist.

Public Goods

A public good is characterized by jointness in supply, in the sense that to produce such a good for one consumer, it is necessary to produce it for all consumers. In many cases, no individual can be excluded from benefiting from the good or service, whether or not he pays for it. Also, the marginal cost of producing public goods is zero.

⁹ Button, K. J., 1982. *Transport Economics*. Heinemann. London.

When several originators and recipients are involved, externalities, such as air pollution and traffic congestion, may be considered as ‘public bads’ and their correction as ‘public goods’. In fact, a public good may be thought of as an extreme case of a good that only has externalities.

In such circumstances, the free market is unlikely to produce public goods, such as cleaner air or reduced congestion, because individuals acting alone are unlikely to pay for them.

Uncompetitive Markets

Even when markets do exist and are very active, there may be market failure in the form of insufficient competition. For markets to be efficient, there should be a large number of buyers and sellers of reasonably homogeneous products with few barriers to the entry of new suppliers. Potential new entrants act as an insurance against monopolistic practices.

In reality, the transport industry is ridden with monopolistic elements. A market is imperfectly competitive if the actions of one or a few sellers or buyers have a perceptible influence on the prices of services. Market imperfections may arise for a variety of reasons. The transport industry is characterized by declining or decreasing costs arising from the indivisibility of the required capital investment. Average costs tend to fall continuously as output increases, and this provides an incentive for production to become concentrated in the hands of one or two service providers – a condition known as ‘natural monopoly’. Governments sometimes bring natural monopolies under public ownership to limit the excessive monopolistic practices.

Other causes of limited competition may be institutional, legal, or political barriers to entry into a particular market. The main problem with monopoly is that prices are normally kept too high and output too low for social optimality. Such practices are not necessarily inconsistent with the objective of conservation, but are not the best way of promoting sustainable development.

Myopic Planning Horizons

Sustainable development ultimately involves the sacrifice of present consumption for the promise of future benefits. A problem arises where the market rate of interest (discount) fails to reflect society’s true rate of time preference for consumption. A combination of poverty, impatience and risk can create a divergence between the private and social discount rates. If private discount rates are significantly higher than the social discount rates, then current consumption will exceed the consumption appropriate for

sustainable development.

Environmental and market uncertainties, coupled with a short and uncertain lifespan, lead people to adopt myopic time horizons and discount rates, which result in short-sighted decisions in pursuit of short-term profits at the expense of long-term sustainable benefits.

Risk and Uncertainty

Conservation and environmental management are related to the future. Risks and uncertainties concerning future costs and benefits of investments or pricing policies aimed at promoting sustainable development are significant and may prevent sound proposals from being adopted and implemented.

Irreversibility

Market decisions about the future (such as consumption vs investment) are made with the best available, though incomplete, information about future developments, on the assumption that such decisions may be reversed in the light of new information. However, this assumption often does not hold for investments in transport infrastructure, due to their scale and longevity.

Policy Failures

Government policies often tend to introduce additional distortions in the market rather than correcting the existing ones. The correction of market failure is rarely the sole, or even the primary, objective of government intervention; other objectives, such as national security, social equity, macroeconomic management, and political expediency may dominate. Policy failures include the introduction of distortions to otherwise well-functioning markets through taxes, subsidies, regulations, inefficient state enterprises, and poorly planned public investment. It is also commonplace for governments to ignore externalities and not attempt to internalize them. Thus, policy failures include both the failure to intervene in an appropriate manner when necessary and to refrain from intervention where it is unnecessary or inappropriate.

Financial Dimension

Financial sustainability has three components, in the sense that such an activity needs to be able to:

- (i) attract sufficient funds to finance the necessary investment and operation;
- (ii) generate sufficient revenue to recover both the operating and capital costs involved; and

- (iii) provide the necessary financial incentives to attract and sustain wider participation in such ventures.

The role of the private sector is fundamental in facilitating financial sustainability. The major challenge, however, is to devise frameworks that will ensure both economic optimality in the allocation of resources and financial viability, consistent with promoting sustainable development.

Social Dimension

Seen from a broader perspective, development encompasses the strengthening of the material income base as well as the enhancement of capabilities and the enlargement of choices.¹⁰ Such a view of development clearly transcends the narrow concept of development as economic growth and emphasizes the importance of social development in the context of sustainable development.

Social issues are regarded as critical to the concept of sustainable development because of the importance of equity considerations. More precisely, inter-generational or intertemporal equity forms one of the cornerstones on which sustainable development is built. Hence, inter-generational equity – covering the whole gamut of social issues in development, such as regional and gender income distribution – need to be rightly considered as an integral part of sustainable development.

To sum up, sustainable development is inextricably linked to economic viability and social justice. However, other objectives may, in addition, demand the financial sustainability of investment in infrastructure and the provision of services.

Section III Sustainable Transport

The discussion of sustainable development has led to the identification of a further concept, namely, that of sustainable transport. The Organization for Economic Cooperation and Development (OECD) has defined a sustainable transport system as one where:

- (i) generally accepted objectives for environmental quality (such as those set by the World Health Organization concerning air pollutants) are met;

¹⁰ UNIDO, 1998. *Sustainable Industrial Development*. United Nations Industrial Development Organization. Vienna.

- (ii) ecosystem integrity is not significantly threatened; and
- (iii) potentially adverse global phenomena, such as climate change and stratospheric ozone depletion are not aggravated.

For policy purposes, an environmentally sustainable transport system can be defined as one where transportation does not endanger public health or ecosystems and meets the needs for access consistent with (i) use of renewable resources below their rates of regeneration, and (ii) use of non-renewable resources below the rates of development of renewable substitutes.

Although it is somewhat difficult to give an unambiguous definition of sustainable transport, it can be argued that current trends in the sector are environmentally, and therefore economically and socially, unsustainable¹¹. Transport pollutes the environment in many ways. Mechanized transport generates noise, toxic fumes, dirt and fears for safety, and often results in community severance, loss of privacy, and a need for people and industry to relocate. Many environmentalists argue for substantial reduction or even total elimination of certain adverse effects of the operation of transport services and the building of transport infrastructure. Indeed, OECD is currently attempting to define minimum standards in respect of certain criteria related, among others, to the control of carbon emissions, nitrogen dispositions, reducing the emission of particulates, and abatement of noise. The problem with these approaches is that they may ignore the costs of removing such externalities and underestimate the benefits of the transport system as also of greater mobility.

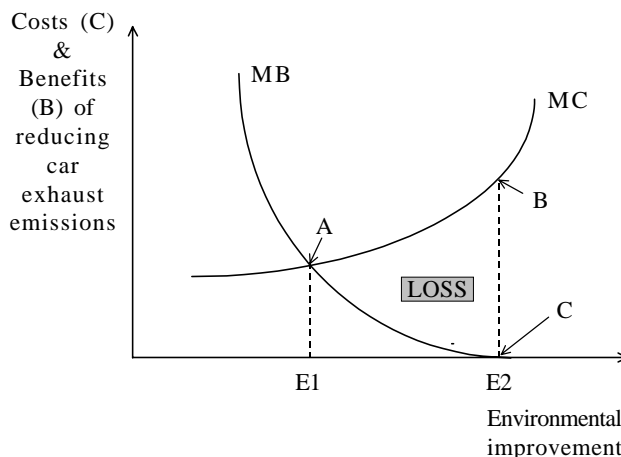
Figure 1.1 depicts the costs and benefits of environmental improvements in different scenarios. The vertical axis shows the costs and benefits of, say, reducing noxious exhaust fume emissions from road vehicles. The marginal cost (MC) of obtaining improvements is likely to rise; in addition, as more fuel refinement occurs, fuel efficiency is likely to decline and the toxic waste generated by refineries will increase. The marginal benefits (MB) of successive environmental improvements are likely to fall, as the public becomes less aware of successive improvements in environmental standards. The diagram shows that there is likely to be an optimal level of environmental improvement (E1), where the marginal benefits arising from further improvement equal the marginal costs of securing those improvements. If an environmental standard was introduced, which secured even greater environmental improvement – say, to point E2, where exhaust fumes are deemed pure and hence further improvements would not produce further environmental benefits

¹¹ Todd, L., 2000. *Issues in Sustainable Transportation*. Victoria Transport Policy Institute. Vancouver.

– then marginal costs would exceed marginal benefits and there would be a net loss in social welfare equal to the area bounded by points ABC.

Consequently, when discussing excessive environmental harm caused by various forms of transport, it is important to remember that this is an excess above the optimal level of pollution, not above zero pollution or some perceived 'pure' environment. Charging and pricing systems have a significant role to play in securing improved sustainability. Also, in view of the fact that it is the concentration of particulate matter in the atmosphere that is important, consideration also needs to be given to fiscal solutions and cleaner fuels.

Figure 1.1 : Costs and Benefits of Environmental Improvements



Economic and Financial Dimensions

If transport itself is to be sustainable, it must be cost-effective and continuously responsive to changing demands, which may be achieved by creating competition in those parts of the sector where a commercial and free market can operate, without significant adverse spillover (externalities) and distributional consequences. In areas where this cannot be achieved, it will be necessary to enhance user participation.

Table 1.1 provides examples of the ways to increase the responsiveness of transport supply to user needs by creating competition, increasing the efficiency of transport financing and management, and enhancing user participation in planning transport provision. These objectives are derived from those proposed by Gwilliam and Shalizi.¹²

Environmental Dimensions

Environmental sustainability requires that environmental issues be addressed as an integral part of transport strategy formulation and project design. Transport activity causes stress on the natural environment in two ways¹³:

- (i) By drawing on scarce resources, such as fossil fuels, minerals and various

non-metallic resources like sand and stone, and by using land.

- (ii) By giving rise to wastes, such as pollutant gases, solid wastes, noise and accidents, which all flow back to the natural environment that acts as a sink to absorb them, and by partitioning or destroying the ecosystem of the neighbourhoods of transport operations, such as farm land, wildlife habitats, and water systems (see Annex 1).

Table 1.2 describes the important environmental effects of the main modes of transport¹⁴.

Reducing loss of life and health threats is of the highest priority. Over half a million people are killed annually in road accidents, accounting for over 1% of GDP in some countries. In many cities, road traffic accounts for 95% of health threatening lead and carbon monoxide in the air. Cost-effective technology is necessary, but not sufficient, for transport to be environmentally sustainable. Strategic action is also required in the form of better directed land-use planning, stricter demand management, and greater incentives to use public transport. Some of the policy measures, which will contribute to the provision of environmentally sustainable transport, are set out in Table 1.3.

Social Dimensions

Social sustainability of transport refers generally to the improvement of standards of living and quality of life. In particular, it embraces poverty reduction as an integral part of transport planning and strategy. Meeting the transport needs of the poor, lays emphasis on the maintenance of rural access facilities and the role of the informal transport sectors, which are more labour-intensive and less motorized. Table 1.4 outlines a number of alternative policies available to address social sustainability of transport. Inherent in it is the assumption that public transport has become a quasi-public good and, as such, there will be divergence between price and marginal cost, the degree of divergence depending on the levels of poverty prevalent in a country.

Sustainability: Synergies and Trade-offs

According to Gwilliam¹⁵, economic, social and environmental sustainability in

¹² Gwilliam, K. M., and Shalizi, Z., 1996. *Sustainable Transport : Sector Review and Lessons of Experience*. TWU 22 10/96 Washington. World Bank.

¹³ Sengupta, R., 2000. *Environmental Sustainability and Transport Pricing*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

¹⁴ ESCAP, 1995. *State of the Environment in Asia and the Pacific*. UN-ESCAP/ADB; ST/ESCAP/1585; p.289.

Table 1.1: Creating Competition

Policy Objective	Methods
To develop the capabilities necessary to devise competition policies for the transport industry	<p>Establish a strategic planning framework to facilitate competitive transport systems.</p> <p>Develop methods of user and community participation in transport development and operation.</p>
To establish an enabling framework for competition and private sector participation	<p>Review, disaggregate and restructure agency responsibilities to enable the sale, leasing, or subcontracting of transport infrastructure provision, operation and maintenance.</p> <p>Create or strengthen regulatory institutions to:</p> <ul style="list-style-type: none"> - monitor performance standards - ensure competition - prevent cartelization - protect public interests <p>Increase the capacity for private/public ownership partnerships by clearly defining risks and responsibilities for private investors.</p>
To develop competitive structures in the transport industry	<p>Explore privatization and corporatization of transport operators and infrastructure agencies.</p> <p>Encourage private provision of passenger and freight transport services.</p> <p>Discourage uncompetitive practices, such as cartels and protectionist transport policies like cargo reservation and flag discrimination.</p>
To increase efficiency in the use, provision, financing, and management of transport infrastructure	<p>Promote deregulation to facilitate ‘competition in the market’.</p> <p>Develop franchising and concessions to ensure ‘competition for the market’.</p> <p>Introduce direct charges based on resource costs, including external costs.</p> <p>Use ‘user charges’ based on earmarking of taxation to pay for infrastructure upkeep and maintenance.</p>

Table 1.2: Selected Environmental Effects of Principal Transport Modes

	Marine and Inland Water Transport	Rail Transport	Road Transport	Air Transport
Air		Air pollution in populated areas; global pollution from thermal generating stations for electric traction	Air pollution (CO, HC, NO _x , particulates such as lead), global pollution (CO ₂ , CFCs)	Air pollution, greenhouse & ozone depletion effects at higher altitudes due to NO _x emissions
Water Resources	Discharge of ballast water, oil spills, etc.; modifications of water systems during port construction & canal cutting and dredging		Modification of water systems by road building; pollution of surface and ground water by surface run-off	Modification of water tables, river courses and field drainage in airport construction
Land Resources	Land taken for infrastructure; dereliction of obsolete port facilities & canals	Land taken for right-of-way and terminals; dereliction of obsolete facilities	Land taken for infrastructure; extraction of road building materials	Land taken for infrastructure; dereliction of obsolete facilities
Solid Waste	Abandoned and laid-up vessels and craft	Abandoned lines, equipment and rolling stock	Abandoned facilities and rubble from road works; road vehicles withdrawn from service; waste oil	Abandoned facilities and aircraft withdrawn from service
Noise		Noise & vibration around terminals and railway lines	Noise around highways	Noise around airports
Risk of Accidents	Bulk transport of fuels and hazardous substances	Derailment or collision of trains carrying hazardous substances	Deaths, injuries & property damage due to road accidents; risk from transport of hazardous goods	Deaths, injuries & property damage due to aircraft accidents
Other Impacts		Partition or destruction of neighbourhoods, farmland and wildlife habitats	Partition or destruction of neighbourhoods, farmland and wildlife habitats; congestion	

transport is often mutually reinforcing. Diagram 1.2 describes some of the key synergies and trade-offs available to transport policy makers when devising strategies to promote sustainable development.

Zone A signifies that trade-offs exist between environmental and economic sustainability. For example, policies designed to improve access and mobility for

Table 1.3 : Environmental Sustainability

Policy Objective	Methods
Improve health and safety in transport	<p>Develop road/transport safety programmes.</p> <p>Introduce cleaner fuel standards to eliminate lead and sulphur emissions, combined with fuel supply and pricing policies to encourage use of cleaner fuels.</p>
Integrate environmental and economic dimensions in transport planning and development	<p>Develop systematic methods for estimating the impact of transport on safety and air pollution, including monetary valuations in economic rate of return calculations.</p> <p>Provide protection against adverse environmental impact of road construction and other activities related to transport development on forests, wetlands, natural habitats, non-motorized transport and heritage sites.</p>
Develop an environmentally sensitive strategic framework	<p>Develop local standards for the provision of environmentally sensitive/non-motorized transport.</p> <p>Develop responsive urban mass transport plans which respond to changing land use.</p> <p>Establish road user charges that reflect externalities (road damage, air and noise pollution, congestion and safety), using fuel taxation.</p> <p>Ensure that the fare, service and finance policies related to public transport services reflect the need to maintain these services as also the true economic benefits of public relative to private transport.</p> <p>Structure transport funding to maintain optimal modal balance.</p>

employment and amenities will generate more motorized traffic thereby increasing environmental degradation.

Zone B indicates that trade-offs may exist between the economic benefits of transport improvements and their implications for income distribution. Examples include schemes, which increase public transport fares in order to reflect the private and social costs involved and which may adversely affect the poor. Similarly, new road schemes may cause some industrial dislocation with consequential redundancies and job losses,

Table 1.4 : Social Sustainability

Policy Objective	Methods
Address the transport problems of the (urban) poor	Improve physical access to jobs and amenities. Promote informal transport methods subject to health and safety standards. Eliminate gender bias in transport provision.
Improve methods of addressing transport problems of the poor	Emphasize access more than speed in transport development. Support labour intensive infrastructure development and operation. Ensure community participation in transport planning.
Protect the poor against adverse changes in transport policies	Develop efficient subsidy schemes for socially necessary services. Plan redundancy schemes arising from increase in the efficiency of transport services and operations. Minimize dislocation and resettlement costs arising from transport schemes.

or give rise to residential resettlement costs for low-income groups.

Zone C shows the potential trade-off between social and environmental effects arising from changes in the transport industry. The development of the informal transport sector and two-wheeled motorized transport to meet the mobility needs of the poor, can give rise to significant amounts of noise and air pollution.

¹⁵ Gwilliam, K. M., and Shalizi, Z., 1996. *Sustainable Transport: Sector Review and Lessons of Experience*. TWU 22 10/96 Washington. World Bank.

The *Zone of Synergy* represents opportunities to improve safety and health, promote efficient pricing, internalize externalities, provide for asset maintenance, develop efficient contract design and operation, and invest in efficient infrastructure.

Gwilliam also suggests that it is possible to exploit the potential synergies and to minimize the adverse effects of transport development and that transport can be managed in a way which promotes sustainable development.

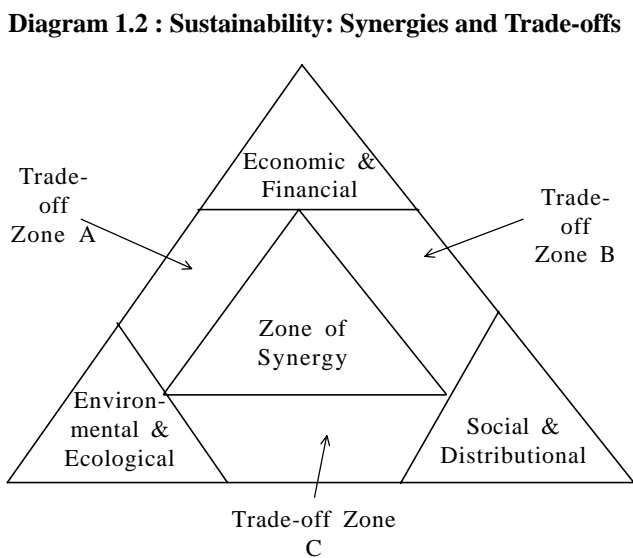
Role of the Government

The implementation of a market-based approach to transport policy implies a radical change in the role of the government. The private sector can increasingly take up the responsibility for providing, financing and operating transport services and also some types of transport infrastructure through concession arrangements and franchising. Thus, the role of the government as the supplier and quantitative regulator of transport should decline. However, its role as enabler of competition and the custodian of environmental and social interests should increase. In the area of investment planning, social cost-benefit analysis is becoming widespread in determining the efficient allocation of resources, both for transport investment and for outsourcing. However, setting efficient charges for the use of publicly provided infrastructure, maintaining a competitive environment in the transport sector, and increasing community and user participation in decision-making, is essential, particularly in those areas where markets do not function adequately.

Sustainable Development and Transport

It may be pointed out that the concept of sustainable development when applied to transport refers to its role in securing a balance between equity, efficiency and intertemporal concerns. This is predicated on:

- (i) the maintenance of high and stable levels of economic growth and employment in order to generate the necessary resources to achieve sustainable development;



- (ii) the protection of the environment – both at the local level and globally – through the prudent and efficient use of natural resources and the development of renewable resources, wherever possible; and
- (iii) social progress which addresses the needs of everyone – through reducing the impact of pollution, poverty, and unemployment on health and the quality of life.

Sustainable Transport Policy

A policy of sustainable development in the transport sector can pursue multiple objectives. Alagh¹⁶ identifies the following:

- (i) securing energy efficiency;
- (ii) fully reflecting the costs of non-renewable resources in the operation of transport vehicles;
- (iii) creation of responsive and effective markets; and
- (iv) adopting environmentally friendly processes which discourage the generation of external diseconomies in a costless manner.

In a review of diesel and gasoline prices and taxes, Metschies¹⁷ has shown that within the Asian countries, the retail price (as of November 2000) of motor fuel ranged from 2 cents/gallon to 80 cents/gallon for diesel and 2 cents/gallon to 146 cents/gallon for gasoline. In addition, in a significant number of Asian countries, the retail price of motor fuel lies below the resource cost and border price (import price) per litre. The justification for providing subsidies usually relates to concerns on grounds of income distribution; however, these subsidies often tend to conflict with the need to promote sustainability.

In summary, a sustainable transport policy would involve intervention in the market system to ensure that:

- (i) the direct or indirect use of natural resources should be such that they can at least be replaced by natural regeneration (e.g. by hydroelectric energy for electric traction), or by discovery of new deposits of the currently used exhaustible resource (e.g. oil or natural gas reserves), or by the use of a new renewable resource (e.g. wind or solar power), or by conserving the use of resources per unit of transport output, or by a combination of these;
- (ii) the damage to the environment should be within such limits that the

¹⁶ Alagh, Y. K., 2000. *Sustainable Development Policies*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

productivity of other economic activities and the quality of life, in terms of health and security against accidents, do not deteriorate over time.

The scarcity value of the natural resources used in the provision of transport infrastructure and services and the external costs due to pollution and degradation of the environment (i.e., the social cost of transport), should be built into the price of providing or using transport facilities and services¹⁸.

Section IV

Transport Pricing and Sustainable Development

In many countries of the region, transport services and infrastructure facilities have been treated as public services or instruments of social policy. In this environment, prices have rarely reflected the cost of provision of these services and facilities, subsidies have been given, and strict commercial and management accounts have often not been maintained. In consequence, it is being increasingly recognized within the public sector that such pricing practices lead to a number of outcomes which will not promote sustainable development, including:

- (i) economic inefficiency with the resultant waste of resources;
- (ii) generation of insufficient funds to develop, operate and maintain transport infrastructure and services;
- (iii) creation of distortions in user's choice of mode of transport; and
- (iv) externalities in production (such as pollution) as well as externalities in consumption (in the form of congestion).

Divergences between prices and costs not only send the 'wrong signals' to the providers and operators of transport infrastructure and services when considering their investment decisions as well as their operating and maintenance plans, they also send the 'wrong signals' to consumers. For example, where railway prices incorporate infrastructure costs but road prices do not, consumer preferences between these two modes will be distorted. External costs, which are not reflected in prices, are another source of distortion. This is particularly the case with pollution and congestion externalities.

¹⁷ Metschies, G. P., 2000. *Present Taxation Levels on Diesel, Gasoline and Vehicles in Asian Countries and the Consequences for Future Taxation Strategies*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

¹⁸ Sengupta, R., 2000. *Environmental Sustainability and Transport Pricing*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

One of the consequences of these practices is that insufficient revenue is generated to develop and/or maintain transport services and infrastructure. In some sub-sectors, governments have turned to the private sector to assist in the financing, development and operation of these services and facilities. However, the maximization of private net benefits by the private sector does not necessarily coincide with the maximization of social benefits of the public sector and, consequently, conflicts arise. It is further recognized that if the private sector is to participate in infrastructure development, then its private benefits (revenues, including profit) will need to reflect its costs. Apart from cases where concessions are awarded to develop adjacent land, most revenues in the private sector are derived from the prices charged for facilities and services provided. In cases where insufficient revenues are derived from these sources, or where it is deemed necessary that other social or pragmatic reasons take precedence over commercial pricing principles, alternative policies will need to be developed.

In view of the enormous financing needs of transport infrastructure, the generation of sufficient funds for its development and maintenance presents a considerable challenge to countries of the region. In some sub-sectors, for example, power generation and telecommunications, the private sector has been playing a significant role in infrastructure development. In most other sub-sectors, there has been low private sector involvement. As a result, there is a need for the governments to continue to identify sources of funds and to mobilize them for development and maintenance of transport facilities and services. User-related charges represent one such source of funds. For some services and facilities, user charges are levied; however, the revenue from these sources does not necessarily find its way back into the concerned sub-sector. Consequently, there are again insufficient funds for development and maintenance of the infrastructure.

In many countries, the development of the transport infrastructure is constrained by the skewed process of capital formation; the ineffectiveness of central planning organizations; and the prevailing political system. Further, it is generally the case that the limited funds which are available are allocated to the construction and development of transport facilities and services, with only limited funds being made available for the maintenance of the assets.

Investment and pricing are relatively straightforward in a 'first-best' world. To promote sustainable development, the investment rule is to invest if benefits exceed costs and the pricing rule is to set prices equal to marginal cost. The transport sector, however, has a number of characteristics, including the joint nature of costs, indivisibilities of supply and demand, durability and externalities, which complicate pricing and investment decisions, and can, if not addressed, adversely impact development and its sustainability.

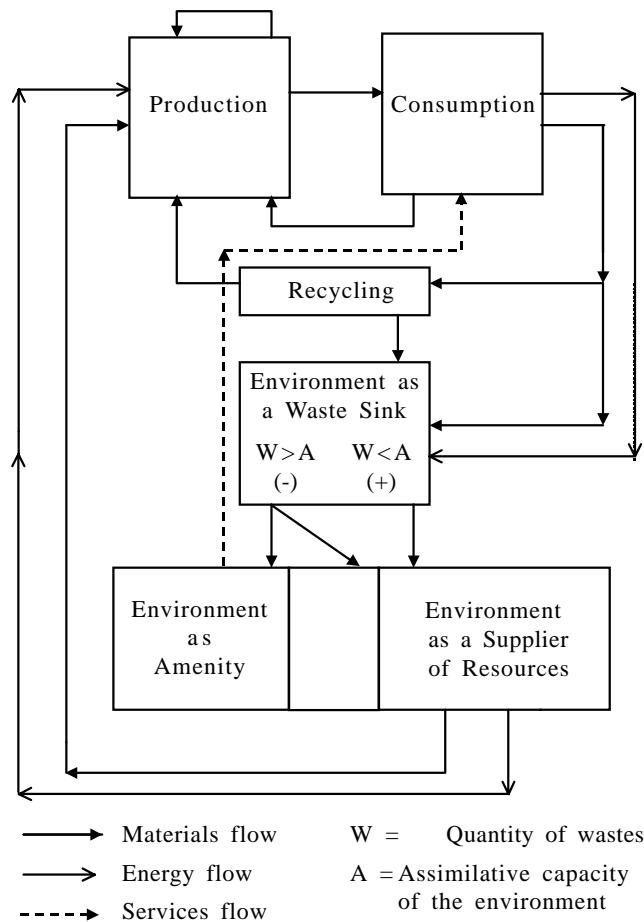
The shortcomings and problems described above give rise to an urgent need to examine existing pricing practices with the intention of developing alternative regimes which promote sustainable development. The issues are complex and interrelated and need to be addressed within the context of the transport sector and its wider challenges. This volume in the subsequent chapters attempts to identify, analyze, and propose solutions for the following major issues:

- (i) *Issues related to the transport market* – Such issues include the nature of the transport market itself, including indivisibilities or peak-load demand and the supply-side causes of market failure in the transport sector. These embrace such characteristics as high capital and sunken costs, non-tradable and site-specific investment, the ‘public goods’ nature of some services, widespread government subsidies, the existence of natural monopoly, government regulation, and the prevalence of externalities and spillover effects generated by the transport sector.
- (ii) *Issues related to the development, operation, maintenance and management of both transport infrastructure and transport services* – The management practices and marketing policies, found in the transport sector, give rise to a number of problems associated with the raising and sourcing of finance for the development and maintenance of transport infrastructure and services.
- (iii) *Issues related to income distribution* – Need often arises to take account of the impact of transport pricing policy on equity and distribution of income.
- (iv) *Issues related to private sector participation* – These relate to experience and advantages of private sector participation and the prevailing constraints concerning such participation.

Ecological Economics

A system of living things in relationship with their environment is called an ecosystem which can be global, regional or local. Ecosystems are characterized by the principles of *materials circulation* and *one-way energy flow*. It is possible to envisage materials as being natural resources that are taken from the natural environment. In Diagram 1.3 the environment is shown as both a supplier of resources and a provider of amenity. Environmental resources, in the form of materials and energy, are shown flowing to the

Diagram 1.3 : Materials Circulation and Energy Flow in the Environment



economic production sector where they are processed to become consumable goods.

The first law of thermodynamics suggests that these resources cannot be created or destroyed, and so, the resources flowing into the production sector must go somewhere, as follows:

- (i) some will be embodied in consumer goods and will pass to the consumption sector;
- (ii) some will be made into capital goods and will remain in the production sector, as physical capital;
- (iii) worn out capital will be returned to the environment as waste;
- (iv) substantial quantities of resources will be discarded as waste during the production process;
- (v) energy resources will be passed on as consumer goods – usually as energy in directly consumable form – or will be ‘used up’ in the production process. However, this does not mean that energy is destroyed; it is merely dissipated as waste heat, gases and noise. Some materials will be recycled within the production sector, as shown in the diagram.

Apart from recycling and storage of capital, Diagram 1.3 shows that whatever is taken out of the environment in the form of physical resources must reappear as consumer goods or waste. Recycling defers materials and energy appearing ultimately as waste.

In the consumption sector, the same proposition is true, namely, that whatever passes from the production sector to consumption sector must also reappear as waste. Even agricultural produce consumed by humans appears as waste or as biological capital. All other goods are either directly consumed and disposed of (e.g. paper, coal, timber) or are held for varying periods of time and then disposed of (e.g. cars, light bulbs). Of course, some consumer waste is recycled, as shown in the diagram.

The combined materials and energy flows from the production and consumption sectors can, therefore, be seen as waste disposed of to the environment. For any given period of time, these waste flows may be less than, equal to, or greater than the resource extraction for that period. It all depends on how much absorption of materials is occurring in various forms of capital in that period, and how much disposal is occurring from past flows. It is reasonable to suppose that the two flows will be approximately equal at any time period.

An essential feature of this extended view of the economic process is that consumption is no longer the final act, because whatever is consumed also appears as waste residuals disposed of in the environment. It is necessary to recognize that ecosystems also have their own waste products which, in steady-state, are recycled to reappear as nutrients that sustain the ecosystem. For non-steady-states, the recycling process still operates, but it is likely to bring about changes in the structure of the ecosystem. Many of the wastes, produced by economic activity, are degraded by the environment's degrader populations; for example, bacteria in water systems often deal efficiently and quickly with sewage. However, it is important to note that, with regard to such wastes, the environment has limited assimilative capacity. The environment cannot degrade all waste, nor can it assimilate any quantity of wastes. If this quantity (W) exceeds the environment's assimilative capacity (A), it will remain as a noxious stock in the environment, eventually preventing the degrader populations themselves from functioning. In the diagram where $W > A$, the outputs of the waste sink have a negative effect on the environment both as an amenity (for example, stagnant waterways cannot be used for fishing or recreation) and as a supplier of resources (e.g. stagnant waterways cannot be used for water supply). Conversely, where $W < A$, the effect on the environment will be positive.

Where waste emissions and pollutants are kept within the bounds of the environment's assimilative capacity, degrader populations will be reasonably stable and the wastes will then be properly degraded and recycled as nutrients. The environment, in such circumstances, may be regarded as sustainable (see Pearce, D.W., 1976. *Environmental Economics*. London. Longman).

So far, the environment has been described as a supplier of resources; however, it is useful to classify resources in terms of their intertemporal characteristics, that is, what their use now implies for their future use. There are basically four categories of resources:

Stock-energy Resources

Resources in this category, such as oil, gas and coal, are fixed in total stock and will be ultimately exhausted as long as usage is positive. Recycling is rarely possible and intertemporal trade-offs are secured by abstention from current consumption.

Stock-material Resources

Stock-material resources – water, copper, lead and other metals – that are subject to fixed supply and intertemporal transformation, can only be secured by deferring or

abstaining from current consumption. However, such resources are usually subject to some recycling.

Flow-energy Resources

The rate of capacity flow for these resources, such as solar power, tidal energy and hydroelectric supply tends to be fixed, but actual usage falls well below the capacity flow. These resources also possess two unique features:

- (i) they cannot be stored for very long and, therefore, intertemporal trading is not possible; and
- (ii) although optimal usage might appear to be infinite, this ignores possible economic-ecological trade-offs and opportunity cost (e.g. hydroelectric schemes may adversely affect irrigation, and tidal barriers may increase siltation and erosion).

Stock-renewable Resources

Such resources are subject to fixed stock at any point in time, but are renewable through biological reproduction. Agricultural produce, marine resources and human labour are examples of stock-renewable resources. The intertemporal trade-off is determined by the reproduction rate and the size of the parent population. Overfishing provides an example of the sub-optimal utilization of such resources.

The above provides a brief outline of the inextricable links between the environment, ecosystems and economic development.

2

Pricing and Charges for Transport Services and Infrastructure

Introduction

This chapter aims to define optimal transport pricing policies and address a range of complexities that arise due to the nature of the cost structure and market distortions generally observed in the transport sector. It comprises five sections: section I defines optimal transport pricing policies; section II deals with the problems of fixed and joint costs, indivisibilities, and peak-load issues; section III analyzes market distortions, such as monopoly and externalities; section IV examines financial targets, taxation, and equity; section V covers financing and private sector participation.

Section I Optimal Pricing Policies

Pricing is a method of resource allocation; there is no such thing as the ‘right’ price, rather there are optimal pricing strategies to permit specified aims to be achieved. For example, optimal price aimed at achieving profit maximization may differ from that needed to maximize social welfare, facilitate sustainable development, or, maximize passenger numbers. One of the major problems in evolving transport pricing policies is to decide exactly what the objective is. Bennathan and Walters¹⁹ have identified a major distinction between the ‘European’ and ‘Anglo Saxon’ doctrines for setting port prices. They suggest that the former set prices to facilitate the economic growth of the port’s hinterland, while the latter attempt to ensure that ports cover their costs and, where possible, make a profit irrespective of the effects on the wider local economy.

The discussion that follows examines the appropriate pricing policy for transport undertakings to adopt when faced with a variety of objectives and a range of market conditions. Whilst the primary objective is assumed to be concerned with maximizing the social benefits of transport, commercial interests and the profit objective are also

¹⁹ Bennathan, E. and Walters, A. A., 1979. *Port Pricing and Investment Policies for Developing Countries*. Oxford. Oxford University Press.

considered.

Pricing Objectives

The rationale and purpose of transport pricing policies are rarely made explicit. However, governments usually seek to maximize the ‘public interest’ when devising their transport policies, where the ‘public interest’ embraces a number of objectives, in particular:

- (i) economic efficiency;
- (ii) profitability;
- (iii) environmental sustainability;
- (iv) income distribution; and
- (v) relationship with macroeconomic policy.

Economic Efficiency

The concept of economic efficiency is derived from the theory of welfare economics, and is concerned with the allocation of resources in an economy. An inefficient resource allocation is one which can be changed in such a way as to make some people better off, and no one worse off, in terms of their own preferences²⁰. An efficient allocation is one where no such change is possible. Economic efficiency also implies managerial and technological efficiency whereby the services are provided at their lowest cost. Most governments have transport policies that adopt allocative or economic efficiency as a core objective. Economic efficiency means that the full cost of transport services is accounted for including social and environmental costs and that the investment evaluations account for the opportunity cost of consumption and investment. The efficiency criteria will be central to much of the analysis that follows.

Profitability

The *gross trading surplus* of an enterprise is the excess of its total revenue over its operating costs. Its *profit* is defined as the excess of its gross trading surplus over interest and depreciation provisions.

It is common for the governments to focus on the gross trading surplus as an indicator of ‘profitability’ for a publicly owned transport undertaking whilst the private sector would centre on profit after meeting the tax liabilities. There are two main reasons for focussing on this connotation of ‘profitability’:

²⁰ Mishan, E. J., 1960. *A Survey of Welfare Economics*. Economic Journal.

- (i) *Finance* : The extent to which public enterprises could contribute to the exchequer would depend entirely on the size of their gross surpluses. In addition to the surpluses of these enterprises, the government has two other principal sources of revenue : taxation and borrowings. Hence, if the surplus gets reduced the government will need to raise more funds through taxation and borrowings. It may be even be forced to lower its spending on other essential areas like health, education and defence, and may cut down subsidies to individuals (e.g. benefits and pensions) and to firms (e.g. investments grants).

The Government is not concerned with how the gross surplus is allocated by the public enterprises between interest, depreciation and profit; all that matters is the total size of the surplus for it alone determines the quantum and flow of funds to the exchequer. In turn, this will also have a bearing on the extent to which taxation and borrowing would be increased and the extent to which other forms of public expenditure would be curtailed.

- (ii) *Motivation* : The pursuit of profitability is perceived as a means of stimulating managerial and technological efficiency. It would, of course, be difficult to accept profit maximization as a goal of public enterprises because they often possess a degree of monopoly power; yet, in the absence of this objective there is little incentive for holding down the costs. In practice, publicly owned transport undertakings, which possess some monopoly power, have usually set profitability objectives, but are subject to a degree of price control or monitoring of pricing policy.

Environmental Sustainability

Protection of the environment has become an important objective for the governments in recent years. Transport, in general, and road transport, in particular, are widely recognized as major generators of pollution which threatens environmental sustainability. Increasingly, governments are introducing measures, including pricing measures, such as pollution taxes, to reduce and control environmental pollution in the public interest²¹.

It is possible to argue that promoting environmental sustainability is consistent with the aim of securing welfare maximization through economic efficiency, where social welfare incorporates environmental social costs and benefits.

²¹ Pearce, D. W. *et al*, 1990. *Sustainable Development: Economics and Environment in the Third World*. Edward Elgar Publishers Ltd. Cheltenham, England.

Income Distribution

A government will generally have explicit views about the distribution of real incomes in society. These will be reflected in the pattern of taxation and public expenditure, which are often regarded as instruments through which a government may bring about desired changes in the distribution of real income. Whilst transfer payments, such as benefits and pensions, are a major means of redistributing income, the provision of services, such as transport at subsidized prices, is often considered to be equally important. Such subsidies permit higher levels of usage by certain income groups than they could otherwise afford. Since such subsidies lead to higher levels of output of particular services, they can lead to benefits to factors of production, such as labour, in the form of greater employment and higher wages. Finally, real incomes can be redistributed through the provision of subsidized factor inputs, such as land for transport projects, which, in turn, will be reflected ultimately in transport prices. Consequently, governments often use transport pricing policy as a means to secure the desired distribution of income.

Macroeconomic Policy

In formulating macroeconomic policies, governments usually focus on four target variables: the level of unemployment; the rate of inflation; the balance of payments; and the rate of growth of national output. The level of investment in, and the pricing of, transport infrastructure and transport services both affects and is affected by macroeconomic policies. Transport infrastructure investment, through the multiplier effect, can influence the level of economic activity and growth; it can also generate employment and improve export competitiveness. Macroeconomic policies can, therefore, impinge, directly and indirectly, on transport pricing policies.

The sorts of objectives set out above are complex and are often incompatible. Table 2.1 sets out examples that are frequently encountered in reconciling multiple objectives. Whilst there are numerous transport pricing policy objectives, economists will usually argue that the pursuance of economic efficiency should take precedence over others for the following reasons:

- (i) Even assuming that clear, consistent and stable objectives are set, they are likely to imply higher costs and lower profits than if the sole aim was to achieve economic efficiency. Securing income distribution goals or macroeconomic targets through controlling the pricing policies of transport undertakings is likely to be at a high opportunity cost in terms of economic efficiency.

- (ii) Where wider objectives than economic efficiency and profitability are pursued, they tend not to be properly defined in a practicable and operational way.
- (iii) Objectives, other than economic efficiency and profitability, usually require extensive and complex political intervention.

Economic Efficiency Defined – A ‘First-Best’ World

Welfare economics postulates that social welfare will be maximized, and, as a result, economic efficiency will obtain, if prices are equated with marginal social cost. The rule is derived from welfare theory which assumes that interpersonal utility

Table 2.1 : Pricing Policy Objectives and Conflicts

Pricing Policy Objectives	Conflicts
Economic Efficiency vs Profitability	Pricing to promote the efficient use of transport capacity may lead to financial losses.
Environmental Sustainability vs Income Distribution	Pollution taxes may adversely affect poorer income groups and lead to unemployment.
Profitability vs Macroeconomic Policy	Pricing for profitability may lead to higher transport prices thereby creating inflationary pressure.
Profitability vs Income Distribution	Pricing for profitability may lead to higher transport prices with adverse effects on the poor communities.
Economic Efficiency vs Macroeconomic Policy	Macroeconomic price restraint policies may conflict with the need to increase transport prices during periods of congestion and excess demand.

comparisons can be made and seeks to maximize the sum of consumers’ and producers’ surpluses. Diagrammatically, assuming linear cost and demand functions, this can be shown as in Figure 2.1.

In the diagram, MC is the marginal cost curve or supply function, and D and MR represent the demand and marginal revenue curves, respectively. A profit maximizing firm would produce the output *og* at the price *oc*. Consumers’ surplus (assuming no income effects) would be given by the area *cde*, and producers’ surplus by the area *acek*, giving a sum of *adek*. Consumers’ surplus represents the benefit to consumers,

as expressed by their willingness to pay, in excess of the cost of providing a particular quantity or level of output. Producers' surplus represents the revenue in excess of the cost of providing that level of output i.e., profit. The sum of the consumers' and producers' surpluses is the social surplus.

If this firm were to set its price equal to marginal cost, output would expand to oh and price would fall to ob . Consumers' surplus would be given by area bdf , and producers' surplus by the area abf , giving a social surplus of adf . This exceeds the previously obtained aggregate by the triangle kef . When price is

set equal to marginal cost, the sum of the two surpluses is at the maximum²². In other words, social welfare is maximized when price is equated with marginal social cost. Traditional theory tells us that such a condition prevails in the long term when perfect competition exists despite the fact that each firm is trying to maximize its own profits i.e., producers' surplus. The ability to exercise any degree of monopoly power, however, permits a firm to set prices above marginal cost, so that it can achieve additional profit at the expense of reduced output and at a cost to the consumer. The price charged by a profit maximizing monopolist will deny some consumers the use of a service, even though they are prepared to pay at least the marginal social costs involved. Indeed, it is the fear of monopoly exploitation that has led to widespread price regulation in many transport sectors by numerous governments.

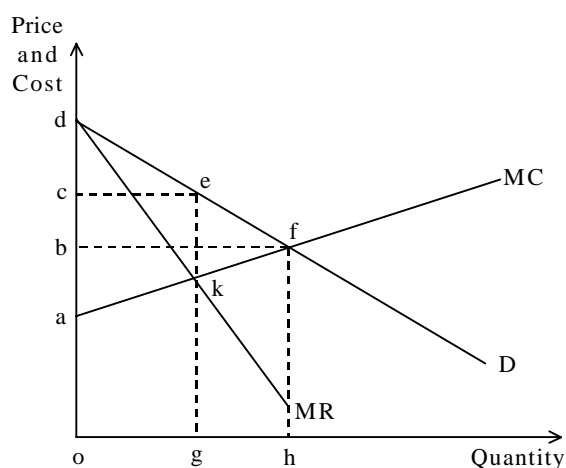
In conclusion, it may be pointed out in a 'first-best world' characterized by perfect competition; no externalities; complete information about future prices, tastes and technology; no subsidies; and no indivisibilities of supply or demand, the market will set prices equal to marginal social cost and thereby maximize social welfare.

Distinction between Long-Run or Short-Run Marginal Costs

In specifying the marginal cost pricing rule, it is important to understand the

²² The assumptions upon which this analysis is made e.g. additivity of consumers' surpluses, may be considered too restrictive; however, a similar pricing rule can be derived from ordinal welfare theory, namely, that prices should be at least proportional to marginal social cost for optimality.

Figure 2.1 : Welfare Maximization



distinction between short-run and long-run marginal costs. The distinction arises because different factors of production, used in providing transport services, have differing degrees of fixedness or variability over various business planning horizons. For example, an airport operator, faced with increasing demand, may be able to increase throughput in the short term by employing more labour, whereas, in the longer term, he may be able to improve productivity through additional capital investment. In the very short term, all inputs and costs are essentially fixed and, conversely, in the long run, all inputs and costs are ultimately variable. Over a particular planning horizon, it is important to identify those costs that can be varied (varied costs) and those which cannot be varied (i.e., fixed costs).

To illustrate the above, consider a transport operator who faces, at any given point in time, two problems:

- (i) to set, at the beginning of year 0, a price that will prevail in that year; and
- (ii) to choose, at the beginning of year 0, an investment programme based on a planned price and output in year 1.

To investigate how this problem may be solved, it is necessary to assume that the operator has some forecasts of demand for years 0 and 1.

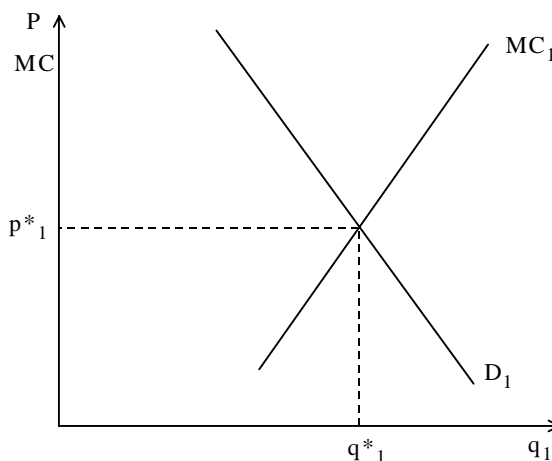
Consider first the analysis for year 1, as depicted in Figure 2.2. D_1 is the demand curve for year 1, and MC_1 is the relevant marginal cost curve, based on the assumption that both capital (K) and labour (L) are variable over that time horizon. The principle of marginal cost pricing implies that planned output should be q^*_1 and planned price p^*_1 , since in that case the market clearing price equals marginal cost. Corresponding to the planned output q^*_1 there will be cost-minimizing input quantities K^*_1 and L^*_1 . Since K_0 is the amount of K available at the beginning of year 0, the enterprise must carry out new investment in year 0, to the extent of $K^*_1 - K_0$.

Consider now Figure 2.3, which analyzes the choice of price and output for year 0. Is the optimal price²³ and quantity determined by the intersection of the demand curve D_0 and MC_1 ?

It should be noted that MC_1 is not relevant to year 0. Only if the level of investment in the previous planning period (t-1), based on a planned output of q_0 , was such that the marginal cost curve in year 0 was MC_0 , would the optimal price and output be p_0 and q_0 , respectively. Alternatively, if the investment in (t-1) had been for a planned output of q'_0 with an associated marginal cost curve of MC'_0 , then social welfare will not be maximized at q'_0 : instead it would be maximized at an output level of q^*_0 with a price of p^*_0 .

The above analysis leads to the important conclusion that marginal cost pricing requires the price, at a given period of time, to equal the marginal cost of varying output within that period of time. Output variations within the current year are constrained by the current level of fixed investment (K), while variations within future years are not. At a given point in time, two sets of decisions have to be taken. Given the demand in the current year, the price and output combination has to be assessed. Given demand projections for subsequent years, future price and output has to be planned, and the corresponding investment programme determined and set in motion. The planned price and output will actually be implemented in the future years, only if the assumptions about demand and costs upon which they are based turn out to be correct.

Figure 2.2 : Long-Run Marginal Cost (year 1)



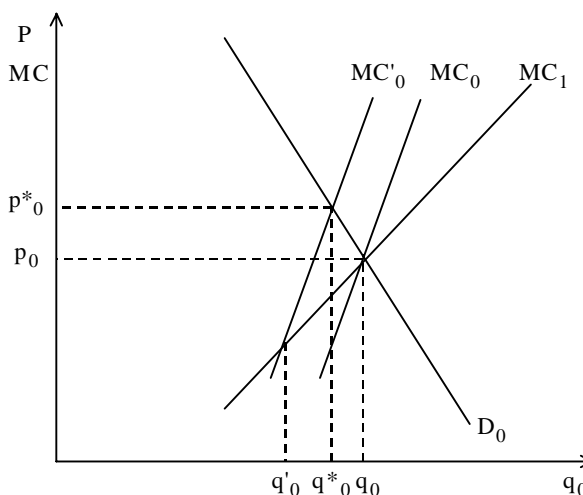
Prices should normally be set in relation to short-run marginal cost, which may lie above, below, or equal to long-run marginal cost.

‘Second-Best’

The conditions necessary for a ‘first-best world’ are rarely found in the real world, which is normally characterized by some, or all, of the following:

- (i) indivisibilities of supply, in the form of short-term fixed capacity constraints;
- (ii) indivisibilities of demand, in the form of peak load problem;

Figure 2.3 : Short-Run Marginal Cost (year 0)



23 Optimum for the long run is not necessarily optimum for the short run.

- (iii) elements of monopoly, instead of perfect competition;
- (iv) externalities, in the form of congestion and pollution; and
- (v) strategic interdependence and uncorrected externalities.

In addition, there is a lack of complete information about future prices, tastes and technology, and subsidies are common. In consequence, it is very unlikely that the market, without regulation, will set transport prices equal to marginal social cost and, therefore, social welfare will not be maximized.

In summary, 'optimal pricing' must balance economic efficiency, equity, and transaction costs. Optimal markets must offer consumers choices and information, minimize distortions, encourage producer competition, and avoid adverse income re-distribution. Investment and pricing are relatively straightforward in a 'first-best' world; however, it is the distortions of the real or 'second best' world that create the challenges in determining an optimal pricing policy²⁴.

Section II

Marginal Cost Pricing in Practice : Fixed Capacity, Peak-Load Pricing, Uniform Pricing, Indivisibilities and Joint Costs

When the principle of marginal cost pricing is applied to the transport sector, it is usually necessary to extend the theory in order to deal with certain industry specific characteristics. In this section, the following are addressed:

- (i) fixed capacity
- (ii) peak-load pricing
- (iii) indivisibilities
- (iv) joint and common costs

Fixed Capacity

In a 'first-best' world, capacity is assumed to be perfectly flexible in the sense that it can be varied, or optimally adjusted, as demand varies at a low variable or marginal cost. In reality, the output or capacity of transport infrastructure is often fixed in the short run²⁵. This means that without investment in extra capacity, there is a physical limitation on output in the short term. This is true, no matter how much variable resource or short-run inputs are applied.

²⁴ Lipsey, R.G. and Lancaster, K., 1959. *The General Theory of Second Best*. Review of Economic Studies.

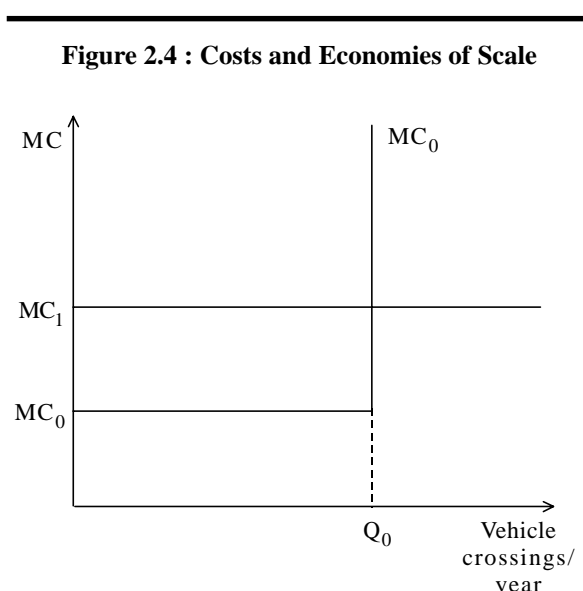
The capital costs of physical transport infrastructure are often much higher than the associated operating and maintenance costs, and can be very long-lasting. Once committed, infrastructure investment usually has few alternative uses and is normally regarded as sunken cost. As a result, infrastructure facilities, such as roads, rail-beds, bridges, piers, runways, locks, harbours, tunnels, etc. give rise to significant economies of scale. In addition, transport vehicles, such as buses and railcars are also subject to scale economies in operation, though they are generally not as expensive as the infrastructure.

Figure 2.4 illustrates the classic pricing problems associated with transport infrastructure, arising from the existence of high fixed costs and significant scale economies (in this example, in building and operating a bridge). The output of the bridge is represented by Q or the number of vehicle crossings per year. Assuming that all costs increase proportionately with traffic flow, MC_0 would measure maintenance and toll collection costs per vehicle crossing and MC_1 would be comprised of this plus the capital cost per vehicle. Before the bridge is built, the capacity flow of vehicles can be varied continuously by varying the size and design of the bridge, but once built, capacity is fixed at Q_0 , with a fixed capital cost of $(MC_1 \text{ minus } MC_0 \text{ multiplied by } Q_0)$.

Thus, when the bridge is being planned, the problem is to decide on the optimal capacity (size and design) and associated price or toll. However, once built, it is simply necessary to decide on the toll.

If tolls are to be set at marginal cost, then a number of possible situations might arise as described in Figures 2.5(a), (b) and (c).

Figure 2.5(a) depicts a position of excess capacity. In this situation, if a toll was set on the basis of short-run marginal cost, then the toll would



25 The actual duration of the “short-run” will depend upon the asset being considered. For example, equipment may be acquired within weeks or months whereas infrastructure may take years to construct.

be MC_0 and output or vehicle crossings per year would be Q_{b0} . The amount of excess capacity would be Q_0 minus Q_{b0} . Setting the price at this level (P_0), however, would result in a large financial loss to the bridge operators, equivalent to the capital costs. This loss will be equal to MC_1 minus MC_0 multiplied by Q_0 .

The justification for setting prices equal to short-run marginal cost under conditions of partial equilibrium is as follows: suppose under these conditions of excess capacity the bridge operators were to set a toll greater than P_0 (MC_0), say, at MC_1 . At such a price, the quantity demanded would fall to Q_{a0} and excess capacity would increase to Q_0 minus Q_{a0} . At this level of demand, the marginal social cost of increased output MC_0 is less than the value that consumers place on the marginal unit of consumption, which is measured by price. Hence, net social benefits are increased by increasing output and utilizing otherwise idle capacity; the capital costs are treated as 'sunken' or as 'bygones' and could be ignored.

The above argument, although acceptable in a 'first-best' world, ignores the fact that the 'policy makers' may have a legitimate interest in the amount of net revenue accruing from the bridge operation. The fact that no interest payments are being made on the bridge investment implies that the government must be servicing the debt issued on the project. In a 'first-best' world, the finance could be raised by lump-sum taxation, and so the deficit could be met in this way – provided that the 'policy makers'

are happy with the resultant redistribution of income. Such a solution would not reduce allocative efficiency, but other welfare concerns related to indirect taxation for non-users and implications for income distribution (over time) are likely to arise.

Figure 2.5(a) : Economies of Scale and Excess Capacity

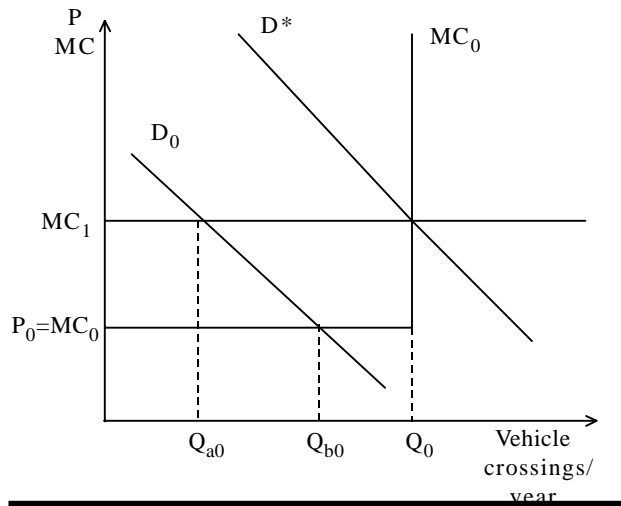
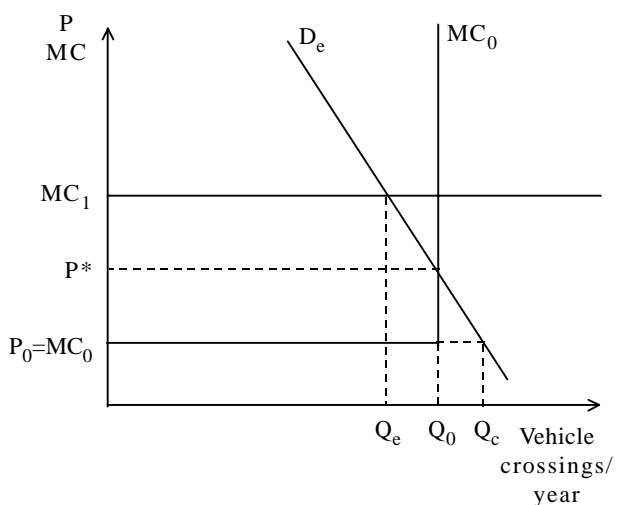


Figure 2.5(b) shows a demand curve that intersects with the vertical portion of MC_0 , the marginal cost curve. A price set at MC_0 would now *imply* excess demand to the extent of Q_c minus Q_0 vehicles per year. If excess demand is to be avoided, a price of P^* will need to be charged. Indeed, whatever price is charged, total utilization cannot exceed Q_0 , or full capacity. The operator must decide how to ration the limited

capacity available amongst users. This could be done either by price or non-price methods. If the toll for crossing the bridge was set at P^* , then the bridge would be used by those who valued it most highly and were prepared to pay at least the marginal costs involved. In addition, using price as the rationing method would maximize social welfare and be relatively inexpensive to administer. If the government or relevant ‘decision-makers’ had concerns about the distribution of income, or the impact of tolls on the poor, the alternative would be to use non-price rationing methods. This means that some users who are not prepared to pay at least the marginal cost would be permitted to use the bridge; that welfare was not maximized; and that resources would need to be devoted to administering the rationing system. In such circumstances, there will be some users who will value the bridge crossing more than other users and there will be scope for ‘gains from trade’ whereby everyone could be made better off, and hence social welfare could be increased. Trade would be worthwhile up to the point where the toll equalled marginal cost, P^* .

Figure 2.5(c) depicts a situation of excess demand where the demand curve D_1 intersects the marginal cost curve on its vertical section but above the level of long-run marginal cost, including the cost of capital. In such cases, the operator will earn profits equivalent to P^{**} minus P_1 ($=MC_1$) multiplied by Q_0 . The

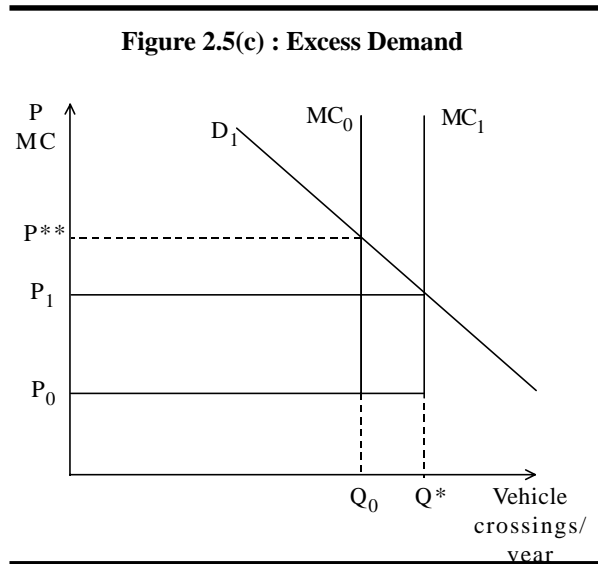
Figure 2.5(b) : Full Capacity Utilization



bridge operator will, therefore, earn excess profits on the capital invested and be provided with an incentive to increase capacity, say to MC_1 . If the investment in additional capacity occurred, then a price of P_1 would equate with both short-run and long-run marginal cost at an annual vehicle flow of Q^* . The total revenue collected would cover the total costs of building and operating the bridge, including the capital charges.

The above analysis leads to the important observation that under fixed capacity conditions with constant marginal costs, setting prices at marginal cost always leads to losses when demand has been overestimated and to profits when it has been underestimated. If capacity exactly matches the demand, then neither abnormal profits nor losses will occur²⁶.

Figure 2.5(c) can also be used to analyze the planned price and capacity decision for year 1. If D_1 is the demand curve which the bridge operator expects in a subsequent period e.g. year 1, and the relevant marginal cost curve for year 1 is MC_1 – it follows that the optimal price will be P_1 associated with the intersection of MC_1 and D_1 . In order to increase output from Q_0 to Q^* , the bridge operator will need to undertake the necessary investment programme in year 0. Then, when year 1 arrives, the bridge operator will have to set price according to the actual level of demand prevailing at that time, as per the analysis in Figures 2.5(a), (b) and (c).



It is possible now to re-state the above analysis. Suppose it is year 1 and the bridge has not yet been built. If the bridge operators estimate the demand curve, in year 0, to be D^* as in Figure 2.5(a), then the toll will be set at MC_1 and the vehicle crossings per year will be Q_0 . The toll will be equivalent to long-run marginal cost and the bridge will break even, in that, total revenue will equal total costs, including capital charges. Long-run and short-run marginal costs will be identical at this level of output, and the bridge's capacity will be fully utilized.

If the demand forecasts are right than the pricing plan in year 0 and in each subsequent year, whereby price was set at marginal cost, then demand would just match the physical capacity available. If demand has been seriously overestimated, as shown by D_0 in Figure 2.5(a), then the toll will be set at P_0 and the variable costs of maintenance and toll-collection will be covered, but there will be no contribution made to the fixed capital charges of building the bridge. If demand was less seriously overestimated, as shown by D_e in Figure 2.5(b), then the toll will be set so as to ration demand to capacity and revenues will make a partial contribution to capital costs. If demand was underestimated, as in Figure 2.5(c), the toll that rations demand to capacity (avoids traffic jams at the approaches to the bridge) will lie above long-run marginal cost and

²⁶ Nelson, J. R., *Marginal Cost Pricing in Practice*. Prentice-Hall. New Jersey. 1964.

will yield profit. In this case, if demand is expected to remain at this level, it is necessary to expand the capacity of the bridge for subsequent years.

Now, consider the financial and political consequences of marginal cost pricing for the bridge. If demand turns out to have been overestimated, as in Figures 2.5(a) and (b), then capital charges will not be covered. The bridge operators will still have met their 'financial obligations' in respect of the bridge, unless the government intends to subsidize the capital costs of the bridge. If this cannot be achieved through lump-sum taxes, then it will be necessary for bridge users to pay tolls in excess of short-run marginal cost by means to be discussed later.

On the other hand, consider the case shown in Figure 2.5(c) where demand has been underestimated. Although an expansion of capacity is being planned, this may take quite some time to complete, and in the meantime bridge tolls and profits will be high. Users of the bridge and their political representatives may well protest. It will be argued that low-income families cannot afford to use the bridge with adverse consequences for personal mobility. There will be thus great pressure to restrict the toll, possibly to the break-even level. The consequence of this will be excess demand, congestion and queuing, with greatly increased time and petrol costs for bridge users. A better solution would be to identify 'the poor' and use some of the profits to subsidize them directly. Nevertheless, the economic inefficiency of non-price rationing might be judged to be an acceptable price to pay to ward off the political clamour or to avoid the income-distributional consequences of a market-clearing bridge toll.

Peak-Load Pricing

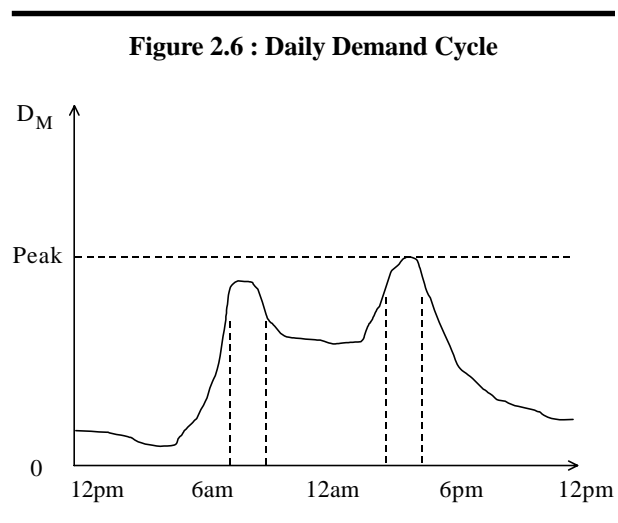
The above analysis has worked with uniform demand spread over an arbitrarily chosen time period of one year, over which a uniform price has been charged to all users. Many parts of the transport industry, however, experience a systematic pattern of demand fluctuations within a given period, the pattern repeating itself from period to period. The duration of the fluctuations is too short to permit capacity to be varied to match them, while it is too expensive to store spare capacity to meet the requirements of the peak demand period. In many transport sectors, the demand cycle usually completes itself within a day, whereas it may take more than a year to increase capacity. Figure 2.6 shows the kind of daily demand pattern that may be faced by bridge, rail or bus operators. Demand is plotted, minute-by-minute, against each successive minute of the day and displays typical peaks and troughs. The problem is that demand at all minutes is met from the same installed capacity, and so, there are corresponding fluctuations in capacity utilization.

If sufficient capacity is provided to ‘meet the peaks’, then the rest of the time varying amounts of it will be lying idle. On the other hand, the demand at each minute will depend on the price that prevails in that minute, as well as the prices set for other minutes. In general, therefore, pricing policy could be used to flatten the peaks and raise the troughs, so as to get a more even rate of capacity utilization and a lower level of required capacity. The indivisibilities of demand, or the peak-load pricing problem, is then that of determining an overall capacity level and optimal values for a sequence of prices over the demand cycle²⁷. Alternatively, the problem may be regarded as one of applying marginal cost pricing to a system with fluctuating demand.

It is now possible to develop the methodology mentioned earlier to provide a solution to the peak-load problem. It is, of course, impracticable to implement a system where, for example, rail fares vary by the minute. However, the day can be divided into relatively few discrete time-periods, within which uniform prices may be set. This suggests that the peak-load pricing problem consists of two interdependent parts:

- (i) To identify the optimal number of pricing periods, given the costs of introducing pricing structures, which are too complex to understand or administer.
- (ii) To determine the optimal total capacity and uniform prices within these periods.

Continuing with our example of the bridge, Figure 2.7 now shows a day that consists of peak and off-peak demand periods comprising two twelve-hour periods within which demand is uniform. Assuming negligible costs of administering a simple two-period price differentiation, there is no need to worry about determining the optimal structure, since the demand pattern itself suggests a two-period structure – it is only necessary to determine the price to set in each period.



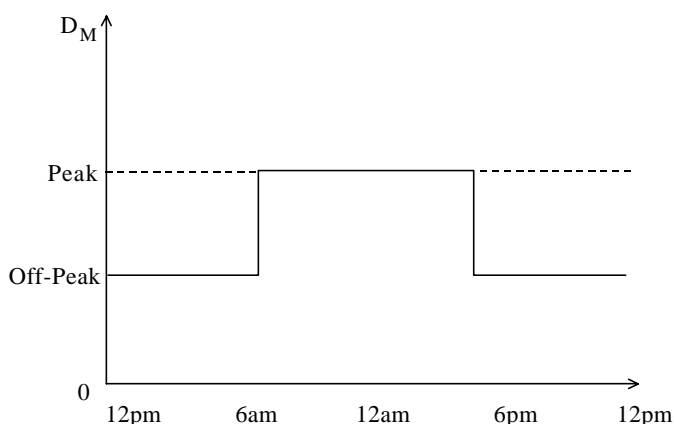
Analysis of this problem, depicted in Figure 2.8, brings out the main elements of peak-load pricing problems, and suggests immediate applications. Q_1 is the total number

²⁷ Boiteux, M., 1960. *Peak-Load Pricing*. Journal of Business.

of vehicle crossings in the 12-hour period with lower demand and Q_2 is that produced in the 12-hour period with higher demand, assuming that the bridge has a fixed capacity and that marginal costs are constant. The problem faced by the bridge operator is similar to that considered in the previous section, with the added dimensions of price differentiation between outputs Q_1 and Q_2 . At the first instant of year 0, he must set prices for Q_1 and Q_2 , which will hold for each respective 12-hour period throughout the year, given the available 12-hour capacity $Q_0 (= Q_2)$. It must also plan the prices and outputs for year 1, in the light of its expectations about 12-hourly demands in that year, and given that it can regard capacity as variable for that year. These planned prices and outputs will then determine an investment programme, if any, to be carried out in year 0.

Figure 2.8 shows the situation in year 0. Capacity is fixed at Q_0 , and D_1 and D_2 are the demand curves for the two periods. On marginal cost pricing principles, the price in period 1 (off-peak) should be set at P^*_1 , implying an output of Q^*_1 , which is below capacity. If it is accepted that capacity should be rationed by price, then price in period 2 (peak) should be P^*_2 , since the demand, at point a on D_2 , is equal to capacity Q_0 . The fact that D_2 is higher than D_1 implies that, the price at peak periods should be higher than

Figure 2.7 : Peak and Off-Peak Demand



at off-peak times, unless D_2 and D_1 both intersected the horizontal section of MC_0 , in which case, the peak and off-peak prices would have been P^*_1 . This would have implied significant excess capacity even during the peak period each day.

The situation depicted in Figure 2.8 is one in which the bridge operators make a financial loss each day. The cost including the capital costs each day of operating the bridge, is $CQ^*_1 + CQ_0$ whereas the daily revenue is $P^*_1Q^*_1 + P^*_2Q_0$. The net loss (L) is:

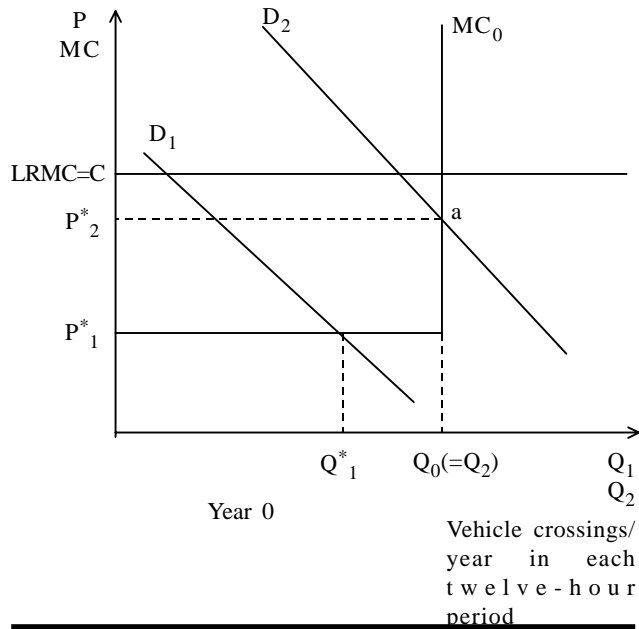
$$L = CQ^*_1 + CQ_0 - P^*_1Q^*_1 - P^*_2Q_0$$

Or

$$L = Q^*_1(C - P^*_1) + Q_0(C - P^*_2)$$

In other words, off-peak users simply cover their running costs and make no contribution to capacity costs, while peak users make a contribution to the capacity costs of the difference between the price they pay and the running costs. These losses are due to past overestimation of demand. It could have been the case that a profit would have been made if the peak demand curve had intersected the short-run marginal cost curve above the level of long-run marginal cost. Thus, the losses or profits arising from marginal cost pricing are due to previous investment decisions which were, in turn, made on the basis of demand predictions which may have been over-estimates or under-estimates²⁸.

Figure 2.8 : Peak and Off-Peak Pricing



Now, it is necessary to determine the capacity and planned prices for year 1. The important concept to remember here is that of the users' willingness to pay for capacity. At any given level of output, price can be regarded as the users' valuation of the marginal unit of output. Users willing to pay a particular price for a transport service must implicitly assess its value as equal to, or greater than, that price. If we deduct from this price the variable costs of providing the transport service, then the remainder must represent the user's willingness to pay for the capital costs of the capacity provided. In Figure 2.8, off-peak users (D_1), at price P^*_1 and output Q^*_1 , are not willing to make any contribution to capacity costs. On the other hand, peak users (D_2) are prepared to make a contribution of $Q_0(P^*_2 - P^*_1)$. Now, a unit of capacity is in place and will remain in place and capable of providing output throughout the entire day, i.e., it is capable of providing user benefits for both periods. It follows that it would be worth installing extra capacity if the sum of user's willingness to pay for that unit of capacity over both periods exceeded its cost. The optimal capacity level is that at which this sum of marginal willingness to pay is just equal to marginal capacity cost.

The situation is similar to that found in the case of *public goods*, since consumption by users in one period (e.g. the peak) does not affect the benefits enjoyed by users in another period (e.g. the off-peak). Consumption is, however, ‘rivalrous’ within each period. The situation is also similar to that of *joint products*. By rearing sheep both mutton and wool are produced, and the value of the sheep is the sum of the value of the wool and mutton. Similarly, a unit of transport capacity is available to provide a service in both peak and off-peak periods, and so, its value is the sum of the benefits to both peak and off-peak users.

The central principle is that the optimal level of capacity is that which equates the sum of the marginal willingness to pay across periods with the marginal capacity cost. From this, it is possible to derive the marginal cost pricing policy. In Figure 2.9, in the lower part (b), the demand curves for the two periods are shown as D_1 and D_2 and the marginal operating cost is shown by C^* . Now, if C^* is subtracted from each price along D_1 and D_2 a ‘willingness-to-pay-for-capacity’ curve is obtained. This is the portion of both curves lying above C^* . These two demand curves can be summed to obtain D_K in the upper Figure 2.9(a). The kink in this total-demand-for-capacity curve occurs at point e , because D_1 cuts C^* at point a – to the right of point a , D_K simply measures $P_2 - C^*$.

Suppose the marginal (daily i.e., peak plus off-peak periods) capital cost is $2z$, as shown in Figure 2.9(a), then optimality occurs at point f , implying an optimal capacity of Q_0 . Then, optimal prices are:

$$\text{For } P_1 = C^* \text{ and for } P_2 = C^* + 2z$$

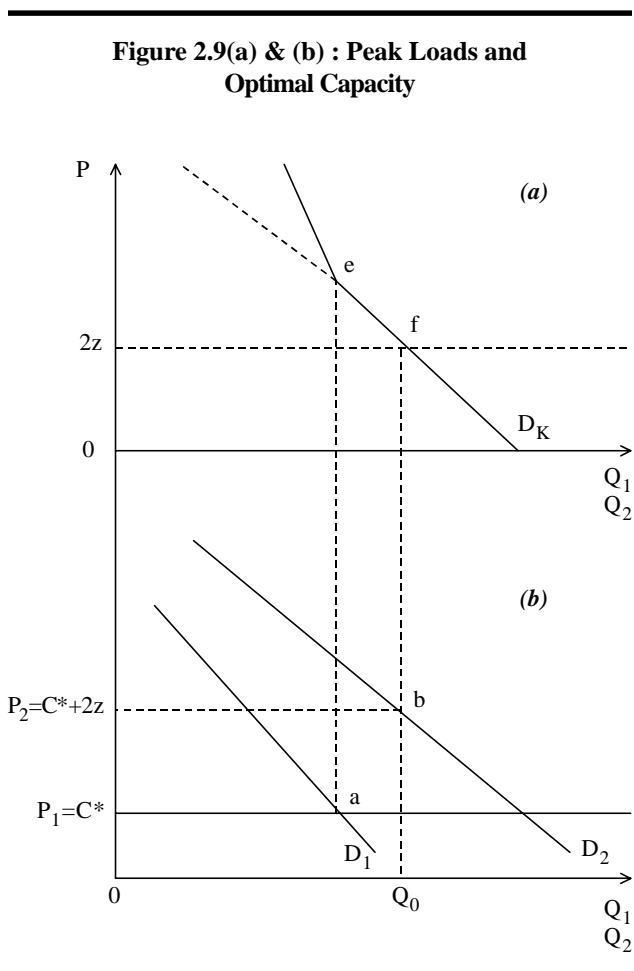
This means that the peak traffic pays all the capacity costs. The reason for this is that, in the off-peak period, demand is below capacity and, therefore, the marginal cost imposed by variations in use during this period is only C^* . On the other hand, an increment to capacity in the peak period can only be met by expanding capacity; so, the marginal cost in this period is $C^* + 2z$ and this is the peak-load price.

The main implication of this analysis is that off-peak users pay only operating costs and make no contribution to capacity or capital costs, while peak users pay all the capacity costs. This may be regarded as ‘unfair’. After all, since off-peak users are also benefiting from the capacity, why should they not pay something towards it? The

²⁸ Williamson, O. E., *Peak-Load Pricing and Optimal Capacity under Indivisibility*. American Economic Review. 1966.

answer is that what matters are the costs and benefits at the margin. Beyond point e , in Figure 2.9(a), further increases in capacity yield no net benefit to off-peak users and they are not prepared to pay for them. These increments are made – in spite of the fact that the capacity lies idle for half the day – because peak users are willing to pay the entire cost. Under these circumstances, to ensure that the optimal capacity is provided, it is necessary to charge peak users the full marginal cost. To charge off-peak users in excess of the operating cost, C^* , would lead to a loss of welfare, since the value of the transport service to them would exceed the cost of providing the service, which is C^* when capacity is fixed at Q_0 . It should also be noted that in this example the operator would also break even in financial terms by charging at marginal cost, since total revenues would cover total costs, including capacity costs. This is essentially a result of the constant marginal costs assumption. However, if marginal costs were increasing, the operator would make a profit; if they were decreasing, the operator would make a loss.

In the situation analyzed so far, the ‘peak-load’ phenomenon does not disappear. Off-peak use is still less than peak usage, so the load curve would still have the general shape in Figure 2.7. Thus, it should be emphasized that ‘flattening the load curve’ is not an end in itself. Rather, as a result of optimal pricing, an optimal load curve can be obtained, though this will not necessarily be flat with constant use at all times of the day. To attempt to further flatten the load curve would lead to welfare losses. However, it is possible that if marginal capital costs were much higher, at say $2z^*$ in Figure 2.10, then the optimal pricing policy may imply a flat load curve, that is, 100% utilization at all times. Note that peak and off-peak usage will be Q^* , but the prices for peak and off-peak users will be very different, that is, C^*+2z^* and C^* ,



respectively. This implies that the optimal transport capacity is Q^* and that the operator will break even. However, there is an excess demand for off-peak at price C^* .

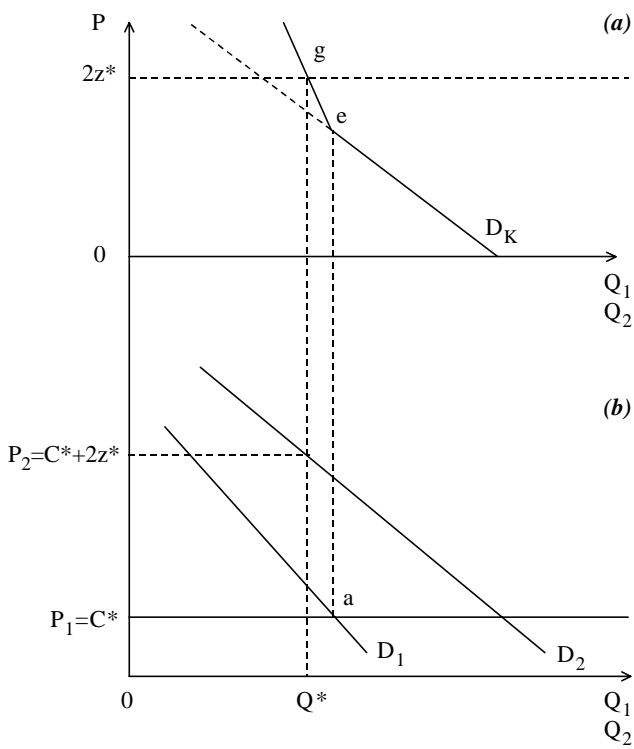
This situation is referred to as that of the 'shifting peak'. The difference between peak and off-peak usage with marginal cost pricing will depend upon the size of marginal capacity costs and the elasticity of demand in the two periods.

An important problem which arises when implementing peak-load pricing is that of trying to determine the optimal number of charging periods within which a uniform price will be charged. To have a large number of different charging periods would be prohibitively costly to implement. It is still necessary, however, to 'decompose' the load curve into a number of charging periods. In simple terms, the optimal structure occurs when the costs of adding a further charging period exceed the benefits from doing so. In addition to any welfare losses arising from uniform pricing, there are the administrative and informational costs to be accounted for. In practice, it is unlikely that more than two or three pricing periods will be justified, depending on the available technology and the relative costs of introducing price changes.

Indivisibilities

Problems arise when applying marginal cost pricing to transport infrastructure and services, because capacity can only be increased in relatively large indivisible units. Examples abound: if the capacity of a railway coach is 'x' passengers, then to carry 'x' plus one more passenger requires another coach; to increase the capacity of a ferry service requires another ship; if existing airports are at capacity, expansion requires a new runway and terminal facilities, and so on. In many cases, it is extremely costly to make small additions to

Figure 2.10(a) & (b) : Optimal Pricing and Flat Load Curves



physical capacity. The question of the appropriate scale of operation, therefore, becomes an economic problem.

Figure 2.11 depicts the situation, again for the bridge operators, with no peak-load demand problems but with indivisibilities in provision of capacity. At the beginning of year 1 the bridge operators must forecast a demand curve for year 1; choose a planned price, output and capacity; and determine its investment programme. D_1 is the estimated demand for year 1, and MC_0 and MC_1 are the current marginal cost curve and the marginal cost curve if capacity was increased by the smallest feasible amount in year 1. The operators can choose whether or not to invest in extra capacity in year 1 – the implications of these two options are shown in Figure 2.11.

Firstly, if capacity was not increased to MC_1 , then, in year 1, usage of the bridge would be limited to Q_0 and the toll would be set at P^*_1 . Now, it is possible to consider the implications of increasing capacity to MC_1 for the bridge users (Note LRMC includes both the running costs and the incremental cost of providing extra capacity). The level of usage will increase to Q^*_1 and the toll should be set at C^* , equivalent to the bridge running costs. The extra benefit, net of the toll to bridge users is area cda . On the other hand, the cost of providing the new capacity is area $efgd$. However, since area $ebad$ is common to both measures, it is necessary to compare areas ceb and $bfga$. Since the latter is greater than the former in Figure 2.11, it can be concluded that it is not worth increasing the capacity of the bridge just yet, because the costs exceed the benefits of so doing.

The issue is essentially one of optimal investment timing since, under conditions of growing demand there will come a point at which an increase in capacity will be worthwhile. Figure 2.12 shows a situation in which there is expected to be a higher level of demand such that the costs of increasing the capacity of the bridge equals the net benefits to users. Note that the areas ceb and bfg are now equal.

This analysis implies that marginal cost pricing could produce marked fluctuations in price before and after capacity adjustments are made. Further, whether or not the bridge makes a profit depends on whether the price lies above or below the long-run marginal cost curve, LRMC. At any given point in time, the bridge might be profit or loss making; however, over its life, the net present value of the investment in additional capacity should be positive if it is to be worthwhile.

Such fluctuations in prices and profits are likely to be undesirable; however, any other pricing pattern will produce welfare losses. If the price floor is raised above marginal running costs during times of excess capacity, then underutilization will occur. If a price

ceiling is set, then, during periods of excess demand, non-price rationing methods will be required. Similarly, premature investment in capacity is likely to represent a waste of capital resources; it will, however, keep marginal cost based prices low.

Joint and Common Costs

The conceptual and practical problems of determining transport prices associated with fixed and variable costs and choosing the relevant time period are often compounded by the fact that many costs may also be ‘joint’ or ‘common’ to a number of users.

Joint costs exist when the provision of a specific service necessarily entails the output of some other service. Jointness is a technical feature and exists at all points in time, that is, both before and after investment or capacity decisions are made²⁹. The classic example of jointness is the return trip or ‘back-haul’, where the supply of a transport service in one direction normally implies the provision of a return service. Joint costs are not allocable to specific traffics and can only be escaped jointly when a service is withdrawn. If the service is withdrawn for one direction, it is automatically withdrawn for the reverse direction.

In a market situation, joint

Figure 2.11 : Indivisibilities

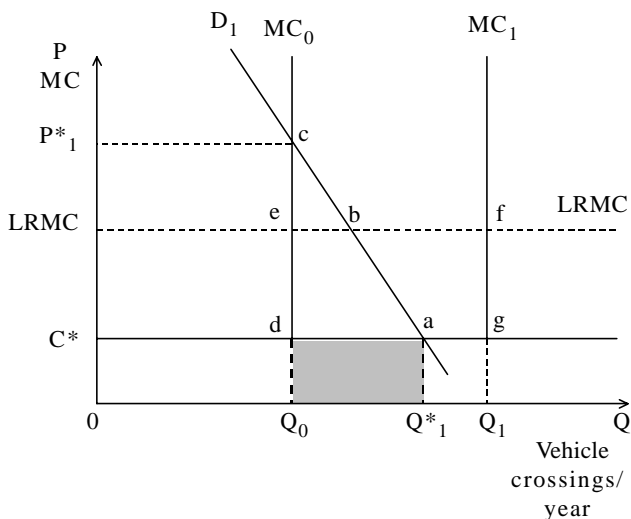
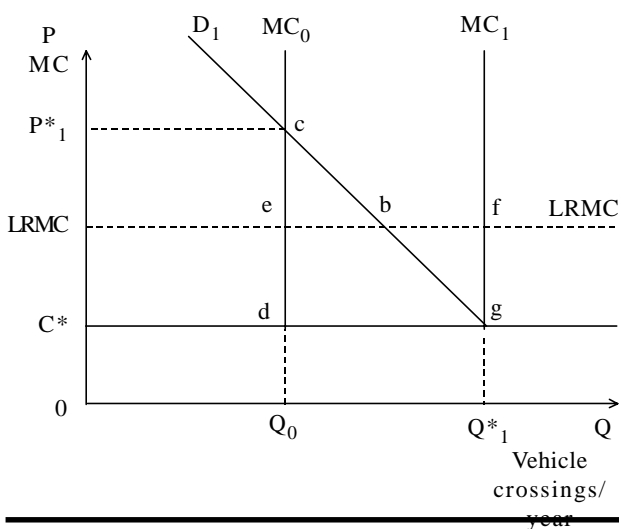


Figure 2.12 : Indivisibilities and Investment



costs pose a few problems in practice³⁰. If there is a competitive road haulage service offering a round trip between *A* and *B* and back again each week using *M* trucks, then equilibrium rates would soon emerge for each service (i.e., from *A* to *B* and *B* to *A*). Although there will be specific delivery, terminal and pick-up costs, there would be little difference between the costs of running the trucks fully loaded or empty and hence prices would primarily be determined by the differences in the demand in each direction. In the short term, the combined revenues, from *A* to *B* and *B* to *A* services, may not be sufficient to cover joint costs, but in such a situation, the number of trucks provided would soon be reduced below *M*, increasing the price of trips in both directions until joint costs are covered. Excess revenue above joint costs would have the opposite effect. The key point is that if differences exist in the demands for the out and return services, then different prices should be charged at each service equilibrium.

Common costs are similar to joint costs, in that they are incurred as a result of providing services to a range of users, but differ, in that the resources used to provide one service do not unavoidably result in the production of other services. Track costs are a classic example of common costs. A road may be used by both cars and trucks, but if use of the road is terminated for trucks, the track costs have still to be borne by the car users. A proportion of track costs may be allocable to certain users, but a large proportion may not be.

It is sometimes said that all fixed costs may be regarded as joint or common costs, but it should be noted that many variable costs may also be joint or common. For instance, fuel costs in the case of the back-haul truck trip are joint and road maintenance costs are, to an extent, common to all users. Indeed, some fixed costs are specific; for example, railway freight wagons are not relevant to passenger traffic.

In summary, cost structures in the transport industry display a number of features that have to be accounted for, including, high fixed and joint costs combined with the peak-load problem that lead to variations in marginal cost, both over time and between different services and users. Such variations have to be identified and accounted for in designing an optimal pricing policy.

Section III

Marginal Cost Pricing in Practice: Monopoly, Externalities, Strategic Interdependence and Uncorrected Externalities

²⁹ Littlechild, S., 1970. *Marginal Cost Pricing with Joint Costs*. The Economic Journal.

³⁰ Mohring, H., 1976. *Transportation Economics*. Cambridge, Mass. Ballinger.

The application of marginal cost pricing to the transport industry raises particular issues due to the ‘second-best’ nature of the real world. This section examines the three main common types of market failure in the transport industry, namely, those of: (i) monopoly; (ii) externalities; and (iii) strategic interdependencies and uncorrected externalities.

Monopoly

An efficient market usually requires that producers face competition so that they encourage technical efficiency and innovation. It is often suggested that the existence of high fixed costs and economies of scale, together with significant indivisibilities in the provision of capacity, have limited competition and give rise to numerous monopolies in the transport industry – particularly in terms of infrastructure. For example, most public transit (and some taxi) services are legal monopolies. Constraints on competition are intended to avoid certain problems, but this tends to increase costs and does not provide incentive for the development of alternatives, such as demand response transport, premium-service commuter buses, or other innovative and niche services. It is often argued, therefore, that a more competitive transport market may result in benefits to consumers.

If the transport market is fairly small relative to the optimal size of capacity, then a ‘natural monopoly’ is likely to emerge, since the market is unlikely to be able to support more than one optimally-sized firm. The existence of monopoly creates a major distortion in the market for transport services. There is every risk that the unregulated market will not provide optimal transport prices and, therefore, will not lead to the maximization of social welfare or facilitate sustainable development. In such circumstances, government often intervenes by either directly providing transport services or by regulating the prices of private operators.

In a competitive environment with many providers of transport services, prices will normally equate to marginal cost. However, this is unlikely to be the case under conditions of monopoly. Figure 2.13 depicts how price and output would be determined under conditions of monopoly and under competition within a transport industry, subject to declining costs.

The monopolist aiming to maximize profit would find it worthwhile expanding output up to the point at which marginal revenue was equal to marginal cost – point *a* in Figure 2.13. At this level of output Q_M , the monopolist could charge a price of P_M and supply the market at an average cost of C_M . The monopolist would make a profit of $Q_M (P_M$

- C_M). Under conditions of perfect competition, price is determined freely in the market place by the forces of supply and demand. In Figure 2.13, if the aggregate marginal cost curves, are assumed to represent the industry supply curve then equilibrium occurs at point *b* with output Q_C . At this point the firms in the industry will make a financial loss equivalent to $Q_C (C_C - P_C)$. However, welfare will be increased to the extent of the increase in consumers' surplus net of cost.

The existence of declining costs in the transport industry is an important reason for the emergence of natural monopolies in many sectors. Concern over the potential abuse of monopoly power in setting price and output has often led to government price regulation and public ownership. Governments have had to provide guidance to publicly owned transport undertakings in respect of their pricing and investment policies.

The existence of monopoly implies that the market mechanism will not achieve, without regulation, economic efficiency and welfare maximization.

Externalities

The essence of an externality is that it involves (i) interdependence between two or more economic agents, and (ii) failure to price that interdependence. Externalities can be positive (benefits), or negative (costs). Transport generates many negative externalities, including noise, accidents, pollution, and congestion. If such costs are not borne by those who generate them, then the market mechanism again fails to allocate resources efficiently. It should be emphasized that optimality does not imply the total elimination of congestion and pollution, but rather the achievement of optimal levels of external cost. Ideally, externalities should be contained to the point where the costs of further reductions exceed the marginal social benefits. Let us analyze the problem of externalities by examining pollution and congestion in general terms.

Pollution – the 'Polluter Pays' Principle

Figure 2.14 depicts the theory of optimum pollution charges in respect of noise pollution – the general principles can be applied equally well to other forms of pollution, such as air or water pollution. In Figure 2.14, MPC is the marginal private cost of transporting goods (in tons) by truck in an area and includes labour, fuel, maintenance costs, and so on. MEC is the marginal environmental cost, representing the money value of the marginal noise nuisance at each level of traffic and D is the demand curve for road haulage services. MSC is the marginal social cost curve and represents the summation of MPC and MEC. If road hauliers take account of only their private running costs, then equilibrium output will be Q_A and price or cost will be P_A . However, if all costs, private and social are taken into account, then Q^* is the optimal level of output and the price or cost is P^* . At Q^* the marginal benefit is equal to the marginal social cost of road

haulage. The ‘polluter pays’ principle³¹ suggests that the hauliers should be made aware of the external costs they generate by their paying a pollution tax equal to the MEC (i.e., t at the optimal level of output). This would reduce the amount carried to the socially optimal level. In practical terms this may mean some goods will be transferred to other modes or that industrial relocation may occur reducing the total transport input.

While it may seem to be impracticable to impose pollution taxes directly on users, a number of indirect methods have been

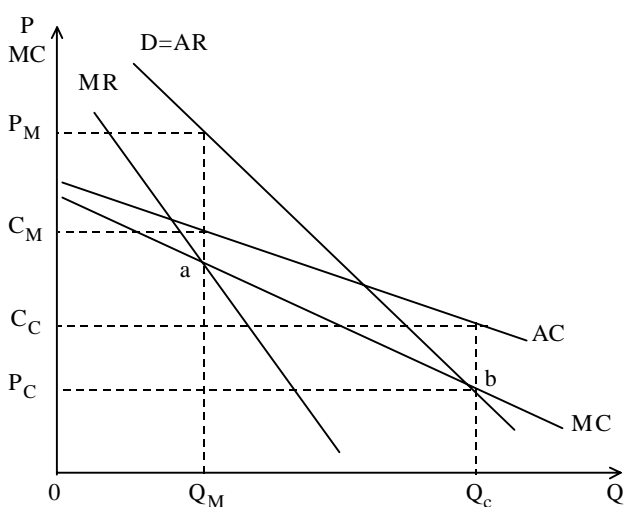
proposed, including pollution taxes on vehicle acquisition or on annual road fund licences. While such approaches do not offer ideal ‘polluter pays’ schemes – in particular, the tax does not relate to the actual use made of a vehicle – they do offer pragmatic devices for pricing pollution externalities.

Some economists argue that it would be preferable to auction licences that actually grant the ‘right to pollute’ or to grant subsidies to hauliers, to the extent of t in Figure 2.14, as an incentive not to pollute. The latter is often preferred on the basis that pollution taxes are regressive and adversely affect those with lower incomes, whereas subsidies which are paid from general taxation are essentially progressive in terms of income distribution.

Congestion

It is not only in the context of pollution that externality pricing has been advocated. One idea for optimising the level of congestion is to use the price mechanism to make road users more fully aware of the impedance they impose upon one another. The idea was suggested first by Pigou³² in 1920. He proposed that, motorists should pay for the additional congestion they create, when entering a congested road. Ideally, as with pollution charges, they should pay the actual road users affected but practically this is impossible, so the idea is that the relevant road authority or agency should be made

Figure 2.13 : Price Determination under Monopoly vs Perfect Competition



responsible for collecting the charges.

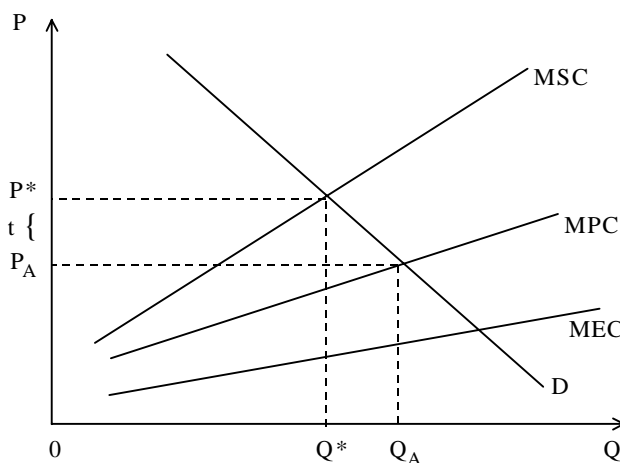
Speed-flow studies show that as vehicle numbers entering a road increase, average speed at first does not significantly change as does flow, measured by vehicles per hour. However, as vehicle numbers continue to increase, a certain point is reached at which average flow and speed start to fall. Speed-flow curves can be used to measure the economic costs of congestion. Generalized travel

costs include the motoring costs plus the travel time cost for a particular journey. Broadly, faster travel in urban areas means cheaper travel in terms of generalized costs – vehicles are used more effectively and travel times are reduced. The average generalized cost of a trip will increase as flow increases up to the maximum flow rate. It will continue to rise as congestion sets in and produces a reduction in the rate of vehicle flow. In Figure 2.15, the average and marginal generalized costs of different flows are shown – indicating all the time and money costs (social costs) borne by road users when trip-making. However, individual users entering the road will only consider the costs they personally bear, but not the external (congestion) costs they impose on the other road users. Consequently, the individual motorist will only consider the average costs experienced by road users and take no account of the impact of their trip on other vehicles. It is frequently argued that the MC curve, therefore, relates to the marginal social cost for the new trip maker and existing road users of an addition to the traffic flow, while the AC curve is the equivalent to the marginal private cost curve – i.e., the additional cost borne and perceived by the new trip maker alone. The difference between the AC and the MC curves at any traffic flow reflects the economic costs of congestion at that flow.

The fact that potential trip makers tend to consider only the private costs of any trip and ignore the external or social costs means that effectively AC is the decision-making curve for private motorists. In Figure 2.15, if the demand curve is D, then road utilization will be Q_A , at which marginal private benefit equals marginal private cost, with

³¹ Pearce, D.W., 1976. *Environmental Economics*. Longman. London.

Figure 2.14 : The Optimum Pollution Charge



an associated cost of P_1 . From a social point of view this is excessive, since at traffic flow Q_1 , the marginal social cost, including congestion cost, is much higher at P_3 . In fact, the socially optimal level of traffic flow is Q_2 with an associated cost of P_2 . At this point, marginal benefits and marginal social costs are equal. To internalize the externality or external cost of congestion, it would be necessary to impose a congestion tax of $P_2 - P^*$.

According to Sharp³³, while the basic theory of congestion pricing is relatively straightforward, its detailed implementation can be problematic. He identifies several reasons for the same including difficulties in devising workable methods for charging and the 'second best' nature of issues.

Strategic Interdependence and Uncorrected Externalities

The transport industry abounds in situations where particular modes or services compete with other modes that do not charge prices at marginal social cost. There could be any number of reasons for this, including the granting of subsidies for failure to correct the externalities. For instance, it is common for railway operators to have to finance both their operating costs and capital costs of land and track, whereas road users rarely pay the full costs of the road infrastructure. To illustrate the latter, consider Figure 2.16(a) where the price of a rail trip from X to Y is P_r . Now suppose, in Figure 2.16(b), that the private cost of a road trip is P_c and that there is no scope to introduce a road tax to cover the marginal social costs of road travel. Therefore, usage would be Q_c at which demand equals marginal private cost AC_c . However, Figure 2.16(b) shows that the optimal usage of the road would be Q_c^* , at which level marginal benefit (demand) would equal marginal social cost, MC_c . Since the optimal level of use cannot be brought about through road pricing it is necessary to consider reducing the price of rail fares from P_r to P^* . This has two effects: firstly, rail usage will increase from Q_r to Q^* ; and, secondly, the demand curve for road usage will fall from D to D^* . It implies that road traffic will be diverted to rail. On the basis of decision-making using private marginal costs the usage level falls from Q_c to Q_c^* - the desired level. At the same time, the price of the road transport is also expected to come down from P_c to P_c^* . The rail price reduction might be achieved either by regulation or by subsidy. Although the desired modal split between road and rail usage is now achieved, it is not as satisfactory as road pricing since the costs of the resources used in providing rail services exceeds their value to users.

Pricing for uncorrected externalities is another example of how to deal with the

³² Pigou, A.C., 1920. *The Economics of Welfare*. Macmillan. London.

problems of a 'second-best' world in transport. International agreements on standards may, however, be required to the extent that environmental impacts have cross-border or possibly global consequences.

In summary, the internalization of externalities is a fundamental requirement in devising transport pricing policies to promote sustainable development.

Section IV Profitability, Taxation and Income Distribution

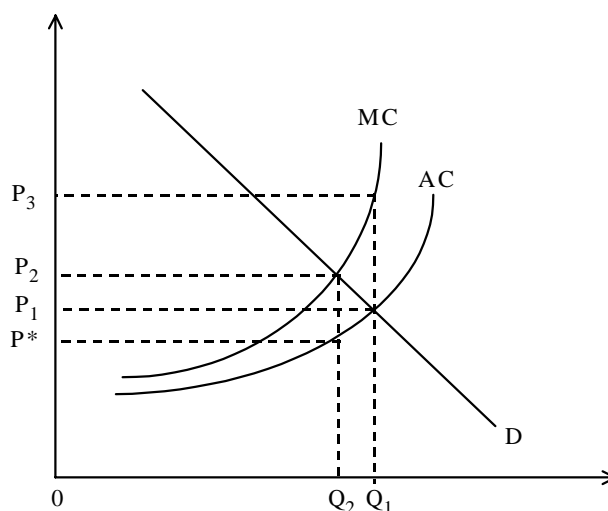
So far, the analysis has been concerned with establishing pricing policies that will achieve allocative efficiency or welfare maximization. The implications of any changes in the transport pricing policy for income distribution have been ignored. The analysis has also assumed that whether the undertaking made a profit or loss was irrelevant to determining the optimal pricing policy. In reality, policy makers are greatly concerned about the distribution of income. Further, unless they can effect income redistribution and finance losses through lump-sum transfers that leave the marginal conditions unchanged, they need to give full consideration to such matters. Indeed, optimal pricing policies must balance efficiency with equity and financial viability.

This section deals with the profitability objective, analyzes pricing methods which may achieve a given profit target, and also examines how explicit equity requirements can be satisfied.

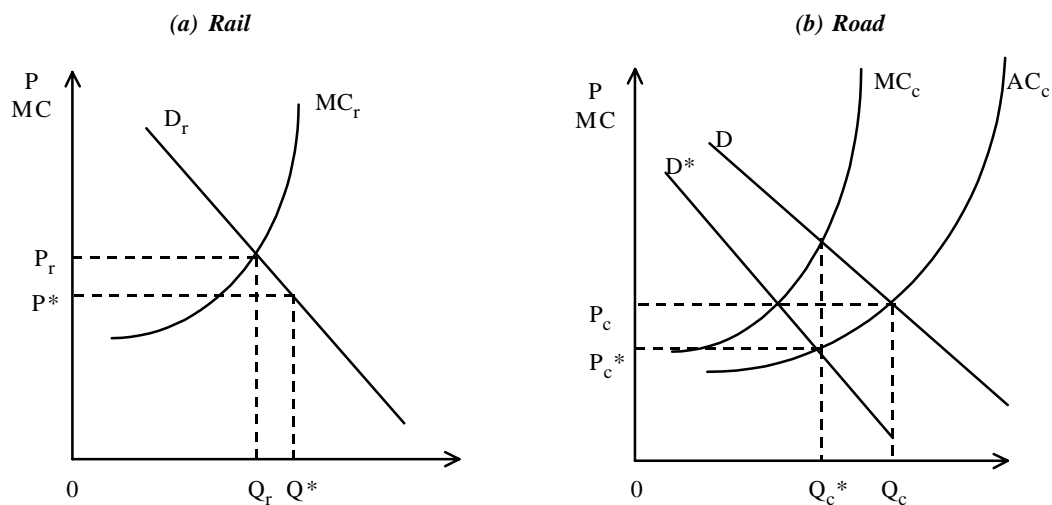
Profit Targets

In the context of publicly owned transport undertakings, prices relate directly to a government's fiscal concerns, since gross trading surpluses are an important source

Figure 2.15 : Optimal Traffic Flows and Congestion Pricing



³³ Sharp, C. H., 1966. *Congestion and Welfare: An Examination of the Case for a Congestion Tax*. *Economic Journal*. Vol.76. pp 806-17.

Figure 2.16(a) & (b) : Pricing with Uncorrected Externalities

of government revenue and public enterprise investment constitutes a heavy claim on public expenditure. Thus, the surpluses that publicly owned transport undertakings generate are often viewed as a form of indirect taxation. In addition, the services offered are often perceived as ‘public utilities’ with benefits for the distribution of income. Further, the labour for such operations is often provided from lower income groups whose earnings the government may wish to protect or increase. Clearly, therefore, publicly owned transport prices, and hence profits, may be seen as important instruments in taxation and income re-distribution policies.

Profitability has a further significance: the self-interest which is at the root of profit maximization leads an entrepreneur to seek to minimize costs at each level of output, and hence to pursue technological efficiency. If the link between efficiency and self-interest is removed, managers may pursue ‘a quiet life’ or self-satisfying goals which may imply technological and managerial inefficiency. It is now widely accepted that profit targets need to be used to stimulate innovation, cost reduction and efficiency. Finally, where market distortions do not exist, profit maximization can be demonstrated to be consistent with allocative efficiency and welfare maximization.

Pricing for Profit

The way in which profitability and efficiency goals are brought together in practice

is through the setting of a 'financial target' or 'profit target'. This is normally expressed as a specified value of the gross trading surplus, or excess of revenue over direct operating costs. Thus, the surplus can be taken to include provision for those costs which can be taken as fixed with capacity. The question now is: how should prices and outputs be chosen, in a way that meets the profit target with a minimum loss of efficiency?

Firstly, consider, for example, a railway operator who produces a single rail service where all the users are broadly equivalent in terms of the value they attach to the service. Figure 2.17 illustrates how the operator might achieve the target of 'break-even', that is, to cover total costs, including fixed capital costs. The simplest way to achieve a break-even position is to set a price equal to average cost. In Figure 2.17, this would involve setting a price of P_M with an output of Q_M . It should be noted that, in terms of social welfare, this is less efficient than charging a price of marginal cost. Whether or not such a price is sustainable depends on the degree of competition in the marketplace and the extent of any under- or over-capacity.

In general, however, transport undertakings usually produce more than one output, and so, it is necessary to determine how to allocate the profit target amongst different outputs.

Consider again the railway operator, but this time when there are two distinct outputs, for example, business passengers and elderly leisure passengers, that are unrelated in demand and cost – i.e., there is zero cross-elasticity of demand and variations in the number of passengers in one group do not affect costs for the other. Further, it is assumed that the railway operator can identify the two groups; that tickets are non-transferable between groups; and that it is possible to distinguish between them in operating a pricing policy.

Figure 2.18 shows how a monopolist might increase profits by selling the same rail service at two different prices. The two groups represent two segregated markets, each of them having a demand curve with different elasticity. It is assumed that the demand curve for business travel will be more inelastic than that for leisure travel and that business travellers will be prepared to pay more than leisure travellers for the same service. In such circumstances, it is possible for the rail operator to attempt price discrimination between the two groups.

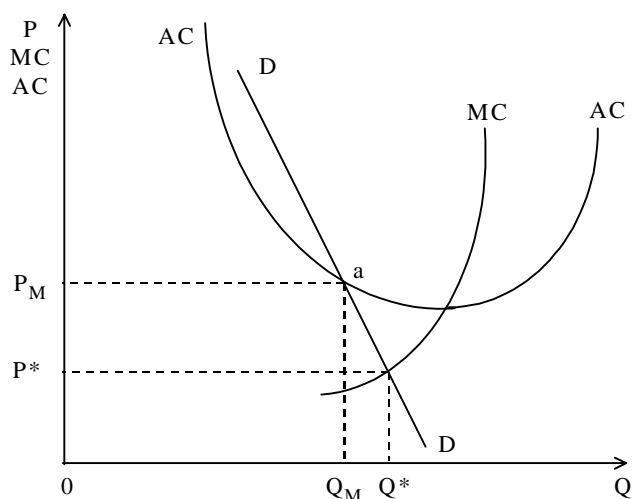
In Figure 2.18, D_1 is the demand from leisure passengers and has a higher elasticity than D_2 , the demand curve of business passengers. D is the aggregate demand curve and MR is the aggregate marginal revenue curve. MC is the marginal cost curve.

The price discriminating monopolist has to decide (i) the total output of rail services, and (ii) how much to sell in each market and at what price so as to maximize profits.

The total quantity to be produced is determined by the point of intersection of the MC for aggregate supply and aggregate MR curve. In Figure 2.18, the two curves intersect at point *k* thus defining a total output of *Q*. If the monopolist charged a uniform price, this would be *P*, and the total revenue would be *PAQO*. The profit would be the difference between the revenue and the cost of producing *Q*.

However, the monopolist can achieve a higher profit by charging different prices to the two markets, if the marginal revenue curves are different due to the differences in demand elasticity in the two markets. In order to take advantage of such price discrimination, he will have to equate marginal revenues in the individual two markets with the marginal cost at the level of aggregate supply. Thus, profit maximization will be achieved at *m* and *n* respectively in the two markets, giving an output of *Q₁* and price of *P₁* for leisure passengers and an output of *Q₂* and price of *P₂* for business passengers. In this case, the policy of price discrimination produces a reduction to profits of *ABCD* from leisure group, and an increase in profits of *PP₂ED*. Clearly, the increase to profit

Figure 2.17 : Average Cost Pricing



significantly exceeds the reduction to profit and, therefore, the operator has scope to increase profits; reduce losses; or, break even through price discrimination³⁴.

Price discrimination is preferred by many economists as being more efficient than average cost pricing in terms of social welfare. Price discrimination can involve more than two charging groups and become quite complex to operate. The main attraction in welfare terms is that those with the highest ‘ability’ or ‘willingness to pay’ make the highest contribution to the fixed costs of providing the rail service. In other words, those who benefit directly from the rail service pay for it, rather than society as a whole

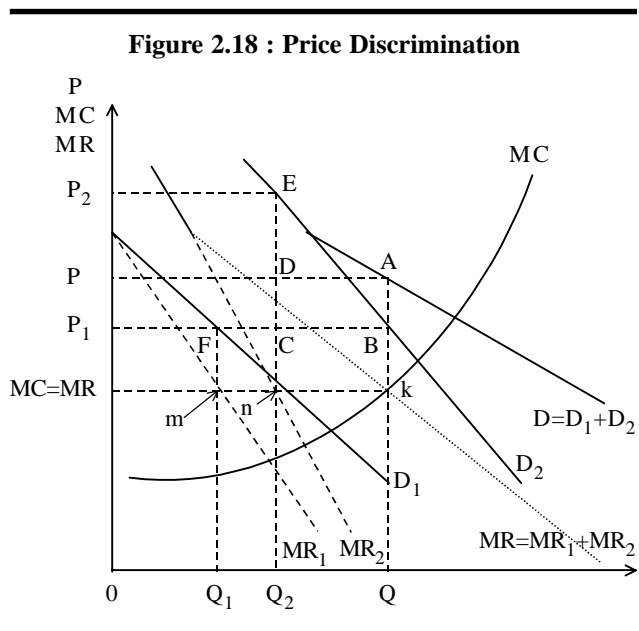
through general taxation and subsidy. Price discrimination can also be criticized on grounds of lack of transparency and fairness. In its purest form, it is sometimes referred to as Ramsey Pricing, or demand differentiated pricing whereby all unattributable fixed and common costs are apportioned among transport users on the basis of the value they attach to a particular service³⁵.

Another method of breaking even or achieving a profit target is to operate a two-part tariff whereby users pay a fixed sum regardless of the amount of travel plus a marginal cost based charge directly related to the amount of travel. The fixed sum represents the average fixed cost of capacity and is calculated to recover the capital costs. Two-part tariffs are effected through railcards or permits. If the fixed sum element is high in relation to the variable element of the price, the method suffers from the same loss of efficiency as average cost pricing.

Equity Considerations and Cross-Subsidization

The above analysis has been concerned with identifying pricing policies that meet profit targets with minimal losses in allocative efficiency without reference to income distribution. In practice, governments are not indifferent to income distribution. However, many governments find that they cannot, due to practical difficulties, effect redistributions through lump-sum taxes and subsidies which create the least economic distortion. As a result, income distributional considerations will be brought to bear on transport pricing policy, and often policies to redistribute income will be effected through direct *ad hoc* intervention, such as:

- (i) rejection by government of proposed price increases;
- (ii) imposition of specific taxes e.g. on fuel or vehicles;
- (iii) perpetuation of loss-making services by subsidization; and
- (iv) perpetuation of overmanned services.



When an operator is required by government to provide transport services at fares which lie below the relevant marginal cost, a mechanism needs to be evolved to fund the gap. If a government faces a severe fiscal constraint and is unable to finance the gap with subsidies, then cross-subsidization often occurs in an attempt to maintain the profitability of the operator. Cross-subsidization involves charging some users above marginal cost to offset losses made on traffics or services where prices are fixed by government at levels which do not cover the relevant marginal cost. However, cross-subsidization within transport pricing systems often acts like a regressive tax and may not target the appropriate income groups effectively. In consequence, it is now widely accepted that, in the interests of economic efficiency, cross-subsidization should be eliminated and, if subsidies are socially desirable, it would be better to shift responsibility for raising revenue and distributing subsidies to the general tax system. Indeed, efficiency gains from removing cross-subsidization may generate sufficient resources to compensate those who are faced with an unfair burden as a result and still make society as a whole better off³⁶. It is important, therefore, to identify the equity implications of particular pricing policies and develop specific strategies to realize economic efficiency within the constraints presented by income distribution requirements.

Section V

Financing and Private Sector Participation

It is widely acknowledged that the role of the private sector is fundamental in facilitating financial sustainability and the major challenge is to devise frameworks that will ensure both economic optimality in the allocation of resources, consistent with promoting sustainable development, and financial viability.

In devising optimal pricing systems, the objective of economic efficiency has normally to be achieved subject to the constraint of securing financial viability. This translates into the need to ensure the financial sustainability of investment in transport infrastructure and in the provision of transport services. Financial sustainability has three components, in the sense that such activity needs to be able to:

- (i) attract sufficient funds to finance the necessary investment and operation;
- (ii) generate sufficient revenue to recover both the operating and capital costs

³⁴ Koutsoyiannis, A., *Modern Microeconomics*. Macmillan. London. 1979.

³⁵ Sengupta, R., 2000. *Equity and Transport Pricing* ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

involved; and

- (iii) provide the necessary financial incentives to attract and sustain wide participation in such ventures.

There is an important interdependency between pricing, investment and financing, and, therefore, any discussion of pricing policy is incomplete without considering financing policy. If the absolute value of the price elasticity of demand for transport is greater than zero, then the choice of financing policy may affect the demand estimates and investment programmes, as the costs of finance may differ by source. In addition, all the three policies are related through the objective function of government, or of the particular transport undertaking. Privatization, as a means of both financing investment and promoting managerial and technological efficiency, is critical to the reform of the transport industry in many countries.

This section considers methods of financing transport infrastructure and services; private sector participation; and the implications for pricing policy.

Sources of Finance

Broadly speaking, it is possible to identify six possible sources of financing transport infrastructure and services. These are:

- (i) *Debt capital from the market* – this method normally involves raising funds with or without the provision of government guarantees.
- (ii) *Equity capital* – this involves issuing equities on the market and raising capital from the public, either directly or indirectly, through financial institutions.
- (iii) *Borrowing from government* – this is a major source of funds for publicly owned transport undertakings and contributes directly to the national debt. Such borrowings are, in turn, financed by government borrowing; taxation; increases in the money supply; or, surplus revenues.
- (iv) *Internal resources* – publicly and privately owned transport undertakings are likely to finance some proportion of their investment needs from retained earnings and depreciation provisions. The actual proportion of funds available will depend on the pricing and investment policies adopted. The setting of prices at the level required to enable an industry to internally finance all its investment programmes is unlikely to be consistent with a pricing policy

³⁶ Anant, T. C. A., 2000. *Price Distortions in the Transportation Sector*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

which would achieve the desired efficiency of resource allocation. It is more desirable to pursue efficiency rather than some arbitrarily determined financial surplus.

- (v) *Borrowing from banks and other financial institutions* – this source is normally used for short-term working capital and may be with or without government guarantees.
- (vi) *Grants from both private and public sectors* – these include such provisions as capital subsidies and the writing-off of capital debt.

It is considered preferable to seek new funding from private sources rather than to rely on government funding, and this may extend to full privatization of either the provision of infrastructure or services.

Private Sector Participation

Many governments are turning to the private sector in developing and delivering transport services. They hope to take advantage of private sector skills and know-how, improve the efficiency of service delivery, and gain access to finance for new investments. Experience in some countries that have entered into arrangements for private sector participation shows that, if well-designed, these arrangements can produce improvements in the quality, availability and cost-effectiveness of transport services. Further, many governments perceive that private sector participation will improve the return on capital employed; reduce the need for government subsidies; and make the industry more responsive to consumers' needs and preferences³⁷.

Private sector participation, on its own, is no panacea for problems in providing transport services. It requires partnership between the government and the private sector, and the governments must:

- (i) devise sectoral policies that introduce and maintain competition;
- (ii) establish a sound regulatory framework for the remaining monopolies, both public and private;
- (iii) maintain transparency and convince investors that their investments are secure;
- (iv) negotiate, monitor, and enforce contracts with private suppliers of services, management and financing;
- (v) ensure that funds raised from privatization are put to productive uses; and
- (vi) manage the inevitable political and social tensions that arise from such reforms, especially if foreign ownership and redundancies are involved.

It has previously been mentioned that for many types of transport provision competition does not come naturally. However, careful design of sector reform and private sector participation can introduce the potential for competition and, consequently, reduce the need for regulatory intervention. The options available to the government in promoting competition in the transport industry, are:

- (i) creating direct competition in the supply of services, sometimes referred to as *competition in the market*;
- (ii) creating competition for the right to supply transport services through concessions or other contracts, often referred to as *competition for the market*;
- (iii) promoting competitive pressures by introducing private sector financing sources; and
- (iv) using benchmarks to assess the performance of different operators.

There are a number of methods of increasing private sector involvement, including:

- (i) *Service contracts* – these secure private assistance for performing specific tasks, such as road repairs, vehicle maintenance, and airport services. The responsibility for management and investment remains with the government. Such contracts need to be carefully specified and monitored. They are a cost-effective way of meeting the specific technical needs of an undertaking that is well-managed and commercially viable. They are not a substitute for reform in public undertakings that are plagued with inefficient management and poor cost recovery.
- (ii) *Management contracts* – these transfer responsibility for the operation and maintenance of government-owned transport services to the private sector. Fees can be either lump sum or linked to agreed performance targets. Management contracts are useful where the main objective is to rapidly enhance an undertaking's technical capacity and efficiency and to prepare it for greater private sector involvement. They transfer neither the commercial risks to the private sector nor the responsibility for investment.
- (iii) *Leases* – these represent a way to pass on the commercial risks. Under a lease arrangement, a private firm leases the assets of a transport operation from the government and takes up the responsibility for operating and

³⁷ Patel, U. R. and Bhattacharya, S., 2000. *The Constraints in Pricing and Financing*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

stream from the service operation (minus the lease payment), it assumes much of the commercial risk of the operation. Under a well structured contract, the lessor's profitability will depend on how much it can reduce costs, while maintaining the specified quality standards, so that it has incentives to improve operating efficiency. Leases leave responsibility for financing and planning major investments with the government and are most appropriate where there is major scope for improving operational efficiency, but little need for new investment.

- (iv) *Concessions or franchising* – these give the private operator responsibility not only for the operation and maintenance of a utility's assets, but also for investments. Asset ownership, including those made by the private partner, remains with the government and full usage rights revert to the government after the contract ends. Concessions are often bid by price: the bidder that proposes to operate the service and meet the investment targets for the lowest tariff wins the concession. Concession contracts usually set out the main performance targets, such as service levels and times; arrangements for capital investment; and mechanisms for adjusting tariffs. The main advantage of a concession is that it passes full responsibility for operations and investment to the private sector and provides incentives for maximizing operational efficiency. Concession is an attractive option where large investments are needed to expand the coverage or improve the quality of services. If concession grants a degree of monopoly, then regulation will be required to ensure an appropriate distribution of benefits between the profits of the concessionaire and the prices and service quality offered to the consumers.
- (v) *Build-Operate-Transfer (BOT) contracts* – these are similar to concessions but are normally used for new or 'greenfield' projects, such as new toll-free highways or bridges. In a typical BOT arrangement, a private firm might undertake to construct and finance a new inter-urban road, maintain and repair it for a number of years, and, at the end of the contract, relinquish all rights back to the government. Monies payable to the contractor would normally be linked to the level of road usage.
- (vi) *Full or partial divestitures* – these involve the sale of assets or shares and can be partial or complete. A complete divestiture, like a concession, gives the private sector full responsibility for operations, maintenance, and investment. But, unlike a concession, it transfers ownership of the assets to the private sector, so the nature of the public-private relationship differs

slightly. The government's role after the sale of an operation is limited to protecting consumers from monopolistic pricing and poor service. This may involve continuing scrutiny of the private company's investment plans.

Financing, Public Participation and Pricing Policy

Pricing policy is inextricably linked to the interests of the stakeholders which, in turn, depends on the structure of ownership. Financial viability will be a prerequisite for expanding private sector participation, particularly for privatization. If pricing depends on the level of capacity provided, relative to demand, then it is more likely to be in equilibrium for private sector firms with clear objectives expressed in terms of profitability.

Role of the Regulator in Price Setting

If governments pursue the privatization of transport infrastructure and services, then they are likely to strengthen their regulatory regime on transport pricing, particularly if natural monopolies emerge. One of the main tasks of a regulator is to control prices and quality of service and to ensure that any service provider who has monopoly power does not overcharge users. Such regulation will be referenced against the need to achieve the kind of efficiency that pricing on the basis of marginal cost would produce in a competitive environment. This means that the regulator will be concerned with efficiency and cost minimization. In addition, the regulator must ensure that service providers earn a reasonable return on capital. Typically, the goals of regulation will be to eliminate shortages of infrastructure services, strengthen the financial position of the sector, bring prices in line with costs, and reduce operating costs.³⁸

Universal Service Obligations, Competition and Income Distribution

It is likely that the desire to meet social objectives, particularly to meet the needs of the poorest people, will remain important even after the introduction of restructuring and increased competition. In consequence, there will be a need to reconcile private participation in the transport sector with certain social objectives. This will require a regulatory regime which ensures that operators have sufficient incentive to invest in activities, including services for poor people, which they may not have otherwise considered profitable. Alternatively, the regulatory regime could include a clear specification of universal service obligations (USO) in the scope of the service providers' obligations. Specific and targeted subsidies may have to be granted for those services that cannot be provided on a commercial basis.

³⁸ Kathuria, R., 2000. *Role of Regulator in Price Setting*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

3

Pricing and Charges for Road and Urban Transport

Introduction

The benefits of an efficient urban transport system are many and varied. It is, indeed, a major contributor to economic growth, employment and competitiveness. However, the present transport scenario is far from satisfactory in most countries. Road traffic congestion has become a serious transport problem in most parts of the world. Increasing traffic delays have brought down travel speeds to levels that prevailed in the age of horse-drawn transport. Air pollution causes many serious diseases and accounts for thousands of deaths. These problems are compounded by the fact that in many cities there is a chronic lack of funds to finance the maintenance and development of urban transport and road systems.

This chapter examines the role of pricing and charges in determining the efficient provision and utilization of road infrastructure in recovering related costs and in paving the way for private funding and participation in road construction and management. It also examines the provision of public urban transport services and their pricing policies. It comprises two sections: section I examines pricing for road infrastructure and its use, and section II deals with pricing for public transport

Section I Pricing for Road Infrastructure

Pricing normally performs three functions. Firstly, it rations and allocates the use of competing resources; secondly, it provides a signal on the need for, and viability of, investment; and thirdly, it helps in generating funds for the development of the related sectors. In the case of urban transport, these functions are not being performed efficiently for the following reasons:³⁹

- (i) the responsibilities for the provision and operation of infrastructure and transport services are usually separated;

³⁹ Gwilliam, K. M., 2000. *Urban Transport Strategy Review – Draft*. World Bank. Washington.

- (ii) there are often a multiplicity of objectives being pursued, particularly in public transport; and
- (iii) infrastructure financing is usually separated from pricing and charging.

The challenge is to address these problems and find ways to improve the effectiveness of government policy in the provision of urban transport through appropriate pricing and financing mechanisms.

While the construction and use of urban roads provides several benefits, these also involve certain allied costs. If a road user has to pay for the use of a road (e.g. the use of energy, infrastructure, etc.), the associated costs can be considered as internal costs. If, on the other hand, the road user affects the well-being of others (e.g. by polluting the air) without paying for this, then the ensuing costs are external to that person. Table 3.1 sets out a simple typology of the costs of road transport.

The typology distinguishes internal or private costs from external costs or externalities. Costs relating to infrastructure, congestion, accidents and environment are

Table 3.1 : Classification of Road Transport Costs

Cost Categories	Social Costs	
	Internal/Private Costs	External Costs
Transport Expenditure	Fuel and vehicle costs; tickets and fares	Costs paid by others e.g. free parking provision
Infrastructure Costs	Tolls and user charges; vehicle taxes; fuel taxes; road taxes	Non-recovered infrastructure costs
Congestion Costs	Personal time costs	Delays and time costs imposed on other road users
Accident Costs	Costs covered by insurance; own accident costs	Pain and suffering caused to others (accident victims)
Environmental Costs	Personal disbenefits	Non-recovered loss of amenity; noise and air pollution costs

examined in more detail in the following paragraphs.

Infrastructure Costs

Building and maintaining road infrastructure entails costs. This raises the question

of how these costs should be recovered from the infrastructure users⁴⁰. In many countries, the charges that are currently levied are not related, or are only partly related to the actual costs of providing or using the infrastructure. In evaluating infrastructure costs, it is essential to make a distinction between:

- (i) *Capital cost* – Road networks represent important assets. The provision of these assets implies real costs. The capital invested in the provision of road infrastructure gives rise to a fixed cost that bears no relationship to the actual use of the infrastructure. The capital value of the road infrastructure increases over time with additions made by new investments. Measuring the asset value of the network requires detailed information on parameters, such as the replacement value of the infrastructure; and
- (ii) *Operating and maintenance cost* – such as the (annual) maintenance expenditure on roads. Some operating costs vary with traffic volumes, but other factors, such as weather conditions, geographical settings and topographical features also play an important role. For example, the pavement of a road will be damaged not only by the axle weight of the vehicles travelling on it, but also by temperature variations, rain, snow and excessive heat.

In addition, the construction of infrastructure often has major land-use implications. It can lead to a disturbance of eco-systems and might create the so-called barrier effects. The associated costs are not well understood, but could be significant and should, among others, be taken into account in the design stage.

Charging for infrastructure costs should ideally meet three criteria:

- (i) The system should link charges, as much as possible, to marginal cost at the level of the individual user. Marginal cost pricing is important for the efficiency of the transport system, since it gives individual users an incentive to reduce the underlying costs as cost savings are rewarded by lower charges. For example, traffic induced road wear and tear can be taxed relatively easily and efficiently by charging trucks on the basis of their axle weights and mileages. Such a system gives hauliers an incentive to use configurations with lower axle weights, to reduce empty runs or, in some cases, to use mix of transport.
- (ii) Ideally, infrastructure charges should recover overall infrastructure costs. If

⁴⁰ Walters. A. A., 1968. *The Economics of Road User Charges*. World Bank Occasional Paper No. 5.

significant parts of total costs are not use dependent – as is the case with capital costs – then marginal cost pricing alone will not lead to full cost recovery. Cost recovery is, however, important for a number of reasons. Firstly, private owners of infrastructure have to recover costs – privately owned toll-roads have developed pricing schemes that relate charges to relevant parameters in order to recover full costs. Secondly, in the absence of full cost recovery in the transport sector as a whole, the general budget would have to fund the sector by imposing taxes/charges elsewhere. It is generally felt that although for reasons of economic efficiency marginal cost pricing is the best, this should be tempered by cost recovery on the principle that ‘those who benefit should pay’. Therefore, in principle, total infrastructure costs should be recovered in the long run.

A number of considerations, however, have to be kept in mind when implementing the cost recovery criterion. First, it is quite common and perfectly legitimate to invest in road infrastructure for non-transport related policy reasons, such as regional balance. It seems unreasonable to require transport users to cover the road infrastructure cost imposed on these grounds. This highlights the need for a clear system of accounting. Second, past decisions on road infrastructure projects that no longer meet present-day transport demands have, in some cases, generated high costs which cannot be recovered from users. This kind of infrastructure investment may have to be given special treatment.

- (iii) The system of charging for infrastructure costs should be transparent and easily intelligible.

Infrastructure Usage and Externalities

Externalities arise where a road user either does not pay for the full costs (e.g. the environmental, congestion or accident costs) of the road journey, or does not receive the full benefits from it. Any transport activity creates benefits as well as costs. However, not all of these costs and benefits accrue only to those who pay for this transport activity (i.e., the transport users). Some of the costs have to be borne by other persons or by society as a whole. One can, therefore, distinguish between the so-called ‘internal’ or private costs, i.e., those that are borne by the persons engaged in the transport activity (e.g. time, vehicle and fuel costs) and the so-called ‘external costs’ i.e., those that accrue to others. The sum of both these types of costs is called ‘social costs’. In general terms, externalities arise whenever the well-being of an individual is affected by the activities of others who ignore this ‘spillover’ when taking their decisions.

The crucial importance of externalities arises from the fact that, in a market economy, (economic) decisions are heavily dependent on market prices. However, when market prices fail to reflect existing scarcities (clean air, absorptive capacity of the environment, infrastructure, etc.), the individual decisions of consumers and producers no longer add up to an outcome that provides maximum benefits to society as a whole. Thus, pricing on the basis of full social costs is a key element of an efficient and sustainable transport system, in general, and road system, in particular.

Externalities impair the efficient distribution of resources across sectors and activities. For example, if the use of a certain vehicle type entails significant air pollution and road damage costs which are not charged, then the demand for this vehicle type will be 'too' high and the demand for cleaner and less damaging vehicles 'too' low. This represents an inefficient use of resources. Moreover, the external costs are borne by others: tax payers implicitly end up footing the bill of road maintenance and health care due to damage from air pollution, whilst damage to buildings and crops resulting from acidification and other forms of pollution is paid by property owners, businesses and farmers. This is unfair and inefficient. To correct this, there is need for the government to take certain measures.

The government measures should aim at curbing these externalities, both for reasons of economic efficiency and equity. A price-based approach ensures that prices paid by the transport users reflect total costs better: this can be achieved by internalizing the external costs i.e., by charging them to users. The internalization approach represents an alternative to the policy of setting standards and regulations. Internalization will be successful in reducing the amount of road travel only if the demand is elastic with respect to price. It may be emphasized that the outcome of such an approach will not necessarily be the complete absence of environmental damage, accidents or congestion. Instead, the negative side-effects of transport activity will be at a level that is 'optimal' from the point of view of society, i.e., the marginal costs of further reducing these side-effects would exactly equal the marginal benefits from doing so. Reducing the side-effects still further would entail higher costs than benefits. In order to devise policies for internalizing transport externalities, it is first necessary to measure them.

Externalities can be measured, in monetary terms, by inferring their value from observed market transactions (e.g. expenditure on damage avoidance, health costs, property value loss, etc.). Alternatively, externalities can be measured by asking people how much they would be willing to pay for the reduction of a specific negative transport externality by a certain amount. Market related valuation approaches tend to systematically underestimate the full amount of external costs by only including those that lead to easily

identifiable changes in prices. Estimates based on willingness-to-pay or willingness-to-accept permit a more comprehensive measurement of costs.

Road Congestion

Clearly, there is an important link between congestion and infrastructure costs, because congestion implies that infrastructure is being used beyond its designed capacity. Imposition of an infrastructure service charge could result in a reduction of traffic volume over the congested sectors and, as a consequence, bring about an increase in speed.

Congestion causes waste of time. It arises when road networks carry more traffic than their designed capacity. In such a situation, every user incurs delays and imposes delays on others. These delays represent economic losses, because people value time. In addition, energy use increases with such delays. As more and more users enter the network, the delays increase disproportionately until traffic comes to a standstill. This is why, in congested networks, even a small reduction in traffic levels can significantly speed up flows.

Congestion deals with infrastructure scarcity through queuing (i.e., quantity rationing), since everybody gets stuck in a traffic jam and incurs time losses whilst imposing further delays on others. When making travel choices, every individual road user only takes his or her own time (and other) costs into account, ignoring those imposed on others. As everybody does the same, there is too much traffic and all infrastructure users waste time. Although, in the resulting situation, all infrastructure users, put together, pay for the total time costs, there is still an externality and an ensuing wastage of scarce resources (time, energy). This results from the fact that there is a 'market failure', because, as stated above, an individual infrastructure user does not compare the private benefits of his/her decision with the total costs this decision imposes on society as a whole (the so-called 'marginal social costs'). A price-based solution to infrastructure scarcity corrects this 'market failure' by ensuring that prices paid by individuals reflect the full costs of transport choices to all other infrastructure users. The result is that trips, which carry higher costs than benefits, are avoided. The ensuing reduction in transport volumes leads to higher travel speeds and time savings, which represent a benefit to all.

An important point to note here is that the value of time differs considerably between different infrastructure users. The cost of delays is likely to be more for commercial vehicles and business travellers than for leisure travellers. Apart from charging for congestion costs, there is no means of allocating the scarce infrastructure capacity to those who derive the maximum benefits from using it. As a result, money is wasted and society, as a whole, does not reap the maximum benefit from its road infrastructure

networks.

An important characteristic of congestion is that it varies strongly across space and in time. Road congestion is largely concentrated in urban areas. Costs during peak hours are significantly higher than during off-peak hours. Whilst traffic in rural areas accounts, in some countries, for more than 20% of total traffic, its share in total congestion costs is estimated at less than 1%. As a result, policies to curb congestion must be differentiated in time and space. Across the board increases in charges are unlikely to be effective, as they would not allow for the required differentiation. Moreover, they would be unfair. Higher charges would penalize rural areas and may still fall short of what is required in heavily congested conurbations. Thus, an efficient and equitable solution would be to introduce highly differentiated charges which vary over time and space. These charges should reflect the cost of congestion to all road users and should give citizens an incentive to base their transport decisions on the full social costs of road travel. This would dissuade them from making those trips for which the total costs are higher than the concomitant benefits and thereby raise welfare by lowering congestion. Pricing is only one element of a comprehensive strategy to curb congestion.

Other policies also have an important role to play. For example, the introduction of telematics based traffic guidance, management and information systems can significantly increase the 'virtual' capacity of infrastructure networks. Similarly, the provision of efficient public transport systems will facilitate the transfer of travellers from private cars to bus and rail. Moreover, teleworking, teleconferencing, and teleshopping can all reduce the need for mobility. Clearly, the different components of a strategy have to dovetail in order to maximize its effects. The introduction of such a strategy would necessitate electronic road charging which has significant advantages in terms of flexibility and non-interference with traffic flows (as vehicles would not have to stop at toll stations). Recent technological developments would fully respect the privacy of motorists by relying on smart-card technologies. Significant progress has already been made in developing sophisticated telematic systems and a large variety of trials are underway in many countries.

Road Accidents

Road accidents are a cause of large number of fatalities. In India alone, 80,000 persons die annually in road accidents. In the European Union, such accidents account for 50,000 deaths annually. The concept of cost in relation to accidents is complicated with regard to the underlying economic theory; it involves problems in making practical estimates and raises the ethical questions. Nevertheless, it is possible to examine the benefits of internalizing the external costs of accidents and accident risks.

It is estimated that a year's road accidents in the European Union cost approximately 2.5% of GDP. In developing countries, these proportions are still higher even with lower vehicle ownership. For instance, in India, these costs are of the order of 3.5% of GDP. In terms of money value, the yearly medical, administrative and damage reparation expenditure on accidents is some US\$ 14 billion in the European Union alone. The future loss of production in European Union is estimated to be an additional US\$ 28 billion (net).

Many argue that the true costing of road accidents will prompt:

- (i) government to allocate resources to minimize road safety risk by providing better infrastructure, improved road maintenance and stringent enforcement of safety regulations;
- (ii) car manufacturers to develop even safer cars, both for those inside and outside the car; and
- (iii) safer driving habits.

Taking into account the differences in risks between different users, vehicle types and roads with different traffic safety performance, will help to ensure that charges are brought closer to costs at the level of the individual who makes the transport decisions. More careful driving can be secured by, for example, a differentiation of the insurance premium with a greater bonus for safe driving. Vehicles with higher total safety to the passengers as well as the unprotected users outside the vehicle, should be rewarded with lower rates. Driving on roads with higher safety standards, such as highways with segregated lanes, should, in principle, ensure a lower rate than in the case of the rest of the road network.

Among the potential economic instruments available to the government are adjustments to existing fuel excise duties, purchase taxes and annual circulation taxes. However, the use of these economic instruments to exert a positive influence on road safety has major shortcomings. Fuel tax, although more or less distance-related, cannot take into account the difference in risk between users and vehicles or networks. Purchase tax and circulation tax can, albeit with difficulty, be differentiated according to the vehicle's and, possibly, the owner's risk characteristic, but will not be distance related or related to the network. The inverse is true of road tolls.

This discussion shows that efficient instruments to internalize the external costs of road accidents should aim at reducing risk-taking in the broadest sense of the word and should, therefore, be introduced at the level of the individual motorist. This suggests that

one should seriously review the possibility of using existing insurance systems and ensure that the premiums, both in level and in structure, reflect risk to society as a whole. The use of insurance premiums has the additional advantage of relying on an existing instrument.

The basic principle underlying such an approach would be to ensure that the insurance liability covers the whole accident cost and the premium is differentiated as much as possible. This approach would imply withdrawing the 'road accident subsidy' that is currently paid by most governments and societies and leaving the true and total costs to be borne by the driving public. The cost of insurance should ideally vary with risk. Assessment of a driver's attitude towards risk-taking could include historical evidence, such as rewarding responsible behaviour through driving offence registrations, or proof that driving behaviour is better than average, i.e., by assessing acquired ability to drive in a safer manner. A scale of charges by offence, perhaps automatically linked to the driving licence point system, could be an equitable way of targeting the higher risk-taker rather than only relying on a blanket charge for high risk groups, such as young male drivers.

This approach would give consumers an incentive to buy safer cars, drive more safely, use safer roads, drive less, switch to other modes where appropriate, or resort to car pooling. In this way, this approach would leave it to the individual users to decide how to reduce accident risk in a manner they deem fit.

Pollution from Road Traffic

In most countries, the major share of carbon monoxide (CO) and oxides of nitrogen (NO_x) emissions comes from road traffic. This sector also contributes a substantial share of non-methane volatile organic compounds and a minor share of sulphur dioxide (SO₂) emissions. Secondary pollutants are formed as a result of complex chemical reactions that the primary pollutants undergo in the atmosphere. The main secondary pollutants attributable to transport activity are nitrogen dioxide (NO₂) and ground-level ozone. Oxides of sulphur and nitrogen also contribute to acidification. Other air pollutants of concern come from fuel substances like lead and benzene in gasoline that are directly emitted from diesel vehicles in the form of particulate matter (PM), or are linked to fuel consumption, such as emissions of carbon dioxide.

The control of transport-related air pollution in most countries has largely followed a regulatory approach aimed at bringing down emissions through product standards and rules to reach air quality standards. In addition, fuel standards usually limit the sulphur content of diesel and lead and benzene content of petrol. Vehicle inspection and maintenance programmes have been introduced in many developed countries to ensure compliance with existing emission standards. Some countries operate a differential between

leaded and unleaded gasoline excise rates.

The nature and causes of air pollution differ, sometimes considerably, from country to country and region to region. For example, regional air pollution differs considerably across Europe where the damage costs of acidification are much higher in northern and central Europe than in southern Europe. Even in the case of local air pollution, the differences are widespread. The Hague, for example, suffers more from summer ozone than from high NO₂ concentration, while the opposite is true in Milan. Lead in gasoline has practically disappeared due to the high turnover rate of the vehicle fleet and the early use of relatively strongly differentiated taxes in Denmark, Austria, Finland, and Sweden, while it is still highly consumed in Portugal and Spain (70%-80% leaded petrol). The average age of the vehicle fleet also differs considerably across Europe along with the composition of the urban vehicle fleet in terms of private/public transport vehicles as well as in terms of gasoline and diesel vehicles.

This high degree of variation suggests that the implementation of blanket physical measures alone is not very cost-effective, because these cannot take account of regional differences in the impact and cost of air pollution.

Given the wide-ranging variations in the problems to be addressed, it seems clear that economic instruments have a key role to play in a broad strategy that provides the necessary flexibility to accommodate the significant differentiation that is likely to be found in practice.

The amounts and proportions of air pollutants emitted from an engine depend on several factors, including the design and size of the engine, the characteristics of the fuel, and the conditions in which the vehicle is used: how the vehicle is driven, its age, and its state of maintenance. A diesel engine, for example, produces much less CO and VOC than a gasoline engine, but it produces more emissions of PM₁₄ and NO_x. A well-maintained and tuned-up engine emits less pollution per unit of travel than a poorly maintained one. New vehicle technology has great potential to bring down emissions. But emissions depend on real driving conditions. As such, in-use compliance testing is important, because it guarantees the continued effectiveness of exhaust emission systems.

Charges that are directly based on emissions would, in principle, be a particularly attractive policy instrument, since they would give the road users an incentive to select the least polluting vehicle and journey types.

Road Noise

In many countries, traffic noise disturbs people more than any other noise source. The continuing growth in traffic volume in all modes and its increased spread over space and time is offsetting the impact of the policy measures implemented to date to address the problem. For example, European Community legislation governing the emissions of noise from vehicles has been in existence for over twenty-five years for passenger cars and heavy vehicles, and for fifteen years for two-wheeled vehicles. Since the implementation of the first directives, the current regulations have achieved a reduction in specific noise levels of 60% for motorcycles, 85% for passenger cars and over 90% for heavy lorries. Despite these measures, there has not been significant overall reduction in road traffic noise. The reasons for the low level of effectiveness have been identified as: the increase in road traffic, a worsening of traffic fluidity and general traffic conditions and a lower threshold to achievable noise reductions caused by the interaction of tyre and road noise. Other disadvantages of relying solely on regulations are: (i) the test procedures do not realistically reflect driving conditions; and (ii) without a regular inspection procedure to ensure maintenance of the acoustical design features, the noise levels may increase over time.

The future impact of legislation limiting vehicle noise emissions on overall noise levels is likely to remain limited and effective noise abatement action will require increased recourse to other instruments, such as land-use planning and economic instruments combined with stricter standards. In order to put greater emphasis on the 'polluter-pays' principle, economic instruments, in particular, should have a greater role.

An overview of the studies conducted by Quinet⁴¹ in 1993 revealed that the estimated costs of noise pollution varied between 0.1% and 2% of GDP. Generally, studies based on the avoidance-cost approach give low values for noise costs – below 0.1% of GDP, while studies using the willingness-to-pay approach give higher values, partly because they are carried out in countries with a high per capita income. Most in line with the 'polluter-pays' principle is the 'willingness-to-pay' method.

The use of economic instruments for noise abatement from transport, especially from road transport, is not widespread in Europe. However, the OECD in its report 'Fighting Noise in the 1990s' concluded that economic incentives for noise reduction have shown their effectiveness in relation to road vehicles in the few cases where they have been used. Noise charges – except in the field of aircraft noise – have been used even less than incentives and, where used, have generally been set too low to encourage noise reduction. Their main function has been to raise funds for noise control measures, such as the noise insulation of buildings.

The possibilities for taxing noisy vehicles include: a tax on new vehicles dependent on their noise category (which may depend on noise emission and type of use/average annual mileage), or an annual tax dependent on noise category. Such a tax may be used in conjunction with in-service checks that a vehicle is still within its designated noise category (this would also open the possibility of operators being able to reduce their annual tax by fitting noise suppression equipment). A third possibility is a charge on noisy vehicles when they are used in an environmentally sensitive area. Noise taxes levied on vehicle manufacturers have the advantage of pressuring them to produce quieter vehicles. In addition, if the users are also required to pay noise tax, they shall be prompted to reduce noise by properly maintaining the vehicle, fitting better noise suppression equipment, and using the vehicle less (assuming that the taxes are made dependent on in-service noise and distance travelled). For example, Austria has introduced, since 1996, a road user charge that differentiates according to the noise (and also emissions) of vehicles.

A consequence of road pricing should be the optimal use of the road network, and this may lead to both increases (e.g. if speeds increase) and decreases in noise nuisance. However, inclusion of a noise cost element in the charge could be an incentive to reduce noise on road networks.

Incentives in the form of grants to purchase low-noise goods vehicles have been in operation in Germany and the Netherlands. For a period of about ten years, operators of heavy goods vehicles in the Netherlands were offered a two-tier subsidy if they purchased and used vehicles fitted with 'hush kits'. Subsidy levels were 7.5% and 5% for noise reductions of 6dB(A) and 3dB(A), respectively. The costs of these measures were borne by the operators. The programme, now abolished, resulted in specified lower noise levels – more than 60% of the lorries now in use in the Netherlands have noise levels 5dB(A) below current minimum standards. Similar schemes exist in Germany. Although of limited scope, this type of initiative is likely to become more widespread in the future and could be extended to include incentives for tyres and road surfaces producing lower noise. Instead of subsidizing reduction in pollution, compensation could be given to those affected by pollution. It may be pointed out that compensation for house price depreciation caused by noise or other environmental impacts is a well-established policy.

Economic incentives in the form of charges or subsidies have been provided in the

⁴¹ Quinet, 1995. *The Social Costs of Transport: Evaluation and Links with Internalization Policies*. ECMT/OECD.

case of road vehicles. Evidence suggests that some reductions in noise levels have been achieved. The further, or future, development of incentive schemes for road vehicles based on annual taxes or road pricing can be a step forward to the extent that it is effective and rests on the three key principles of cost-relatedness, non-discrimination and transparency.

Road Pricing and Recovering Infrastructure Costs

There are several different systems in operation for allocating and recovering road infrastructure costs, which all rely on such instruments as annual vehicle taxes, fuel excise duties, road tolls, road user charges, and distance or kilometre taxes. In some countries, annual vehicle taxes are based on overall weight and, sometimes, axle weight of the truck. However, the relation between total taxes paid and damage caused is generally quite weak. This is because usually there is no link with mileage in the annual vehicle tax systems. Moreover, the relation between fuel consumption (taxed through fuel excises) and road damage is also imprecise, especially for larger trucks. Nevertheless, it is widely recognized that an efficient charging system for road damage costs will have to be highly differentiated by vehicle type.

Efficient charging implies that charges should be linked as closely as possible to costs. As infrastructure and congestion costs vary significantly across vehicle characteristics, both in time and in space, efficient charging systems will have to differentiate accurately in many respects. Such a high degree of differentiation would require the introduction of telematics based pricing mechanisms to implement a system of road pricing.

Whilst full electronic road pricing will take time to implement, it is still worthwhile investigating whether, as a first step towards an efficient system of charging, a so-called 'electronic kilometre charge' – based on axle-weights and other characteristics – could be introduced for heavy goods vehicles. Such a system is mileage based and can differentiate very finely across different vehicle types. A primitive version of this system, which relied on proven technology, was used in Sweden prior to its accession to the European Union. Essentially, an electronic odometer would keep track of the mileage driven and charges would be imposed on the basis of a registration card stamped in the vehicle's meter. A major advantage of this system over using diesel excises, which are not strongly related to costs, is that charges could be brought close to road damage and other costs. The system does not, however, differentiate across time and space. It would, therefore, be useful to investigate whether improvements, such as an electronic tachograph, can be introduced which would further enhance its attractiveness.

Road Pricing and Congestion

It is sometimes argued that the best remedy against congestion is to simply provide more infrastructure. Notwithstanding the need for additional infrastructure in most countries for other reasons, this statement is generally untrue: as motorists are discouraged from using a congested road, there is a 'latent' demand which is triggered once extra capacity becomes available. In the long run, therefore, congestion will persist. Apart from introducing bans – which suffer from a wide variety of disadvantages – the only way to curb congestion in the long run is to set an explicit price for road infrastructure capacity. The introduction of congestion charging would also bolster the efficient provision of infrastructure. Efficient road infrastructure provision basically entails making decisions on capacity by comparing the benefits from capacity extensions (e.g. time savings) with the costs of construction and maintenance. Optimal infrastructure provision requires infrastructure to be extended to the point where the costs of doing so are no longer outweighed by the benefits. However, in the absence of congestion charging, road speeds will, in the long run, always be too low because of the persistence of congestion, and traffic volumes will be higher than those which are desirable. Therefore, in the absence of congestion pricing, it will generally be attractive to build more infrastructure than that which is socially desirable. Congestion pricing could thus lead to important savings in resource costs of building road infrastructure⁴².

If infrastructure is provided efficiently and capacity is priced through congestion charging, both the quantum and the use of infrastructure would be such that it would be impossible to accommodate more traffic and higher speeds on the network at a cost lower than the associated benefits. The joint use of congestion charging and efficient investment rules is thus an essential precondition for arriving at a balanced transport system.

Congestion charging would also raise significant revenues. Firstly, these revenues could go a long way towards recovering the capital costs of the network and might – if certain conditions are met – ensure full recovery. Secondly, these revenues would allow reduction of other taxes, which are currently used for the public financing of road infrastructure, but which are largely unrelated to the costs of infrastructure use. Thirdly, these revenues could be used for financing other parts of a comprehensive strategy to deal with congestion (e.g. route guidance systems, public transport, etc.). In the long run, congestion charging holds out the prospect of reforming transport taxation in a manner that would greatly increase the efficiency and equity of our transport system. This is of great importance to the financial viability of public-private investment partnerships; the introduction of congestion charging would provide a stable revenue source which could cover large parts of the costs⁴³.

In practice, in most countries, developed and developing, urban roads are provided to their users without any direct charge. The only payment from the private user to the public supplier comes indirectly in the form of those taxes (primarily on fuel) which vary directly with road use. But this does not reflect the costs of congestion. So, the cost perceived by the marginal road user, his own private cost, does not take into account the extent to which he slows down all other road users. This has several adverse effects. First, as rail and some other public transport infrastructure is paid for directly through fares, there is a distortion in the choice of mode. 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 prices can be charged with varying degrees of precision through a variety of techniques of different levels of technological sophistication and cost. While, theoretically, different prices could be set for each link in the network for each period of time, in practice, a cruder approximation may be used both for reasons of practicability of application and predictability of response from drivers. Three principal forms of congestion pricing have been developed to date:

- (i) *Cordon pricing, or area licences* – 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.
- (ii) *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. Among developing countries, it has been used on a number of tunnels in Hong Kong, the Namsam tunnel in Seoul and the expressways in Singapore.
- (iii) *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

⁴² Newberry, D., 1990. *Pricing and Congestion: Economic Principles Relevant to Pricing Roads*. Oxford Review of Economic Policy, Vol.6. No. 2.

⁴³ Morrison, S., 1986. *A Survey of Road Pricing*. Transportation Research. Vol. 20A. No.2.

much more attractive and several large test applications have been undertaken. The only comprehensive system actually installed is that in Singapore, where it replaced the Area Licensing Scheme and tolls on a few major access roads in 1998 and is gradually being extended.

Pollution Taxes

Although direct emission charging is currently not possible because of very high transaction costs, a number of European countries have experimented with a variety of economic instruments to complement the existing regulatory approach. Sweden, for example, has long recognized the advantages of using market-based incentives as an environmental policy tool. Other examples are vehicle or sales taxes based on engine powers (found in a number of countries) or on emission standards (Finland and Sweden). In the past, several countries (Austria, Finland, Greece, Netherlands and Germany) introduced temporary tax reductions to promote the introduction of cars with catalytic converters in the vehicle fleet.

Many alternatives have been proposed over the years to internalize the environmental costs of transport – from the charging of advanced emission fees based on actual emissions to the adjustments of the level of the existing fuel, vehicle or purchase taxes to take into account an approximation of the emissions. Tradeable permits have been discussed as a means of setting levels. Even road pricing and scrapping fees have been examined as possible measures. Clearly, the closer the charge is to actual emissions, the more efficient it will be in ‘tapping’ all the different sources of emission reduction. However, implementation decisions will have to be based on a trade-off between an instrument’s link to emissions and its implementation costs. In addition to emission fees, which require advanced metering technology, the following instruments could be considered:

- (i) Different approximations of the actual emissions based on distance driven and emission per kilometre for each vehicle type can be developed. In practice, this could consist of modifying existing annual vehicle taxes to include environmental charges. In the somewhat longer run, it could be assessed as to what extent a kilometre based element in these charges could be introduced, either through annual checks (e.g. inspection and maintenance control), or through an electronic device. Such systems would, in addition, give vehicle owners incentives to reduce fees by maintaining the car properly, or by driving less. Obviously, implementation and enforcement issues would have to be analyzed carefully.

- (ii) The option of increasing fuel taxes is often promoted as an effective solution because of the direct relationship between fuel consumed and distance travelled, as well as its low administrative costs. However, the relation between fuel consumption and emissions is generally weak (with the exception of CO₂), as a result of which increased fuel prices do not tend to trigger effective responses, particularly on emission controls. Moreover, increases in fuel efficiency generally imply a cost. They are, therefore, likely to be neither an effective nor a cost-effective way of reducing emissions (other than CO₂) from road transport. However, differential fuel taxes can be used to promote the consumption of cleaner fuels. The difference should, however, be based on an approximation of the actual emissions and the same value per unit emission should be used for all fuels. Examples are: the differential taxation of leaded and unleaded gasoline; fuel price surcharges based on the sulphur/heavy oil content of diesel; or lower taxes on clean fuels, such as compressed natural gas. The tax differential in favour of diesel fuel, which is found in most countries, should also be re-considered in the light of its potential environmental impact. Such a tax differential in favour of diesel in Europe has contributed to the growth of a significant market for diesel-powered passenger cars. The advantages of diesel engines regarding CO and HC emissions is, however, substantially offset by the diffusion of three-way catalytic converters. Moreover, poorly adjusted diesel engines are an important source of black smoke and fine particulates. In addition, CO₂ emissions from motor vehicles are directly proportional to the amount of fuel consumed and its carbon content. While diesel engines are more fuel-efficient than gasoline engines, diesel has a higher carbon content per litre than gasoline. The conclusion is that tax differentials in favour of diesel cannot be justified on environmental grounds.
- (iii) Older vehicles are responsible for a disproportionate share of air pollutant emissions. Vehicle scrappage programmes aim to eliminate the most polluting units of the vehicle fleet by means of inducing their destruction and/or replacement with less polluting units. A carefully designed early retirement programme, targeted at cities or regional areas that are out of compliance with air quality standards, could potentially achieve environmental benefits at costs equal to or lower than those of other emission-reduction options. These programmes can also achieve gasoline savings as a by-product. Such an early retirement programme involves equity issues and should, therefore, be based on non-mandatory participation brought about by incentives and economic instruments. Experience in some countries shows that incentives

directed towards new and cleaner cars through differentiated purchase/registration taxes can also be an efficient means to reach environmental goals. The incentive is more perceptible for the user and could, therefore, have a relatively larger impact.

The internalization of air pollution costs presents a major opportunity for improving the air quality policy. Emission fees are, in principle, the most attractive instruments to internalize the costs of air pollution in the road transport sector. Direct emission metering and charging in accordance with regional differences in environmental costs is likely to be feasible in the near future. Until then, the instruments, which should be examined more closely, include:

- (i) adjustment of the relation between taxation on gasoline and diesel fuel to better represent each fuel type's environmental performance;
- (ii) differentiated fuel taxes reflecting differences in fuel qualities;
- (iii) differentiated vehicle taxes in accordance with the environmental performance of the vehicle;
- (iv) a kilometre tax based on a vehicle's environmental characteristics; and
- (v) differentiated user charges and road tolls according to the environmental performance of vehicles.

Section II

Public Transport Pricing

Public transport in developing countries is essential for the urban poor who have to rely on walking, cycling and road based public transport to meet most of their travel needs. Urban public transport is provided mainly by buses. Competition best guarantees the efficient supply of public transport services, and through franchises and concessions can mobilize low-cost operations to provide the best quality of service and the best price for any resource capability. The informal sector can also contribute effectively to satisfy demand in competitive markets. Mass transit can contribute both to city efficiency and to the needs of the poor in the larger cities, but can impose heavy fiscal burden, and should only be adopted within an integrated planning and financing structure ensuring effective coordination of modes and affordable provision for the poor.⁴⁴

This section examines how competition may be used to procure public transport and suggests fare structures and pricing policies that meet the multiplicity of objectives.

Buses are the main mechanized form of public urban transport. The World Bank has estimated that there are 6.5 trillion passenger kilometres per year in 3 million vehicles, of which over 2 million operate in cities. In addition, there are over 2 million para-transit vehicles operating in these cities. Railways are important in the larger cities. As cities increase in size to the point at which walking can no longer satisfy the major trip requirements of citizens, public transport and cycling become the preferred modes of transport of the poor. If adequate public transport is not available, the relatively poor will shift first to bicycles, then to motorcycles, then to taxis and ultimately to cheap cars. The failure of conventional public transport may also generate a burgeoning small vehicle para-transit sector. This will also ultimately have adverse effects on congestion, air pollution and urban structure.

Urban Bus Services

In many developing countries, bus services were at one time provided by private regulated monopolies. In many cases, these monopolies were taken over into national ownership, but continued to operate as protected monopolies which have now mostly collapsed. In Latin America, they have been replaced by smaller privately owned companies operating under permissions granted by the municipal authorities. In Africa, they have largely been replaced by a fragmented small vehicle para-transit sector, while in eastern Europe and central Asia, a similar process of decline is at various stages of completion. In China, where operations remain in public ownership, the operators are adopting increasingly commercial approaches to business. Only in major cities of eastern Europe and India municipally owned bus monopolies have survived.

Gwilliam⁴⁵ argues that, although the circumstances vary from country to country, the major problems faced by the bus sector arise from the multiplicity of conflicting objectives. The primary objective, if public transport is to be sustainable, is for pricing to generate sufficient revenues to ensure an adequate, efficient and continuing supply of public transport services. Here it may be mentioned that public transport also contributes to the reduction of congestion and environmental impact of road traffic, efficient co-ordination between various public transport modes, and reduction of poverty. It is commonly argued that if urban public transport is to satisfy these objectives, it cannot be expected to cover its full costs. Urban public transport is consequently subsidized in many major cities. However, many of these cities are no longer in a position to fund such policies, and their public transport sectors are facing deterioration as a consequence

⁴⁴ Gwilliam, K. M., 2000. *Urban Transport Strategy Review – Draft*. World Bank. Washington.

of cash starvation.

This process of decline is compounded by the fact that many governments have attempted to use the public transport industry as an instrument of social policy by simultaneously constraining fare levels and structures and by guaranteeing favourable wages and working conditions to the employees.

Procurement and Competition

A possible solution to the problem of promoting sustainable public transport may be found in introducing competition into urban public transport provision. Competitive pressures may be introduced in various forms, both within the traditional monopoly and between firms, either 'for the market' or 'in the market'. Given the inherent defects of the traditional uncontested monopoly and the demonstrable potential of competition to generate cost reductions and service quality improvements, the critical issue is to establish the best ways of organizing competition to secure the strategic objectives of the urban transport system.

At its simplest, competition can make public sector operations more efficient through the competitive procurement of equipment and a range of support services thereby reducing costs and improving product quality. At the other end of the scale, privatization of bus services and competition between independent commercial companies will produce well-motivated organizations. This can be managed by creating 'competition for the market' or 'competition in the market'.

There are several ways in which firms can 'compete for the market':

- (i) *Gross cost service contracting* – involves the procurement of specified services from an operator by a public authority at a price determined through competitive tendering. Contracts are usually for 3 to 5 years. The operator passes all on-bus revenues to the procuring authority and does not take any revenue risk.
- (ii) *Net cost service contracting* – is similar to gross cost contracting except that the operator keeps the revenue and hence incurs both the revenue and supply cost risks.
- (iii) *Management contracting* – involves operator responsibility for the management of the operation of a system, possibly including service

⁴⁵ Gwilliam, K. M., 2000. *Urban Transport Strategy Review – Draft*. World Bank. Washington.

specification within agreed parameters. Physical assets are normally owned by the procuring authority, though the operator may be responsible for their procurement and maintenance as well as for negotiating labour wages and conditions. Inter-modal co-ordination is relatively easy to achieve through this mechanism, and provided the payment arrangements are well structured, there is a major incentive to provide high quality of service to attract customers. The problem is that the competitive pressure may be fairly weak, trade union power relatively strong and costs comparatively high.

- (iv) *Franchising* – involves the grant of an exclusive right to provide a service meeting a number of general quantity, quality, and price standards laid down by the authority in a universal service obligation (USO), usually as a result of competition. The franchise may be for a self-contained area, such as a town or a sector of a larger city, but it is also possible to have route franchises – especially with fixed track systems. They differ from service contracts in allowing the contractor a greater degree of freedom to develop the system. The franchisee may have to be paid by the authority to provide services and fare combinations that are not commercially viable.
- (v) *Concessions* – these involve the granting of an exclusive right to provide a service but without payment by the authority, although the authority may attach conditions, such as maximum fares or minimum service requirements. Contracts are usually for longer periods, often ten years or more, to allow the contractor to benefit from his development of the market.

The most extreme form of ‘competition in the market’ is that of a totally open market in which there are no entry barriers for transport operators. Even in such circumstances, the open market is usually associated with some form of quality licensing, which specifies minimum conditions for entry, such as vehicle specifications, environmental performance and maintenance standards. In some cases, the qualitative conditions may also cover the type of service to be operated (including stopping places), fares and trading practices.

A restricted form of competition in the market may occur where, although there may be several operators providing services in competition with one another, the total number of vehicles allowed to operate is limited by the authority. This is a very common form of regulation for taxi markets.

In circumstances where bus services suffer from chronic financial deficits, the main

options, if socially desirable services are to be maintained, are:

- (i) to deregulate the industry, where it is in public ownership;
- (ii) to introduce competitive tendering for licences to operate stipulated routes or receive a subsidy to do so. This involves the government, or an appointed agency, specifying the route and required service standards;
- (iii) a combination of (i) and (ii).

Deregulation is likely to lead to the maximum amount of route mileage being operated at a profit; the elimination of cross-subsidization; and the minimization of the subsidy requirement. However, where traffic levels are limited, a degree of market concentration and area monopolies will emerge. Subsidies could be justified on social grounds – for example, in rural areas – or on grounds related to the correction of externalities generated by other modes, such as private cars. Internalization, however, is usually preferable to compensation through subsidization.

Competition in the market gives suppliers the greatest degree of freedom to respond to consumer demand, and gives to the consumer the most direct instrument – his willingness to pay – to influence what is supplied. But market competition is not responsive to several important types of ‘market failure’. First, if there is insufficient demand to meet the costs of supply, then there will be no service, irrespective of the importance which society attaches to the provision of some basic minimum service level. Second, the market is not responsive to various ‘external effects’, such as congestion and environmental impact unless they are directly charged for. Third, because of information asymmetry and the difficulties of ‘shopping around’, the process of competition may result in a combination of price and quality of service supplied, which is not what the majority of consumers would prefer. Fourth, it will not be in the interest of the individual bus operators to adapt their services and fares to promote modal integration.⁴⁶

Having created a competitive urban public transport market, it is still necessary to understand the cost structure and establish a fares policy.

Cost of Urban Bus Services and Pricing of Services

The costs of providing scheduled bus services comprise the costs of providing both the infrastructure and operating services over that infrastructure. Once the schedule is determined, a high proportion of the costs is essentially fixed. Detailed analysis will be required to identify costs incurred for specific traffics or users. Peak users will be subject to much higher costs than off-peak users, particularly if there is excess demand.

The analysis of costs can be used to determine the minimum or 'floor rates' below which charges and fares must not fall. Total revenue collected, on the basis of such rates, would probably be insufficient to cover total costs. If it is necessary to recover capital costs, in full or in part, then market or demand-based prices need to be employed.

The costs of using the road infrastructure may comprise a combination of road license fees, tolls and fuel taxes. These are unlikely to be determined on the basis of economic efficiency and are not likely to cover social costs and externalities.

The basic characteristics of the urban road passenger markets are likely to be:

- (i) *Relatively high usage* – this arises from the relatively high density of population which makes it possible to offer a better quality of service than, say, in rural areas. Further, restraints on private car usage to reduce congestion and pollution in urban areas, such as parking charges and road pricing, combine to increase the attractiveness of public transport in general.
- (ii) *Demand is likely to be peaked around journeys to and from work* – this leads to the problem of low utilization of assets and staff in off-peak periods.
- (iii) *Relatively low journey lengths* – these produce a relatively high cost per bus mile.

Inter-urban road passenger demand patterns will vary from country to country depending on the state of the infrastructure available, population size and density, and distances involved. As journey distance increases, competition from railways and air transport will increase. Buses and coaches will primarily compete on price in an attempt to compensate for the slower journey times and higher time costs of travel by road.

If the objective of the commercial bus or coach operator was simply to maximize profits, then the decision rules would be fairly clear. The operator would raise fares, whenever demand was elastic, so long as this enabled a bigger reduction in cost, through reducing service frequencies, than bear the loss of revenue. The operator would also attempt to avoid cross-subsidization between passenger groups. However, profit maximization is not likely to be permitted where the operator has a perceived monopoly or where entry to the market is difficult. In such situations, government or municipal

⁴⁶ Gwilliam, K. M., Kumar, A. J. and Meakin, R. T., 2000. *Designing Competition in Urban Bus Transport: Lessons from Uzbekistan*, TWU Discussion Paper - TWU-41. World Bank, Transport, Water and Urban Development Department, Washington DC.

regulation of the market is likely to focus on price or fare controls.

It is realistic to assume that a bus operator will have a degree of area monopoly and will not just seek to maximize profits, but rather to behave commercially, that is:

- (i) to be financially self-supporting, subject to any subsidy provision;
- (ii) to earn a target rate of return on capital; and
- (iii) to self-finance all or part of the capital requirements.

Given the cost characteristics described above, there are a number of alternative fare structures to be considered.

Bus Fare Structures

The implication of the cost structure is that, unless prohibited by law, price differentiation or 'Ramsey Pricing' is likely to be the best means of recovering fixed costs. There are generally five main fields of differentiation in urban bus fare structures:

- (i) Time of day
- (ii) Length of trip
- (iii) Route or area
- (iv) Service type
- (v) Passenger type

Given that differing demand elasticities are the basis for differentiation in the fares charged, following are the most practical methods of discriminating:

- (i) *Time of day* – since peak elasticities are generally much lower than off-peak elasticities, there is a clear case for differentiation on the basis of time of day. Off-peak discounts might be applied only to the more elastic types of demand – for example, holders of day return tickets, shoppers, persons undertaking evening trips, etc.
- (ii) *Length of trip* – there is rarely any justification for urban bus fares to be a constant rate per mile/kilometre regardless of trip length. In the peak, heavy costs may be incurred in offsetting overcrowding on the shorter inner city sections of urban bus routes. Indeed, some buses will only make one trip at full capacity in the peak and then only for certain sections of the route. This implies that the extra costs of longer trip lengths may be confined to fuel and wear and tear costs. Generally then, a case can be made for a fare structure that tapers with trip length, where the degree of taper is determined by the variation of demand elasticities with trip length and fare.

- (iii) *Route or area* – it is feasible to differentiate fares on the basis of relative income levels in different suburban areas. If, for example, areas with higher income levels have greater car ownership and more transport options, the elasticity of demand with respect to fares may be high. Theoretically, there could be a case for lower fares for these areas. However, varying fares by route or area is not likely to be politically acceptable, particularly if it is perceived that poor passengers are cross-subsidizing more affluent passenger groupings.
- (iv) *Quality of service* – where services providing different degrees of speed, comfort or reliability operate between the same origin and destination, the difference in quality of service may lead to a difference in demand elasticities and hence in the optimal fare. Thus, it is essential to assess the cross-elasticity of demand between different modes of travel and varying quality of services on parallel routes. Examples include differentiation between suburban express or limited stop buses and stopping services.
- (v) *Type of passengers* – it is possible to segment the market by passenger types, provided this is not prohibited under anti-discrimination legislation. In terms of practical implementation, it is necessary to be able to identify designated classes of travellers, for example, by use of travel cards bearing photographs. The main passenger types for such purposes tend to be students and young people, retired people, families travelling together, and groups.

Public Transport Pricing and Equity

Public transport pricing policy is often used as an instrument of poverty alleviation. The fares are regulated in order to provide an affordable service to the poor who may have no alternative mode of travel. The mechanisms generally used in this regard are the fare levels, fare structures, fare concessions and fare discrimination.

Fare levels – If the per capita income of public transport users is significantly lower than the users of private urban transport, then it is often proposed to control fares as a means of re-distributing income. If, however, fares are set below cost and if tax revenues are not used to cover the financial deficit, then the result will be a reduction in the quality of service provided and also a reduction in, or withdrawal of, such service. It is unlikely that this will ultimately benefit the poor. Decisions on the control of fares should, therefore, be taken in the context of an assessment of the effects of the control on the sustainability and quality of the service.

Fare structure – Sometimes, municipal authorities justify a flat fare structure

throughout the city on equity grounds. The problem with such a policy is that if it requires a high fare, it may force shorter distance travellers to seek alternative modes, and hence to undermine any 'within-mode' cross-subsidization. Similar problems arise where flat fares apply across different modes. For example, identical flat fares for bus and urban rail services will lead to cross-subsidization from bus to rail and possibly a re-distribution of income from poorer to richer social groups, if the former primarily use bus services. Therefore, modal coordination, within urban public transport systems, requires an integrated fares and charges strategy which accounts for the effects on the poor.

Fare reductions or exemptions – The distribution of subsidies should ideally be explicitly related to the income levels of service users. Particular, and deserving, categories of passengers (e.g. schoolchildren and senior citizens not in receipt of a full income) can be easily identified and charged a lower price on all services. It is, however, more difficult to manage and justify the practice of providing free or reduced fares to public servants. The main problems with such a policy are : firstly, if there are large numbers of such passengers, it may lead to fare evasion by other passengers; secondly, it is the equivalent of implicit rather than explicit subsidization; and thirdly, such passengers may not be the poorest and, therefore, not the most deserving.

Fare discrimination – Where residential locations are highly segregated by income groups, specific services, particularly for journeys to work, may also be identified for subsidy on income distribution grounds. To be effective, such an approach needs to be well-targeted.

In circumstances where road congestion and other externalities, generated by road users, are widespread, it is often suggested that public transport should be subsidized to attract commuters from private to public transport. However, such a policy, as a counter-congestion measure, needs to be viewed with caution due to a number of difficulties:

- (i) Subsidies can lead to managerial inefficiency;
- (ii) Where congestion is limited to a particular time of day or particular locations, it becomes increasingly difficult to target public transport subsidies as a response to the problem;
- (iii) Public transport subsidies may lead to the generation of extra demand for travel by public transport at peak periods, thereby adding to congestion costs;

- (iv) Since the cross-elasticity of demand for private car use, with respect to public transport fares, is very low, subsidized fares are unlikely to divert traffic from private cars to public transport unless service quality is also improved;
- (v) Public transport subsidy to counter undercharging for private transport will mean that all transport is subsidized. This, in turn, will tend to generate excessive travel and sprawling land-use; and
- (vi) The fiscal cost may be significant.

Gwilliam⁴⁷ has analyzed the more complex argument for public transport subsidy, which arises when there are two transport modes with different cost structures operating within an urban system. For example, in typical bus operations, over 90% of costs vary with respect to either the number of vehicles employed or the number of bus miles operated. Besides, there are no significant indivisibilities, and short-run marginal cost pricing would, therefore, be consistent with full cost recovery. However, this is not the case for rail systems where typically only 50% or 60% of the costs are directly related to the service provided and the costs are consequently decreasing in the short run. Since the relationship between average cost and marginal cost differs between the modes, setting prices equal to marginal social costs would yield different levels of cost coverage and some cross-subsidization between the modes would appear to be appropriate to secure an efficient modal split within the public transport sector, i.e., the modal split which would occur if both modes set fares at short-run marginal cost. If this fails to recover long-run costs, and if cross-elasticities between alternative transport forms are zero, cost recovery could be achieved by adding mark-ups to the marginal social cost in proportion to the reciprocal of the price elasticity of demand. This is the Ramsey inverse elasticity mark-up rule.⁴⁸

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 road pricing ensured that the road users paid the marginal social cost of their journeys, including infrastructure costs and externalities, there would be no case for subsidization of public transport on grounds of 'second best'. In such a situation, public transport operators should be asked to charge all passengers a price equal to marginal cost and recover long-run costs in an economically efficient manner. In the circumstances, where the road infrastructure is not properly priced and subsidies to public transport exist, the question arises as to how best to improve the situation. If subsidies are removed

without introducing efficient road pricing, modal split will be further distorted. However, if due to fiscal problems, a government cannot maintain the subsidization of public transport, then price controls should be relaxed and measures to internalize the externalities created by road users introduced, if operators are to be financially sustainable. Price differentiation offers urban transport operators a means of securing both an economically efficient level of output and a specific financial objective.

Strategic Planning and Policy

Urban transport infrastructure and public transport pricing have strong interdependencies and, therefore, any 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 different modes and the widespread 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. Indeed, it may be optimal, for reasons of 'second best', to facilitate financial transfers between the exchequer and public transport services, or, between roads and public transport services, or, between modes of public transport. However, in order to secure the efficient supply of public transport services, providers should operate competitively, with purely commercial objectives, financial transfers being achieved through contracts between municipal authorities and operators for the supply of services.

If any non-commercial objectives are imposed on operators, then they should be compensated for directly and transparently imposed objectives. In case public transport cannot be subsidized to compensate for the inadequate road pricing policies, then financial sustainability of the public transport service should take precedence over price or fare regulation.

⁴⁷ Gwilliam, K. M., 2000. *Urban Transport Pricing and Financing*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

⁴⁸ With non-zero 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. 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. But that would cause some travellers to shift from peak to off-peak. The structure of consumption would then have diverged from that resulting if price were everywhere equal to marginal cost.

4

Pricing and Charges for Railways

Introduction

Railways display most or all of the complexities that are encountered in devising efficient transport pricing policies. These include the problems of high fixed and joint costs, indivisibilities of supply and demand, peak loads, and externalities, particularly congestion. In addition, the infrastructure comprises expensive and durable assets which provide a significant barrier to entry as also an incentive to natural monopoly formation. In consequence, the industry is highly regulated and usually in competition with other modes, particularly roads, which are subject to entirely different regimes in respect of cost recovery and which generate numerous uncompensated externalities.

This chapter examines the competitive context of the railways and the scope for private sector participation in both the provision of infrastructure and services. It also includes an analysis of the problems of allocating track costs or calculating access charges to train operators and identifies methods of charging passengers and freight users for rail services. The chapter comprises three sections: section I describes the competitive context of railway pricing decisions; section II examines pricing methods for railway infrastructure; section III identifies pricing methods for railway passenger and freight services.

Section I

Railway Pricing : The Context

The railway industry is characterized by chronic financial deficits, growing operating subsidies, excessive costs, inefficient pricing structures, severely congested services, low operating efficiencies, inadequate funds for investment and deficiencies in physical infrastructure. Other notable features are widespread state ownership and monopoly in provision of infrastructure, operations and services. These characteristics are often the cause of economic and financial unsustainability of the rail systems. Table 4.1 identifies the generic causes which give rise to these characteristics.

Table 4.1 : Railway Problems and Causes

Problems	Causes
Chronic financial deficits	Constraints on charges imposed through government regulation; persistent excess capacity; provision of guaranteed service levels at fixed prices or with 'excess' competition; provision of services at below marginal cost; failure to understand or identify costs; ineffectiveness in collecting revenues; low productivity; unduly high operating costs; overmanning.
Growing operating subsidies	Chronic financial deficits; lack of corporatization; inadequate distinction between the role of the government and that of the railway operator; unsustainable subsidy policies.
Archaic pricing structures	Prices are not related to marginal costs; costs not properly identified or measured; inadequate financial and management accounting systems; inadequate or non-existent objectives of pricing policy.
Lack of equitable fare structure and excessive fares	Lack of user or community representation in service and price decision-making; public or private monopoly.
Excessive costs; low managerial and technical efficiency; low productivity	Lack of competition or existence of a 'natural' monopoly; overmanning; inadequate funds for investment.
Low service quality; congested services; services failing to respond to needs.	Lack of competition; no peak-load pricing; inadequate cost recovery in pricing policies; inability to reinvest operating surpluses or raise funds for investment.
Deficiencies in physical infrastructure; insufficient investment funding; lack of proper maintenance of assets	Failure of pricing policies to recover capital costs; structural inability to retain/reinvest surplus funds; constraints on investment or borrowing.
Widespread state ownership of infrastructure and services; low private sector participation.	Lack of policy or strategic commitment to competition/corporatization/privatization.

The solution to these numerous problems could possibly be found in creating a competitive 'market-based' railway industry.

'Market-Based' Railway Industry

Since absence of competition is one of the major causes of railway's problems, it can be argued that the best way to align consumer needs and demand with the provision of railway services, in a manner which promotes economic and financial sustainability, is to create a competitive 'market based' railway industry.

Since there are significant barriers to entering the market for railway services and the efficient scale of operation is large, it is relatively difficult to create 'competition in the market'. One possible way forward is to create 'competition for the market', which can be described as developing private operations within a framework of public regulation and control.

There are a number of reasons why it may be sensible to retain a degree of public control over the right to supply railway services:

- (i) Duplication of rail operators on a given route could be wasteful or impractical. The existence of indivisibilities in capacity provision could lead to the emergence of a 'natural monopoly' with its associated adverse consequences.
- (ii) Unregulated competition could lead to undesirable practices, such as frequent timetable changes and volatile fares.
- (iii) Direct competition could lead to the discontinuance of particular services designed to benefit poorer communities, as they would not be viable without cross-subsidization or government grants. In such circumstances, it may be desirable to create competition for the right to provide subsidized services at least cost.

Such imperfections give rise to the need for control, but do not justify continued state operations or the granting of monopoly franchises. As a matter of fact, there is considerable scope for private sector management in the railways.

Competition in service provision can be effected through the selling of route franchises for both profitable and unprofitable railway routes. Regulation over safety, service quality and prices can be retained, while using competition to secure the lowest cost operator for a fixed time period. Further, introducing different operators

on the same or competing routes and encouraging competition with alternative modes can produce significant benefits.

The creation of competition in railway infrastructure provision is more problematic. The competitive award of long-term concessions, licences, or leases of facilities, such as stations and permanent way is the primary means for introducing market forces into the provision and management of railway infrastructure. This can be structured with the objective of stimulating efficiency by transferring risk to the private sector.

The effectiveness of railway infrastructure concessions depends on the skills of governments in designing and implementing contracts. Auctioning concessions to the highest bidder will give an incentive to the most cost-efficient and market-oriented operators, but is likely to provide the successful firm with a monopoly position. The alternative is to invite bids on the basis of the lowest price for the provision of a specified quality of service. This will require a high degree of sophistication in bid evaluation.

There is scope for introducing private sector financing and funds into the building and maintaining of railway infrastructure by allowing this sector to assess and retain income flows from train operators, who themselves may or may not be in receipt of government subsidies. To some extent, therefore, it is likely that private sector participation in railway infrastructure provision will involve some form of private and public sector partnership.

Finally, management efficiency can be increased by corporatizing the agencies responsible for providing the railway infrastructure. The discipline brought about through acquiring a commercial remit with accountability for prices and hence revenue and expenditure, can drive costs down and quality up dramatically.

Section II

Railway Infrastructure

The operators of railway services may either be directly responsible for the provision and maintenance of railway infrastructure or, alternatively, buy access rights to use the infrastructure. Infrastructure includes stations, terminals, track and signalling, which are characterized by longevity, joint use, scale economies and indivisibilities.

As a consequence, the pricing of railway infrastructure is difficult both in conceptual and practical terms. The problems to be addressed are the same regardless of who is responsible for supplying the infrastructure. In cases where a single entity provides both infrastructure and train services, the task is one of cost allocation. On the other hand, in cases where separate entities are involved, it is a matter of determining the charges to be levied by the entity responsible for infrastructure for access to the rail tracks and stations by train operating companies.

Economic Principles

Any charging framework should have the following key characteristics:

- (i) *Comprehensibility* – the structure should be understood by the industry participants whose behaviour it is meant to influence, and should not impose undue transaction costs to identify the appropriate information.
- (ii) *Transparency* – the structure should provide clear information to industry participants on the make-up of charges, and, hence, not confer undue advantage on particular industry participants, e.g. through information asymmetry.
- (iii) *Stability* – charges should not fluctuate or alter in arbitrary or unpredictable ways, except where significant short-term cost changes are being signalled – if congestion (scarcity) pricing is introduced, short-run prices could be unstable, but predictability about future average levels could be given in some cases by establishing a long-run avoidable (marginal) cost around which short-run prices might be expected to fluctuate;
- (iv) *Measurability, cost-effectiveness and objectivity* – the data required to derive charges should be objectively measurable, cost-effective to collect and unambiguous to apply (for billing purposes).
- (v) *Cost-reflectivity* – in order to meet the objective of economic efficiency, charges will need to be cost-reflective.

In principle, the costs that underpin infrastructure charges should consist only of those elements which are relevant to the specific pricing, investment, or operating decisions under consideration. Relevant costs can, in general, be divided into variable costs which vary with output, such as maintenance, operations and replacement, and fixed costs, which are incurred whether more or less traffic is carried. Use of parts of the network by more than one user can result in costs that cannot be directly attributed to any one particular user, the ‘common cost problem’.

A distinction needs to be made between price signals based on variable costs which guide operating decisions – for example, whether to operate six car trains instead of four car trains – and charges to recover fixed costs, which do not vary with output and which need to be levied in the least distortionary way.

Price signals for the efficient production and allocation of railway infrastructural resources should be based on the avoidable (marginal) costs of changes in the use of the existing network and changes in the network itself. There are a number of characteristics of the railway network which result in avoidable costs, varying according to the place and the time period in consideration. In particular, railway infrastructure is intensive in assets which are ‘lumpy’ to install and renew, with long economic lives. This means that, in practice, charges may need to signal the corresponding avoidable costs associated with significant and sustained changes in demand, in order to generate appropriate practical measures of incremental costs which at the same time provide meaningful investment signals and incentives.

However, since capacity is indivisible and fixed in the short term, in case the charges were to be based on long-run costs and, specifically, where these are lower than those based on short-run costs, this would lead to demand exceeding supply in the short run. Where there is excess demand, the price mechanism by itself may not be able to balance supply and demand without very high charges in the short term.

Efficient costs should be forward-looking to reflect what is anticipated about the future results of a decision to change prices or operations, or undertake investments. Efficient use of the network could require decremental investment where, for example, the benefits of keeping a line in use are outweighed by all the costs, private and social. The kinked nature of the supply curve could mean that there is a large difference between the incremental costs of enhancing capacity and the decremental costs of reducing capacity.

Other characteristics that need to be taken into account are:

- (i) existence of multiple users of parts of the network with different requirements (e.g. access right qualities);
- (ii) lack of any direct relationship between access charges and final demand, i.e., passenger fares which are determined by the train operating company; and

- (iii) the fact that there is no market based mechanism for valuing some elements, such as congestion costs and access right qualities.

The starting point is that charges should reflect the incremental (or avoidable) costs involved. Incremental costs are the avoidable costs of an incremental change in output. Both terms are used synonymously, but it should be noted that incremental costs can be different from decremental costs because of the kinked nature of the supply curve. Depending on the particular decision under consideration, these can relate to the costs of changes in the use of the existing network or to the costs of changes in the network itself. Avoidable costs may cover such items as:

- (i) increases or decreases in operating costs (e.g. signalling staff);
- (ii) increases or decreases in maintenance and renewal costs;
- (iii) costs of change and disruption, and compensation payable to other parties during construction; and
- (iv) capital costs, including the value of additional land used, and an appropriate return on the capital costs.

These costs are forward-looking and represent the costs an efficient network would incur. They are not simply avoidable, they are the most efficient costs that can be avoided. In the short run, some elements of the first three costs may be avoidable, but in the long run all costs are avoidable. This is an important distinction because short-run avoidable costs will differ from long-run avoidable costs, and the use of one rather than the other will give different price signals.

Indivisibility (lumpiness) of assets with growing and changing demand can lead to:

- (i) congestion costs – as the network becomes more crowded, there is less flexibility to recover from the effects of delays; and
- (ii) opportunity costs – the costs of having slots occupied by lower valued services in place of higher valued services .

These costs need to be accounted for if the railway network is to be used efficiently, but, if fully reflected in charges, they could lead to a wide divergence between short-run and long-run costs and the associated price signals, particularly where capacity is lumpy.

One further point about congestion pricing is that it may, in the absence of capacity enhancement, yield supernormal profits, essentially pure (scarcity) rents to the ownership of the congested facility. It is necessary to ensure that these profits are used to offset other charges.

Route-Based Charges

Incremental costs are a function of the capability and condition of the existing assets, and are likely to vary systematically by route on the network. For example, variable costs can vary by route depending on the quality of the track; and fixed costs, that are user specific (i.e., not common), can also vary by route. There are costs that do not vary with output, but can be attributed directly to specific users where there is no shared use – for example, single use of a branch line. Route-based charging or other geographically specific charging signals would potentially be appropriate to reflect cost variations and avoid undue discrimination.

Short-Run or Long-Run Costs

Rules can be identified for determining whether the appropriate charge is based on short-run or long-run costs. These include:

- (i) using the higher of short and long run costs in all circumstances; or
- (ii) reflecting the nature of the underlying demand by using contract length as a proxy, e.g. long-run costs for long-run contracts and short-run costs for 'spot' contracts.

It can be argued that the appropriate cost to be used should be the greater of short and long-run incremental cost. This is because, if long-run costs are below short-run costs and the charge is based on long-run costs, excess demand will occur in the short run. In the absence of any price rationing, some form of quantity rationing will be needed which could result in efficiency losses (if users do not, or, are not able to, express their preferences). However, if the charge is based on short-run costs, unless all are aware that the costs are short-run, there could be over-investment resulting in excess capacity with potentially damaging re-adjustment costs.

Alternatively, the charge could attempt to reflect the persistence of the underlying demand, the forecast of future use of the route over the long term, as opposed to contracted use for the remaining part of a franchise. Short-run incremental costs could apply to short-term contracts, and long-run incremental costs could apply to long-term contracts (or medium-term contracts with trading

rights) according to the needs of the customer. Short-run incremental costs could apply to demand management (use of the network), while long-run incremental costs could apply to capacity enhancement (development of the network).

If the system of franchising had contracts that did not reflect the nature or persistence of the underlying demand, it would be restrictive to relate the type of costs to the contract length and this could lead to ‘gaming’ of contract length by operators seeking lower charges.

In order to prevent sending out signals that could be misread, it is better that, irrespective of the length of the contract, only costs that reflect the nature of demand are taken into account, whether they are short-run or long-run. In this case, congestion costs would reflect short-run avoidable costs for short and long contracts, while capacity costs would reflect long-run avoidable costs, and either would be charged depending on the level of demand, given existing supply.

There are a number of options for the charging structures appropriate for conveying these signals. One option would be to publish cost-based short-run charges, with a robust forecasting methodology that enabled users to chart their expected charges over time as the underlying short-run costs concerned, including congestion and opportunity costs, would vary as demand and capacity changes. Another option would be to publish indicative long-run incremental costs. Charges based on short-run costs would be expected to average over the expected lives of the assets concerned. These long-run costs would be relevant for long-run access contracts. Hence, if the evolution of short-run incremental costs could be predicted, they could be reflected in the published indicative long-run incremental costs drawn from an optimized investment plan. A third option could be to base charges on the average expected short-run incremental costs discounted over the life of the access contract. This could fit well within any system in which contracts of different lengths could be entered into and where charges for a contract of a given length would reflect the short-run incremental costs discounted over the life of the contract. This would offset the gaming problem mentioned above, if the prices in the contract reflect the smoothed equivalent of the relevant short-run avoidable costs, and if future prices are stated and believed to be based on short-run avoidable costs or their smoothed equivalent for a relevant contract length.

Charges based on incremental costs, including capital costs annuitized over the life of the assets, could completely replace short-run costs as the cost floor for variable charges. Alternatively, they could be used to underpin the cost floor of

charges fixed for the duration of access agreements (as they relate to investment commitments to support demand levels expected to be sustained for at least the duration of most access agreements). The choice of mechanism should depend on the assessments as to how elastic demand will be in response to the price signals concerned in the short versus the medium term.

The size of the increment is an important issue. Too small an increment will result in most costs being treated as non-avoidable, while too large an increment will result in all costs being avoidable. Where routes have more than one user, it is likely that the incremental cost of all users of the route will exceed the total of the incremental costs of each user resulting in a common resource problem, i.e., the allocation of common costs. The appropriate size of the increment needs to be considered in the light of the circumstances, such that the common resource problem least affects final demand.

Incentives for Investment

Investment should only take place where there is an expectation of a reasonable rate of return. This will largely depend on demand signals and how persistent they are. Changes in congestion costs would reflect levels of existing demand relative to the capacity of the network at a particular point, but they may reflect only short-term variations in demand.

Where capacity can be increased and a persistent demand is forecasted, there will be a need to identify the likely value of additional capacity to determine whether there is a case for enhancement to the network. Where some high value flows are currently not able to operate because of insufficient access rights, this value may exceed avoidable costs, including congestion costs. Unless there is some form of secondary trading of access rights, this demand is unlikely to be expressed in terms that would stimulate changes in capacity.

Section III

Passenger and Freight Services

In devising pricing policies for railway passenger and freight services, it is necessary to understand both the nature of the cost structure and the characteristics of the market⁴⁹. It is also important to recognize that significant differences exist

⁴⁹ Bhandari, L., 2000. *Rail Pricing and Efficiency*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

between the urban and the inter-urban markets. The key features of the fare structures of these markets and pricing policies that may meet various objectives, are similar to those discussed in the preceding chapter dealing with road and urban transport.

Passenger Services

The cost of providing railway passenger services comprises the costs of providing both the infrastructure and operating services over that infrastructure. Detailed analysis will be required to identify costs incurred for specific traffics or users. Peak users will be subject to much higher costs than off-peak users, particularly if there is excess demand.

Urban Railways

The basic characteristics of the urban rail passenger markets are similar to those of the urban road transport, already described in the preceding chapter. To repeat, these are: relatively high use because of high density of population; low utilization of assets and staff in off-peak periods; and high cost per train kilometer because of relatively low journey lengths.

The pricing objectives of the railway operators are also no different than those of the road transport operators. On account of their monopoly (partial or full) in the market, they would be inclined to raise fares, whenever the demand is elastic, in order to maximize profit. However, this is not likely to be permitted, and the government would step in to regulate the market through price or fare controls. The train operators will, therefore, have to act commercially and aim at a target rate of return so that they can be financially self-supporting.

The fare structure of urban railways also tends to follow the same pattern as that of the urban road transport – given the elasticities of demand, fares could be differentiated on the basis of ‘time of day’, ‘length of trip’, ‘route or area’, ‘quality of service’, and ‘type of passengers’. The characteristics of each of these structures have been elaborated in the chapter dealing with road and urban transport.

Inter-Urban Railways

Most of the world’s railways operate a fare structure for inter-urban rail services based on distance, usually with a tapering charge per mile as distance increases. Sometimes, a degree of differentiation by route is operated with ‘notional’ distances being employed on specific routes. Some countries, however, operate a

policy of price differentiation as a means of recovering fixed costs. Differential pricing is consistent with the public interest in the economical utilization of resources.

The three most important dimensions in designing a differentiated fare structure are: time of travel (if load factors vary by time of day or day of the week); journey purpose; and personal characteristics (income and/or socio-economic group). Whilst the first of these is straightforward, the other two are not. It is, therefore, necessary to seek proxy measures, which may reflect journey purpose and personal characteristics and may also estimate the elasticity of demand with respect to each of them.

The main factors to be taken into consideration for the purpose of designing the fare structure are:

- (i) time and/or day of travel;
- (ii) time and/or day of return travel (leisure trips tend to take place across weekends);
- (iii) quality of service (fast and convenient services are likely to attract a higher proportion of business travellers than other types of passengers);
- (iv) quality of competition (strong competition will lead to higher price and service elasticities);
- (v) type of persons travelling (a limited degree of discrimination by age is possible for children, students and pensioners); and
- (vi) number and characteristics of persons travelling together (e.g. family groups, including children).

Such factors provide scope for differentiating rail fare structures for inter-urban rail travel. The highest fares can be charged to business passengers travelling alone in first class during peak times on weekdays. On the other hand, the lowest fares would be charged to students or children, travelling second class, at off-peak periods, perhaps in groups. This is the logic behind the range of tickets offered by operators of inter-urban rail services. The range can include day, weekend, and period return fares, family tickets, student and pensioner reductions, and special off-peak bargains. Operators have to determine how complex the structure needs to be. For example, it may be that differentiation by time of day, etc. is sufficient to achieve the pricing objectives. Alternatively, it may be necessary to use the

purpose of travel and personal characteristics as the key factors in determining the structure. It may perhaps be sensible to combine both the approaches.

A number of complexities arise in determining fare structures. Judgements need to be made on the number of classes of travel to be offered with their varying load factors. Further, the elasticity of demand needs to be carefully assessed; for example, the leisure traveller making a day return trip during off-peak periods may be regarded as quite demand elastic with respect to price. However, over longer distances, there may be little or no effective alternative mode of travel available to the traveller and this might increase the inelasticity of demand. On services where there is a capacity constraint, the marginal cost of expanding capacity will influence the fares in the long run; so too will demand elasticities. It is quite possible that a high-quality inter-urban rail service would have lower marginal costs but higher peak fares than a low-quality one. As such, differentiated fare structures need to be carefully designed.

Railway Freight Services

Railway freight services can be provided as a common carrier on the basis of a scheduled service; as bulk carriage for a single client; or on an own account basis. When assets are valued at replacement cost, the dominant element in trainload freight costs (i.e., bulk freight or where one consignment makes up a complete train) is almost certainly the capital charge on locomotives, wagons and terminals. Thus, asset utilization is crucial. The most favourable type of traffic, therefore, comes in regular, large consignments in sufficient frequency for rolling stock to be permanently committed to the traffic, even if this means empty back-hauling. Costs are usually minimized by operating the largest and heaviest trains permitted by the infrastructure. Length of haul has less influence on unit costs where there is no transshipment between modes as, for example, on port to industrial plant merry-go-round services. Irregular or low frequency trainload traffic will give lower asset utilization, and if traffics have to be combined, there will be additional marshalling and empty-running. Nevertheless, such bulk traffic will probably be unattractive to road hauliers despite their lower fixed costs and greater flexibility.

Where consignments are too small to make up a complete train, the only way a railway can handle the traffic is by combining consignments for different origins and destinations. This involves trip working and marshalling which reduces utilization and increases delays and hence unit costs. Wagon load costs will vary according to the amount of trip working, marshalling and route length, since the proportion of the former will decline quickly as route length increases. Economies

of scale are critical to competitive advantage in rail transport. As traffic flow on any particular route increases, it can be handled in longer trains, with less intermediate marshalling or changing of trains and with better utilization of resources, especially track and signalling.

Provided the market is large enough, there is nothing inherently wrong, in terms of cost structure, in promoting a competitive rail freight industry. In the bulk freight sector, there is no real evidence of economies of firm size. Only where firms offer common carrier services, carrying goods for different customers, is there a major risk of wasteful competition. It is, thus, in the parcels business with its high fixed costs that competition could lead to a duplication of services and low load factors. For bulk traffic, there is no real objection to allowing shippers to choose between alternative rail freight operators, with the alternative of using their own trains that are available. Competition should lead to the minimization of average costs of providing the desired standard of service. Prices will be based on average cost which, under conditions of constant returns to scale, will reflect marginal cost. Government regulation may be necessary to ensure that cost pressures do not compromise safety standards either in terms of maintenance or in working practices.

Common carrier rail services suffer much higher fixed or inescapable costs and may, thus, lead to the emergence of a natural monopoly, though the extent of potential monopoly power may be limited by competition from other modes. Because of economies of scale, efforts to cover total costs may lead to a divergence between price and marginal cost. This problem is most marked in the case of infrastructure costs, particularly where certain services regularly have spare capacity to meet peak demand or maintain service quality. A similar problem is posed by the long time-scale of cost avoidability; traffic may be worth retaining for a substantial period of time, even though in the long run it cannot pay for the renewal of the assets involved.

There appear to be three solutions to this problem for the rail freight operator:

- (i) he could base prices on marginal costs with the resulting deficit being financed by government subsidy;
- (ii) he could base prices on average cost, but this could only be sustained with a degree of protection from competition vis-à-vis other modes; and

- (iii) he may seek to discriminate between traffic flows according to the shipper's willingness to pay, in order to cover the difference between average and marginal cost.

The advantages and disadvantages of each approach are as follows. In case (i), the government needs to be able to estimate the volumes and corresponding marginal costs conforming to the optimal allocation of traffic in order to determine the appropriate subsidy. If it does not do so, then it will revert to deficit financing with no real ability to exert financial discipline on the rail freight operator. The financial deficit will have to be met either by taxation or by diverting other forms of government expenditure. The opportunity cost of this expenditure will need to be assessed and investment in the rail system evaluated using social cost-benefit analysis.

In case (ii), the government will need to become involved in the day-to-day administration of the freight service and will have to employ arbitrary methods to allocate trains to traffic and establish priorities. Again, investment in the rail system will need to be evaluated using social cost-benefit analysis. By contrast, approach (iii) largely eliminates the regulatory role of government. This is because in this case the allocation of freight is determined by the market and long-term planning and investment can be made on the grounds of profitability. The basis for discrimination is usually the class of freight determined by value and the extent of competition and alternative modes of transport.

In summary, price differentiation offers train operators a means of securing both an economically efficient level of output and a specific financial objective. In the real world, no clear relationship between prices and average or marginal costs should be expected. Most fare decisions involve assessing the extra traffic and revenue from a particular discounted fare offer against the loss of revenue from existing passengers transferring to the lower fare, thereby diluting the benefit.

5

Pricing and Charges for Maritime and Air Transport Sectors

Introduction

This chapter examines the factors that determine pricing policies in the maritime and air transport sectors. Both terminal infrastructure in the form of ports and airports as well as shipping and airline services are covered. There are many parallels between the two sectors, but there are numerous differences or unique attributes as well. The chapter comprises four sections: section I examines port and inland waterway pricing for ship and cargo services and facilities; section II covers maritime transport pricing in the charter and liner markets; section III examines airport pricing, including that for air traffic related externalities and section IV deals with air transport pricing, including yield management.

Section 1 Ports and Inland Waterways

Ports

Ports constitute a composite system comprising of facilities and services designed to provide an interchange between water and land transport systems. The facilities or infrastructure comprise the physical assets, such as breakwaters and quays, while port services cover the efficient transit of ships within the port and the transfer of passengers and goods between ships and inland transport (road, rail and inland waterways). The main facilities and services provided by ports include:

- (i) *Ship Arrival/Departure* : Navigation aids; Approach channel; Pilotage from outside the port; Lock (if any); Protected water; Port pilotage; Towage; Berthing/Unberthing.
- (ii) *Quayside* : Opening/Closing of hatches; Breaking out/Stowage; Cargo handling on board.
- (iii) *Cargo/Container transfer to/from quay*

- (iv) *Cargo Arrival/Departure* : Cargo handling on quay; Transport to/from storage; Storage; Delivery/Receiving.

Similar facilities are provided in respect of passenger services. Ports also provide a number of additional services. Services to ships include radar surveillance and traffic management; water, telephone, stores and fuel; police and security; repairs; fire-fighting; waste disposal; and medical services. Services to cargo include warehousing, security, weighing, lighterage, and rent of equipment.

This section examines the main constraints on ports in devising sound pricing policies, defines possible port pricing objectives, and identifies the factors to be accounted for in devising a rational port pricing system.

Constraints on Port Pricing

The provision of port facilities and services gives rise to flows of costs, benefits and revenues. The costs refer to the real or economic costs of the resources employed to provide port facilities and services. The resources include capital, land and labour. The measurement of port costs can be quite complex, particularly if labour would be otherwise unemployed or if the land has no alternative use. Further, once built, breakwaters and quays can be used for little else and have no economic or opportunity cost, particularly if excess capacity exists.

The benefits of port facilities and services accrue to ship and cargo owners. The former enjoy greater profits, at a given freight rate, when port improvements occur, and the latter benefit from lower shipping costs to the extent the freight rates fall. Once expressed as a financial flow, the benefit can be re-allocated through the pricing system.

The revenue to the port authority is that part of the benefit created for cargo owners and shipowners which the port authority can tap and retain. Port revenue cannot exceed the extent of the benefit generated.

The main constraint on creating a sound port pricing policy is the need to recover costs through revenues. This constraint may apply at the level of the port entity as a whole, or at the level of individual facilities and services. The costs may be defined either as the combined capital and operating costs or as only operating costs. In practice, it may not be possible to recover total costs; for example, where overcapacity exists due to inaccurate demand forecasts, low initial utilization, or indivisibilities in capacity provision.

A further constraint on pricing policy may arise from the structure, ownership and

accountabilities of the port authority and the organizations providing facilities and services within the port environs. Ports possess a degree of locational monopoly, which may give rise to the potential for monopoly pricing. Although some countries use this to justify state ownership and operation of ports, many countries have managed to create strong private sector involvement in the provision of both port infrastructure and port services.

Objectives of Port Pricing

For promoting sustainable development, a sound port pricing system should have three main objectives: (i) to recover the costs incurred in providing the services; (ii) to promote economically-efficient utilization of port assets; and (iii) to ensure that a fair share of benefits derived by customers can actually be captured by the national economy. Specifically, port pricing can support these requirements in three ways:

- (i) By measuring the aggregate demand for each service and its social profitability. This will determine whether effective demand warrants the continuation, expansion or contraction of port services and facilities.
- (ii) By encouraging the use of excess capacity and rationing capacity when excess demand exists.
- (iii) By providing information to support the coordination of long-run and short-run decisions so that total system costs are minimized and the surplus of value over cost is maximized.

To achieve these objectives, port prices need to be related to marginal social cost of the resources used to provide each service.

There are many other possible objectives which may be assigned to a port pricing system. Prominent among these is the need to retain the benefits of port improvements within the country that undertakes the investment. A further pricing objective which deserves attention is that of building up financial reserves. This implies charging prices above resource costs which, in turn, may produce lower utilization than is desired⁵⁰.

Port Pricing Systems

There is a wide range of port pricing systems in operation around the world⁵¹. These can be broadly divided into 'port dues' which are charges for the use of the port facilities as a whole, and 'specific port tariffs' which are charges payable either by

⁵⁰ Juhel, M. H., 2000. *Port Pricing and Charges*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

shipowners or cargo owners for specific services.

Port Dues

Port dues comprise dues on cargo and on ships. Dues on cargo are generally calculated on the basis of the volume or weight of the cargo. Dues on ships are usually calculated on the basis of gross registered tonnage, net registered tonnage, or length of ship.

Specific Port Tariffs

Specific charges are many and varied, mainly including the following:

Berth Occupancy – may be charged on the basis of tonnage or ship and quay length. Normally, the charge is levied on a time basis, such as per day.

Aids to Navigation – are normally charged on the basis of ship's size and are made for a given period of time or number of visits.

Berthing/Unberthing – are normally charged on the basis of ship's size or per operation.

Pilotage – is charged on a variety of bases, including vessel draught, ship size, or a combination of tonnage and distance piloted.

Towage – may be charged on the basis of the characteristics of the ship, such as its size, or that of the tug, such as its power. In the case of the latter, the charge may be defined either per operation or per unit of time (e.g. per hour).

Pilotage – may be charged on the basis of ship size; per operation; draught; distance piloted; or, a combination of these.

Storage and Warehousing – most ports offer a free period for cargo or container storage awaiting transit. Thereafter, the charge is normally derived on the basis of length of stay combined with either the characteristics of the cargo or area occupied.

Cargo-handling - most cargo-handling firms charge on the basis of weight or, occasionally, by volume. In addition, the tariff is often classified by cargo type.

⁵¹ UNCTAD, 1975. *Port Pricing*.

Due to inertia and lack of proper accounting records, many port pricing tariffs bear little relation to costs.

A Cost-Based Port Tariff

Two major problems arise in relation to devising a cost-based port tariff. The first is to determine which expenses are to be covered by prices, and the second is to decide how these should be covered. Deciding the method of recovery also involves choosing a unit (such as, ship type, gross registered tonnage (GRT), net registered tonnage (NRT), or freight tons handled) upon which the charges are to be based.

It is possible to categorize expenses into five categories:

- (i) immediately escapable costs (short-run marginal cost);
- (ii) joint costs;
- (iii) common costs;
- (iv) inescapable costs (fixed costs);
- (v) social costs (externalities).

Social costs can actually be sub-divided into items (i) to (iv). They represent the difference between the usual market price and social opportunity costs, including an allowance for indirect costs and benefits to third parties and amenity values. The most common divergences relate to port congestion and difference between wages to port labour and the social cost of unskilled labour.

Immediately escapable costs should form the basis of the minimum charges in a cost-based port tariff. Port users should be required to meet those expenses which could be immediately avoided if they did not receive the service.

Although the joint and common costs can be recovered in a number of ways, there is a rationale in employing price discrimination to allocate the indivisible expenses 'according to what the traffic can bear'. This principle is consistent with the notion of 'he who benefits, should pay'; it recovers the indivisible expense from beneficiaries without the need for cross-subsidization or general revenue subsidies. It also encourages the maximum utilization of capacity; the individual users contribute to these expenses in proportion to the magnitude of their individual benefits.

The principle of price discrimination can also be applied to inescapable costs. The

main difference is that whilst joint and common costs are escapable, inescapable costs are fixed even in the long term. Therefore, in the case of joint and common costs, the relevant assets should only be renewed if users are prepared to pay sufficient revenues to meet the costs of renewal. However, fixed port costs are inescapable and should continue to be used towards pricing, whether or not the revenue paid by users covers the historic or replacement cost of the port assets. Such assets may have been made on the basis of inaccurate assumptions about demand and revenue.

The issue of social costs is more problematic. The theoretical solution of the problem is to calculate port dues and charges on the basis of social accounting prices, with due allowance for indirect costs, such as congestion, and for the government to meet any resulting financial deficit. However, such an approach creates a number of major difficulties, including the following:

- (i) it adversely affects the government's budgetary planning;
- (ii) it poses complex management accounting problems;
- (iii) it distorts competition between private and public sector infrastructure and transport operators; and
- (iv) it could discourage rigorous investment and disinvestment decision-making.

Since the demand for port services is generally inelastic, it is preferable to adopt a policy based on price discrimination. It will, however, be necessary to trade-off the need for tariff simplicity against the need to differentiate extensively charges by cargo type. Indeed, ports may use the concept of a 'promotional due' to provide temporal variation in the tariff and market-testing.

Calculation of Port Charges

The first step in calculating port charges is to determine port costs. In this respect, the various port services and facilities which give rise to port costs have to be classified in a way that facilitates cost analysis. The identification of appropriate 'cost centres' is the best methodology. A 'cost centre' is an accounting device for allocating costs and building up the pricing structure. To avoid cost misallocation, the definition of the cost centre should be based on the following criteria – the service provided; the location where the service is given; the duration of the service provided; and the user of the service.

The second step is to calculate the specific costs for each cost centre using the principles set out above. In particular, a time horizon relevant to the pricing decision

needs to be determined – this will normally be one year. Besides, it will be necessary to identify whether costs are escapable or inescapable within that period of time. There will be a significant amount of overheads, which cannot be allocated to specific cost centres and which will need to be recovered in some way.

The third step is to identify ‘revenue centres’ – these are accounting devices which allow the grouping of all revenues of the same nature. Cost centres and revenue centres should be linked together so that the extent of cost recovery can be observed.

The fourth step is to collect information on the utilization of assets corresponding to a given cost centre. Thereafter, the desirable level of such utilization over the port’s life should be considered and the extent to which port charges can contribute to the improvement of asset utilization should be assessed. For example, periods of excess demand may be managed by surcharges based on the period of occupancy or use of particular facilities.

The final step will be to determine the charges and charging basis by cost centre. It is probable that an element of price discrimination will be required to cover total port expenditures.

Cost-Based Dues on Vessels

It should be recognized, however, that about 90 per cent of port services are commercial and privately provided. If competition in these areas is limited, a port authority may have to discharge a regulatory function.

To cover those immediately escapable costs, which do not vary with the size of the ship (or classes of ships), a fixed charge – or one related to the duration of service where this varies and affects escapable costs – would be appropriate. This applies mostly to pilotage and towage. When immediately escapable costs vary with the ship size, alternative charging bases would be GRT (a measure of the volume of enclosed spaces), length, beam, or draught.

For joint and inescapable expenses, the relevant basis for charging will be the ship’s willingness or ability to pay. Joint expenses cannot be avoided unless the other elements of the joint service are also discontinued; inescapable ones cannot be escaped under any circumstances. It is appropriate to try and recover from each user as much as possible without losing the traffic.

Common costs, such as those related to the maintenance of existing quays and transit sheds, generally vary with ship size and duration of stay. The recovery of costs according to willingness to pay is probably best linked to NRT, a measure of a ship's cargo carrying capacity, as a proxy indicator of ability to pay. An alternative basis for charging could be the tonnage of cargo worked in port, which also provides a proxy measure of the duration of a ship's visit.

Port dues on ships to cover harbour and quay costs are difficult to allocate. Harbour dues are intended to cover the costs of maintenance dredging and the much higher capital dredging costs. Similarly, the quay dues are intended to cover the significant fixed costs of locks, lock basins, and quays. All these costs vary with the ship size in terms of length and weight. However, although such expenditure is incurred for the largest ships, once built, the escapable costs of usage are almost zero. This would suggest charging on the basis of 'willingness to pay' by using NRT as the charging unit. The only further consideration is that capital dredging has to be undertaken to accommodate the larger ships. This leads to the suggestion of charging according to the size of the vessel. Similar considerations apply to quay dues.

During periods of congestion, or excess demand, rates should be upwardly adjusted to reflect the associated increase in short-run marginal social costs.

Table 5.1 sets out a possible port charging system for dues on ships:

Cost-Based Dues on Goods

Dues on goods are generally based on weight or volume depending on the type of cargo. This means that the tariff structure can vary according to base classification. The charging system can be flat rate or progressively increasing rates. Port charges can generally be classified as : dues on cargo, cargo handling on board, cargo handling on quay, storage, and warehousing. The nature of charge may relate to utilization of quaysides and facilities, cargo operations between the ship's hold and the quayside, cargo operations between quay and shed, use of transit sheds or container yards, and use of warehouses. It is necessary that the dues levied should cover immediately escapable costs, and cargo classifications should broadly be based on homogeneous cost characteristics and a uniform ability to contribute towards joint, common and inescapable costs.

Table 5.1 : Port Charges on Ships

Type of Charge	Nature of Charge	Charging Base	Basic Units	Charging System
Port dues on ships	For utilization of general port maritime facilities and services e.g. dredged channels/ quays	Two-part tariff: size of ship; type of ship	Nrt and/or length	Flat rates for different sizes and types of ships
Pilotage	For piloting the ship	Size of ship	Nrt and/or length	Flat rates for different sizes or lengths of ships
Towage	For towing the ship	Size of ship	Nrt and/or length	Flat rates for different sizes or lengths of ships
Berthing	For line-handling during berthing and unberthing	Size of ship	Nrt and/or length	Flat rates for different sizes or lengths of ships
Berth occupancy	For occupation of a berth	Three-part tariff: size of ship; length of ship; berthing time	Nrt and/or length	Flat rates per day for different sizes and types of ships

Table 5.2 sets out a possible port charging system for dues on cargo:

In summary, basic infrastructure charges in respect of ports should be based on long-term marginal social costs, which would take into account new investments and externalities relating to environment, congestion and accidents. It is recognized that cost accounting techniques that split cost between shipping and cargo cost are not precise. Also, any formula-embedded⁵² rationale has to reckon with the respective bargaining powers of the usually highly organized and powerful shipowners on the one hand, and shippers who are typically scattered and often much less able to effectively negotiate with the port authorities, on the other.

Inland Waterways

Transport on rivers and canals with barges or lighters and in small vessels along

the coast and between inland centres, plays an important role in the transport system

Table 5.2 : Port Charges on Cargo

Type of Charge	Nature of Charge	Charging Base	Basic Units	Charging System
Port dues on cargo	For utilization of the port quaysides and facilities	Two-part tariff: weight; nature of cargo	Metric ton	Flat rates per ton for different groups of cargoes according to ability to pay
Cargo-handling on board	For all operations from the ship's hold to the quayside and vice versa	Two-part tariff: weight; presentation of cargo e.g. bulk, container, pallet	Metric ton	Flat rates per ton for different groups of cargoes according to cargo-handling method
Cargo-handling on quay	For all cargo operations from quay to shed or delivery and vice versa	Two-part tariff: weight; presentation of cargo e.g. bulk, container, pallet	Metric ton	Flat rates per ton for different groups of cargoes according to cargo-handling method
Storage	For use of transit sheds or container yards	Three-part tariff: weight; volume or stacking condition; time	Metric ton and day	Progressively increasing rates per ton per day for different groups of cargoes with the same volume
Warehousing	For use of warehouses; no free period	Three-part tariff: weight; volume or stacking condition; time	Metric ton and day	Flat rate per ton per day/week for different groups of cargoes with the same storage characteristics

of many countries. In an international context, the system also provides an important means of feeding and distributing cargoes to and from deep-sea vessels. The issues involved in the pricing of waterway infrastructure facilities are very similar to those for ports, described above.

⁵² Juhel, M. H., 2000. *Port Pricing and Charges*. ESCAP-AITD Regional Seminar on Transport Pricing and Charges for Promoting Sustainable Development, New Delhi.

Inland waterway transport services can be provided as a common carrier on the basis of a scheduled service; as bulk carriage for a single client; or on an own account basis, among others.

Provided the market is large enough, there is nothing inherently wrong, in terms of cost structure, in promoting a competitive waterway sector. In terms of bulk freight, there is no real evidence of economies of firm size. Only where firms offer common carrier services which carry goods for different customers, is there a major risk of wasteful competition. It is, thus, in the parcels business, with its high fixed costs, that competition could lead to a duplication of services and low load factors. For bulk traffic, there is no real objection to allowing shippers to choose between alternative waterway operators, with the alternative of using their own vessels also available to them. Competition should lead to the minimization of the average costs of providing the desired standard of service – prices will be based on average cost, which, under conditions of constant returns to scale, will reflect marginal cost. Government regulation may be necessary to ensure that cost pressures do not compromise safety standards either in terms of maintenance or in working practices. In addition, government intervention may be required to ensure the internalization of externalities and the recovery of infrastructure costs.

Common carrier inland waterway services bear much higher fixed or inescapable costs and may thus lead to the emergence of a natural monopoly, though the extent of potential monopoly power may be limited by competition from other modes. Secondly, because of economies of scale, efforts to cover total costs may lead to a divergence between price and marginal cost. This problem is most marked in the case of infrastructure costs, if charged for, particularly where certain services regularly have spare capacity in order to meet peak demand or maintain service quality.

There appear to be three solutions to this problem for the barge operator:

- (i) the operator could base prices on marginal costs, with the resulting deficit being financed by government subsidy – this is unlikely, but may be justified to compensate for non-internalized externalities created by other modes;
- (ii) the operator could base prices on average cost – but this could only be sustained with a degree of protection against competition from other modes; and
- (iii) the operator may seek to discriminate between traffic flows according to the shipper's willingness to pay in order to cover the difference between average and marginal cost.

The advantages and disadvantages of each approach are as follows. In case (i),

the government needs to be able to estimate the volumes and corresponding marginal costs conforming to the optimal allocation of traffic in order to determine the appropriate subsidy. If it does not do so, then it will revert to deficit financing with no real ability to exert financial discipline on the operator. The financial deficit will have to be met either through taxation or by diverting other forms of government expenditure. The opportunity cost of this expenditure will have to be assessed and investment in the waterway system will need to be evaluated using social cost-benefit analysis.

In case (ii), the government will need to become involved in the day-to-day administration of the freight service and will have to employ arbitrary methods to allocate vessels to traffic and to establish priorities. In this case also, investment in the waterway system will need to be evaluated using social cost-benefit analysis.

By contrast, approach (iii) largely eliminates the regulatory role of government. This is because the allocation of freight is determined by the market and long-term planning and investment can be made on grounds of profitability. The basis for discrimination usually is class of freight determined by value and the extent of competition from alternative modes of transport.

The above assumes that the providers of inland waterway services compete with operators on other modes who charge prices equivalent to marginal social cost including externalities. This is unlikely in practice. In particular, the costs of road infrastructure, congestion, pollution, accidents and noise are rarely fully internalized. In many developed countries, governments subsidize waterways through the provision of grants towards the costs of building new berths and termini, and/or tax road hauliers, in order to promote the diversion of freight traffic to the waterways.

Section II

Maritime Transport

Maritime transport comprises three principal forms of operation:

- (i) Industrial transport
- (ii) Charter (non-liner) shipping
- (iii) Liner services

This section describes the nature of the alternative forms of operation, market structures, and methods of price determination in each of these sectors.

Forms of Operation

Industrial shipping are services provided mainly to meet the well-defined transport needs of large industrial enterprises, such as oil companies and energy firms. The ships

are either owned and operated by such firms, or are contracted by them on a long-term basis. The fleets are usually engaged on regular routes with shiploads of bulk and homogeneous cargoes.

Ships operating in the charter markets usually operate on a non-scheduled basis with no fixed itinerary or fixed sailing schedule. Users or charterers either pay charter rates, which are related to the amount of cargo shipped, or fixed prices, normally per day or per deadweight ton, for the hire of the vessel.

Liner shipping services are offered through ships which operate to a prefixed sailing schedule between fixed ports on a regular basis. In liner operations, ships carry cargo, as common carriers, to many different shippers. Each shipper pays freight in accordance with a tariff based on the volume, weight, or value of the cargo.

Shipping Cost Structures

Maritime transport costs comprise company overheads, vessel depreciation, vessel operating costs, voyage costs, and cargo costs.

Company overheads cannot usually be allocated to a specific ship, although agency fees and warehouse charges may be allocable to specific liner services. The capital costs of using a ship usually comprise the opportunity cost of capital tied up plus the depreciation or loss in the value of the asset over the period it is used. Operating costs include crew costs and costs related to repair and maintenance, insurance and stores. Voyage costs include fuel costs, port dues and agency fees. Cargo costs are those related to loading, unloading, storage and ancillary charges. Payment of the respective costs will depend on the nature of the contractual or charter terms upon which a ship is operating.

Costs may be regarded as either fixed or variable (inescapable or avoidable), depending on the nature and time horizon of a particular decision. For example, where a ship is to be chartered for a single voyage, the cargo costs, voyage expenses, and at least a proportion of the operating costs will be variable and could be allocated to the specific voyage charter under consideration. Alternatively, where a container ship is operating on a fixed route to a definite timetable, voyage costs and operating costs are largely fixed while spare capacity exists. One complication which arises in the short term, when assessing avoidable costs, is that the owner can always lay up a ship and reduce operating costs for a limited period of time. This can affect the determination of the opportunity cost of ship's time when evaluating chartering decisions.

Shipping Markets

In charter shipping, there are many suppliers of services with fairly homogeneous services to offer. In the various sub-sectors, such as tanker or dry cargo, the ships may vary with regard to their size and equipment, but basically they can all do the same job of moving bulk cargoes⁵³.

The charter markets are open markets, in the sense that here it is easy to buy or build a ship and put it into operation. There are a large number of buyers and sellers of shipping services, and no individual shipowner is large enough to be able to influence the market charter rate. There is also an active second-hand market and, as such, it is relatively easy to leave the business. In consequence, perfect competition exists with prices being determined by the forces of supply and demand. Shipowners are essentially 'price-takers' and have to make operational decisions on the basis of market prices. Since price elasticity is relatively low, changes in freight rates have only a marginal effect on demand. Therefore, small changes in supply and demand can produce large changes in price or freight rates over a short period of time.

By contrast, the liner markets are characterized by monopoly and oligopoly. In the liner trades, individual routes are often controlled by a single operator, who sets the freight rate or tariff. On some routes, a limited number of firms provide parallel services and collude over the setting of rates, sometimes by forming conferences or price-fixing cartels.

Freight Rates and Tariffs

In the charter markets, the shipowner, as a matter of principle, will aim to obtain rates which will cover all costs, including the opportunity cost of capital tied-up in ships, and yield maximum profits. In the short run, while a ship is still operable, the shipowner has considerable fixed costs: depreciation, company overheads and operating costs. These costs will have to be borne whether the ship is sailing or idle and, therefore, will not influence the decision on whether, or not, to accept a particular freight rate. A rational shipowner will accept rates provided the marginal or avoidable costs involved are covered. Indeed, a profit maximizing firm will operate up to the point where marginal revenue equals marginal costs. All income above marginal costs will contribute towards fixed costs and the owner, by continuing to operate, will be better off than refusing cargoes, in the very short term. In the short run, say, over the period of one voyage, the shipowner's marginal costs comprise the voyage and cargo-related costs. In the

⁵³ Stopford, M., 1988. *Maritime Economics*. Unwin Hyman. London.

longer term, operating costs may also be viewed as variable and these too will need to be covered by charter rates. In the very long run, to remain in business, the shipowner will need to cover all costs from charter revenue.

The charter markets are subject to considerable variations in prices, which are always equal to short-run marginal cost (SRMC). SRMC may lie above or below long-run marginal cost (LRMC) depending on the extent of excess demand and supply in the charter markets which are highly efficient and sustainable. Attempts by governments to regulate the markets through protectionism and cargo reservation are likely to lead to losses of social welfare.

Liner shipping services are offered through ships that operate to a prefixed sailing schedule between fixed ports on a regular basis. In liner operations, ships carry cargo as common carriers for many different shippers. Each shipper pays freight in accordance with a tariff based on the volume, weight, or value of the cargo. The liner markets are characterized by monopoly and oligopoly. In the liner trades, individual routes are often controlled by a single operator, who sets the freight rate, or tariff. On some routes, a limited number of firms provide parallel services and collude over the setting of rates, sometimes by forming conferences, or price-fixing cartels. Liner vessels transport many different types of cargo of widely differing values. Since marginal costs associated with some specific cargoes can be very low, liner operators practice price discrimination, or ad valorem pricing to cover total costs. Often the same service is sold to different shippers at different freight rates, according to the principle of 'what the traffic will bear'.

Conferences, and a degree of concentration, exist in the liner trades due primarily to the higher entry costs and the fact that fixed costs are very high for scheduled services. It is often claimed that competition can be wasteful in the sense that extreme rate volatility could result from open competition. However, the ownership of shipping fleets and control of operating routes vests in the hands of a few dominant players denying marginal players, like shipping lines of developing countries, due share of the shipping business. Therefore, while liberalizing market access, liner conferences need to be brought within the purview of regulatory and anti-trust legislation, even though intra-conference competition may exist.

Many different types of cargo are transported by liner vessels. Since there are widely differing values to the cargoes shipped, and marginal costs associated with specific cargoes can be very low and well below marginal cost, price discrimination is prevalent. Price discrimination, or ad valorem pricing, implies that the same service is sold to different shippers at different freight rates. High value cargoes tend to pay higher unit

charges than low value ones, thereby contributing relatively more to the fixed costs of the service than cargoes with lower unit values. It is sometimes suggested that this may amount to cross-subsidization. However, since fixed costs cannot be allocated to specific cargoes in a rational manner, this argument is difficult to sustain. Indeed, it is argued that price discrimination, as operated by liner firms, tends to reduce welfare losses to society as a whole. Without it, probably less cargo would be transported.

Deregulation, Competition and Pricing

An argument in favour of administered pricing for liner operations is that as long as space is available aboard the ship, it costs a shipowner very little to load extra cargo, since most costs are fixed for scheduled shipping services. In periods of little competition, a conference may obtain rates that cover full costs even for low value cargo. During periods of competition, however, this may not be possible. If the rates for these cargoes still cover their marginal costs, the conference will accept it. As a consequence, some firms would go out of business until price and capacity were again in equilibrium. Defenders of the conference system argue that competition would produce too much volatility in the market with the result that neither freight rates nor services would be offered with any certainty. It is also argued that conferences are needed in order to provide stable and continuous transport services.

Proponents of competition suggest that it will force market prices down to the relevant short-run marginal cost of the service, and that, when surplus space exists, these rates will not be sufficient for operators to cover their costs and break even in a financial sense. This will not be a stable position and the most expensive lines will go out of business first. As a result of competition, shippers would be forced to compete with each other for the remaining space available. This would drive rates up to the point where profits returned which, in turn, would provide an incentive for new firms to enter the route and offer services. Rates would again cover average costs, but these would perhaps be lower than those that would exist with conferences or other cartel arrangements in place. It is debatable whether any form of price discrimination would continue under competitive conditions.

There has long been general international acceptance of shipping conferences as a means of ensuring a stable liner shipping environment. However, the exemption of conferences from anti-trust legislation has led the competition-based market mechanisms to pursue their abolition. In the USA, the Ocean Shipping Reform Act of 1998 has led to the demise of many conferences and to an increasing role of independent carriers and consortia, although current antitrust immunity is still maintained within a regulated framework. The recent years have witnessed decline of conferences and emergence of

carrier alliances on a global basis. The size of investments being made by global shipping services, not only in new tonnage but also in port infrastructure and equipment, has led to vertical integration and a degree of concentration in the liner shipping. The economies of scale have benefited the shippers by way of improved service reliability and reduced freight rates, these developments have made it increasingly difficult for the developing countries to keep pace and for their shipping lines to remain competitive. The liner shipping services, therefore, need to be brought within the purview of regulatory and anti-trust legislation to allow marginal players a due share of business.

Section III

Airports

Airport charging systems cover a wide variety of charges related to different airport facilities and services. These include landing, lighting, parking, refuelling and storage facilities as well as aircraft, passenger and freight services. This section examines the main constraints on airports in devising sound pricing policies, defines possible airport pricing objectives, and identifies the factors to be accounted for in devising a rational airport pricing system.

Current Airport Pricing Structures

Existing pricing systems at international airports have evolved over years in a unsystematic manner. Charges range widely in structure and levels, but most systems have one thing in common: the main aircraft charge is normally based on the maximum all-up-weight of the aircraft, often with break points in the scale. This is, in many cases, supplemented by a passenger charge which may be paid directly by the passenger to the airport and may vary from one type of passenger to another type of passenger. More specifically, the typical charging criteria include:

- (i) the origin or destination of the flight, with perhaps a distinction between domestic and international flights, for landing, take-off, passenger and lighting charges;
- (ii) the mass of the aircraft, often the maximum take-off weight, for landing and parking charges;
- (iii) the noise category of the aircraft for the noise charge or, if no such charge exists, for the landing charge when modulated according to the noise emissions of the aircraft;
- (iv) the parking time, sometimes modulated in accordance with the flight schedule, for the parking charge;

- (v) the number of passengers, their age and sometimes the distance flown, for the passenger charge;
- (vi) the freight tonnage loaded or unloaded, for the freight charge.

The International Civil Aviation Organization (ICAO) recommends that airport charging systems should be such that each type of traffic should bear its fair share of airport costs. But the ICAO does not specify how airport costs are defined or measured, or how a fair share of such costs should be assessed.

Some airports interpret this requirement to mean that airport charges should be related to long-run marginal cost. However, this is extremely difficult to measure; it is unlikely to lead to an optimal usage of airport capacity, provide guidance on investment requirements, or allow a particular financial objective to be met.

Airport Pricing Objectives

Airport pricing objectives will depend on the ownership structure and government policy. Nevertheless, it is possible to suggest objectives which will be consistent with the promotion of sustainable development. The aims of airport pricing system should be to achieve optimum use of existing capacity and guide investment decisions within the framework of financial sustainability. Specifically, airport pricing can support these requirements in three ways, namely, (i) by measuring the aggregate demand for each service and its social profitability which will determine whether effective demand warrants the continuation, expansion or contraction of airport services and facilities; (ii) by encouraging the use of excess capacity and rationing capacity when there is excess demand; and (iii) by providing information to support the coordination of long-run and short-run decisions so that total system costs are minimized and the surplus of value over cost is maximized.

To achieve the above objectives, airport prices need to be related to marginal social cost of the resources used to provide each service.

The issue of social costs, however, is problematic. The theoretical solution is to calculate airport charges on the basis of social accounting prices, with due allowance for indirect costs, such as congestion and noise.

The main constraint on creating a sound airport pricing policy is likely to be the need to recover costs through revenues. This constraint may apply at the level of the airport as a whole, or at the level of individual facilities and services. The costs may be

defined either as the combined capital and operating costs or as only operating costs. In practice, however, it may not be possible to recover total costs where overcapacity exists due to inaccurate demand forecasts, low initial utilization, or indivisibilities in capacity provision.

Airports possess a degree of locational monopoly which may give rise to the potential for monopoly pricing. Although some countries use this to justify state ownership and operation of airports, many countries have managed to create strong private sector participation in the provision of both airport infrastructure and airport services.

Airport Pricing Principles

The principles which apply to airport pricing are those of marginal social cost pricing in conditions of lumpy investment where long periods of less than full capacity operation are inevitable, complicated by periods of congestion and existence of externalities, particularly arising from aircraft noise. If the airport is required to be self-financing, then it is likely that it will need to raise additional revenue above that which would result from efficient prices. This suggests that the airport should devise a pricing structure which:

- (i) always charges enough to avoid excess demand;
- (ii) never charges less than short-run marginal cost, including social costs; and
- (iii) charges 'what the market will bear' until the financial target is achieved.

Cost-Based Airport Charges

Two major problems arise in relation to devising a cost-based airport tariff. The first is to determine which expenses are to be covered by prices, and the second is to decide how these should be covered. Deciding the method of recovery also involves choosing a unit (such as aircraft type, aircraft weight, passenger numbers or freight tons handled) upon which the charges are to be based.

As in the case of ports, it is possible to categorize airport expenses into five categories, namely, (i) immediately escapable costs (short-run marginal cost); (ii) joint costs; (iii) common costs; (iv) inescapable costs (fixed costs); and (v) social costs (externalities). These categories have already been discussed in Section I of this chapter under the heading 'A Cost-Based Port Tariff'.

It may, however, be pointed out in this context that using measures of aircraft size and passenger numbers for working out airport charges may be deemed fair on the grounds that it is the largest and heaviest aircraft that determine the capacity and

strength of runways, stands and termini. However, they also indirectly measure 'ability to pay'. For example, larger aircraft have larger payloads and hence larger revenues. Further, they tend to be employed on long-haul routes with higher unit fares.

In calculating charges, it will also be necessary to recognize that the relationship between demand and capacity may differ between terminals and runways. It is possible that the peak periods for these will be slightly different each day and that while one may suffer from excess demand, the other may not. Thus, peak charges may apply to different periods for runways and other aircraft related services compared with passenger facilities. In this regard, it may be appropriate for an airport to determine the capacity limitations for each type of airport facility and to invite airlines to 'make bids' for user rights against the declared capacities for particular periods during the year.

Externalities

The most significant cost that does not appear in the accounts of airlines or airports is that of noise. Predominantly, aircraft noise remains an externality which affects people residing in an airport's flight path. In principle, it is possible to estimate the cost of noise or the willingness to pay for less noise. In an ideal world, airlines and their users should compensate those adversely and directly affected by aircraft noise. In practice, however, airports or governments can seek to tax noise pollution in an attempt to moderate its impact and internalize this externality.

Calculating Airport Charges

As in the case of port charges, there are five steps involved in the calculation of airport charges, namely, (i) to determine airport costs; (ii) to calculate the specific costs for each cost centre; (iii) to identify 'revenue centres'; (iv) to collect information on the utilization of assets corresponding to a given cost centre; and (v) to determine the charges and charging basis by cost centre. These steps have already been discussed in Section I of this chapter under the heading 'Calculation of Port Charges'.

Section IV Air Transport

In recent years, scheduled air passenger transport services have been subject to deregulation in many parts of the world. The process of deregulating air transport has resulted in greater competition leading to reduction in fares, which is seen as an advantage for the traveller. Some people, perhaps adopting a somewhat critical view of the deregulation policy, maintain that this reduction in fares is to the detriment of aircraft maintenance and thus of safety.

This section examines airline pricing policies, competition, and the resulting service/price mix. In particular, it examines airlines' commercial policies under competitive conditions, that is to say, their overall fares and promotional strategy and whether or not these are to the traveller's benefit.

The International Civil Aviation Organization (ICAO)

Article 1 of Convention on International Civil Aviation (1944), which established the International Civil Aviation Organization (ICAO), also established the principle that every state must obtain the agreement of other state or states to operate international air services, normally through bilateral agreements, though there are also multilateral agreements.

Bilateral agreements provide for:

- (i) reciprocity and equality of opportunity in sharing of traffic rights;
- (ii) procedure for approval of fares and rates and entry and exit from the routes;
- (iii) restrictions on route operations;
- (iv) limitations on the development of capacity;
- (v) designation of airlines, substantial ownership and effective control of the designated airlines; and
- (vi) taxation.

In the current climate of increased liberalization and globalization, ICAO's general policy is for greater freedom in the provision of international air transport services through 'open skies' agreements that provide for unrestricted traffic rights, capacity and tariff flexibility. However, for domestic services, every state still has sovereignty to decide its own policy. Such policies have led to greater competition among airlines with regard to the quality of service and other fringe benefits provided and also in respect of fares. Airlines often offer discounts on published fares to attract traffic and fill the seats during off-peak periods, and governments and regulatory authorities tend to condone such practices. Thus, it is the level of demand and supply that effectively determines prices. Short-run marginal cost plays an important role in determining the minimum fare, while passenger characteristics influence actual fares in terms of determining 'willingness to pay'. This has led to the argument that airline prices need to be regulated to avoid cut-throat and wasteful competition among airline operators.

Airline Tariff Structures

Tariff structures in air transport are referred to as fares for passengers and rates for cargoes. Passenger fare structures have developed a wide variety of different types of tariffs that are fashioned to respond to different markets as well as different segments of the same market. This variety also reflects, in part, an air carrier's effort to maximize revenue by varying capacity and price. For example, by allocating more seats at discount fares to routes, flights or time periods with weak demand and fewer seats where there is a strong demand, airlines seek to have as many passengers as possible to fly at normal economy or premium fares, while concurrently stimulating discretionary travel⁵⁴.

A normal economy fare is the lowest priced fare⁵⁵, which allows a passenger maximum flexibility in terms of fare combination, refund, changes in itinerary and routing, etc. Such fully flexible and unrestricted fares often serve as a reference point for determining the price levels of other types of fares, such as premium class fares and restricted economy fares. There are several types of restricted fares, depending on the type and extent of restrictions. Moreover, there are other types of special fares, such as excursion fares, standby fares, budget fares, incentive, affinity and non-affinity group fares, individual and group inclusive tour fares, youth, family, military, pilgrim, local resident, student, and teacher fares. Charter fares are different and generally lower than the fares on scheduled services.

Normally fares per passenger-kilometre vary considerably with distance. The worldwide average fare per passenger-kilometre for a distance of 16,000 km is only about 20 percent of the average fare for a distance of 250 km. This reflects the lower cost of operation on long-haul flights. However, the proportion differs from route to route and region to region.

Cargo rates have some of the same distinctions as in the case of passenger fares and these are determined using similar pricing practices and concepts, but with a different terminology. The general cargo rate varies with weight by applying different prices per kilogram depending on whether the weight of the shipment falls above or below a breakpoint. General cargo rates do not vary with the nature or values of the cargo transported and are used when the property being shipped does not qualify for any other cargo rate. As a reference for calculating other rates, the general cargo rate serves a similar purpose as normal economy class passenger fare.

A rate that combines the pricing features of both premium and special fares is called a class rate. Class rates are determined by applying a discount or a surcharge to a general cargo rate for certain commodities (for example, a discount for newspapers

and a surcharge for commodities requiring special treatment during shipment, such as livestock, gold and securities). A rate, which has a similar purpose to special or discount fares is a specific commodity rate, used for certain types of cargo, which is generally lower in price than the general cargo rate at comparable weights. Container rates are generally different from normal cargo rates, and the rates for owner-packed containers are generally lower than the normal cargo rates.

It thus appears that airline passenger fares and cargo rates are not directly related to either average costs or long-run marginal costs but mainly to 'what the traffic can bear', taking into account the short-run marginal cost which acts as the minimum price. The overall objective is to cover total costs. The fare policy normally adopted by airlines, in a competitive environment, is designed to maximize the profitability or yield of each individual flight. This is due to the fact that once the annual flight timetable or schedule has been determined, a significant part of airline operating costs becomes fixed or inescapable over the relevant planning horizon.

In general, airline fares policy is part of a mix of instruments that cover the range between two extremes: quality and extent of service on the one hand, and price on the other. This mix of instruments is normally tailored on the basis of the characteristics of each category of traveller so as to maximize the revenue from each individual category. This means that most airlines, under conditions of competition, will tend not to increase the market share in each market segment to the greatest possible extent, but to maximize revenue from each individual segment. Some companies, however, adopting a very aggressive commercial policy based on medium and long-term objectives, sometimes offset the revenue from one segment against another, thus accepting that expansion in one area will involve an initial loss.

Fares and Yield Management

Fares policy decisions cannot be taken in isolation of service, and, in this respect, the differentiation that exists in this area, depending on the class of travel, is of particular importance. The commercial policy for the First Class category of travellers, who often travel at the expense of the companies they work for, is geared primarily towards service. The promotion of this class of travel is focused on the conditions and comfort of travel and on the services provided on the ground. Moreover, the Frequent Flyer Programmes, as discussed below, tend to be targeted at this category of travellers. The obvious intention

⁵⁴ Kesharwani, T. R., 2001. *Pricing and Charges in Civil Aviation*. Asian Institute for Transport Development. Economic Issues Number 5, New Delhi.

⁵⁵ International Civil Aviation Organization (ICAO), 1996. *Manual on the Regulation of International Air Transport*. Quebec.

in this case is to maintain high fares, maximizing turnover per passenger/kilometre. In the context of an aggressive commercial policy, First Class may be regarded as the category which, in a climate of fierce competition, compensates for the significantly discounted fares in the Economy Class.

An airline's fares are drawn up on the basis of the rates determined by IATA and adjusted on the basis of the fares offered by competing airlines on specific routes. Continuous monitoring of competitors' fares is prevalent. This is how published fares are arrived at, i.e., the fares that an airline charges without any discounts and with the minimum of restrictions on travel. On the basis of these, the airlines formulate the other discounted fares that are normally subject to certain restrictions. For example, fares for groups are examined on a case-by-case basis, depending on factors, such as the number of passengers; how full the flight is at the time of the request; the final destination; and the fares mix for the flight in question. Therefore, the crucial factor for drawing up fares is yield management⁵⁶.

The principle of yield management is simple: each class of seats is divided into subclasses, each of which is assigned a fare, starting with 0, which represents free seats, e.g. for Frequent Flyer Programme members, all the way up to full fare. Yield management is administered centrally (for each company, there is a single office which manages the seats on all the flights); depending on demand, seats are assigned to each subclass and, apart from seats already reserved, a seat may be transferred at any time from one subclass to another. With yield management, the market sets the price at which a seat is sold. As a result, two seats in the same class may be sold on different days at different prices, and the seat sold first may not necessarily cost less. If there is limited demand for that class, it may very well mean that prices will fall. In general, different prices will mean different restrictions, but even this is a decision for which the only governing factor is the market situation obtaining at the time.

It would seem, then, that the passenger, or buyer, is in a situation of very limited transparency, in which all the data on the market is known only to the airline company. In reality, even if travellers cannot see exactly where they stand in terms of the fares policy decisions of a single company, they can at least compare the offers (at a given point in time) of the various airline companies operating on the same route. The real problem occurs on those routes where there is no competition, where competition is limited, or where the route is operated as a public service. In this case, the absence of transparency is not compensated for by the passenger being able to choose between a number of carriers.

In summary, on most routes, it is possible to classify fares into three groups. First of all, there are the *more flexible fares*, also referred to as full fares, which include those in economy class without restrictions; then there are *promotional fares*, that is, discount fares or fares with restrictions, and finally, there are the *special fares*, which are limited offers. These fare categories apply to all classes of travel. Yield management, as operated by the airlines, is a sophisticated kind of Ramsey Pricing or price discrimination aimed at extracting the maximum consumers' surplus consistent with maximizing utilization. Since marginal costs are very high at times of excess demand, very few or even no passengers will be accepted at special or promotional fares. On the other hand, since marginal costs are very low when excess capacity exists, some passengers will often be permitted to travel at low (special) fares or even zero fares (say, in the case of Frequent Flyer Programme members). Yield management requires sophisticated CRS computer systems and airlines perceive that there is also competitive advantage in owning one of these systems.

Frequent Flyer Programmes (FFPs)

These are promotional schemes, with significant impact in terms of competition, aimed at gaining the customer's loyalty by offering him or her travel incentives, usually based on journeys already completed and targeted essentially at business customers. FFPs have gained ground particularly in the USA with the deregulation of air transport. FFP members are normally travelling at the expense of their employer, or passing on the cost of travel to third parties. The philosophy behind these programmes is that of making the airline seem more attractive in terms of price, the price/quality ratio and other factors of value to passengers, by taking advantage of the natural predisposition people have to receiving gifts. The incentives may be free flights or free upgrades. As a matter of fact, complex tariffs have now emerged for both earning and spending rewards.

In some FFPs, loyalty is accentuated by the reduced number of miles necessary to obtain a further advantage; for example, to obtain the first free flight, one may need 30,000 miles, while the second requires only 20,000 miles with the result that those participating in the programme are encouraged not to leave the scheme. However, various management and financial problems also accompany FFPs. Since the management of turnover per seat is critical to yield management, the obligations arising out of FFPs can reduce the profitability of a seat, and this is all the more serious because of the difficulty of forecasting when, and on which flights, this entitlement to free travel will be exercised. For this reason, methods to control FFP commitments have been introduced to avoid, or reduce to the minimum, the lost profitability per seat resulting from free flights; for example, by seeking to prevent free travel entitlement from being exercised on the better

covered flights.

Overbooking

Yield management techniques have evolved to the extent that managed overbooking often occurs, particularly in the USA. On overbooked flights, auctions are held at the departure gate whereby passengers can surrender their seat and gain compensation in terms of money and/or upgrades for deferring their flight. There is an incentive for the airline to undertake this practice where the fare received from the additional passenger exceeds the costs, including compensation to the passengers surrendering their seats. There is, however, scope for the airline to misjudge the required compensation and incur losses through overbooking.

Airline Deregulation and Competition

Initially, it was expected that airline deregulation would lead to a general reduction in fare levels through the entry of new airlines on particular routes⁵⁷. However, the early route expansion strategies developed by a number of the former trunk carriers in the USA quickly proved to be unsuccessful. The high traffic density routes, which were the focus of expansion, were the target of other airlines and were suddenly subject to excess capacity and plummeting fares. The former charter airlines, with low operating costs, competed aggressively on price. By contrast, the incumbent high-cost operators with their wide-bodied aircraft were at a competitive disadvantage as hub-and-spoke route networks developed. Such networks were the main method of achieving expansion through the generation of new traffics. Hub-and-spoke networks require smaller aircraft operating frequent services and supplying feeder traffic. The former trunk operators were not well-placed during the initial phase of route expansion.

The competitive responses of the incumbent operators, following deregulation, were many and evolved in three discrete phases, as follows:

- (i) *Phase 1 : Defensive Tactics : Cost-reducing Activities*
 - Improving labour productivity : new wage structures; reduced demarcation; new management; de-unionized set-up
 - Aircraft : investment in more efficient aircraft; downsizing
 - Network : operating a hub-and-spoke network
- (ii) *Phase 2 : Revenue Generating Activities*
 - develop a CRS (Computer Reservation System)
 - introduce Frequent Flyer Programmes

- vary travel agent commission levels
 - increase service frequency
 - enter code sharing alliances
 - improve in-flight service
 - increase advertising
 - benefit from price discrimination
- (iii) *Phase 3 : Offensive Tactics: Transforming the New Competitive Environment*
- control of CRS
 - active Yield Management
 - tie in commuter feeder services

Deregulation brought about a fundamental change in the point of sale of airline seats, with passengers turning to independent travel agents for ‘impartial’ advice. Agents were sought by passengers due to the emergence of a wide range of airline and fare options. Airlines developed competitive advantage by offering differential and sales target based commissions to agents and by installing on-line CRS (Computer Reservation Systems) giving instantaneous advice on fares and seat availability.

Another important means by which airlines sought to enhance their revenue was price discrimination. This was an unexpected outcome of deregulation, as policymakers had anticipated that the ensuing competitive environment would make the practice unviable. The peaking demand characteristics of airline markets are such as to allow a considerable variation in price, partly due to the inability of carriers – even in competitive markets – to vary supply to the same extent. The key to achieving the full benefits of such a policy lies in an ability to minimize revenue dilution. Ordinarily, discriminatory pricing necessitates the existence of monopoly or a highly collusive oligopoly, as it is only in these types of markets that firms are able to exercise the necessary control over their customers. The non-storable nature of the service, however, alters this situation.

The vast amount of information gathered by CRSs has enabled their owners to fine-tune their price discrimination activities, allowing them to extract even more economic rent; their non-CRS owning rivals earning less as a consequence. Without the ownership of such equipment, airlines had little option but to use one of the four CRS vendors at very high fees.

⁵⁷ Williams, G., 1993. *The Airline Industry and the Impact of Deregulation*. Ashgate.

**Airline Mergers, Alliances,
Privatization and Transnational Ownership**

Airline mergers may lead to an increasingly monopolistic industry and build up the potential for fare increases. Proposed mergers between airlines and transnational ownership of airlines are being reviewed by the regulatory authorities, especially in the USA, to avoid monopoly or near monopoly exploitation of the market.

Similar concerns have been expressed about the increasing dominance of airline alliances. Such alliances increase the scope of air services to different parts of the world and offer improved frequencies and better connectivity for the passengers. Another marketing arrangement to emerge in recent years is that of 'codesharing'. This is a practice whereby one carrier permits another carrier to use its airline designator code on a flight or where two carriers share the same designator code on a flight. The practice is intended to benefit from better utilization of the rights under the bilateral agreements, cost savings, economies of scale, and increased net revenue. Regulatory authorities are likely to prevent the abuse of market power by such groupings.

Privatization and transnational ownership are currently growing mainly because of financial problems faced by many governments and state-owned airlines. These institutional changes may lead to improvement in management and operation of airlines. If such changes reduce operating costs, then competition would keep downward pressure on airfares.

