5

SUBSIDIES IN CANADIAN PASSENGER TRANSPORTATION

Trevor D. Heaver

July 1991

I. Introduction

The Royal Commission on National Passenger Transportation was established in October 1989. Its mandate is to make recommendations for an integrated national passenger transportation system that will serve Canadians well into the 21st century.

The Commission provides an opportunity for a much needed review of a wide range of practices and policies that affect the characteristics and costs of passenger transportation services in Canada. This paper is a part of the background research of the Commission. Its purposes are to provide, in a concise readable form, a comprehensive survey of the nature and role of subsidies affecting passenger transportation services in Canada.

The Commission has a unique opportunity to examine the passenger transportation system of Canada as a whole. It is important, therefore, that research which reports on the historical record of passenger services in Canada look beyond the modal orientation of past practices. This paper considers the general characteristics of subsidies, examines the attributes of subsidies in the individual modes and, finally, draws general conclusions.

OBJECTIVES OF THE PAPER

The terms of reference for the paper encompass the following objectives:

- identification of the linkages, if any, between the types of subsidy, the mechanisms for the provision of subsidies, and the institutional characteristics of the services involved;
- assessment of the effectiveness of subsidy types and mechanisms; and
- assessment of the potential role of subsidies in the future.

OUTLINE OF THE PAPER

The paper is divided into four sections. Following the Introduction, Section II describes the various characteristics of subsidies, including the types and purposes of subsidies, the economic effects of subsidies and the desirable attributes of subsidy mechanisms. Indeed, there is such a variety of subsidies that the second section of the paper begins with a definition of "subsidy."

The frameworks used to describe subsidies provide structures for the survey of the role of subsidies in the various modes of passenger transportation in Canada. Policies and programs exist on a modal basis for a variety of reasons. For convenience, the survey is presented by mode in Section III, concluding with a synopsis of the record of subsidies.

Section IV presents conclusions based on the Canadian experience to date which provide a basis for recommendations on the direction of policy and on further research to assist the framing and application of this policy.

II. CHARACTERISTICS OF SUBSIDIES

This section provides an overview of the nature of subsidies affecting passenger transportation services in Canada. Because of the great diversity of specific subsidies, they have been placed into broad categories. At times, this has involved making some somewhat arbitrary allocations.

DEFINITION OF "SUBSIDY"

The Commission is considering the impact of a variety of government programs and policies on the roles of the various modes of transport in serving the passenger transportation requirement of Canadians. This paper deals with those programs and policies that involve subsidies.

The breadth of perspective adopted here is appropriate to the broad mandate of the Commission. The Commission needs a perspective on the influence of subsidies in passenger transportation equivalent to that being developed by the ministers of agriculture in various countries who are trying to develop a common international measure of the effect of subsidies on agriculture. Their solution, the Producers Subsidy Equivalent Measure, places a wide range of policies into a consistent framework which allows the effect of diverse programs on agriculture to be measured. Faced with a wide range of subsidies in passenger transportation, the Commission has a similar task.

The term "subsidy" carries a variety of meanings. Subsidies are most frequently thought of as grants or payments from the government to aid a particular group. This interpretation is the basis of many studies of subsidies. Subsidies, however, may take a variety of forms. They may be implicit as well as explicit. They may also be accidental as well as deliberate. They may result in the absorption of private costs by the public rather than an actual transfer of public resources to private interests. All subsidies, however, increase the value of benefits relative to costs accruing to the providers or users of a service. Therefore, they affect the output of individual services and the relationship between competing services. They affect the role of the modes of transportation in serving Canadians' needs.

For this paper, subsidy is defined as "a transfer of benefits to or costs from an affected party by an implicit or explicit program or policy." This definition is sufficiently broad to encompass the wide variety of types of subsidy considered below. The result of such programs or policies is usually a greater output of the subsidized good or service than warranted in the absence of the subsidy.

The immediate expense of a subsidy is a cost to society, whether it falls on some or all members or whether its effects are channelled through government. A subsidy program should only be continued willingly and knowingly, if the total value of benefits exceeds the total value of costs. A subsidy might be made available to all industries — for example, a grant in relation to research expenditures — or it might be made available selectively. A subsidy will often involve different treatment of one firm or industry from another. For example, the management of international shipping has been given tax-free status in Canada under particular conditions, effective March 1, 1991. This is classified as a subsidy here. The effects of the subsidy and whether it has merit are separate issues to be considered in the light of domestic and international conditions of this industry.

TYPES OF SUBSIDY

Several means exist by which income may be transferred with the result that the output of a good or service is larger than would otherwise be the case. There are four categories of such means, each requiring an explicit or implicit public decision. The public may pay for services; the public may provide services; the public may protect services; or the public may absorb costs created by services. Exhibit 1 is a summary of the types of subsidy discussed here.

This classification structure encompasses a broader range of subsidies than are normally considered. Most studies focus on direct subsidies involving payments; these are most easily identified. However, the full effect of government intervention requires a broader approach. A report by the Organisation for Economic Co-operation and Development on government intervention places subsidies in six specific classes.[1] They are cash expenditures; preferential credits; tax expenditures; subsidy equivalents of regulatory measures; possible subsidy elements in public purchasing contracts; and subsidy equivalents of tariffs and non-tariff barriers to trade. The classification used here places subsidies in fewer, generic classes. It has the advantages of highlighting the general nature of the promotional policies and of avoiding detailed technical listings.

Exhibit 1
SUMMARY OF TYPES OF TRANSPORT SUBSIDY

A. Subsidies which the public pays
- *ex post* compensation for unremunerative services
- *ex ante* compensation
- support for research and development
- tax concessions

B. Subsidies when the public provides facilities or services
- unremunerative services provided to travellers
- operation of unremunerative facilities and related services
  - routes
  - terminals
  - vehicles (not applicable to passenger transport)
- implicit capital subsidy of Crown corporations

C. Subsidies associated with the protection of services
- cross subsidies by regulated monopolies

D. Subsidies by the absorption of costs
- loan guarantees
- environmental externalities

Subsidies Which the Public Pays

Subsidies which the public pays may come in a variety of forms. Some may involve actual payments, others may enhance the net cash flow of undertakings by providing tax concessions. The programs are described in four categories: *ex post* compensation, *ex ante* compensation, payments for research and development and tax concessions.

Ex post *compensation:* Payment to a company for an imposed public service provided at a loss, but which is required in the interests of society, is the most obvious form of a subsidy in transport. It is compensation to the provider of the service for the losses incurred. The primary beneficiaries of the subsidy are the users of the service. After the *National Transportation Act* (NTA, 1967) and until the formation of VIA Rail in 1978, CN and CP received compensation for 80 percent of the losses they incurred in the provision of intercity passenger services. This was an *ex post* payment system.

Ex ante compensation: The *National Transportation Act, 1987* (NTA, 1987) has provision for an *ex ante* subsidy system should it be desirable to maintain airline services existing in 1987 but which are no longer remunerative. Section 85 of the NTA, 1987 requires that unremunerative services that are to be maintained should be contracted following competitive tendering.

The payment of compensation to carriers for losses resulting from the imposition of a public duty was an important recommendation of the MacPherson Royal Commission in 1961. It was incorporated in the statement of national transportation policy in the NTA of 1967 and is unchanged in the NTA, 1987. Paragraph 3.(1)(f) of the latter Act states: "each carrier or mode of transportation, so far as practicable, receives fair and reasonable compensation for the resources, facilities and services that it is required to provide as an imposed public duty."

Compensation is appropriate for two reasons. First, as transportation markets become more competitive, it is no longer possible for carriers to earn a sufficiently high profit in some markets to offset losses in another. The effect of competition on the railways made this abundantly clear to the MacPherson Commission. Second, even if the cross subsidy of one service by another is possible, it is inappropriate for one sector of society, identified by chance through the importance of a particular service to them, to pay the subsidy decided appropriate by society at large. This would amount to selective taxation.

Using the railways as an example, and supposing that cross subsidies were possible, it would most likely be the shipper of bulk natural resources from remote locations that would pay the high rates to cover the cross subsidy requirement. A cross subsidy is possible when an enterprise is able to earn some economic rents or excessive profits in a market with limited competition. An inefficiently small output is produced in the "taxed" market.

Payments for research and development: Subsidy payments may be made to carriers for the support of research and development otherwise unattractive to a company. No actual service to the public may be provided immediately. However, on occasion, a service may be involved, as with the experimental short-take-off-and-landing (STOL) service operated by Air

Canada between the island airport in Montreal and Ottawa between 1973 and 1976. Users will benefit coincidentally from the greater range of service available (while some other suppliers may be losers).

Tax concessions: Various forms of tax concessions may be used to enhance the cash flow of tax-paying organizations and, thereby, stimulate certain activities. Examples of concessions could include exemption from consumption taxes otherwise payable, for example, a fuel tax; tax credits for certain expenses incurred, for example, research expenditures; and acceleration of capital cost allowances on particular classes of assets for an industry.

Tax concessions may be viewed as "tax expenditures," a term used to denote the tax revenue foregone. Subsidies provided by tax concessions are less visible than direct payments. They are also less readily measured. The amount of tax foregone on a consumption tax may be easier to estimate than the amount foregone through accelerated capital cost allowance. The latter only applies when firms are earning profits.

Subsidies When the Public Provides Facilities or Services

Subsidies when the public provides facilities or services arise in a variety of situations when the revenue generated by the facilities or services do not cover their costs. The immediate beneficiaries are the users of those facilities and services. If these beneficiaries are suppliers of services in a competitive market they may be forced to pass on the benefits of the subsidy to consumers. The subsidies that exist in the provision of facilities and services to carriers are in spite of the principle advanced by the MacPherson Commission and reflected in the NTA, 1987, paragraph 3.(1)(e), namely, that "each carrier or mode of transportation, so far as practicable, bears a fair proportion of the real costs of the resources, facilities and services provided to that carrier or mode of transportation at public expense."

Unremunerative services provided to travellers: The public provides some services directly to the public without covering the costs from user revenue. Ferry services are an example. They range from the extensive deep-water services in Atlantic Canada to short crossings of rivers throughout Canada. VIA Rail is also an obvious example, although it is appropriate to note that VIA purchases a significant component of the transport operation from CN and CP Rail.

Unremunerative facilities and services for routes: The public also provides infrastructure and services which must be used in conjunction with other assets to create transportation services. Public expenditure provides routes both in the form of physical structures and of services. The largest public investment in the nation is the network of roads provided for the operation of private and for-hire vehicles. As well, navigation aids and a variety of services are provided to facilitate the operation and safety of air and marine services.

Unremunerative facilities and services for terminals: Passenger terminals are currently provided at public expense in the case of airports, and, in the special case of cruise ships, in ports. For both routes and terminals, the losses incurred may be visible in the sense of shortfalls in revenues against costs, although they may be difficult to quantify accurately. Hidden subsidies may also exist — for example, the absence of taxes on public land. This is a subsidy in the sense that competing land uses have to pay taxes.

Implicit capital subsidies of Crown corporations: The public provision of services by government may involve subsidies which are hidden in the capital structure. The cost of capital to a government agency is reduced by the assurance provided by state backing. The cost of capital does not reflect the true risk associated with the investment undertaken. The effect of risky projects on the cost of capital is a continuing concern for private firms, even though the effects of independent projects are averaged across the activities of a firm.

Subsidies Associated With the Protection of Services

The output of firms may be made more profitable by providing protection against competition. This may be achieved by tariffs or quotas that keep out foreign goods, thus protecting domestic goods or services by the regulation of entry into the industry.

The transportation industry has been subject to a wide range of economic regulation, including entry control for a variety of reasons. They include the avoidance of destructive competition, protection of "infant industries," and the interest in preserving the ability of carriers to meet public service or common carrier obligations. The latter category has really meant that carriers sustain unremunerative services by cross subsidizing unprofitable services from profitable ones. The cross-subsidy structure has been sustained by precluding competition on the profitable routes. The effect is to support a subsidy mechanism through selective taxation achieved through the regulatory process. A recent high-profile example of this policy is the regulation of entry into the telephone industry and the cross subsidization between long distance and local calling services in that industry.

An example of this policy in passenger transportation is in the regulation of bus services in Quebec and Saskatchewan, where there is strong concern to preserve services to small communities. The latter concern had also been reflected in the service mix operated by Air Canada when it was a Crown corporation.

Regulation may be introduced and administered to achieve cross subsidization. However, it may also simply create an environment in which cross subsidization evolves because of the passive role of authorities in regulating rates and because of the diminished market discipline faced by carriers.

Cross subsidies most often become evident when firms are faced with a loss of earnings from their protected markets. For example, cross subsidies have become issues as competition from other modes erode the earnings of main bus routes and as potential competition stimulates "rebalancing" in telephone rates.

Subsidies By the Absorption of Costs

The costs of doing business may be absorbed by society in two quite different ways. The first is frequently recognized as a form of indirect subsidy. The second is less often thought of as a subsidy.

Loan guarantees: Loan guarantees are frequently used by government to enable capital to be available to companies, or available to companies at lower rates than otherwise would apply. The result is that the costs associated with the possibility of the venture failing are shifted to government.

Environmental externalities: Externalities are the consequences, positive or negative, of an activity on others and for which the supplier does not receive or pay compensation. Externalities have long been recognized as important in transport. For example, the consequences of added traffic on a route, whether highway, airline or canal, can be delay to other vehicles. The immediate effects of congestion are internal to the particular transport system. They result in inefficiency but not directly to the subsidization of the mode. Air and noise pollution, however, are externalities exacerbated by congestion; they are examples of costs imposed on society. These externalities fall on society in a general sense, not on government. The incidence of externalities within society varies with their specific nature.

Differences in accident rates among modes of transportation are also associated with an externality to the extent that the health care costs resulting from accidents are borne by government and not by the affected travellers.

The environmental consequences of transportation are becoming more widely recognized. While the costs are being reduced by a variety of measures, such as non-leaded gasoline and mandatory seat belts, they are not eliminated. Their values are difficult to estimate but differ among the modes. The effects of their absorption by society on modal output is uncertain. (It should be noted here that the environmental performance of modes must be judged on the basis of their effects in relation to actual passenger transportation performed and not on the basis of their potential performance.)

PURPOSES FOR SUBSIDIES

In view of the various types of subsidies, it is not surprising that while all subsidies have explanations, some have more explicit purposes than others. Before these purposes are discussed, it is helpful to recognize important attributes of subsidy programs in practice.

First, not all subsidies are the result of deliberate programs. While many subsidies are intentional, others are coincidental and still others are accidental. An example of a deliberate subsidy is payment to VIA, but the saving in the cost of capital attributable to Crown corporations is coincidental. Such a coincidental subsidy could reasonably have been anticipated, but the subsidy flows from the program of Crown ownership which was introduced for other reasons.

Some subsidies have come into existence "by accident." An example of such an accidental subsidy is in the failure of the St. Lawrence Seaway to come up to early expectation and cover its costs. This situation produces a "subsidy" in the sense that the users will not be able to meet the system costs. However, where a loss arises from an investment decision which, with the benefit of hindsight, was inefficient, no "subsidy" should be considered to exist. The sunk costs are now irrelevant. In the long run, there will be questions about the ability of the service to cover future costs.

Second, subsidy programs frequently lack clarity of purpose and recognition of costs. Given the variety of political considerations surrounding many subsidies, obfuscation may be politically expedient. However, as will be argued later, it can be a major contributor to the high cost and ineffectiveness of a subsidy in achieving desired objectives.

Subsidies may be introduced to achieve a variety of purposes. This is particularly true of the transportation industry which affects so many aspects of society. Although individual subsidies may contribute to several purposes related to various broad goals of society, the reasons for subsidies to passenger transportation are restricted to three broad categories: contributions to nationhood, change in income distribution and increase in economic efficiency; these are listed below in Exhibit 2.

Exhibit 2

Summary of the Purposes for Subsidies

A. Contributions to nationhood
- defence, justice and social services
- idependence
- political unity
- prestige
- preservation of culture

B. Income redistribution

C. Economic efficiency
- public goods and services
- externalities
- decreasing costs
- inadequate information

Contributions to Nationhood

Transportation contributes to nationhood in various ways. The relative importance of these contributions changes over time and varies between countries, but collectively they exercise continuous and strong influences.

Defence, justice and social services: An independent, sovereign state must have the mobility as well as the resources to ensure the administration of justice and the performance of national defence. Historically instrumental to the construction of early trunk roads, these fundamental requirements still have their relevance. The establishment and maintenance of Canadian sovereignty in the Arctic islands require a sufficient Canadian presence — a presence which still must be supported from government budgets.

In the social area, a current transportation-dependent service which Canadians might expect to be able to provide to all communities is access to health care. While not a necessary attribute for a nation, it may be a fundamental service expected for communities, irrespective of their economic opportunities. Such accessibility is provided by air transport.

Independence: Transportation played well-known roles in ensuring Canadian independence and sovereignty. Each mode of transportation has its example. Observation of Canadians' behaviour in selecting their travel routes today suggests that the need to preserve independence by "buying Canadian" no longer seems relevant. However, in political debates, the preservation of Canadian alternatives, be they companies or routes, is a well-known argument. How important independence is in this argument and what it currently means are not clear. It is certainly in the minds of some. However, the protection of their own capital and of their own jobs may be in the minds of others.

Political unity: The basis for political unification of disparate geographical units into one Canada has been achieved, in part, by offsetting the economic cost of distance with subsidies. Railway service and, later, the reduced freight rates for Atlantic Canada, and the promise of a railway to British Columbia were the early forms of this. The later extension of the Maritime freight subsidy to trucking in Atlantic Canada, and the operation of buses for rail passenger service in Newfoundland have been efforts to make the subsidies more cost effective.

Prestige: National symbols are important. Sports heroes such as Pelé and Gretzky, and participation in the Olympics are evidence of this. So too, apparently, are national airlines, given the number of countries that have one. While Canada, unlike many developing countries, has resisted the need for a national shipping line, we are now testing the importance which we attach to Canadian-owned airlines. (Related issues in the current airline policy debate include the competitiveness of markets and the equality of opportunity for Canadian and foreign firms to behave and compete in comparable ways.)

Preservation of culture: Assertion of community cultural values is a strong phenomenon of our times. Within Canada, it may take the form not only of greater identity within various ethnic and cultural groups, but also of sustained concern for preserving Canadian values in the face of the Free Trade Agreement and broader trends to globalization.

The environment gives rise to concerns for the preservation of "things Canadian," quite possibly including the continuation of transcontinental rail passenger service. The argument might go: "Canada would not be the same without it; the railway laid the foundation of the country." If unremunerative rail service is to be retained, the Commission may have to decide whether rail service has cultural value in an evolving and dynamic society. Of course, there may be other reasons for subsidizing rail service. (This does not deny that there may be economic as well as cultural value in operating museums, as the number of steam-engine rail services and paddle-wheel cruises testify.)

Income Redistribution

Regionalism is strong in Canada for geographic and cultural reasons. The country early used railways as an instrument to attract remote regions into Confederation. Subsequently, the subsidization of transportation was used to promote regional economic development. In effect, this is a way of taking wealth from the "haves" and using it to enhance the economic opportunity of the "have nots." The Hudson's Bay railway is a costly example.

Transportation may also be subsidized to provide enhanced mobility for particular sectors of society. For example, it is argued that the maintenance of subsidized rail service facilitates the mobility of poor and physically disabled people. (Rail service can offer persons with disabilities better mobility while travelling than can bus service.) The requirement that carriers provide extra facilities for disabled persons could result in services at less than cost. In the absence of public compensation, this would require cross subsidies.

Economic Efficiency

There are many arguments for the provision of subsidies to increase the efficiency with which a nation's resources are used. Subsidies may be used to enhance the efficiency of the economy rather than to redistribute wealth for public reasons but at an economic cost. However, subsidies introduced for efficiency reasons will have some distributional consequences. They may also have some offsetting side effects on efficiency in the economy, as discussed in the next subsection. Subsidies intended to enhance efficiency have their costs as well as benefits.

The economic arguments for such subsidies can be placed in four categories. They are: inefficiencies in the provision of public goods and services; the influence of externalities; decreasing costs; and inadequate information.

Public goods and services: Public goods and services cannot be efficiently marketed. A consumer can derive a benefit from the service without adversely affecting consumption by others. The classic example is the service provided by a lighthouse which is available to all ships without any effect on lighthouse costs. The marginal cost is zero; therefore, the appropriate price is zero.

The view of roads as public goods was important during the early development of roads when they primarily provided property access and were, in general, little used. The increased use of roads and their use for long-distance travel rather than access, have changed their nature. The costs of providing additional road capacity for users have become significant, and control of access to divided highways through tolls or similar means has become practical.

Externalities: Economists' models of perfect competition assume that suppliers and consumers experience the full consequence of their actions. This is not the case in reality. For example, transportation gives rise to a number of effects felt beyond the individual supplier or user. Externalities, as noted earlier, are relevant here as they may give rise to conditions warranting subsidies.

Where externalities are positive, a subsidy may be appropriate to encourage an expansion of output to an optimum level. A frequently cited example is the case for a subsidy to orchard owners to reflect their contribution to the value of honey produced by separate beekeepers. An example may occur in transport when an increase in the number of users of a scheduled service enables service frequency to be increased, to the benefit of all users. The result for a service charging marginal cost, taking into account producer and user costs consistent with economic theory, is that marginal cost is declining and that a subsidy is necessary for the service to be financially viable.

Transport services produce a number of negative externalities. Most widely recognized are air and noise pollution. These externalities may be dealt with directly through taxes or regulations, or by encouraging other activities or modes of transport which do not produce these negative externalities. Subsidies are suggested for bus and train to offset the negative externalities of the automobile.

This is the appropriate place to discuss the issue of energy conservation. There are four points to make. First, if energy prices reflect scarcity, there is no reason for specific intervention. Second, energy is only one of a number of inputs. Forced conservation of energy will likely result in waste in other resources such as capital. Third, energy conservation at one stage in production may result in higher energy consumption at others. It is necessary to look beyond the immediate process. Fourth, energy consumption must be related to the work performed, not the technological possibilities. For example, a full aircraft may require less oil per passenger trip than a train with 20 percent occupancy.

Decreasing costs: Charging an efficient price equal to marginal cost will result in a loss when a firm's cost per passenger or ton falls as the total amount of traffic carried increases. This condition may arise when there are increasing returns to scale — for example, larger vehicles are inherently more efficient — or if there is a surplus of capacity which may be caused (temporarily) by lumpy investment. Losses may be avoided in two ways. One method is for the public to subsidize the service. An alternative method is to allow the carrier to practice price discrimination, that is, to charge different users different prices. This is the approach used by airlines in seat management and by VIA in the pricing of services.

A particular form of the increasing-return argument is for infant industries which may not currently be efficient or viable because of small size but subsequently should grow into viable and profitable firms. The infant industry may justify "start-up" subsidies; the early protection of Trans-Canada Airlines might have fallen under this category.

Inadequate information: A final condition that may justify a subsidized activity is inadequate market information. For example, consumers may lack knowledge which may justify government support of consumer groups.

THE ECONOMIC EFFECTS OF SUBSIDIES

Whatever the reasons for subsidies, it is important that their effects be continually assessed. From the time that they are inaugurated, it is appropriate to consider the benefits and the costs, and how the effects vary with the level of subsidy. Questions should be asked not only about the existence of a subsidy but, if justified, whether it is at the right level.

The economic effects of subsidies are derived from the obvious: subsidies change behaviour. The major difficulty with the analysis of subsidies is the impracticality of forecasting all of the effects *ex ante* or of identifying and measuring them *ex post*. In a dynamic world this is of little surprise, but it is of no comfort.

A study of government intervention in OECD countries[2] notes that it is impractical to carry out theoretically correct comprehensive analyses of subsidies. The absence of detailed data requires levels of assumption that makes partial analysis more appropriate.

The effects of subsidies depend on attributes of the buyers and sellers of goods and services. These attributes are discussed in the following pages before an examination of the nature of the economic costs of subsidies and of the unintended and, generally, indirect effects of subsidies. The section concludes with a consideration of some issues raised by the incomplete recovery of costs for publicly provided facilities and services.

The Importance of Buyer and Seller Responses

Subsidies are designed to change the consumption of particular goods or services. The extent to which they achieve this is dependent on the elasticity of supply and demand, that is, the extent to which the behaviour of the buyer and seller is influenced by changes in price. If the amount supplied and demanded are both highly responsive to price, a significant change in output will result from a subsidy. If neither are responsive to price, little change in output will be achieved.

Further, the extent to which a subsidy is consumed by suppliers' cost increases rather than realized by buyers in increased consumption is dependent on the relative elasticity of supply and demand. In a situation where suppliers face sharply increasing costs, much of the subsidy will be consumed by the supplier. If supply could be increased without limit at a constant cost, the benefits of a subsidy would accrue to users.

Assessment of the effectiveness and incidence of subsidies is critically dependent on a knowledge of the elasticity of supply and demand for the service concerned. The elasticity of demand is affected by the availability of substitute services. The elasticity of supply is affected, also, by the amount of competition.

When buyers or users of a transportation service benefit from a subsidy, it is important to distinguish between those that realize a "windfall" gain, and those that benefit from the subsidy shift. Travellers that would have been using a service anyway, get the windfall gain. Those that use the service because of the subsidy are the intended targets of the subsidy. (Given the general willingness of foreign tourists to pay a much higher [probably commercial] rate for rail service, it would be interesting to know the proportion of foreign tourists that use VIA service between British Columbia and Alberta!)

The Direct Economic Costs of Subsidies

In theory, the economic cost of a subsidy is not measured by the expenditure of the subsidy, whether in the form of cash payment or tax revenue foregone. Such expenditures are merely transfers from one group in society to another. The economic cost should be measured by the value of output foregone as a result of the transfer. That depends on conditions in the economy and the mechanism by which the revenue is raised. As a practical matter, payments (as distinct from tax expenditures) are presumed to be economic costs. They may overstate the economic costs; for example, the social cost of labour may be overstated during a period of high unemployment. However, they are more likely to understate the *total* economic cost.

Costs are incurred not only by the need to transfer resources from one sector to another by taxes, but also by shifts in the consumption of substitute services. For example, subsidies to rail service results in transfer of passengers from air, car and bus services. The substitution effect is believed to be greatest on bus service. The effect is to lose the net value of some bus service, together with the cumulative effect that this has on the value of bus service frequency. A transfer of travel will also take place from car to rail, which may be associated with an immediate net gain because of the negative externalities associated with car travel. In the example, the additional loss in value from bus travel might be offset by the reduction in externalities of car travel.

However, perhaps a more important source of cost, than the inefficiencies resulting from marginal shifts in resources from more to less productive uses, is the behavioural response of buyers and sellers. Presumption that the economy is efficient when prices are set on the basis of appropriate concepts may be seriously in error when suppliers are protected from market forces. This is now a matter of increasing concern to the World Bank.[3]

Three aspects of behavioural response are important. First, because of subsidies, suppliers may be protected from the rigours of the market and become less innovative and efficient. Second, those providing goods or services to suppliers may be able to achieve some economic "rents," that is, returns higher than competitive market conditions would warrant. This might be true for various inputs provided to VIA. Third, consumers may be lulled into the acceptance of service and productivity conditions not otherwise acceptable. This is currently evident in the response of users to government proposals to increase cost recovery in the marine and air modes. Faced with new charges for facilities and services, carriers are demanding greater efficiency in supply. The collective result of these three responses can be a significant detraction from the efficiency of supply. The indirect costs of subsidies can be large.

The Indirect Effects of Subsidies

The behavioural response of suppliers and consumers just outlined are unintended. Other unintended ramifications of subsidies are of concern here.

Subsidies have indirect ripple effects, the magnitude of which are not easily identified with subsidies. The well-known Canadian example is the effect of subsidies on the movement of grain by rail. The effects include a shift in meat processing out of the Prairie region; concentration on grain monoculture in the Prairies with various economic, environmental and social effects; a stultified trucking industry; and a high-cost grain logistics system. These effects of holding down the rates for rail transportation are well known. Getting rid of the inefficiencies is not easy because groups develop vested interests.

Subsidization of passenger transportation can have comparable effects. Take the case of subsidizing rail and bus services in southern Ontario, to offset the externalities of the car. The result? Low-cost transport and the consumption of more transport than is warranted. Social and economic choices are geared to the low apparent cost of transport. Community growth and travel behaviour are more transport-intensive than would otherwise be the case. Long work trips by rail, as well as by car, may become more common, while low-density living will result in reliance on the automobile for local and other trips.

The examples of grain and passenger transportation demonstrate the important interrelationship that exists between land-use patterns and transportation. The effects vary with the subsidy from local to national in scale. The magnitude of the effects depends on the relative importance of transportation in location choice. When transportation subsidies affect locational decisions, they have deep and long-lasting effects.

Subsidies and Cost Recovery

Some subsidies exist in transport because fees collected from users of publicly provided facilities and services may not cover the costs those users impose on the system. This applies in each of the modes of transport and gives rise to issues worthy of special attention.

Unfortunately, the reasons for less than full cost recovery are not generally known. It may be that subsidization has been intended for reasons of nationhood, income distribution or efficiency. However, it may also be that practices followed when the national infrastructure was being developed have been continued because of the various technical and political difficulties in implementing user charges. Subsidies may have evolved by default.

The role of infrastructure has changed as the economy has matured (and as technology has changed). Infrastructure which was capital-intensive and little used originally required public support, for example, roads and railways. Traffic on roads has increased greatly; in fact the issue has often shifted to one of dealing with congestion.

Second, economic growth and technological change have diminished the importance of specific facilities to nation-building, and alternate pricing has become practical. The volume of traffic is sufficient to pay for facilities, and practical methods exist or are becoming available to implement charges. This is not the place to examine the various issues and mechanisms for implementing user-charges to achieve cost recovery. It is sufficient to note here that Canada has not and is not pursuing these aggressively! Witness the exclusion of recreational boaters on the recent government proposals for user-charges on waterways. Developments in road pricing and private funding for roads are examples of the technological and attitudinal changes affecting highway finance, especially in other countries.

In theory, efficiency is maximized if tax revenues are raised from sources that have least effect on the (assumed) efficient allocation of resources in the economy. Following this principle, taxes would be increased on goods and services with inelastic demands. This argues for general revenue funding, which provides government with the greatest flexibility in raising and allocating tax revenues. These advantages must be weighed against the tendency toward inefficient supply when users and suppliers are not subject to pricing discipline. The combined result is that arguments are arising for the collection of user-charges sufficient to cover facility costs and for these revenues to be placed in dedicated funds.

SUBSIDY MANAGEMENT

The management of subsidies covers many tasks, including deciding whether a subsidy is warranted, selecting the type of subsidy and managing the subsidy program. A summary of management considerations is presented in Exhibit 3.

Exhibit 3

SUMMARY OF SELECTED SUBSIDY MANAGEMENT CONSIDERATIONS

A. Desirable attributes of subsidy programs
- provide direct subsidies
- subsidies should allow market competition
- subsidies should not distort input costs
- subsidies should be transparent
- subsidies programs should be monitored and results publicized

B. Deciding on a subsidy program

C. Financial subsidy administration
- competitive tendering
- operating tenders
- negotiated service contracts
- *ex post* subsidies

Desirable Attributes of Subsidy Programs

The management of subsidies should be guided by generally desirable attributes of subsidy programs. Subsidy programs are considered effective when resources are used efficiently to accomplish the program objectives and when the cost and effectiveness can be monitored over time. Unfortunately, many subsidies do not have these attributes, and the beneficiaries of subsidies would often rather hide the existence of the subsidy, perhaps for fear of losing the benefit. Five program attributes are recommended to ensure that services are provided efficiently and that desired results are achieved.

Provide direct subsidies: Subsidies are most effective if provided as directly as possible to the intended beneficiary. First, the process requires explicit recognition of beneficiaries, a process likely to cause a critical assessment of the subsidy's merit. (For example, the subsidy per passenger given to tourists going on the Algoma Central tour train might cause taxpayers and competing tourist attractions to take exception.) Second, the effectiveness of the subsidy could be enhanced by leaving the beneficiary with an unbiased choice of transportation to be used, rather than by subsidizing a mode or carrier. If the "pay-the-farmer" policy makes sense for efficient grain transportation, so does "pay-the-traveller" policy for efficient passenger transportation, if a decision to provide a subsidy to this group has been made.

Subsidies should allow market competition: The effectiveness of subsidy expenditures will be enhanced by mechanisms favourable to competition. Maintaining competition reduces the risk of a subsidy inhibiting innovation and efficiency. An example is the use of competitive service tendering as a means to ensure that, even if a specific service is subsidized, competition is effective periodically, that is, at the time of tendering. Subsidies, including service tendering, should be structured to allow intermodal as well as intramodal competition as far as possible. When subsidies are introduced to benefit the mobility of travellers, the competitive balance in the transportation market and the effects on innovation are affected least by following the "pay-the-traveller" policy.

Subsidies should not distort input costs: Subsidies which distort management decisions on resource inputs should be avoided, unless they are specifically needed to correct otherwise inefficient input costs. They should not be used as a general form of subsidy. For example, providing tax relief through accelerated capital cost allowances, or providing low-cost capital, provides incentives for an organization to use capital more intensively, perhaps in place of labour or fuel. When capital costs are subsidized, it may be necessary to introduce offsetting investment rules to avoid waste, as evident in the controversy which has surrounded the use of capital by the transit systems in Ontario.[4]

Lump sum payments, whether paid after tendering or negotiation with a selected provider, are least likely to distort supply efficiency. The duration and monitoring of contract provisions are important to the sustained efficient performance of suppliers.

Subsidies should be transparent: Subsidies should be visible and their purposes clearly stated. It is unfortunate that many subsidies lack these important attributes.

Visible subsidies involve explicit financial transfers. Subsidies achieved through concessions and protection from competition are difficult to identify and even more difficult to quantify.

The purposes for subsidies should be spelled out with care. Transport is not subsidized for its own intrinsic value. Subsidies may be provided for transport to remote regions, but it is not "the regions" that are subsidized, but the particular travellers and interest groups.

For example, a rail service into a remote area may serve three broad clientele groups. The first are the local residents who may or may not have alternate means of travel and who live in the area for a variety of reasons. The second group is made up of the travellers who might come to such regions for the fishing or hunting and to stay at recreation lodges and facilities. Rail service may be one form of access providing a benefit to these travellers as well as to the lodges they visit. The third group may be thought of as wilderness travellers who use the rail service to get to a remote location, where they are then self-reliant. When the rail service is subsidized, and low rates are provided to everyone, all travellers are subsidized by the taxpayers; is this intended?

The reasons for subsidizing mobility for residents are likely different from the reasons (if any) for subsidizing other travellers. In order to examine alternatives, the explicit reasons for subsidies must be revealed. (An interesting condition sometimes applies in remote regions served by rail, under which existing lodges do not want lower-cost road service provided, since the presence of more vacationers would detract from the "wilderness" assets that they enjoy.)

Subsidy programs should be monitored and results publicized: Careful, explicit rationales for subsidies provide the basis for effective monitoring programs. The more direct and explicit a program, the easier and the more likely is the conduct of program evaluation. The effectiveness of the subsidy in relation to program goals and the incidence of costs and benefits should be estimated and reported.

Experience with monitoring subsidy programs is not good. Programs involving expenditures are reported through records of departmental and agency budgets and records. However, except where an agency, such as the National Transportation Agency, has responsibility for a subsidy program, reporting is subject to limited visibility. Critical assessment of the effectiveness and incidence of subsidy effects is usually not carried out.

Subsidy program assessment raises difficult questions about whether assessments should be carried out by program officials or by staff with special expertise in evaluation methods. Difficulties also surround the way results are reported. Recipients and their elected representatives are normally reluctant to see subsidies analyzed and the results publicized. Their perspective is commonly dominant, to the detriment of taxpayers in general and, often, to the detriment of subsidized interests themselves as the effectiveness of subsidies diminishes. Subsidies may even become counter-productive.

Deciding On a Subsidy Program

The management of subsidy programs is about choices. For choices to be made well and for programs to be managed efficiently, transparency and explicit decision processes are essential.

To decide whether or not a transportation subsidy is warranted cannot be a technical process. Judgement will always be important in weighing intangibles and making trade-offs between heterogeneous considerations. However, guidelines can help in the selection of effective programs. They are intended to ensure sufficient information to support rational and effective choices, and include:

- Clear identification of program goals, whether related to nationhood, income redistribution or economic efficiency;
- Consideration of transport and non-transport alternatives;
- Consideration of alternate amounts of subsidy;
- Evaluation of alternatives by such techniques as cost-benefit analysis or cost-effectiveness analysis; and
- Detailed description of the expected distribution of costs and benefits.

The guidelines reflect several important aspects of program choice, emphasizing the different means of achieving goals. There is a right size as well as type of subsidy. A knowledge of the distributional effects of a subsidy is different from and must be weighed with other costs and benefits.

Allowances must always be made for uncertainty in subsidy effects, especially in unintended side effects. Directness of subsidies, the preservation of competition as far as possible, and good monitoring programs help to increase the effectiveness of programs.

Financial Subsidy Administration

Various methods used to provide direct financial support to transportation firms reflect the structure of the industry and government policies. For convenience, four categories of methods are used.

Competitive tendering: Competitive tendering has become more popular for the provision of subsidized services, in conjunction with the move to private provision of public services, as is evident in the United Kingdom, for example. It is most common in Canada in the provision of transit services in small communities. Competitive tendering requires the careful specification of service requirements, itself a beneficial discipline for public (and private) organizations involved in contracting out. The establishment of successful service standards is evident in the contracting of emergency health-care services.

The tendering process is intended to provide the advantages of competition in terms of innovation and efficiency among alternate suppliers. The competitive advantage is realized each time a tender is opened, although the life and versatility of the assets affect the optimum life of the contract period. While an incumbent normally enjoys some advantages on retendering, competition still remains an important influence.

Tendering may allow for some versatility in service design. For example, a community may allow applicants to use large or small vehicles with different service frequency or, alternatively, may define a service level, including the type of equipment to be used. The former approach allows greater scope for innovation.

Services such as bus or air service lend themselves to the tendering process. In rail transportation, the normal service regime provided by the company owning the infrastructure is not consistent with competitive tendering. However, if a separate contract is issued for passenger services to use a railway track, competitive rail tendering would be possible. The practicality of independent rail passenger service is evidenced in Western Canada where one private rail service to Calgary is already approved and a second is being planned.

Operating tenders: While service tendering requires the provider both to supply capital stock and to assume operating responsibility, an operating tender requires only the latter. Capital equipment is provided by the government. This is the case in Ontario, for example, where capital subsidies are used to provide buses, which in some municipalities are operated and maintained by private contractors. While this may result in some advantages derived from centralized buying power and standardized equipment, it has the disadvantage of curtailing innovation and responsiveness to local conditions.

Negotiated service contracts: Where competitive tendering is not possible, negotiated service contracts may be used. Such is the case for BC Rail's passenger services which are subsidized by the province. Negotiated contracts can involve careful service definition and agreed levels of capital investment and operating subsidy. In the case of a "captive" Crown corporation, the quality of the contract is very much dependent on the attitude and ability of the individuals involved. Competition does not provide any check.

The contracts, like tender arrangements, can include penalties for service failures and can provide incentives for the achievement of additional revenues.

Ex post *subsidies:* Subsidies may be provided to public services on an "as-needed" basis. Under such regimes the level of service may be tightly or loosely defined. In either case, the shortfall will be made up by government. Budgets, however, would normally be subject to government approval. The system operates as a cost-recovery process. Consequently, incentive for innovation may be low but the risks of innovation may also be low, at least for the organization, if not the individuals involved. Overall, the system removes the pressure of bottom-line results and has generally been seen as conducive to waste.

III. The Subsidy Experience

The breadth of purpose, types and effects of subsidies outlined in the previous section makes a general description of subsidy programs difficult. Therefore, while certain themes are relevant across the modes of transport, a modal approach is used to review the role of subsidies in Canadian intercity passenger transportation.

HIGHWAY TRANSPORTATION

Since the automobile dominates intercity passenger transportation in Canada, its use must be given careful consideration by the Commission. Relatively small percentage changes in car use can have large absolute values and large potential effects on other modes which are small by comparison. Therefore, the first part of this discussion deals with issues of the provision of highways and of automobile use in general. Characteristics of subsidies involving buses are considered separately.

The discussion does not deal at length with the operation of trucking, although the efficient provision and pricing of roads for passenger transport is unavoidably linked with the efficient use of roads by trucks. The linkage between freight and passenger services must be dealt with by the Commission in each mode of transportation.

Highways and the Automobile

The major component of the transportation system is under provincial not federal jurisdiction. This began in the 19th century when highways were perceived as local in nature and so came under the provinces. Roads under federal jurisdiction are still few, for example, roads in national parks and roads which are a part of federal port or airport complexes. There are only two examples of national highway programs: the Trans-Canada Highway and the roads-to-resources program of the 1960s. The effect of segmented jurisdiction among the modes of transportation is a particular issue that the Commission must address.

The early development of roads was the responsibility of the local community through the use of statutory labour. Local roads were important for property access. Early trunk roads, needed chiefly for defence and the administrative of justice, were considered a public charge. Some turnpike roads were allowed, as in the United Kingdom and the United States, but toll collection delays were of concern in the provincial road systems. Tolls were not practical for roads used dominantly for access to property.

Consequently, the pattern grew up of financing roads from general revenue, with only a share of the revenue coming from users. For many years, about two thirds was collected from users in most provinces, in the form of licence fees and fuel taxes. Fuel taxes, which provide the bulk of revenue, were seen as user-fees varying with use, both by mileage and vehicle weight. However, the increasing fuel efficiency of large trucks with diesel power and improved engine performance have offset higher taxes for diesel fuel. In several American states, a weight-distance tax has been introduced to increase the revenue collected from large trucks.

Governments have found that fuel taxes are an excellent way to raise revenue. In addition to provincial taxes, which now approximate provincial expenditures on roads, the federal government levied an excise tax on fuel, and some local communities levy a transit subsidy tax on fuel. Total taxes paid by users in forms traditionally regarded as user-fees, now exceed highway expenditures.

Thus, taxes on fuel may have various purposes: to increase the cost of highway transportation to reflect some road costs and, thereby, to place road use, in aggregate, at a more efficient level; to act as a "pollution tax" on fossil fuel use (in transportation); to encourage a shift from automobile travel by raising its costs and, possibly, using the resulting revenue to subsidize alternate transport services; to serve as a source of general government revenue.

The latter objective appears important to the federal government. It may cause the Commission to consider the economic, social and political ramifications of policies which identify transportation, like tobacco and alcohol, as a source of high contribution to general revenue. It is interesting that the government may now see transportation as a special source of general revenue, even though, during the development of the country, special attention had been given to reducing the costs of overcoming distance. The taxes affect transportation as both an industrial input and a consumption service.

Unfortunately, the simple comparison of highway tax revenues and expenditures does not adequately compare revenues raised from users with the value of the road resources they use. Annual tax flows measure revenue; that portion (of provincial charges, especially) above normal taxes might be regarded as a user-charge. However, highway expenditures on maintenance, operations and capital are not accurate reflections of true, aggregate highway costs. They ignore the cost of capital, and do not reveal whether capital is being consumed by inadequate maintenance and replacement or being built up by advanced construction. The 1989 study by the National Highways Policy Steering Committee suggests a substantial under-investment in highways.

The current method of financing highways is associated with a number of issues:

- What is the relationship between vehicle types and road costs?
- Would the supply of highways be more efficient if user fees were placed in a dedicated highway fund?
- Do fuel taxes result in user-fees less than or more than relevant costs in aggregate?
- How do costs and revenues compare by road type and by type of user?

As the Canadian transportation system matures, the major issue that has emerged is the efficient use of roads. The most frequently expressed concerns revolve on questions associated with the allocation of costs among users. What wear and tear are associated with which vehicle types, and do vehicle types pay for their share of road costs?

However, issues of paying for the existing system are only part of the picture. How to avoid excessive congestion? How to provide the right highway capacity? These questions are prominent because of the need for different approaches to the cost of additional highway capacity. They arise because of the general absence of road pricing — there are few toll roads.

Fuel taxes raise revenue for governments and they raise vehicle operating costs. However, they apply system wide and do not reflect the costs of specific facilities. While the appropriateness of user-fees on low-density main roads is an issue, the paramount issue for the Commission is the supply of roads to meet the demands of high traffic.

The cost of additional traffic on high-volume roads is high for two reasons. First, the traffic imposes congestion costs on the system. Second, the relevant cost of accommodating more traffic becomes not just the wear and tear on the existing highway but the cost of providing new capacity. The absence of specific road pricing systems results in the absence of the usual measure of the need for new capacity, that is, consumers paying the price. For highways, the volume of traffic on routes is in response to the general regime of highway finance, not the cost of specific routes. The cost of new roads is much higher than the cost of existing routes. Highway users should be prepared and expected to pay the high cost of new roads.

Ensuring that highway users bear the real cost of high-volume routes is a major concern in many countries. Electronic technology is creating new opportunities to introduce road pricing. However, even traditional toll systems are being reintroduced, for example, the Coquilhalla Highway in British Columbia. The economics of road pricing require careful consideration by the Commission.

Not only does the provision of additional highway capacity need much capital, it will also have important implications for the demand for public transport services. Because of the large volume of car travel, a small percentage change can have major implications for the size of public transport.

The Bus Industry

The bus industry is affected by a number of issues similar to those of automobile users. These include the level of highway taxes; the sharing of highway costs among users; and efficient approaches to traffic congestion. Intercity services are affected by urban traffic as well as trunk road conditions. However, the bus industry stands to benefit from a finding that automobile (and truck) traffic is not making sufficient contributions to counteract the cost of highway congestion or to investment costs.

The development of bus services has also been influenced greatly by provincial regulation, which remained unchanged even with the passage of the NTA, 1987. Provincial governments have licensed major carriers who then were expected to cross subsidize services to small communities with the profit from major routes. It is not evident that the program has been successful. The regulated regime may have inhibited technology and service improvements, while returns in the industry were quite high. Today, increasing competition is making cross subsidy impractical.

The position of the bus industry could be significantly affected by an integrated passenger policy dealing consistently with all modes of transportation.

The regulation of the bus industry seems anomalous. The favoured position of VIA for subsidized passenger services needs review.

RAIL SERVICES

Rail passenger services in Canada have deteriorated over the last 40 years in the face of increased competition from automobile and air transportation. In spite of occasional, well-intentioned initiatives by railway managers and governments, the contribution of rail services has declined.

Symptomatic of the problems of passenger service has been the lack of clear statements about the purposes to be served by rail services. In 1967, an amendment to the *Railway Act* provided that railways would receive compensation for 80 percent of losses approved by the Canadian Transport Commission. The incentive left with the railways to diminish the losses did not have the effects the government had hoped. CN introduced marketing initiatives which stimulated some traffic, but mounting losses and dissatisfaction led to the formation of VIA to take over the intercity passenger services of CN and CP. The passenger services provided into northern Ontario by the Ontario Northland Transportation Commission and Algoma Central continue to be subsidized under the *Railway Act*.

VIA has not been given a clear mandate, nor are the reasons clear for operating specific services. Faced with operating services at a heavy loss, VIA has attempted to increase occupancy by using discount fares which reflect the incremental cost of (subsidized) seats otherwise remaining empty. The resulting low fares led to complaints from the Montreal-based bus line, Voyageur Colonial Limited, and hearings before the National Transportation Agency. The second set of hearings was terminated following appointment of the Royal Commission. Figures such as the subsidy per passenger on routes are available but they have not been used in conjunction with detailed passenger profiles nor compared systematically with the cost of providing services by other means. While the subsidies for VIA are explicit, the reasons, effectiveness and incidence of them are obscure and muddled.

AIR SERVICES

Three types of subsidies are found in air services: incomplete cost recovery, payments for unremunerative services and cross-subsidized airline services. The main subsidies have been created by partial cost recovery.

The contributing factors to this are a scarcity of cost-recovery initiatives — for example, in connection with air navigation services; low traffic volumes at many airports; a national, uniform and inefficient system of landing fees; and limited innovation in airport development and in pricing airport services. Those matters are being addressed currently in the context of the cost-recovery initiative and airport reorganization.

It should be noted that an important issue in the proposed scheme of user-charges is the extent to which charges are designed to achieve cost recovery rather than efficiency. Other issues are whether funds should go to general revenue or an aviation fund, what role new airport authorities may play in the national system and whether air services are to be funded as a system or whether individual airports are to be treated separately.

Although aviation comes under federal jurisdiction, the provinces provide subsidies of two types. Some provinces, such as British Columbia, fund remote airports; others, such as Ontario and Quebec, support services to remote communities. In Ontario, the services are provided on an operating contract basis for the Ontario Northland Transportation Commission. They are not tendered services as would be required under the NTA, 1987.

Before airline deregulation, it was expected that national carriers, especially Air Canada, would provide some unremunerative services to small communities. Meeting this public service obligation was a responsibility that Air Canada accepted. When the forces of competition made cross subsidy impractical, the public responsibility was a concept that some managers (as well as some politicians) were slow to give up. However, innovation under deregulation has resulted in more frequent services to such small communities but with smaller aircraft matched to traffic volumes and operated by regional rather than national carriers, or by regional subsidiaries of the main carriers.

MARINE TRANSPORTATION

Large subsidies are provided for ferry services in Atlantic Canada as a result of constitutional obligations. In British Columbia, ferries are primarily a provincial responsibility, although a modest federal subsidy is provided for remote services.

The level of service and financial performance of ferry services are affected significantly by the seasonality of vacation travel. Explicit policy statements about the level of services and expected cost recovery from tourism and other major traffic segments apparently do not exist.

Where ferry services are short, as in river crossings, they may be treated as parts of provincial highway systems. These services may be provided without charge.

Although marine services can be costly for taxpayers, they do not appear to be controversial at a national level. Two conditions may account for this. First, the main need is constitutional in nature. Second, they have little effect on other modes.

IV. Assessment

Subsidy programs have evolved in the individual modes of transportation in keeping with traditional practice. No coherent approach to the roles of the modes and of the subsidies affecting them is emerging. Notwithstanding the overall policy statement in section 3 of the NTA, 1987 applicable to passenger transportation, the reality is that modal behaviour has simply evolved in a fragmented way.

MODAL SERVICES AND SUBSIDY TYPES

Subsidies are provided in transport under each of the categories recognized in Section II of this paper. However, there is significant variation across the modes. Exhibit 4 is a summary of subsidy types by mode.

Exhibit 4
Summary of Subsidy Types by Mode

	Payment	Provision	Protection	Absorption
Auto	—	Yes	—	Yes
Bus	Minor	—	Minor	—
Rail	Yes	—	—	—
Air	Minor	Yes	(ended)	Yes
Marine	Minor	Yes	—	—

Automobile

Subsidies provided to the automobile are the least well understood. This itself is a major problem, because they are hidden in society's absorption of environmental costs and in the public provision of roads.

The technical amount and economic significance of the environmental costs absorbed by society are important issues in our time. They are slowly being addressed by environmental control measures. Such direct measures are appropriate. While the magnitude of the environmental effects of the automobile are uncertain, their existence is not doubted. The general types of response warranted are known. The environmental literature deals extensively with the role of regulations and taxes to mitigate pollution effects efficiently. Pollution is best dealt with directly, not by subsidies to alternative modes.

The most controversial highway subsidy issue is whether travellers collectively pay for highways. An equally thorny question is which parts of the highway network are subsidized. Presumably low-density roads are subsidized; this is not at debate. It can be justified on two grounds. First, these highways provide basic mobility and access, partly funded out of property taxes. Second, the subsidy can be justified economically when the cost of automobiles on roads with limited traffic is low.

The main highway issue is whether users are subsidized by the provision of high-capacity roads. This may be viewed in another way: Are users actually willing to pay for upgraded roads? (Highway users are always anxious to see road improvements which save them travel time and operating costs; they get more and pay less.) The most effective policy mechanism to deal with the issue is a move in the direction of road pricing.

Road pricing provides an opportunity to move away from the supply-driven policies of the past to those in which demand plays a prominent role. "Pay for growth" is a politically saleable principle. Electronic technology is facilitating more complex road pricing schemes than would have been possible otherwise. However, simple approaches can also work — witness toll roads and the Singapore "road pricing" system.

Effective road pricing would have the desired effect of enabling transportation costs to be seen more clearly by travellers. The commercial prospects of public transport services would thus be enhanced. (The level of road charges in Japan and Europe certainly contributes to the viability of rail services.)

Bus Services

Bus services are provided with some intercity subsidies for unremunerative services, for example, in Newfoundland. However, the main issue in the bus industry is the hidden effects of regulation, retained, in part, to maintain cross-subsidized services. The record in other countries and in Canada suggests that cross subsidies are not sustainable, do not provide services best matched to small community conditions, and lead to less innovation and efficiency in protected services. Cross subsidization has none of the attributes of a good subsidy program, but tendered services would be possible and appropriate. New mechanisms of grants to those needing travel assistance could also be used.

Railway Services

While railway services have the merit of being subsidized mainly through explicit payments, the program is bogged down with historical legacies.

The first is our love of trains. It is necessary to recognize that the purpose of subsidies is not to subsidize trains but to subsidize particular passengers. Can better methods be found to do this? To answer this, it is necessary to have more precisely articulated statements of the objectives of subsidies and to have better information on train users. Data on subsidies per passenger are known, but who are these passengers? How many people, as compared with trips, are subsidized? How many tourists? If passengers were given the dollar subsidy implicit in rail use, what mode of transport would they use?

The second legacy is the reliance on a single main provider, VIA Rail, which is handicapped by the lack of a clear mandate. The innovative approach to tourist service between British Columbia and Alberta is evidence of the innovation possible under an alternate policy regime. If subsidy objectives are defined more precisely by route, could subsidy levels and delivery mechanisms be developed more selectively and precisely? Should tourists be subsidized to the same extent as those travelling on "essential" services? Are similar services required across markets? Should services be provided by a single supplier? To what extent can competitive tendering and service contracts be used?

Rail services have also been impeded by the hidden subsidies seemingly enjoyed by highway users in high-density corridors and by air travellers.

Air Services

Air services have had subsidies in most categories. The regulation that enabled cross subsidies for unremunerative services has ended. The environmental effects of air services are most evident in noise pollution around airports. In Canada, affected property users have, generally, not been compensated as they have been in some countries. In Japan, for example, compensation has been provided to property owners for sound-proofing buildings. In Edmonton, property owners adjacent to the city airport have been given a reduction in property taxes. However, compensation does not remove the subsidy to the air services unless the revenue is raised from them.

The noise frequency at airports is increasing because airport traffic is growing. However, the introduction of new third-generation, quieter aircraft, is greatly reducing the area of the noise imprint of flight paths and reducing noise levels within the affected areas.

Since noise pollution caused by air services continues to be dealt with by regulations on aircraft operations, some inefficiencies may result from excessive or too low standards. However, change to a system of industry payment and compensation for those affected is a change that goes beyond transportation to compensation in society generally.

The subsidization of air services through less than full cost recovery for airport and navigation infrastructure and services is being addressed in the current cost-recovery initiative. This program raises questions about the extent to which efficiency and cost recovery can be achieved; economic efficiency may require deviation from annual accounting cost recovery. The complexity of efficient user-charges has led to special studies of the subject. It has raised questions, also, about the extent to which airports should be treated as parts of a system to be financed collectively.

No air services are provided under subsidy under the NTA, 1987. However, some services to remote communities are subsidized by provinces. In Ontario, for example, the Ontario Northland Transportation Commission has provided services through norOntair since 1973. norOntair continues to serve a mix of profitable and unprofitable northern routes, and receives an *ex ante* negotiated subsidy. Several provinces provide emergency air health-care services under contract with private operators.

Marine Services

Marine services are subsidized primarily through Crown corporations which operate at a loss. The largest losses are incurred on federal services in Atlantic Canada. The services vary by community type and traffic type, and communities vary from small and remote to large urban areas. Passengers, which include residents and business people travelling occasionally, commuters and tourists, exist in different proportions on different routes. As with rail services, the rationale and amount of subsidy incurred on behalf of the individual groups remain unclear.

Some ferry services are provided without charge as a part of a provincial highway system. While these generally are for river crossings, the apparently arbitrary distinction between services provided free and those for which there is a toll gives rise to controversy.

COMPARISON OF SUBSIDY PRACTICES AND PRINCIPLES

There is a significant contrast between the principles recommended for the management of subsidies and the attributes of subsidies found in Canadian passenger transportation. An important contribution to the effectiveness of passenger transportation in Canada would be to bring about a greater coincidence of actual with recommended practices.

The reasons for the gap between reality and recommendation are not hard to find. First, Canada started by using transport as an instrument of nation-building. Subsidies for the construction of roads and railways were to meet the basic community needs of mobility for defence and justice as well as of trade. In spite of changing needs, practices have not changed.

Second, the nature of transport services has involved government in ways that hide subsidies. These include the public provision of facilities and services, as in public roads, airports and ports, and in the regulation of for-hire services.

Third, the limited role of competitive markets has restricted the influence of commercial pressures. Disciplines of the market have been stronger in the freight mode.

Finally, politicians have an understandable but unfortunate reluctance to reveal the level and incidence of subsidies for fear of upsetting their constituents.

The desirable attributes of subsidy programs are:

- Provide subsidies as directly as possible to the intended beneficiaries;
- Provide subsidies in ways that do not distort competition;
- Avoid subsidies which distort input prices;
- Make subsidies transparent, that is, visible and with a clear purpose; and
- Monitor and publicize results.

The most important of these attributes is that subsidies be transparent, a quality that is essential for effective monitoring. Clarity of purpose must go beyond statements of provision of service to a particular community and measures of total ridership, revenues and costs. It must identify particular types of users and beneficiaries.

Detailed technical studies on the effectiveness and incidence of specific passenger service subsidies would be pathbreaking. They would make contributions comparable to the grain costing work of the MacPherson Commission, which provided an example of technical analysis and provided a basis for important policy recommendations.

The attributes of desirable programs do not point directly to one type of subsidy over another. However, the nature of subsidies as they exist in Canadian passenger transportation suggests that those which are least likely to meet the desired criteria are absorption and protection, followed by subsidies through the provision of facilities and services without full cost recovery. The payment method of subsidization is likely to be able to meet the criteria. Unfortunately, as practiced in Canada at present, it does not do so.

The absorption of costs is undesirable as it encourages the inefficient use of resources. In transport, it is mainly associated with environmental effects which are being dealt with directly, albeit late and slowly.

Subsidies by protection, such as in bus services, are hidden, often ineffective for those subsidized and result in less efficient services overall. Subsidized services would be more effectively provided if service contracts or competitive tendering arrangements were used. The extent of competitive tendering may be influenced by the market structure and policies respecting Crown or private operations. Increasingly, the latter is preferred and competition more likely. The provision of direct "payment" to travellers may also be considered.

The provision of infrastructure by governments creates situations in which subsidies may be hidden and purposes unclear. Less than full cost recovery for low volume airports or roads may be efficient and consistent with other policy objectives.

Subsidies may exist also in the provision of intensively used facilities. Congestion costs and the high cost of new facilities must be considered. The similarity of the economically correct policy of having users pay marginal costs with the politically acceptable policy of users "paying for growth" may provide an important basis for policy recommendations. This principle may be particularly helpful in advancing the case for more direct charges than fuel taxes for roads with high traffic volumes. Various schemes are making road pricing a realistic option for the future.

User-fees designed to raise the level of cost recovery will likely raise concern for the efficiency with which facilities are currently used. That is good. For example, the increase in user-charges for air and sea modes has raised awareness about the inefficiency of supply of facilities for those modes. A similar response from road users would probably increase interest in the more efficient use of road space, for example, by high-occupancy vehicles.

The development of new policies for user-charges raises questions about the use of dedicated modal funds. Such funds would not be recommended in theory, but perhaps they are desirable in practice. It should be noted that the discipline that such funds create may not be liked by politicians!

The payment of subsidies to VIA is the most visible passenger subsidy in Canada. The program does not meet a number of criteria for a desirable subsidy program. The target travellers are not identified clearly so that results cannot be measured effectively. The subsidy distorts competition among modes by being provided to a supplier, VIA, rather than to travellers. Competitive tendering of the transportation services does not take place.

Overall, the provision of subsidies in Canadian passenger transportation has evolved over time and under different governments without a coherent framework to ensure that the subsidies would work to achieve a safe, adequate, economic and efficient transport system making the best use of all modes of transportation. A significant shift in subsidy practice is needed to achieve the overall policy goals.

Current practices and policies reflect the historic need for governments to ensure the provision of basic infrastructure and services to build a new nation. Attitudes and institutions have not yet adapted to the requirements of a mature nation in which resources must be allocated among alternate programs, including coping with congestion and expansion. The implication is that less attention should be given to engineering-driven supply initiatives and more attention given to demand-side elements. This includes a stronger role for market mechanisms, such as cost recovery, including road pricing, and a greater attention to particular travellers for whom assistance may be warranted. The desirable attributes of subsidy programs outlined previously are consistent with these recommendations.

Passenger transportation policies in the future need to provide a long-term vision which can guide the development of promotional policy based on the requirements of a mature economy. This shift will be as significant as the change achieved in regulatory policy since the MacPherson Commission first recognized the need to respond to the changing role of competition in transport markets.

ENDNOTES

1. Organisation for Economic Co-operation and Development, *Transparency for Positive Adjustment: Identifying and Evaluating Government Intervention,* Paris, 1983, pp. 32–33.

2. Ibid., p. 249.

3. Rachel E. Kranton, *Pricing, Cost Recovery, and Production Efficiency in Transport: A Critique,* Infrastructure and Urban Development Department, The World Bank, Working Paper 445, Washington, D.C., 1990, p. 45.

4. Mark W. Frankena, "Capital-Biased Subsidies, Bureaucratic Monitoring, and Bus Scrapping," *Journal of Urban Economics,* 21 (1987), pp. 180–93.

6

TRANSPORTATION INFRASTRUCTURE POLICY: PRICING, INVESTMENT AND COST RECOVERY

David Gillen* and Tae Hoon Oum**
May 1992

1. Introduction

Transportation policy issues were, for much of the last several decades, loosely integrated with national and regional economic policies in Canada. Transportation was viewed as a tool of government to promote the economic development and growth of different regions, groups and industries. This position changed somewhat with the passage of the *National Transportation Act* (NTA) of 1967. Based on recommendations of the MacPherson Royal Commission, the Act broke with tradition and established the policy that efficiency within the transportation sector was of fundamental importance in securing a firm foundation for the economic growth and development of Canada. To achieve this goal, competition among modes was selected as the mechanism. The MacPherson report also recommended that any region, group or industry deemed sufficiently deserving of subsidization was to receive direct payment rather than subsidization through the underpricing of either or both transportation services and infrastructure. This was generally adopted as part of the NTA of 1967.

* School of Business and Economics, Wilfrid Laurier University and Institute of Transportation Studies, University of California, Berkeley.
** Faculty of Commerce and Business Administration, University of British Columbia.

Since 1980, external and internal economic and political pressures have resulted in deregulation in the airline and trucking industries, the introduction of intra-modal competition and the privatization of carriers. This policy shift was formalized in the new *National Transportation Act, 1987* which took effect January 1, 1988.

Allowing market forces to determine the structure and behaviour of different modes (primarily carriers) has been the new policy direction. It has also meant a shift away from setting infrastructure prices below costs. Transport Canada's recent proposal of a cost-recovery policy for air and marine infrastructure is a small step toward this goal. A similar shift has not taken place for highways, likely because of the fractured jurisdictional responsibility for roads among the different levels of government. While air is the responsibility of a single agency (the federal government) highways are a combination of provincial, regional, county and municipal responsibilities with the proportions of roadway in each of these levels of government varying across provinces. This spread also characterizes the United States and distinguishes North America from Europe where there is more integrated transportation planning at the central level.

Over this same period, infrastructure policy has been subjected to a number of exogenous stresses including the Free Trade Agreement with the United States and the globalization of industries. Both factors have resulted in growth and shifts in demands which have generated pressure for additional and improved infrastructure as well as for new institutions to manage them. These demands come at a time of restraint with governments under severe economic pressures to be fiscally responsible. Governments are reluctant to commit themselves to significant expenditures even for maintaining the existing stock of transport infrastructure let alone adding to it.

Given the realities facing the Canadian economy, how should one establish pricing principles and investment guidelines which satisfy both the objectives of the *National Transportation Act, 1987* and the needs of Canada and the Canadian economy in the years to come? The fact of the matter is that transportation infrastructure, an important component of the nation's capital stock and an important factor in its economy and welfare, is a victim of both fiscal restraint and the failure to treat it as a scarce economic resource. In the past, transportation policy at all levels of government focussed on capacity expansion (supply) rather than the management of demand for infrastructure. Sensible policy reform must consider the costs imposed by users on the system and on others when using infrastructure, including congestion and other externality costs. Infrastructure shortages resulting from both underpricing and inadequate investment limit the realization of the efficiency gains promised by deregulation and privatization in the transportation sector.

Pricing and investment planning of Canada's transportation infrastructure cannot ignore the forces and pressures developing in the United States and other international markets. Links must be established between investment in infrastructure and the pricing of services delivered by that infrastructure. Indeed, socially optimal modal pricing requires the inclusion of modal air and noise pollution and congestion externalities. Congestion externalities are emphasized in this study but the models and concepts are applicable to air, noise and other externalities. Economic welfare will be lower if these factors are not considered in modal prices because demands for massive public investment in infrastructure will continue unabated, infrastructure will deteriorate prematurely, and the distribution of traffic across modes will not reflect the real modal costs. The solution is not necessarily more investment but rather smarter investment. Smarter investment must start with efficient pricing.

The main purpose of this study is to discuss the principles and methods by which Canada can ensure the optimal use and efficient provision of transportation infrastructure services in all modes of transportation. Emphasis is also placed on the cost-recovery issue. There are a number of reasons for this, including the fiscal constraints faced by government, the motivation to reduce the efficiency costs of financing infrastructure deficits through taxation and the general movement to decentralization.

Cost recovery is treated in two alternative ways. First, cost-recovery conditions associated with optimal pricing and investment in infrastructure are studied and compared to the actual cost-recovery situation. Second, the methods of achieving an exogenously given cost-recovery target which minimizes efficiency loss is discussed. Although the principles and methods used can be applied to all modes of transportation, the emphasis of the discussion (with empirical examples) is on roads and airport infrastructure. The study also emphasizes that allocative and productive efficiency gains are not accomplished through simply financing a cost recovery but rather through achieving cost recovery with efficient pricing.

Section 2 presents a survey and assessment of the principles and alternative methods of infrastructure pricing. It includes a discussion of the potential difficulties for implementing some of the ideal principles. Also reported are the empirical results of some studies which apply the optimal pricing principles. Section 3 examines the literature on cost structures for transportation carriers and infrastructure providers. The focus is given to the empirical results on economies of scale, traffic density and scope. An analysis of the optimal user-charges for airports and roads is presented, and some suggestions for achieving the optimal cost-recovery targets are made in Section 4. This section also reviews current user-charges and cost-recovery status by major user-groups in both air and road modes. Alternative sources of financing road infrastructure currently in use are examined and discussed in Section 5. A summary of the findings is given in Section 6.

2. Infrastructure Pricing: Theory, Issues and Applications

This section presents a survey and assessment of the principles and alternative methods of infrastructure pricing. Although it addresses infrastructure pricing principles, the characteristics and consequences of each pricing principle will also be discussed in a general framework for all goods and services. Potential difficulties in implementing some of the ideal principles are identified and discussed. The empirical results of some studies which apply the optimal pricing principles are also presented.

2.1 THE ROLE OF PRICES AND WELFARE

Transportation infrastructure is like any piece of capital — it represents a stock which yields a flow of services. The services from infrastructure, such as roads and airports, reflect economic scarcity because their construction and continued servicing require the use of scarce resources. Concern about the efficient use of resources in the economy necessitates finding some mechanism and criteria to determine both the distribution of resources between transportation infrastructure and other parts of the economy and resource distribution among modes within the passenger transportation sector. How many miles of roads should be built? How many airports? How big should they be? These, among others, are questions which pricing principles seek to answer.

There are essentially two perspectives as to how to answer these questions. One is the market approach. It uses the price system to allocate scarce resources to their most valued uses on the basis of willingness to pay. If the market mechanism is to be used as a method of maximizing social welfare, all private and public goods and externalities (such as air or noise pollution and congestion) must be "valued." This value or price should reflect the resources used and be a measure of the cost to produce the good or service. This cost is to be the marginal cost, and no buyer or seller of a good or service should have sufficient power to drive a wedge between the price charged and the marginal cost of production. These conditions assure maximization of allocative and productive efficiencies or, equivalently, of social welfare. Allocative efficiency is the measure of performance of scarce resources allocated to end uses, goods and services, that best accords with the pattern of consumer demand. Allocative efficiency is at an optimum when the price of each product equals the lowest resource cost of supplying the marginal unit of the product. Technical or productive efficiency refers to the minimum cost of producing a given output.

The other approach to resource allocation is to use direct planning methods characteristic of some centralized economies and which include the use of various non-market mechanisms. These non-market mechanisms can take a variety of forms including administrative rationing, random allocation and queuing. Some authority must decide how to allocate available goods or services to competing uses and choose the goods and services to produce. However, since there is no efficient method of inducing users to reveal their preferences, the authority has no way of accurately knowing who values a given good or service more. Consequently, the authority may end up providing a good or service to those who do not value it the most. This can lead to significant allocative inefficiency. In most cases of direct market intervention, the objective is rarely to improve efficiency. Efficiency costs, however, are still important as measures of the costs of abandoning a market approach. Furthermore, because there are no signals for the capital market, there will not necessarily be an optimal investment in capacity, except by chance.

In a market economy, prices perform two functions. In the short run, they act as a signal to ensure that scarce goods and services (airport capacity, for example) are allocated to those who value them the most. This ensures that the social benefit, from the utilization of the fixed capacity, for example, is maximized. In the long run, prices provide a signal to the capital market to move capital into those activities which yield the highest return, and thus guarantee an optimal investment in capacity.

Given the different outcomes associated with the alternative policy proposals, efficient pricing or non-market mechanisms for resource allocation, one must use some criteria to assess their relative merits. Economists have generally not considered both economic welfare *and* income distribution outcomes when evaluating alternative policy proposals or pricing methods. The level of economic welfare is defined as the sum of consumer surplus and producer surplus. Consumer surplus is the additional value a consumer derives from consuming a good or service over and above the price paid and is generally measured by the difference between the value revealed by the demand curve and the price paid. Producer surplus is a producer's net revenue over and above the cost of production. A dollar saved by a low-income person is regarded as having exactly the same value to society as a dollar saved by a high-income person. This means evaluation of alternative policy proposals or pricing methods by the welfare criterion alone ignores the income distributional consequences. In other words, an economic reorganization or change is considered beneficial if those who benefit collectively gain more than the total losses incurred by those who lose. Use of this welfare criterion alone is based on the compensation principle which implies that those who gain can compensate those who lose without incurring any redistribution costs. Although the change in the level of welfare (economic efficiency) is generally regarded as more relevant for evaluating policy, the effect on income distribution cannot be ignored because, for most cases, income redistribution is not without cost. The various pricing principles and non-market allocation procedures reviewed are judged primarily in terms of their consequences on economic efficiency (welfare). Attempts are made, however, to make a preliminary assessment of the consequences of some pricing principles for income distribution (fairness).

2.2 ALTERNATIVE PRICING METHODS

Pricing is a method of allocating resources. There is no such thing as the "right" price irrespective of issues and objectives. Rather there are "optimal prices" or pricing strategies with particular objectives to be achieved. Prices can be established to maximize profit, welfare or revenues. They can be used to achieve a particular share of the market or a desired distribution of demand across products (for example, mode-split in transportation). From a society's viewpoint, however, one of the most important goals for pricing goods and services is to maximize economic welfare by optimally allocating scarce resources and goods or services across competing needs in the short run and to ensure optimal investment in capacity in the long run.

Much of the discussion and debate surrounding the pricing of runways and roadways seems to confuse efficient pricing and cost recovery or financing. Pricing transportation and financing transportation are very different concepts. Financing requires only knowledge of costs, and user-charges are set to achieve full or a desired level of cost recovery. Pricing, on the other hand, requires knowledge of both demand (which provides a measure of economic value) and costs since pricing tries to optimize the use of the resource and to balance the revealed value of transportation with the resource cost of providing it. In analyzing various pricing strategies it is, therefore, essential to have a clear understanding of various cost concepts since each pricing principle is related to a specific concept of cost.

Before proceeding with the detailed discussion of the various pricing methods, it is important to clarify the use of subsidies from general revenues to meet revenue shortfalls under some pricing approaches. There are two points. First, it has been argued that any deficits resulting from marginal cost pricing be financed from general tax revenues. Second, it is also argued that full cost recovery is not desirable if it results in prices that do not reflect marginal costs. First-best pricing has prices equal to marginal cost and second-best pricing has prices which deviate from marginal cost in a way which minimizes efficiency losses. Both arguments have an implicit assumption that the efficiency loss is less from raising a dollar for general tax revenues than by raising revenues in a specific market. This is not necessarily true. Jorgenson (1992) noted that the cost of public funds is indeed high. He reported on research by Jorgenson and Yun (1990) which shows that the marginal cost of a tax dollar is $1.46. In other words, for every dollar of public spending, the cost is $1.00 in tax revenue and 46¢ of loss in efficiency of the private sector. Ballard, Shoven and Whalley (1985) produce a marginal cost of $1.33. These numbers seem to justify a lower level of subsidy to some transportation modes than is implied in the current transportation literature. In essence this means that, with a subsidy from general revenues, second-best pricing may result in a lower efficiency loss than does first-best pricing.

The following subsections describe various types of pricing which may be relevant for transportation infrastructure pricing. These are: average cost pricing, marginal cost pricing, social marginal cost pricing, Ramsey quasi-optimal pricing, peak/off-peak pricing and multi-part pricing. They are followed by a subsection on allocation by non-market mechanisms, such as slot allocations at airports.

2.2.1 Average Cost Pricing

The average cost price is obtained by taking total costs and dividing them by the relevant measure of output. For example, if total annual airfield costs are $1 million, and if 10,000 landings per year are expected, then the landing fee would be set at $100. If price is set equal to average cost, total costs will be recovered. Average cost pricing is sometimes termed "full cost pricing."

If an industry's production technology is not characterized by constant returns to scale over the relevant range of output, average cost pricing leads to economic inefficiency by making the value (to society) of producing another unit deviate from the cost of producing it (that is, marginal cost). It is easy to demonstrate that if average cost is falling (increasing returns to scale), setting price at average cost lowers economic well-being due to the underproduction of services. Producing an additional unit of output at a cost equal to marginal cost and selling it for a price that exceeds marginal cost (but is less than average cost) could provide an increment to profit and potentially make society better off. Average cost pricing prevents such a desirable action from occurring.

Unless constant returns to scale prevail in the relevant range of market demand, average cost pricing leads to an incorrect level of output, both in the sense of social well-being and in the sense of profit maximization. Thus, except in the special case of constant returns to scale, average cost pricing is not a desirable basis for establishing a pricing strategy.

2.2.2 Marginal Cost Pricing

Marginal cost pricing maximizes the economic benefit to society by ensuring that a socially optimal volume is traded in the marketplace, and the optimal quantity is allocated to those who value it the most. Since marginal cost pricing maximizes social benefit without any constraint, it is often referred to as "first-best" pricing. For transportation services, this means that the service output is extended to the point where the marginal cost of serving an additional unit equals the price the user is willing to pay for the service. This marginal cost can be quite different from the average cost of production. Suppose for example, that the total cost of servicing 100 aircraft movements is $1,000 and servicing 101 movements is $1,005. This means the average cost of servicing 100 movements is $10, but the marginal cost of providing an extra movement (when the current service level is 100) is only $5. Average cost pricing would set the price at $10 per movement, while marginal cost pricing would set the price at $5 (for all movements). This pricing principle resolves the problems of inefficient production and consumption associated with the average cost pricing principle. Under conditions of constant returns to scale, unit costs are not rising or falling in the relevant output range. In this case, marginal cost equals average cost. Thus, marginal cost pricing and average cost pricing would be the same.

When unit costs are either rising or falling, average and marginal cost pricing strategies produce different results. Preference is generally given to marginal cost pricing based on the assumption that the efficiency loss resulting from any deviation from marginal cost pricing is greater than the efficiency loss of raising revenue through taxation. While average cost pricing would always lead to a financial break even, marginal cost pricing would require a subsidy for a firm to break even in the case of a declining unit cost industry. This is because price (that is, marginal cost) will be less than average cost. Marginal cost pricing will result in break even in the case of constant cost and a profit in the case of rising unit cost.

Another problem with marginal cost pricing is in the difficulty of measuring appropriate marginal costs. In the case of average cost pricing, one simply sums all relevant costs, operating and capital, variable and fixed, direct and indirect, and divides the sum by the anticipated output level to obtain price. For marginal cost pricing, however, it is not easy to identify the costs which vary with output even for the case of a single product. For instance, capacity costs may be fixed in a very short run but may vary over a longer run. Therefore, the measure and variability of marginal cost will depend on the time frame one chooses. This study argues, and this is the consensus in the literature, that the short run is the appropriate time frame for establishing efficient prices.

2.2.3 Social Marginal Cost

When the production or the use of goods or services results in negative (positive) externalities, such as the imposition on others of noise or air pollution or congestion delays, social marginal cost deviates from private marginal cost (cost incurred by the producer or user). In this case, marginal cost pricing based on the private marginal costs will generate too much (too little) output compared to the socially optimal level of output. When this occurs, social marginal cost pricing maximizes welfare (economic efficiency) as it internalizes the externality costs in a user's decision making. Social marginal cost pricing extends the marginal cost pricing principle to the situation where externality costs exist. It recognizes the costs a user imposes on others (externality costs) in addition to the privately borne costs of the user. For instance, an extra vehicle that uses a congested highway imposes costs on the cars and drivers following it by imposing additional delay on them. It also imposes air and noise pollution. In the case of airports, a user of the airstrip may impose a congestion externality, as well as noise and air pollution costs. Social marginal cost pricing internalizes these externality costs by charging users the full social cost. By internalizing the externality costs, the user is induced to make decisions consistent with social benefit maximization.

The application of this pricing principle has been shown to result in the optimal use of a facility (given fixed available capacity) and in the optimal level of facility investment (see, for example, Morrison, 1983). From as early as 1920, this principle was presented in the context of transportation, specifically road congestion (Pigou, 1912; Knight, 1924). In 1961, Walters formalized the peak-load pricing work of Boiteux (1960) and Steiner (1967) using cost functions. About the same time, Strotz developed the same ideas but with the use of utility functions. The basic model has been elaborated upon and extended by numerous researchers including Mohring and Harwitz (1962), Vickery (1965, 1968), Mohring (1970, 1976), Keeler and Small (1977), DeVany and Saving (1980) and Jordan (1983a, 1983b). Morrison (1986) showed that the conventional homogeneous users model needs only slight modification to handle heterogeneous users.

2.2.4 Ramsey Quasi-Optimal Pricing

As discussed earlier, there are circumstances where marginal cost pricing would not be sustainable in the long run. For example, if the production function is characterized by increasing returns to scale, marginal cost is less than average cost and, therefore, total revenue would be insufficient to cover total cost. In this case, there are generally three options to choose from:

- use marginal cost pricing in conjunction with a government subsidy;
- use some form of second-best pricing, such as Ramsey pricing; or
- use multi-part pricing.

The features of Ramsey pricing are described below. Multi-part pricing is considered later.

Economists have demonstrated that a deviation from (social) marginal cost pricing may reduce social benefits and misallocate resources by over or underproducing the service. The essence of Ramsey pricing is to minimize the loss of economic efficiency caused by the deviation of prices from their respective marginal costs while allowing a financial break-even position to be achieved. Ramsey pricing makes use of an inverse-elasticity rule to mark prices up over marginal cost while ensuring that the quantity of service supplied deviates by the least amount from the optimal quantity under marginal cost pricing. In doing so, Ramsey pricing makes use of the willingness of the segments to pay. In other words, Ramsey pricing maximizes social welfare subject to the constraint that the firm achieve a break-even financial position. Thus, it is usually referred to in the literature as a "second-best" strategy.

The Ramsey pricing principle states that when a revenue constraint exists, the ratio of the markups (the excess of the selling price over marginal cost) must be proportional to the inverse of the price sensitivities for the product in question. In other words, the different groups of users may pay different prices depending on their own price sensitivity, even for an identical product or service. If applied to airfields, this might imply that landing fees would be dependent on the length of the flight, aircraft size or type of use because each of these (demand) characteristics can be expected to lead to different price elasticities of demand. A simple approach to covering the fixed costs is to charge all types of users an equal price which exceeds the price based on marginal cost pricing. Although simple, this approach is inefficient and leads to a larger loss of social benefit than the allocation based on Ramsey pricing. According to the Ramsey pricing principle, these fixed costs are recovered by allocating proportionally more of the fixed costs to those who have a lower price elasticity of demand than those with a higher elasticity.

Ramsey pricing relies on the existence of different market segments with a different willingness to pay for the same good or service.[1] Ramsey prices must cover variable costs. It minimizes the total loss of social welfare by allocating the excess of total costs over variable costs to various market segments. For Ramsey pricing to achieve social optimality, it is essential for the producer to have some market power; otherwise, the set of prices established under a Ramsey rule cannot be sustained.[2]

Ramsey pricing can be generalized as a method of finding the optimum set of prices under any revenue constraint. This constraint can take the form of being confined to a given level of subsidy, a requirement to break even, or even to achieve a surplus of revenues over costs of a given magnitude.[3] Whenever the constraint becomes non-binding, the set of optimal prices derived from Ramsey pricing will become identical to those emerging from marginal cost pricing.

2.2.5 Peak/Off-Peak Pricing

Peak-load pricing is a widely used method in pricing public utility services. Peak-load pricing means that peak-period users are charged a higher price than off-peak users. First, peak-load users impose higher costs on the service provider than do off-peak users because they generate the need for capacity expansion, and thus should be prepared to pay all of the capacity costs. Second, peak-period users may also impose congestion costs and thus should be charged their social marginal costs. Third, since peak-load users generally have less price-elastic demands than off-peak users, charging higher fees to peak-period users (who value the service the most) is consistent with the spirit of Ramsey pricing. In other words, Ramsey pricing serves as a rationale for peak-load pricing.

Whatever the economic justification chosen, the end result is essentially the same: charging higher fees to peak than to off-peak users. Charging higher prices to peak users enhances economic efficiency by inducing them to make rational choices as well as helping solve financing problems for capacity expansion. Finally, peak pricing is not, in principle, unfair or inequitable. It assigns costs to those who are responsible for them. It makes no economic sense to restrict the use of a facility in off-peak hours because all that results is underutilization of an existing facility. With socially efficient pricing, peak users are no worse off in terms of what they pay, provided there is a low off-peak price, and off-peak users pay at least their variable costs.

2.2.6 Two-Part Pricing[4]

Another approach which has been developed to price infrastructure and to recover fixed (including overhead) costs is to have a multi-part price. One part, an entry fee for access to a facility (or infrastructure) is fixed and confers the right to use the facility. A second part, a usage fee is a price per unit of use of the facility. A price per kilometre would be an example.[5] The intuitive appeal of the two-part price is the ability to distinguish between the value of the potential demand for access to use and the demand for the actual use of a product or network. The prices of access and usage must be correlated with the proportions of fixed and variable cost while at the same time satisfying a revenue constraint. A person can be charged an access fee even if there is no use of the system. For example, people may benefit from having a roadway because others can visit them. They may at some future date decide to use it, and it may reduce transactions cost to have continual access to a facility rather than to contract for access each time they choose to use it. The ability to match prices with demand, representing different valuations of access and of usage, results in a higher level of economic welfare than if only a single price is charged. Train (1991) summarizes a broad literature which demonstrates that a tariff structure with N+1 tariffs *will always* Pareto dominate a structure with only N tariffs when price exceeds marginal cost and marginal cost differs from average cost.

With a two-part tariff the level at which each part is set (and the ratio of the revenues from each of the two parts) depend on a number of factors. If the demand for access is completely inelastic, the ideal approach is to set the access fee equal to the fixed cost and the usage fee equal to the marginal social cost of use. In effect, the access fee acts like a lump sum tax. When access demand is price sensitive and therefore not independent of the level of the access fee, the access fee cannot be treated simply as a mechanism to cover fixed costs. Raising the access charge involves a loss in consumer surplus as well as revenue from usage because some consumers will choose to forego access to the service (phone line or vehicle licence, for example) in the face of the higher access price. First, as the access charge rises more users are foreclosed from the market. This reduces consumer surplus and revenue from the access charge, if access demand is at all elastic. Second, with fewer users the demand for usage declines; the demand curve shifts to the left. This means a lower level of consumer surplus and revenue from usage. This fact must be incorporated into the determination of optimal prices.

When the number of users of the network, facility or system is affected by the level of the access fee, the institution or authority which has been charged with the responsibility of setting the prices must balance the relative welfare gains and losses as the relative access and usage fees are raised or lowered. If the access fee is lowered to encourage more users, the usage fee must be increased to compensate for the loss in revenue. The amount by which it will have to be increased with the objective of maximizing economic welfare while covering all costs (breaking even) increases with the absolute access price elasticity of the number of users and decreases with the absolute usage-price elasticity of the amount of usage (Ng and Weisser, 1974).

The absolute and relative values of the two elasticities are important in determining the levels of the access fee and usage fee. The access fee does limit the number of users, and it covers some or all of the fixed costs. The usage charge allocates facilities while covering the variable and part of the fixed costs. If the ratio of fixed to total costs is high, it is desirable to have as many members as possible implying a lower access fee and a usage charge greater than usage marginal costs. However, this cannot continue unabated since as the variable charge is increased above variable costs both revenue and welfare are reduced. The ratio of revenue contributions from the variable charge to the fixed charge depends on both the relative and absolute values of the elasticities of access and usage.

When access demand is price sensitive and a financing constraint is in place, the Ramsey concept can be used to compute the second-best tariffs for access/usage services. The Ramsey rule leads to setting the usage and access fees above their marginal cost. Train (1991) derived the following Ramsey rule for determining the optimal access and usage fees that allow the firm to break even.

$$\frac{P_a - MC_a}{P_a}(\varepsilon_a - \varepsilon_{ua}) = \frac{P_u - MC_u}{P_u}(\varepsilon_u - \varepsilon_{au}) \qquad (2.1)$$

In the equation a refers to access and u refers to usage. ε_a is the elasticity of demand for access with respect to the access fee, ε_{ua} is the cross elasticity of demand for usage with respect to the access fee and ε_{au} is the cross elasticity of demand for access with respect to the usage fee. The rule in this situation states that the percent by which the access fee is raised above the marginal cost of access, multiplied by the "net" elasticity for the access fee, is equal to the percentage markup of the usage fee multiplied by the "net" elasticity of demand for usage. The practice of setting the access fee equal to zero and increasing usage fees above marginal cost, as is done for example at airports and electrical and gas utilities, is optimal only if the demand for usage is fixed and the marginal cost of access is zero. Usage demand is usually more price sensitive than access demand because usage is conditional on access. Some economic efficiency gains can, therefore, be expected if there is a move to some reliance on the access fee (Train, 1991).[6] Such a change has been introduced in New Zealand in the pricing of air traffic control services.[7]

A two-part tariff can be used if users are relatively homogeneous. It is possible, however, that a facility, such as a roadway or an airport, may have a number of groups of users and that preferences may vary significantly across groups of users. It may, therefore, be desirable to have the two-part tariff and access/users fees vary for these different groups. For example, some groups may have low access and high usage fees and others high access and low usage fees. Generally, welfare is improved by offering consumers a menu of choices of two-part tariffs.

In sum, when access demand is price sensitive, the optimal access fee is lower and the optimal usage fee is higher than when access demand is fixed.[8] Unlike the situation with fixed access demand, the first-best outcome is not attained when access demand is price sensitive and the firm is required to break even. The reason for this is clear. When access demand is price sensitive, the access fee cannot serve simply as a subsidy mechanism since it also affects access demand and, indirectly, usage.[9]

2.3 NON-MARKET MECHANISMS

In the absence of pricing mechanisms, there are non-market or administrative instruments which are used to allocate scarce goods and services and to undertake new investment to expand capacity. A variety of these have long been used for the allocation of goods, services and resources. These include administrative rationing, random allocation by lottery and queuing. Despite the variety, all non-market mechanisms share the common characteristics that they do not use prices to allocate resources, goods and services. The outcome of both the resource allocation and the level of investment bears little resemblance to what would occur with efficient pricing. Measured in terms of economic efficiency, such allocative mechanisms are not generally as good as those obtained through the market system.

In general, non-market methods lead to inefficient outcomes due to their inability to adequately distinguish between high- and low-valued uses. The prices individuals are willing to pay directly signal such values. The inability to distinguish between high- and low-valued uses often leads the non-market methods to allocate some resources to those users who do not value them as much as others or who do not value the product or service as much as it costs. This results in a reduced level of welfare for society as a whole. Another problem with non-market mechanisms is that there is no built-in pressure which signals optimal timing and amount of capacity expansion. As a result there may be too little or too much capacity. Finally, the lack of a "market discipline" can, and generally does, lead to higher costs than would otherwise be the case.

Administrative allocation has two additional problems. First, the allocative principle tends to be arbitrary and could change with the decision-making authority and the political climate. Second, the resulting allocation of slots or rights to use a facility at a given time could create some monopoly power in secondary markets (such as airline services) by restricting new entrants especially during peak times. In the case of market allocation, such discriminatory allocation problems do not arise as the prices are set to clear the market and optimal capacity is likely to be in place. The problem of restricting potential entrants becomes particularly serious when the allocation committees are controlled by the existing users as in the case of the airport slot allocation committees at major Canadian airports.

Social welfare can be enhanced by allowing slot sales after the initial allocation. Welfare is improved because those who value particular slots the most end up using them. However, this results in windfall income gains to those who are initially allocated the high-value slots. This creates a severe inequity problem. An alternative is to let the airport auction its slots so that it gets to keep the windfall income gains.

Either type of slot market also has the advantage of signalling to the airport when additional capacity should be added. When the slot price is greater than the cost of incremental capacity, then investment will make society better off.[10] Even competitive slot auctions, however, do not always guarantee that social marginal costs are being charged as bidders ignore externality costs.

2.4 POTENTIAL DIFFICULTIES OF IMPLEMENTATION

Thus far various pricing methods which may be applied to pricing transportation services and/or infrastructure services have been presented. Some potential and practical difficulties of implementing some of the pricing methods are identified and discussed in the following subsections.

2.4.1 Pricing a Single Mode Versus Several Modes

All transportation services are capable of being supplied by more than one transport technology or mode. Each mode has different costs and quality characteristics. One of the aims of transport policy is to implement appropriate charging schemes for different modes in order to maximize the overall economic efficiency of the transportation sector. For example, if certain buses, which compete with railways, pay more than they should for the costs they are responsible for, such as the use of highways, then this may lead to an inefficient allocation of passenger traffic between bus and rail modes. An efficient allocation of traffic among competing modes requires marginal cost pricing by all modes as this induces users to make "socially optimal" choices among competing modes.

Problems arise if marginal cost pricing results in deficits for one or more modes of transportation. In such a case, each mode can be subsidized from general tax revenue or, in principle, the combined deficit of all modes can be allocated using the Ramsey pricing method. The former solution may be impractical as governments try to become fiscally conservative. On the other hand, it is difficult to implement the latter as the application of the Ramsey method intermodally requires the centralized control of all transportation infrastructure by a single agency. Currently different levels of governments have jurisdiction over different modal infrastructure.

Even if various governments were to cooperate for the implementation of Ramsey pricing in a multimodal context, it would require knowledge of each mode's combined cost of modal services (for example, airline services) and infrastructure (for example, airport services) as well as price elasticities of the demands. To date, there is no definitive empirical work which examines the structure of the combined cost of service and infrastructure for any mode. This makes it difficult to actually compute optimal prices which maximize economic efficiency in the multimodal context (see Oum, 1981; Winston, 1985 for imperfect attempts for such computations).

2.4.2 The Indivisibility Problem — Lumpy Capacity Expansion

Typically, economists assume that capacity is divisible when they investigate optimal pricing, the provision of capacity and cost recovery issues. For example, Mohring and Harwitz (1962) and Mohring (1970, 1976) showed that optimal congestion toll revenue exactly equalled the capacity investment cost when highway construction (and expansion) was perfectly divisible *and* was characterized by constant returns to scale. The debate regarding the divisibility of capacity expansion, however, remains unsettled, and various theoretical results obtained under the assumption of perfect divisibility require some refinements for cases where capacity expansion is characterized by lumpy investment, such as airport terminals and runways, and expressways.

There are two opposing views in the road pricing literature concerning the divisibility of capacity expansion. Some economists believe that increases in road capacity are characterized by lumpy investments (Walters, 1968; Kraus, 1981; Starkie, 1982) while other economists argue that the capacity of a road can be expanded relatively smoothly by simply improving some design features or by improving the traffic control system. Kraus (1981) showed that the cost-recovery theorem of Mohring and Harwitz required modification under the condition of indivisibility of capacity construction and expansion. The financial performance of roads under optimal pricing and investment policy depended on the type of road: that is, high-capacity roads or low-capacity roads. He found that, in the case of major highways, the cost-recovery ratio was higher when indivisibility was taken into account than for the case of perfect divisibility.

Oum and Zhang (1990) investigated the long-run relationship between congestion toll revenue and the capital cost of an airport. Their model incorporated the lumpy nature of capacity expansion and the demand fluctuations within a given day and over time. They showed that the cost-recovery ratios realized under social marginal cost pricing and optimal investment depend on both the time pattern of traffic growth and the amount of initial capacity in place. The most significant of all empirical results, developed for Pearson International Airport, Toronto, was that the larger the existing capacity, the higher the ratio of congestion toll revenue to capacity expansion cost.

The assumption of divisibility of capacity construction and expansion affects the setting of cost-recovery targets and user-charges. The relevant empirical question is to what extent the less-than-perfect divisibility of capacity expansion makes the cost-recovery results deviate from the classic result of Mohring and Harwitz. No one has a definitive answer to this question.

These results reflect a general problem of pure marginal social cost pricing with fluctuating or changing costs. Prices will be less useful in acting as a signal to users and to capital markets if there are significant and frequent fluctuations in costs because there will be a higher variance and hence more uncertainty as to their value at any given time. Lumpy investment is not the only culprit in causing costs to fluctuate. Scope economies and capacity utilization economies also result in costs changing with output or investment levels and hence cause fluctuation in prices based on marginal cost. The prospects of cost recovery with socially efficient pricing depend on how costs change with expansion and contraction of output and capacity. Lumpiness in capacity growth results in fluctuating prices and may lead to over or underfinancing. An assessment must be made of the welfare costs deviating from strict social marginal cost pricing relative to the welfare gains arising from more stable prices which create less uncertainty and have lower transactions cost.

2.4.3 Infrastructure Pricing and the Issue of Equity

Fairness of Efficient Pricing

The economics literature is paying increasing attention to the consequences of selective pricing methods on the equity of various groups of users and non-users; that is, the income distributional consequences of a pricing method. Problems arise when a pricing method, which maximizes economic efficiency, does not necessarily yield what is considered a fair or equitable outcome. The most obvious case is the implementation of congestion (or peak-load) pricing for (urban) transportation. While the introduction of congestion tolls on urban roads would improve the efficiency of resource allocation, such tolls make those commuters with less-flexible work schedules (which some identify as low- to middle-income groups) pay higher congestion tolls while those with flexible work schedules (generally higher-income groups) may be able to avoid the tolls by shifting commuting time to off-peak hours. Peak-period prices for urban transit services have similar unfavourable impacts on income distribution.

Small (1983) investigated this issue and estimated the net effects of congestion tolls on income distribution. He found that in the absence of any redistribution of toll revenue, low-income groups lose more than high-income groups because the higher value of time savings by high-income groups more than compensates for the toll payment. However, when the way in which the toll revenues are spent is taken into account the distributional impact of toll charges changes. Small examined the income distribution effects of distributing toll revenues to reduce taxes, to subsidize transit or to be used in some other way. The results varied but the basic conclusion was that it is misleading to characterize a congestion or peak-load pricing policy as "regressive."

Peak-period pricing, deemed necessary to achieve the goal of economic efficiency, has been viewed as having negative income distributional effects among various users. Winston (1991) argued that congestion and peak pricing could benefit all income classes if the toll revenues were used to do any of the following: lower property taxes, invest in public transit or replace registration fees or fuel taxes. Therefore, as pointed out by Foster (1974, 1975), the effects of an infrastructure price, such as for a road, runway, or terminal, on income distribution depend on how the collecting agency uses the toll revenue. Congestion tolls and peak-period pricing schemes can be designed to improve both economic efficiency and equity among various groups.

Income Distribution Consequences of Ramsey Pricing

Ramsey pricing allocates fixed or common cost on the basis of willingness to pay. This willingness is a measure of the value to the user and is measured by the price elasticity of demand. This pricing principle has been recommended in circumstances where full or partial cost recovery is desired for underutilized airports, roads and port facilities with economies of traffic density and/or economies of scale (see Morrison, 1982).

The Ramsey principle is used to change prices from marginal (social) cost in such a way as to minimize the efficiency loss associated with departing from first-best pricing. Ramsey pricing does involve price discrimination since it places a greater fiscal burden on those demanders with less-elastic demands. Some would argue that this is unfair and constitutes an inequity. Economists have no comparative advantage over others in establishing what is or is not fair. One can, however, make the following argument. Having everyone pay exactly the same price is no more fair or equitable than charging different prices. If users are charged the cost for which they are responsible, this seems quite equitable. Different cost responsibility would justify different prices. Furthermore, if those paying the higher prices are as well off or better off, measured in terms of economic welfare, as a result of allowing those paying the lower prices to participate in the market, there is also no inequity.

Fairness in Financing Capacity

The issue of generating funds for capacity expansion raises the question of fairness (equity). At first glance, the generation of funds to cover the financing of capacity through average cost pricing, a common method, appears to be a fair method since each user pays an equal amount. In fact, this is an inequitable (as well as inefficient) method as it entails a subsidy to peak-period users from off-peak users. Short-run, marginal cost pricing in which peak-period prices are set higher than off-peak prices is a more equitable method of raising funds to finance capacity.

Fairness also requires that both fixed and variable costs be allocated to those users responsible for them. Some parts of infrastructure are designed for a sub-group of users, and it is unfair to spread this cost across all users except to the extent that benefits which may accrue to other users, directly on indirectly, may be considered in a Ramsey-type pricing scheme.

The question of financing capacity from users versus the general revenue fund (all taxpayers) is an issue of both efficiency and equity. If marginal cost is less than average cost, efficient pricing will result in a deficit. This deficit must be covered. A partial equilibrium approach to this problem would support a policy of funding the deficit out of general revenue while a general equilibrium approach would argue that the relative welfare losses of funding the deficit users versus the general taxpayer should be considered. As already stated, recent evidence (Jorgenson, 1992; Jorgenson and Yun, 1990; Ballard, Shoven and Whalley, 1985) has shown that the marginal cost of public funds is between $1.33 and $1.45. This means that between 33¢ and 45¢ per dollar are lost through a private-sector efficiency loss. It is, therefore, not obvious that proposing a policy of marginal social cost pricing and ignoring the costs of public funds to finance a deficit will necessarily improve economic welfare. A second-best pricing scheme in which there is a cost-recovery constraint may lead to a higher level of economic efficiency.

2.5 APPLICATIONS OF PRICING PRINCIPLES

Although economists have long argued for the introduction of efficient pricing of transportation infrastructure, little progress has been made by governments. This has been due in part to the view by governments (and of the general public) that infrastructure should be provided at no or a very low charge and that investment in roads and airports promotes economic development and, therefore, excess capacity provides a public good. As such, the allocative and productive efficiency with which infrastructure is provided and operated has not been of concern in most countries. Another important reason for the lack of efficient pricing schemes is the failure by the analysts to provide visible and measurable benefits of efficient pricing.

Until recently, the instances of significant capacity shortfall have occurred only at limited times of the day and at a limited number of locations. However, the recent growth in air travel and the reorganization of the aviation industry together with a failure to invest and price efficiently have led to significant shortages of both runway slots and gates. During the last two decades or so, severe congestion has also occurred on many urban and some intercity roadways. Policy makers could, therefore, rightly ask the questions: What happens if we do implement efficient road or runway pricing? Who gains? Who loses? What is the overall net benefit? These are legitimate questions, and economists have sought to provide answers in a growing applied transportation pricing literature. The outcome of the analysis can be generally divided into two areas: the effect on the use of infrastructure and the impact on investment in infrastructure.

There are few examples of governments introducing prices to achieve the efficient use of roads or runways, even in cases of severe congestion. The introduction of road pricing in Singapore (Holland and Watson, 1978) and the experiment in Hong Kong (Harrison, 1986) are two examples frequently cited. More recently, road pricing has been introduced in Norway (Larsen, 1988). Bergen, Oslo and Trondheim charge fees for vehicles entering a specified area of the city. Oslo and Trondheim use manual and electronic collection systems respectively while Bergen's is a manual system. In

Britain, peak-load pricing has been in effect since the early 1970s to ration airport terminal and runway capacity at London's Heathrow Airport. It was extended to other airports including Gatwick and Stansted in the early 1980s to encourage efficient use of the facilities. The peak and off-peak fees (the combined landing and passenger fees) at the three London area airports are illustrated in Table 2.1. It is important to note from the table that the peak-period fee at Heathrow Airport is 2.5 to 4 times the off-peak fee, depending on the type of aircraft.[11]

Table 2.1
LANDING FEES AT THE BRITISH AIRPORTS AUTHORITY'S (BAA PLC.) LONDON AREA AIRPORTS (ALL FIGURES ARE IN UK £, 1991)

Airport	Peak	Off-peak	Weighted average
Heathrow			
B 757[a]	1680	658	844
Shorts 360	654	153	318
B 747	6259	1747	3336
Gatwick			
B 757	1122	450	634
Shorts 360	444	111	211
B 747	4866	1867	2709
Stansted			
B 757	734	365	507
Shorts 360	123	71	88
B 747	3807	1806	2810

Source: Arthur Reed, "The Unlocking of Heathrow," *Air Transport World* (September 1991), pp. 28–33.

a The Boeing 757 seats 140 passengers and is a Stage III aircraft, and thus it has a lower noise charge. The Shorts 360 seats 22 passengers. The Boeing 747 seats 270 passengers and is a Stage II aircraft which carries a higher noise charge.

There have been numerous simulation studies using artificial data which have attempted to determine the impact of introducing infrastructure pricing. What would happen to the use of the facility by different user-groups? Would the infrastructure be self-financing? What would happen over the longer term if additional capacity were needed? What is gained (or lost) in undertaking such a policy? These studies use information on existing demand and costs in a modelling framework to address the questions just raised.

Levine (1969) was one of the first to recommend efficient pricing of airport runways. He argued that not pricing efficiently had led to shortages, the excessive peaking of traffic and hence waste of resources such as labour, fuel and time. Carlin and Park (1970), Borins (1978, 1982), Likens (1976), Morrison (1982, 1983, 1987), Gillen, Oum and Tretheway (1988), Morrison and Winston (1989), and Oum and Zhang (1990) all examined various aspects of non-optimal pricing and investments at airports. Newbery (1988), Lee (1982) and Small, Winston and Evans (1989) undertook similar studies but for roadways.[12]

The conclusions of these studies, with their different time periods and data bases, form a consensus in several areas.[13] First, the existing fee structure has led to a misallocation of traffic and hence the inefficient use of infrastructure (airport and roadway) over the day. For airports, the fee structure, based on aircraft weight, is reasonably consistent with net benefit maximization at airports with sufficient or excess capacity. It is inefficient, however, at congested facilities (Morrison, 1987). The existing fee structure benefits some groups over others. In particular, general aviation, commuter and regional (local) carriers are beneficiaries while international and trunk carriers are losers (Morrison, 1983, 1987). Similar results hold for roadways. Newbery (1988) noted that vehicles impose four costs on society: road damage costs, congestion costs, accident externalities and environmental pollution costs. The current system of licence fees and fuel taxes has, as with the case of airports, resulted in a misallocation of traffic over the day and a misallocation between types of vehicles. Rural automobiles are charged too much while urban-peak automobiles and large four-axle trucks on a rural road pay too little.

Lee (1982) provides a set of efficient user-charges on the U.S. intercity roads for a sample of vehicles.[14] These are reported in Table 2.2. Gillen, Oum and Tretheway (1988) have undertaken a similar type of analysis but for airport infrastructure (Pearson International Airport, Toronto). Their results are reported in Table 2.3. These two studies illustrate the magnitude of the current underpricing to different user-groups and the magnitudes of the changes anticipated under socially efficient pricing. Which user-classes are gaining under the present system and which are losing is also evident from these studies. Furthermore, this illustration provides an intuitive feel for the notion that the current pricing system creates inefficiencies from the choice of vehicle type and the time of travel.

Table 2.2

EFFICIENT USER-CHARGES ON THE U.S. ROADS FOR A SAMPLE OF VEHICLES (VALUES EXPRESSED IN CENTS/VEHICLE-MILE OF TRAVEL IN 1981 U.S. DOLLARS)

Vehicle type	Location	Key parameter	Pavement repair	User-costs	Admin.	Excess delay	Air pollution	Noise	Total	Existing average user-fee
Car (3000 lb)	rural	v/c=.05			0.3	0.3			0.6	1.3
Car (3000 lb)	urban	v/c=.85			0.7	11.2	1.5	0.1	13.5	1.7
Single unit, 3 axle truck (40,000 lb GVW)	small urban	v/c=.35 PCE=1.2 ESAL=.8	25.6	7.5	0.5	2.2	0.2	0.2	36.2	4.8
Combination truck, 5 axle (72,000 lb GVW)	rural interstate	v/c=.15 PCE=1.2 ESAL=1.6	8.0	5.9	0.3	0.4			14.6	9.0
Combination truck ,5 axle (72,000 lb GVW)	urban interstate	v/c=.35 PCE=1.2 ESAL=1.6	24.0	16.3	0.3	1.4	3.0	4.0	49.0	9.0
Combination truck, 4 axle (100,000 lb GVW)	rural arterial	v/c=.05 PCE=3.0 ESAL=27.2	408.0	95.2	0.3	0.3		0.2	504.0	5.0

Source: Reproduced from Lee (1982).

Notes: v/c — volume/capacity; PCE — Passenger Car Equivalent; ESAL — Equivalent Standard Axle Load

Table 2.3
Comparisons Between Social Marginal Costs and 1985 Landing Fees for Selected Aircraft Pearson International Airport, Toronto (1988 CAN $)

	Aircraft type					
	B747-200	DC10-40	B737-200	Dash 8	Business jet	Piston
1985 fees						
Domestic	$521	$355	$ 80	$ 12	N/A	$ 0
International	$769	$524	$ 85	$ 15	N/A	$ 0
Social marginal cost						
High months						
High-peak	$426	$376	$269	$213	$211	$161
Low-peak	$246	$196	$154	$ 98	$105	$ 55
Low *months*						
High-peak	$271	$221	$170	$114	$120	$ 70
Low-peak	$226	$176	$142	$ 86	$ 93	$ 43
Off-Peak	$206	$156	$129	$ 73	$ 81	$ 31

Source: Gillen, Oum and Tretheway (1988).

Gillen, Oum and Tretheway (1988) investigated efficient pricing at Pearson International Airport. They developed a set of socially optimal peak-period prices based on measures of variable costs with runway use, measures of congestion externalities, noise externalities and fixed costs for airfield operations. A comparison of the calculated social marginal costs and the current landing fee for selected aircraft are described below. They pointed out that the fee structure used by Transport Canada is common across all Group I airports regardless of differences in available capacity and demand variation. Furthermore, the fees are weight-based and discriminate against non-domestic flights. Both generate inefficiencies.[15]

The implementation of this pricing structure was simulated to reduce annual light-plane movements from 63,000 to 28,000 with a proportionately higher reduction in the peak periods. The social marginal cost pricing of runways (noise externalities and congestion were included) would have generated approximately $50 million instead of the actual airfield revenue of $26.6 million in fiscal 1985-86. They also developed a set of terminal user-charges which varied with the peak. These were calculated to be $2.82 and $6.49 per passenger for peak domestic and international flights respectively and $1.54 and $3.54 in the off-peak. (The current charge is $1.00 per seat for domestic flights and $2.30 per seat for international flights.) The gainers and losers are the same as those identified by Morrison.

The second area of consensus in the applied pricing literature is the financing of infrastructure. Under the status quo, there is a significant deficit for roadways and to a lesser extent for airports and airways in a number of jurisdictions. The move to socially efficient or efficient second-best pricing improves the financial condition of the infrastructure agency. If social marginal cost pricing is applied to an uncongested facility, there will be a deficit. Thus, rural roads or uncongested airports would require a second-best pricing scheme to break even or a subsidy if marginal cost pricing is applied. For uncongested airports, for example, Morrison (1982) illustrated the set of Ramsey prices which yield the maximum social benefits subject to revenue covering all costs. On the other hand, a congested facility is likely to be self-financing. Oum and Zhang (1990) provided a useful example for airports. A corresponding example for roads was provided by Small, Winston and Evans (1989).

The third area in which there is some consensus is that the benefits from introducing social marginal cost pricing are significant. Borins (1982) reported that efficiency losses due to non-optimal pricing were significant for Pearson International Airport and also for the access roads to the airport. Losses were measured in terms of economic surplus defined as the difference between what individuals are willing to pay for transportation services and the social costs of providing those services. Small, Winston and Evans (1989) showed that efficient pricing of all U.S. roads with the current investment would yield an annual net welfare gain of $5.4 billion (1988 dollars), and if efficient investment policy was coupled with efficient pricing policy, the gain would increase to $7.75 billion annually.[16] Most of the gain would come from savings in pavement costs (maintenance and capital costs) which were reduced from $20 billion to $13 billion in 1982. The sources and distribution of the efficiency gains are listed in Table 2.4.

Morrison and Winston (1989) provided similar results for all U.S. airports. The welfare gain from efficient pricing would be $3.8 billion annually. Efficient pricing in conjunction with efficient investment would raise this amount to $11 billion annually. This is relative to the status quo. The values varied with assumptions regarding the elasticity of demand and value of time. The values reported here are conservative values for both demand and time. The distribution of the benefits is reported in Table 2.5.

Table 2.4

ECONOMIC EFFECTS OF EFFICIENT INFRASTRUCTURE PRICING FOR THE U.S. HIGHWAYS (IN BILLIONS OF 1982 U.S. DOLLARS)

	Efficient pricing and investment (first-best pricing)	Efficient pricing with current investment	Efficient investment with current pricing
Investment costs			
Maintenance savings	9.426	6.441	8.536
Capital savings	–1.276	0.0	–2.236
Total savings	**8.152**	**6.441**	**6.300**
Trucking firms' and shippers' gains	0.134	–5.586	0.0
Gains from modal shifts	0.040	0.615	0.0
Government revenue	–0.574	3.884	–5.190
Total benefits	**7.752**	**5.354**	**1.110**

Source: Small, Winston and Evans (1989).

Table 2.5

ANNUAL ECONOMIC EFFECTS OF EFFICIENT INFRASTRUCTURE POLICY FOR AIRPORTS (CHANGE RELATIVE TO CURRENT PRACTICE, IN BILLIONS OF 1988 U.S. DOLLARS)

Item	Efficient pricing and runway investment	Efficient pricing with current runway investment
Consumer surplus change from landing and takeoff fees	1.10	-12.53
Reduced delay to travellers	7.91	3.62
Carriers' operating cost savings	2.77	1.23
Airport revenues less costs	–0.77	11.50
Total welfare change	**11.01**	**3.82**

Source: Winston (1991).

The airport pricing literature makes it quite clear that the failure to introduce landing fees, which more closely reflect the real costs of usage and externalities of the facilities, has generated significant economic waste. This waste is measured in real resource costs to carriers who use crew time, fuel and capital in congested facilities.[17] It is also measured in wasted time to travellers. These delay costs can mount into the billions of dollars.

2.6 SUMMARY

There is a broad literature which has examined almost every aspect of infrastructure pricing from simple congestion models with homogeneous traffic to infrastructure pricing with several user-groups, uncertainty, varying demands and intermodal congestion. Despite the varied approaches and level of complexity in the models, all studies find that optimal prices for peak-period use are well above those currently charged. Calculated net benefits from measures which introduced some "demand management" were significant. This applies to all modes of transportation in the intercity passenger system despite differences between modes.

The motivation to develop socially efficient prices arises from the need to limit the use of existing facilities and to maximize economic efficiency. The latter objective is not some idol but rather is based on the argument that such pricing would make society better off by having both the efficient use of a facility and an optimal amount invested in infrastructure. Gillen, Oum and Tretheway (1988), for example, have shown that the current landing fee structure in Canada (and elsewhere), which is based on the weight of the aircraft, is inefficient in a situation where congestion exists. Indeed, they demonstrate that the largest proportion of marginal social costs is made up of movement delays to other aircraft and noise externalities. Thus, current landing fees, which ignore the congestion externality, undercharge aircraft in inverse proportion to their weight.

The air transport sector is subjected to problems not unlike those found in urban areas — peak-period demands with multiple user-groups. The road and airport runway pricing literature demonstrate that an effective solution to the problems of airport congestion is efficient pricing *and* investment policies. This is an important point. In the past, supply management was the single strategy. Some economists, as well as other transportation experts, have focussed almost entirely on demand management through pricing. This study argues that it takes a combination of the two, pricing and investment, to yield an efficient long-run outcome.

The basic message is quite clear. Continuing the supply-side approach to solving the economy's transportation problems is not only difficult in a situation of fiscal restraint, it foregoes economic gains which are available independently of whether fiscal restraint applies. Demand management is part of an efficient approach to infrastructure management and involves setting user-charges for roadway, airport or waterway infrastructure to reflect their social costs.

Such a pricing scheme has three significant impacts. First, economic welfare increases as the demand for infrastructure is rationed more efficiently in a way which reflects the costs of using the infrastructure. Second, intermodal distortion is reduced, even with a second-best Ramsey pricing. Under the current regime of infrastructure fees the distribution of traffic among modes can be expected to result in an efficiency loss. Finally, the current pricing policy results in some infrastructure being financed out of general tax revenues. This is not necessarily inefficient if the community receives benefits as a result of the infrastructure (if there are positive externalities sufficient to justify such a subsidy). The current system also ensures that the recurring infrastructure problems of the past will continue into the future.[18] For congestion-prone facilities, basing infrastructure prices on social costs would generate sufficient funds to provide an optimal facility and have the infrastructure more likely to be self-financed.

Winston (1991) summed it up best when he said:

> Surprisingly, the belief of most economists that public infrastructure spending should be substantially increased is not based on efficient pricing and investment principles. Instead, it appears to be based on either personal observations or on a suspicion that because uncongested infrastructure is a public good, society has tended to invest too little in it. Both perspectives have diverted many economists and policymakers from realizing there are surprisingly large but plausible benefits from *efficient* infrastructure pricing and investment. (p. 114)

Marginal social cost pricing is also not incompatible with fairness or equity. Congestion or peak-load pricing in conjunction with a strategy to use the revenues can generate net positive benefits for society. Marginal social cost pricing corrects distortions rather than introduces them. Some groups are made better off and others worse off but this should not justify rejection of this pricing policy. As Hau (1991) noted, it is perhaps asking too much of a pricing mechanism to solve the pricing, investment *and* income distribution problems. One way of viewing this is to use road and airport funds from efficient pricing to both invest efficiently in capacity (new roads or runways) and to use any residual to satisfy the principles of equity taxation. Thus, a transportation fund could potentially invest in public transportation, rural roads or small airports. The magnitude of the fund would depend on scale economies and indivisibilities.

Social marginal cost pricing does not necessarily yield a surplus or break even. When the total revenue from social marginal cost pricing is not enough to cover total cost, there are three options open to the infrastructure authority:

- to continue to use social marginal cost pricing with subsidization from the general tax revenue;
- to use Ramsey pricing which provides a second-best solution with a break-even financial performance; or
- to adopt two-part pricing (access/usage tariffs).

The Ramsey pricing minimizes the loss of economic efficiency caused by deviating prices from the respective marginal costs in order to achieve financial break even. In effect, it charges a higher markup to the less price-elastic product or market segment by setting the markup in inverse proportion to the price elasticity demand.

Two-part pricing consists of a flat fee for the right to access a facility (for example, vehicle licence fee to access the road system), and a usage fee (for example, charge per kilometre of road usage). Two-part pricing leads to first-best optimality if the demand for access is not price sensitive, and if usage charges are set at the marginal costs of usage and the access fee is set at a sufficiently high level to allow the firm to break even. Furthermore, under these conditions, the regulation of the access fee alone can induce the monopoly firm to charge marginal cost as the usage price. When demand is sensitive to price, the two-part tariffs are computed by applying the Ramsey rule to the access and usage demands as if they were two separate products with interrelated demands. This then becomes a form of second-best pricing.

The use of first-best pricing with subsidies from general revenues has been long advocated in the transportation literature. There is increasing evidence, however, that the efficiency costs of raising revenue through taxes may more than offset the relative efficiency costs of moving from first- to second-best pricing schemes.

3. Carrier and Infrastructure Cost Structures: A Literature Summary

3.0 INTRODUCTION

This section summarizes the theoretical and empirical literature on the cost structure of modal services (carriers) and of the provision of infrastructure. Since the infrastructure planner must establish user-charges and make capacity investment decisions to maximize the economic welfare of society, understanding the behaviour of the combined cost of a carrier's service provision and infrastructure provision is essential for the development of a set of socially optimal prices for infrastructure.[19] If short-run costs fall because of increased capacity utilization but long-run costs exhibit constant returns to scale, it is still possible to have marginal social cost pricing and fully cover costs. If, however, long-run costs are characterized by some economies, a second-best pricing approach will be required to have total revenues cover costs and minimize the efficiency loss of deviating from first-best pricing. A prerequisite to understanding the structure of the combined cost is to understand each component, that is, carrier cost structure and infrastructure cost structure.

Social benefit maximization is the objective of infrastructure pricing and investment. The relevant cost is, therefore, the full social cost a user of transportation services imposes on society as a whole. The full social cost includes not only all private resource costs such as payments to labour, capital and energy but also external costs the user imposes on others, such as congestion costs and noise and air pollution costs. The remainder of subsection 3.0 describes several cost concepts which will be used in the sections following. Subsection 3.1 summarizes the literature on several key aspects (economies of scale, economies of traffic density and economies of scope) of the carriers' cost structure for each mode. The empirical and theoretical literature on the cost structure in the provision of infrastructure services is presented in subsection 3.2. Finally, subsection 3.3 discusses the implications of the empirical results for infrastructure pricing.

External Costs and Internal Costs

Internal costs, sometimes referred to as private costs, are the costs borne by the supplying agency. These are normally, but not always, financial costs incurred as a result of purchasing factor services in the market and would include labour wages, interest on capital and the price of fuel. Although external costs represent genuine resource costs, they do not directly influence decisions of transport suppliers in their provision of transport services. The external costs include general social and environmental impacts and the cost of congestion delay imposed on third parties. It also includes the effects on non-users. There is a clear distinction between internal costs that influence the optimal choice of the transportation supplier and the external costs which affect others, but do not directly influence the supplier's decisions. However, both must be considered in the socially optimal design and pricing of infrastructure. For example, in selecting the frequency and timing of flights and type of equipment use at an airport, an airline normally does not consider the noise and air pollution resulting from such decisions.

Short-Run Versus Long-Run Costs

Long-run costs, using the standard economic definition, are all variable; there are no fixed costs. In the short run, the ability to vary costs in response to changing output levels and mixes differs among the various modes of transportation. Since some inputs are fixed, short-run average cost is likely to continue to fall as more output is produced until full capacity utilization is reached. Economies of traffic density are another potential source of cost economies in transportation. Unit cost per passenger-kilometre decreases as traffic volume increases over a fixed network. The fixed network means this is a short-run rather than long-run cost. Density economies are a result of using a network more efficiently. The potential for density economies depends on the configuration of the network. Carriers in some modes, such as air, have reorganized their network, in part, to realize these economies.

In the long run, additional investment is needed to increase capacity and/or other fixed inputs. The long-run average cost curve, however, is formed by the envelope of the short-run average cost curves. For some industries, the long-run average cost often decreases over a broad range of output as firm size (both output and capacity) expands. This is called "increasing returns to scale" or "economies of scale." The presence of economies at the relevant range of firm size means that the larger the size of the transportation firm, the lower the per-unit cost of output. These economies of scale may potentially take a variety of forms in transportation services and may vary significantly according to the mode of transportation involved.

Common and Joint Costs

The production of transport services in most modes involves joint and common costs. A joint cost occurs when the production of one good inevitably results in the production of another good in some fixed proportion. For example, consider a rail line running only from point A to point B. The movement of a train from A to B will result in a return movement from B to A. Since the trip from A to B inevitably results in return trip costs, joint costs occur. Some of the costs are not traceable to the production of a specific trip so it is not possible to fully allocate all costs nor to identify separate marginal costs for each of the joint products. For example, it is not possible to identify a marginal cost for an i to j trip and a separate marginal cost for a j to i trip. Only the marginal cost of the round trip is identifiable.

Common costs occur when the facilities used to produce one transport service are also used to produce other transport services. For example, track or terminals used to produce freight services are also used for passenger services. The production of a unit of freight transportation does not, however, automatically lead to the production of passenger services. Thus, unlike joint costs, the use of transport facilities to produce one good does not inevitably lead to the production of some other transport service since output proportions can be varied. The question arises whether or not the presence of joint and common costs could prevent the market mechanism from generating efficient prices. A substantial literature in transport economics (Mohring, 1976; Button, 1982; Kahn, 1970) has clearly shown that conditions of joint, common or non-allocatable costs do not preclude economically efficient pricing.

3.1 CARRIER COSTS

There are three key indicators of the cost characteristics of a firm. They include economies of scale and economies of scope which are long-run concepts. The other is economies of density. How do they influence costs? Why are they important to this discussion of transport infrastructure pricing? These questions will be addressed in the following subsections.

Economies of Scale

There has been some confusion in the literature between economies of scale and economies of density. The distinction is important since scale is a long-run concept and thus affects industry structure whereas density is a short-run concept and is more meaningful for industry behaviour such as pricing. Density economies are said to exist when a 1 percent increase in all outputs (holding network size, production technology and input prices constant) increases the firm's cost by less than 1 percent. In contrast, scale economies exist when a 1 percent increase in output[20] and size of network increases the cost by less than 1 percent, with production technology and input prices held constant.

Economies of scale refer to a long-run average cost curve which slopes down as the scale of the transport firm increases. The presence of economies of scale means that as the size of the transport firm gets larger, the average or unit cost gets smaller. Since most industries have variable returns to scale cost characteristics, whether or not a particular firm enjoys increasing, constant or decreasing returns to scale depends on the overall market size and the organization of the industry.

The presence or absence of scale economies is important for the industrial structure of the mode. The presence of significant scale economies implies that fewer larger carriers would be more efficient and this, under competitive market circumstances, would naturally evolve over time. Scale economies are important for pricing purposes since the greater the scale economies, the more average and marginal costs deviate. It would, therefore, be impossible to avoid a deficit from long-run marginal (social) cost pricing.[21]

Economies of Traffic Density

Although they have a different basis than scale economies, economies of density can also contribute to the shape of the modal industry structure and affect the way a carrier organizes the delivery of its service, spatially. The magnitude of density economies also depends on the size of the market. In the air market, for example, deregulation has allowed carriers to respond to market forces and obtain the available density economies to varying degrees. Canadian carriers have been less successful than their counterparts in the United States in realizing the maximum density economies because the market is smaller in Canada.

Keeler (1974), Harris (1977), Friedlaender and Spady (1981) and Levin (1981) have all shown that there are large increasing returns to traffic density in the U.S. railroad industry. They show that allowing all factors of production to vary, including information but not route mileage, a railway producing 10 million revenue tonne-miles per mile of road, for example, will have substantially lower average costs than a railway producing only 5 million revenue tonne-miles per mile of road. Harris (1977) estimated that approximately one third of density economies were due to declining average capital costs, and two thirds due to declining fixed operating costs, such as maintenance and administration.

Gillen, Oum, and Tretheway (1985, 1990), using data from the airline industry in Canada, illustrated that unit costs would decrease for all carriers, except Air Canada, if they carried more traffic within their given network. The authors also noted that a large portion of the density economies resulted from fixed costs associated with a network — costs which are independent of level of output.

Economies of Scope

Typically, the transport firm produces a large number of conceptually distinct products from a common production facility. In addition, the products of most transportation carriers are differentiated by time, space and quality. Because a number of distinct non-homogeneous outputs are being produced from a common production facility, joint and common costs occur. The presence of joint and common costs gives rise to economies of scope. There has been some confusion in the multi-product literature among the concepts of sub-additivity of the cost function, trans-ray convexity, inter-product complementarity and economies of scope. Sub-additivity is the most general concept and refers to a cost function which exhibits the characteristic that it is less costly to produce any amounts of any number of goods in one plant or firm than to subdivide the products or service in any proportion among two or more plants. Trans-ray convexity is a somewhat narrower concept. It refers to a cost function which exhibits the characteristic that for any *given set* of output vectors, the costs of producing a weighted average of the given output vectors is no greater than the weighted average of producing them on a stand-alone basis. Economies of scope refers to the cost characteristic that a single-firm, multi-product technology is less costly than a single-product, multi-firm technology. It, therefore, is addressing the issue of the cost of adding another *product* to the product line. Interproduct complementarity is a weak test of scope economies. It refers to the effect on the marginal cost of one product when the output of some other product changes. It, therefore, is changing the *amount of output* of two or more products and not the *number of products*. Whether scope economies exist, and the extent to which they exist, depends on both the number of products and the level of each output. No reliable empirical estimates of economies of scope for transportation modes exist which are both based on reliable data and undertaken in a theoretically consistent fashion.

3.1.1 Air Carriers

A considerable number of studies, Douglas and Miller (1974), Keeler (1974), Caves, Christensen and Tretheway (1984), Caves, Christensen, Tretheway and Windle (1985), McShan and Windle (1989) and Gillen, Oum and Tretheway (1985, 1990), have been directed at determining the functional relationship between total per-unit operating costs and firm size in airlines. All studies have shown that returns to scale are roughly constant; thus, size does not generate lower per-unit costs. Generally, however, the measures of economies of density illustrate that unit cost would decrease for most carriers, in both Canada and the U.S., if they carried more traffic within their given network.[22] In other words, the industry experienced increasing returns to density. The results also indicated that the unexploited economies of density are larger for low-density carriers.

Caves, Christensen, and Tretheway (1984) have shown that it is important when measuring costs to include a network size variable in the cost function, along with output, which would allow for the distinction between returns to scale (RTS) and returns to density (RTD). McShan and Windle (1989) used the same data set as that used by Caves et al., and explicitly account for the hub-and-spoke configuration that has evolved in the United States since deregulation in 1978. They estimated a long-run cost function which employs all the variables included in Caves et al., and found the returns to density of about 1.35. The hubbing variable indicates that, *ceteris paribus*, a carrier with even 1 percent more of its traffic handled at hub airports expects to enjoy 0.11 percent lower cost than other similar carriers.

3.1.2 Intercity Buses

According to Gillen and Oum (1984), the hypothesis of constant returns to scale could be rejected for the intercity bus industry in Canada where they found decreasing returns to scale at the mean of the sample (0.91). Large firms exhibited strong decreasing returns to scale, and small- and medium-sized firms exhibited slight departures from constant returns. These empirical measures may be biased, however, since no measure of network size and market density was included in the estimation. This exclusion would have a differential impact on the measure of scale economies depending on the route mix of each firm. No cost complementarities were found to exist between the three outputs, namely, number of scheduled passengers, revenue vehicle miles of charter, tour and contract services and real revenue from freight. These results, however, are biased since no network measure was included in the estimating equations. The scale economy measure, therefore, contains some of the influence of available density economies.

Since deregulation of the intercity bus industries in the United States and the United Kingdom, the number of firms has been reduced significantly. In the absence of scale economies, the forces leading to this industry structure would include density economies. Route reorganization, for example, has been observed to approximate hub-and-spoke systems and the use of smaller feeder buses on some rural routes.

The industry reorganization is similar to what occurred in the airline industry. The consolidation of firms was driven by density and not scale economies. One significant difference between these two industries, however, is that airline demand has been growing while intercity bus demand is declining.

3.1.3 Trucking

Several empirical analyses of the trucking production function have shown that the long-run marginal cost and average cost curves are relatively flat with respect to changes in the level of output. In other words, economies of scale are either absent or very small. Koenker (1977) showed that there are very small economies of scale present up to relatively low levels of output and, thereafter, unit costs tend to rise gradually. Friedlaender and Spady (1981) also found mild diseconomies of scale in the production of trucking services as well as cost complementarities in the production of the multiple outputs trucking firms produce. Friedlaender and Bruce (1985) found that, for the period from 1974 to 1979, larger firms experienced greater diseconomies of scale than smaller firms. In 1979, however, the situation was reversed and, in their study, the authors suggested that the larger carriers were reaping the benefits from longer hauls.

Friedlaender and Chiang (1984) examined the effect of various network variables on trucking costs. Their study found that at the mean of the data there were constant returns to scale, while, as with other modes, there were significant gains from better network utilization, hence economies of density. Ying (1990) examined the impact of deregulation on the productivity in the United States Class I and II trucking industry. He found that at the sample mean, a 1 percent increase in output caused total cost to increase by about 1.073 percent, suggesting very slight decreasing returns to scale. It is unlikely, however, that given the size of the standard error this result is statistically significantly different from 1.00 or constant returns to scale.

3.1.4 Railways

The structure of railway costs is generally characterized by high fixed costs and low variable costs per unit of output. The essential production facilities in the railway industry exhibit a significant degree of indivisibility. As with other modes, the production of railway services gives rise to economies of scope over some output ranges. For example, track and terminals used to produce freight services are also used to produce passenger services.

Caves, Christensen and Tretheway (1980) found that the United States railway industry was characterized by constant returns to scale over the relevant range of outputs. However, their sample did not include relatively small railroads, firms with less than 500 miles of track. Griliches (1972) and Charney, Sidhu and Due (1977) found increasing returns to scale for such small U.S. railroads. Friedlaender and Spady (1981) suggested that there may be very small economies of scale with respect to firm size. Keeler (1974), Harris (1977), Friedlaender and Spady (1981) and Levin (1981) all found large and significant economies of traffic density in railway services. Friedlaender and Spady (1981) estimated a short-run cost function with five variable inputs, one quasi-fixed factor (structures) and two outputs which took the form of hedonic functions, accounting for factors such as low-density route miles and traffic mixes. The study found constant returns to scale but increasing returns to traffic density. Caves, Christensen, Tretheway and Windle (1985) examined economies of scale and density in the United States railroads. Their basic result demonstrates that there are substantial returns to density in the U.S. railway operations.

The economies of traffic density and economies of scale estimated by various studies are compared in Table 3.1.

Table 3.1
Returns to Density and Scale in U.S. Railways

Study	Density	Scale
Friedlaender and Spady (1981)	1.16	.88–1.08
Caves, Christensen and Swanson (1981)	—	1.01
Harmatuck (1979)	1.92	0.93
Harris (1977)	1.72	1.03
Keeler (1974)	1.79	1.01
Caves et al. (1985)	1.76	0.98

Source: Table 4.3 in Caves, Christensen, Tretheway and Windle (1985).

Note: Estimates of returns to scale are for fixed length of haul and trip length.

3.2 INFRASTRUCTURE COSTS

As early as 1962, Mohring and Harwitz demonstrated that the financial viability of an infrastructure facility, under optimal pricing and investment, depended largely on the characteristics of its cost function. To quote Winston (1991):

> If capacity and durability costs are jointly characterized by constant returns to scale, then the facility's revenue from marginal cost pricing will fully cover its capital and operating costs. If costs are characterized by increasing returns to scale, then marginal cost pricing will not cover costs; conversely, if costs are characterized by decreasing returns to scale, marginal cost pricing will provide excess revenue. (p. 115)

The objective of this subsection is to summarize the theoretical and empirical literature on the cost characteristics of modal infrastructure. The discussion deals with the following types of infrastructure: airports, highways and railways.

In developing a set of socially efficient prices for modes of intercity transport, it is not just the carrier's cost structure which is important. Airports, roadways and harbours all represent public capital which is used by the carriers in the different modes to produce and deliver their modal services. This capital must also be priced in an efficient way to achieve the economic welfare gains available from economically efficient pricing. As with the carriers, the ability to apply first-best pricing principles to infrastructure and still satisfy cost-recovery constraints depends on the cost characteristics of building and maintaining the infrastructure.

Cost characteristics include scale economies, scope economies, density economies and utilization economies. Scale economies refer to the size of a facility. For example, is it cheaper per unit to build three runways than it is to provide two runways? If so, there are economies of scale in the provision of runways. Scope economies encompass similar concepts as with carriers. Small, Winston and Evans (1989) referred to scope economies in highways when both volume and durability are supplied. Volume refers to the number of lanes while durability refers to the ability to carry heavier vehicles. A similar concept would apply to airports (small and large aircraft, VFR and IFR traffic) and harbours (large ships and small ships). Although rail infrastructure is currently supplied by the railways, there have been moves to separate infrastructure and carrier services. This separation would mean pricing the track and terminals separately from carrier services.

Density economies should also, in principle, be evident in the provision of infrastructure. It is, for example, possible to expand outputs and all inputs for highways while holding the size of the network fixed. There are no empirical estimates of these types of economies for any mode.

Utilization economies refer to the short-run cost function. They describe how quickly average and marginal costs fall as capacity utilization approaches capacity. Although not of direct interest, they are important to consider in any cost estimation since failure to consider capacity utilization can bias upward the measures of both long-run average and marginal costs.

3.2.1 Airports

Economists have typically assumed that capacity expansion is divisible. In his analysis of the optimal pricing and investment in airport runways, Morrison (1983) showed that airport capacity construction is characterized by constant returns to scale. Therefore, under perfect divisibility of capacity expansion, the revenue from tolls would be exactly equal to the capital cost of capacity investment (Mohring and Harwitz, 1962). Morrison's results, however, were based on a sample of 22 of the busiest airports in the United States and did not include any small airports. In the literature, there is no empirical evidence on the cost characteristics of capacity construction of new small airports or capacity expansion of existing small airports (for example, one runway).

3.2.2 Highways

In general, highways produce two outputs, traffic volume which requires capacity in terms of the number of lanes and standard axle loadings which require durability in terms of the thickness of the pavement. Before determining economies of scale in this multi-product case, the measure of economies of scale for each output, or the product-specific returns to scale, must be examined. Small, Winston, and Evans (1989) reported the existence of significant returns to scale associated with the durability output of roads and the ability to handle axle loads. This is because the pavement's ability to sustain traffic increases proportionally with its thickness. They also found evidence that there are slight increasing returns to scale in the provision of road capacity; that is, the capacity to handle traffic volume. However, they reported diseconomies of scope from the joint production of durability and capacity because, as the road is made wider to accommodate more traffic, the cost of any additional thickness rises since all the lanes must be built to the same standard of thickness. They concluded that these three factors together result in highway production having approximately constant returns to scale. In other words, the output-specific scale economies are offset by the diseconomies of scope in producing them jointly.

3.2.3 Railways

An important difference between rail and other modes of transportation is that most railroads provide the infrastructure themselves, and the pricing is undertaken jointly for carrier services and infrastructure. In a few cases, however, ownership and/or management of the trackage has been separated from carriers. Sweden is a good example but even in the United States there have been joint running rights on tracks. This creates a situation where one firm may be responsible for the provision of trackage and another for carrier services. It is, therefore, legitimate to ask if there are any scale economies in the provision of railway infrastructure. There are no empirical estimates but it may be possible to use some of the Small, Winston and Evans (1989) work for roads to shed some light on the issue.

Small et al. argued that road infrastructure produces two outputs, durability and capacity. The former refers to the thickness of roads and the latter to their width. They found increasing returns with respect to durability. This is less likely to occur with a rail line since there would be a relatively broad range of rail car axle loadings for a given level of durability of rail, ballast and ties. Thus, there may only be minor economies. The authors found some minor increasing returns to scale in the provision of capacity. One would expect these same economies would exist for rail since doubling track more than doubles capacity (Gillen and Oum, 1984). Small et al. found diseconomies of scope from the joint production of durability and capacity for highways. These diseconomies are less likely to be evident in rail due to the broad range of durability noted above and the ability to allocate traffic to specific tracks. On balance, there may be generally constant or minor increasing returns in the provision of rail line infrastructure. This conclusion, however, is reached from an intuitive discussion of capacity expansion costs and not from empirical estimates.

The output-specific scale economies seem to be minor as do the diseconomies of producing them jointly. In Canada since VIA Rail leases track from CNR and CPR, the relevant question is, what is the optimal price for VIA's use of tracks?

3.3 SUMMARY OF THE COST STRUCTURE FOR CARRIERS AND INFRASTRUCTURE

The full costs of a mode are the sum of infrastructure costs and modal services costs. Since the choice of a particular basis for infrastructure pricing will influence the modal choices of the end users, optimal pricing strategies and cost recovery should consider the combined cost of infrastructure provision and carrier (or user) costs in order to maximize social welfare. If markets for carrier services are competitive and there are constant returns in the provision of infrastructure for the mode, marginal (social) cost pricing will yield a socially efficient outcome and full cost recovery. If there are economies, from whatever source in the provision of infrastructure, first-best pricing may result in a deficit requiring a subsidy from general revenues which has consequences for economic welfare. Second-best pricing to recover costs may also lower social welfare. The issue is which minimizes the loss.

3.3.1 Air

A number of studies have been directed at determining the behaviour of an airline's cost function with respect to changes in the level and composition of output. The studies have shown that the long-run average cost curve is relatively constant over a wide range of output associated with different scales of plant; that is, there are no economies of scale in the airline industry. This means that the size of a carrier does not generate lower per-unit costs. Gillen, Oum and Tretheway (1985, 1990), however, found that the airline industry in Canada experienced increasing returns to traffic density; that is, the unit cost would decrease for all carriers (except Air Canada) if they carried more traffic within their given network.

Studies concluded that airport capacity construction is also characterized by constant returns to scale. This implies that the combined cost of carriers and infrastructure is also characterized by constant returns to scale.[23]

3.3.2 Road

There are somewhat different results for intercity bus and truck. Several empirical studies of the trucking industry found constant returns to scale in the industry while studies on the intercity bus industry found the hypothesis of constant returns to scale rejected in favour of decreasing returns to scale (0.91). The research also found no economies of scope between the three outputs, namely, scheduled passenger, charter and contract services. There is no empirical evidence on density economies. Observing the parallel mergers which have occurred in the United States and the United Kingdom bus industries after deregulation, however, one might hypothesize that there are density economies.

Road infrastructure yields two outputs, namely, traffic volume which requires capacity (measured in number of lanes) and standard axle loadings which require durability (measured in thickness of pavement). Small, Winston, and

Evans (1989) reported the existence of significant economies of scale with respect to the durability of road and mild returns to scale with respect to traffic volume. However, they reported diseconomies of scope from the production of both durability and traffic volume because, as the road is made wider to accommodate more traffic, the cost of any additional thickness rises since all the lanes must be built to the same standard of thickness. The final outcome of these three factors at work is that highway capacity construction is characterized by approximately constant returns to scale. In other words, the output-specific economies of scale are offset by the diseconomies of scope for having to produce them jointly. Since they included both infrastructure costs and the costs incurred by road users (individual drivers and transportation carriers) in the total cost of highway modes, their result on the overall constant returns to scale is for the combined cost of highways and users.

3.3.3 Rail

Currently, an important difference between the railway mode and other modes is that rail infrastructure is provided by carriers and thus the infrastructure cost is reflected in the freight rates and passenger fares. Since railway companies provide their own infrastructure (VIA Rail in Canada is an exception and Amtrak in the United States is a partial exception) the carrier's cost structure represents those of the combined carrier and infrastructure costs.

Several studies in the United States have shown that the railway industry is characterized by constant returns to scale over the relevant range of output. Studies have indicated, however, that economies of scale are present for small-sized firms. On the other hand, all studies have shown that there are large and significant economies of traffic density in railway services.

4. A Framework for the Analysis of Optimal User-Charges and Cost Recovery

4.1 GENERAL FRAMEWORK FOR USER-CHARGES AND COST RECOVERY

In this chapter, the general framework for integrating optimal user-charges and investment is presented and applied to airport and road infrastructure. A similar approach to that proposed by Small, Winston and Evans (1989) is taken to model the integrated pricing and investment decision. This approach has an advantage of treating the choice of price, capacity and a durability simultaneously. This model can be applied to all types of infrastructure but this study presents the analysis using roads as the example.

The social benefit from infrastructure, defined as the difference between what ultimate consumers are willing to pay and the combined cost of providing modal services and infrastructure, can be maximized by choosing price, capacity and a durability standard which maximizes the following expression:

$$\underset{Q_t, W_t, D}{\text{MAX NB}} = \sum_{t=1}^{T} \left(\int_0^{Q_t} P_t(Q)dQ - Q_t \bullet AC_t(Q_t, W,D) - M(Q_t, W; D) - rK(W,D) \right) \quad (4.1)$$

where $P_t\,(Q)$ represents the inverse travel demand function expressed in real (price) present value terms; AC_t is the average cost function in year t expressed in present value terms and includes all of the expenses of users including users' value of travel time; the road authority expenses are contained in M, the present value of the average total maintenance cost per year including the cost of resurfacing and K the capital cost of road construction; Q_t the traffic volume in year t; W the width of the road; D the durability standard (thickness of the pavement) for roads; and r is the real interest rate. The following optimality rules for pricing, investment in capacity and durability standard can be obtained from the following first-order conditions:

$$P_t = \left(AC_t + Q_t \ \frac{\partial AC_t}{\partial Q_t} \right) + \frac{\partial M}{\partial Q_t} \quad (4.2a)$$

$$T \left(\frac{\partial M}{\partial W} + r \frac{\partial K}{\partial W} \right) = \sum_{t=1}^{T} Q_t \ \frac{\partial AC_t}{\partial W} \quad (4.2b)$$

$$T \left(r \frac{\partial K}{\partial D} \right) = \sum_{t=1}^{T} \left(Q_t \frac{\partial AC_t}{\partial D} + \frac{\partial M}{\partial D} \right) \quad (4.2c)$$

The optimal pricing rule in (4.2a) indicates that the total road charge paid by the user should be equal to the sum of private user costs plus the costs of congestion and road damage repair. This is essentially the social marginal cost of road use which varies over time. The cost paid for infrastructure would be equal to the sum of the last two terms in 4.2a, $\left(Q_t\frac{\partial AC_t}{\partial Q_t}\right)+\frac{\partial M}{\partial Q_t}$. The optimal road capacity rule in (4.2b) indicates capacity should be constructed to the level where the marginal benefits from reduced congestion delays (the right side of the equation) become equal to the increased capital cost of construction and maintenance (the left side of the equation). Finally, the optimal durability condition in (4.2c) indicates that the durability of infrastructure must be set at the level where the marginal benefit from increasing investment in durability (terms to the right of the equal sign) equals the additional capital cost.

These rules for optimal pricing, capacity and durability can be applied to any modal infrastructure including roads and airports. The optimal pricing and capacity conditions determine prices and investment levels for airports since the cost of increasing the durability of an airport runway is negligible relative to the cost of congestion delays. On the other hand, since road maintenance costs are significantly influenced by the thickness of the road, all three conditions (pricing, investment and durability) need to be determined simultaneously in the case of roads.

4.2 PRICING AND INVESTMENT IN A NETWORK

This subsection examines the issue of transferring funds between parts of a network (aviation, road or rail) in response to the potential sources of so-called "network externalities." It also discusses arguments that suggest the (allocative efficiency) welfare loss from failing to cross subsidize is minor compared to the (productive efficiency) welfare loss from cross subsidizing.

The essence of the question of pricing and investment rules for multiple roads or airports is whether, for pricing and investment, they should be viewed as independent or as a linked network. If the set of airports, for example, is considered as a network, there may be a price for runway and terminal use which is averaged over the network, similarly to roads. On the other hand, if each airport or roadway is considered independently the price set would consider the cost and demand conditions on that link or at that airport. Many argue that complementarities exist between links (roads) or nodes (airports) and socially efficient pricing must reflect these positive externalities. Averaging prices over the network is one practical way of reflecting such externalities in what is charged. Another method would be to transfer revenue from one facility to another after establishing socially efficient prices.

Consider a system which contains several links and nodes. This could represent a road, air or rail network. Should the setting of user-charges at i be affected by some other node such as j? For example, if j is a small airport and i is a major airport, can a cross subsidy to j from i be justified on economic grounds? Some would say yes if there is a consumption externality between i and j, or if there are decreasing costs at j and constant returns at i, economic welfare is improved by setting $P_j = MC_j$ and $P_i > MC_i$. It is not clear, however, that either network externalities or scale economies are sufficient to justify a cross subsidy.

Consider a situation in which i and j are operated with optimal productive efficiency (an assumption returned to later) or equivalently they are operating on the most efficient cost function. Set runway prices at i are equal to social marginal cost such that the sum of revenues equals costs. At j, setting price equal to marginal costs results in a deficit. What are the options? If there is some lumpiness at i, the Oum and Zhang (1990) result suggests revenues will be available to subsidize j. A second alternative is to raise fees at i to generate revenues for node j such that the surplus at i equals the deficit at j. In this circumstance the welfare loss would be borne by all routes including movements from j to i. Rather than "tax" users at i, the operator at j could employ Ramsey pricing to achieve full cost recovery at j. Suppose ij traffic is the least price elastic, this group will pay proportionately more than, say, ji passengers. The welfare loss, in this case, would be restricted to the j and i markets. In the former case the welfare loss is spread over i,j,k,n,m and e markets. Which approach has a greater welfare loss will depend on the values of the elasticities and the amount by which price changes. This, albeit, simple analysis leads to the following questions: Is there a rationale for the transfer? Why is there a deficit? Should the set of airports or links be treated as a system rather than as individual operations to be priced separately *and* independently?

First, is there a rationale for the transfer? Some argue that there is complementarity between j and i since j feeds passengers to i. While this is true, the cost savings represent a gain to the airlines providing the service through density economies and are not a gain to the airport. To the extent it is a gain to a private firm, it should be reflected in the landing fees airlines are willing to pay at j. In other words, the density economy gained by the airline would be fully internalized through higher landing fees. It makes little economic sense to transfer the rents to the airline and not have j exploit the fiscal capacity it has available. This means landing fees should be increased above marginal cost at j in a price-discriminating fashion until the deficit is covered.

Second, why is there a deficit? What about increasing returns to scale or some other form of cost economies which lead to average cost being greater than marginal cost? While this may be a legitimate source of potential gain, it will likely rest with the smaller airport since evidence suggests constant returns to scale for larger airports (Morrison, 1983). Marginal cost pricing results in a deficit at the small airport and breaking even at the large airport. But this is the argument discussed above regarding the relative values of Ramsey pricing at j or increasing all prices at i to cross subsidize j, and there is no evidence one way or the other as to which welfare loss is greater.

Economies of scale at j are not a sufficient reason to cross subsidize from i nor are they a sufficient reason to subsidize from general tax revenues. Indeed, as argued elsewhere, subsidizing from general tax revenues may also result in a net welfare loss. Thus, in the case of airports, there does not seem to be a demand side argument for cross subsidy. The choice is between Ramsey pricing at j or a subsidy from the general taxpayer, whichever has the lower welfare loss. What about roads? The same arguments would seem to hold here as well. Trucks, buses or even cars could have density economies but these should be internalized through normal commercial transactions.

Should the links and nodes be treated as a system or individually? This depends on a number of factors including the second question raised above, why is there a deficit? Other factors include the practical ability to price separately, to have the information to price efficiently, to have some idea of the welfare losses of moving from link- or node-specific prices to some "average" system price and the extent to which there is substitution between links, nodes or routes. The greater the extent of substitution the stronger the case for individualized facility pricing.

Airports have no apparent demand side externalities. They are also sufficiently distinct that it is possible to identify demand and supply at each facility and establish a set of efficient prices. Furthermore, the set of efficient prices leads to optimal investment decisions at each airport to reflect the need for capacity. If there were increasing returns at one or more airports the decision to "tax" major airports to subsidize smaller airports would have to weigh the welfare gains associated with such a subsidy at the recipient airport against the welfare losses at the taxed facilities. It also seems to imply that there is no competition between airports. To the extent there is competition, the ability to raise revenues for cross subsidy would be reduced.

Moving from a specific to a system fee is probably best illustrated by the system used in Canada until recently; fees were set to be identical for Group I airports regardless of differences in demand and costs between facilities.[24] The greater the differences, the larger the welfare loss from system average pricing. Perhaps the greatest argument for individual pricing is to provide the incentives to individual facilities to achieve least-cost production (productive efficiencies) and to exploit available "scope" economies.

There is strong and convincing evidence that system average pricing and a lack of market discipline has led to cost inefficiencies at Canadian airports (Hamilton, 1991). Furthermore, if they are explicitly or implicitly subsidized, airports have little incentive to exploit available opportunities which contribute to commercial success and break-even financial performance. Many small- to medium-size airports in the United States exploit their commercial and concession revenues to a greater extent than similar airports in Canada. They also go beyond producing the products of aircraft movements and enplaned/deplaned passengers; they exploit commercial potential and scope economies. One could argue that the welfare gains available from achieving least-cost operations far outweigh any welfare losses resulting from forcing small airports to be self-financing. Two important characteristics of airports which must be kept in mind are that they are substitutes as well as complements to other airports and they can produce many products besides those conventionally identified with airports.

Similar arguments may be made for roads, but there may be a stronger case for treating roads in a network framework. Like airports, there does not seem to be a convincing case to be made for demand side complementarities which would support an efficiency argument of cross subsidy between links. Different roads have different capacities and thus face different costs and demands. This would favour a pricing structure which reflected these characteristics. It is obvious that the distribution of traffic over the system, and the attendant welfare level, will be quite different if a single price is charged for the use of any part of the system than if the price is set equal to the marginal cost of each route and node to reflect the cost and demand conditions unique to that part of the system.

The strongest arguments for treating roads in a system are the transactions costs associated with separate pricing of each link and the lack of information regarding demand and costs on each link. It is important, however, that systems be defined to be as homogeneous as possible since, like airports, the greater the divergence in costs and demand across links the greater the welfare loss resulting from system average pricing. Prices will generally be too high on large, high-quality facilities and too low on low-density, low-quality facilities. We may thus end up with too much capacity in the latter because prices greater than marginal cost will attract capital and too little in the former since the low return will discourage capital investment. It is also important to have roadway services, produced in the most efficient way possible, achieve productive economies. This means that some links may need to be abandoned and their freed-up resources devoted to other links yielding a higher return.

4.3 SECOND-BEST PRICING AND SECOND-BEST INVESTMENT

The conventional first-best rule for optimizing the level of investment in transportation infrastructure is the equalization of marginal benefits and costs. If there are institutional constraints which prevent prices from being set at marginal cost in either input or output markets, the optimal investment and pricing rules must be modified. Transportation infrastructure can be financed with tolls (output taxes), fuel taxes (input taxes) or from general revenues. If there are constant returns to scale in production and divisibility in capacity investment, there is no difficulty in efficiently pricing and financing capacity. In the absence of these two conditions, however, and given the constraint that users should pay the full cost and finance the facility without a subsidy from general revenues, optimality requires a second-best investment rule in conjunction with second-best prices (input and output taxes). As Friedlaender and Mathur (1982) have shown, these investment levels and rules depend on the nature of the price distortions and financing constraints.

A second-best investment rule differs from the first-best rule. With first-best pricing and investment, prices are set equal to marginal cost. There are no output or input taxes, and investment is carried to the point at which the marginal cost savings resulting from the investment in infrastructure equals the marginal cost of the infrastructure investment. In a second-best situation the investment rule depends on the way in which the infrastructure is priced and whether there is a financing constraint. Two quite different cases can occur.

In one state, the problem may not be one of financing the facility, insuring total revenue equals total cost, but rather one in which there are output market distortions such as congestion or vehicular-generated pollution. This means that in the absence of an externality charge the mix of prices will not yield an economically efficient outcome. If there is underpricing, for example, the price set for roadway use is less than marginal social cost. A second-best pricing and investment strategy would set a fuel tax and reduce the size of the infrastructure relative to a first-best level. On the other hand if an "arbitrary" fuel tax is in place, as when governments use fuel as a source of general revenues, the second-best rule leads to setting infrastructure prices at less than marginal social cost yielding an implicit subsidy, as well as again reducing the investment in infrastructure capacity.

In this second-best situation, the reduction in the supply of capacity is used to reduce the efficiency loss resulting from the distortion in prices. This is true as long as fuel and infrastructure are substitutes (Friedlaender and Spady, 1981). If fuel prices are too high, people substitute towards infrastructure since its relative price has fallen. To offset this distortion, the supply of infrastructure is reduced to make it relatively more scarce and increase its price. Furthermore, to the extent that the elasticity of facility use with respect to the size of the facility is greater than zero, a reduction in investment brings about a reduction in demand. The same arguments hold when there is underpricing.

A second case, and one perhaps more relevant to the discussion at hand, occurs when there is a financing constraint. The long-run pricing and investment problem is to define an optimal structure of infrastructure prices and fuel taxes and determine the optimal investment in capacity. The difference in the level of investment between the first- and second-best investment rule depends on the sensitivity of revenue with respect to the amount of capacity and output taxes. Revenue will be sensitive to the level of capacity because, as capacity increases, the number of users increases and revenue from output prices will rise. If usage is highly elastic with respect to the level of capacity revenue will also be elastic with respect to infrastructure.

Friedlaender and Mathur (1982) showed that if revenue and usage are not particularly elastic with respect to investment, it is desirable to reduce investment relative to the first-best level because distortions caused by second- best user-charges, needed to finance the capacity, generate a deadweight loss burden. If revenue is elastic with respect to capacity, as it may be at some hub airports, it is desirable to expand capacity beyond the first-best level.

The second-best investment rule provides a linkage between:

- the optimal difference between the marginal benefits and costs of the infrastructure and the sensitivity of revenues to output prices and fuel taxes; and
- changes to the level of infrastructure capacity.

The simulations which have been undertaken to examine the welfare differences of first- and second-best investment rules in the presence of pricing distortions or financing constraints have consistently illustrated that the welfare loss of using first-best investment rules is relatively small. Friedlaender and Mathur (1982) undertook the simulation for rail and road and Borins (1978) performed a similar analysis for air. In both cases, the welfare surface is relatively flat primarily because the benefit function is relatively flat. Therefore, second-best pricing rules in practice can rely on first-best investment rules in establishing the optimal investment in capacity.

4.4 USER-CHARGES AND COST RECOVERY

Moving from the principles, outlined in Section 2 of this paper and formalized in subsection 4.1, to practice in pricing transportation infrastructure requires the development of pragmatic schemes which approximate or build upon social marginal cost pricing while simultaneously providing predictability to users and capital markets. Practical pricing schemes are discussed below, and some information of the type of changes which the various user-groups in the modes might expect with a move from current pricing principles to more economically efficient ones is provided. The illustrations are not meant to be taken as prices which would be implemented tomorrow if there is agreement to change the pricing principles. Rather they are designed to be illustrative of the magnitude and type of changes for different facilities and different user-groups with a shift to more efficient pricing.[25]

4.4.1 Airport User-Charges

Runways at large airports are generally built to handle the largest and heaviest aircraft currently in use by commercial air carriers. The additional construction costs for increasing the durability of a runway is small relative to the congestion delay costs.[26] Therefore, the condition of optimal durability in equation (4.2c) can be ignored in discussing airport pricing and investment.

Major Airports

The 1991 landing fee schedule for most Canadian airports is reported in Table 4.1. The same fee schedule is applied to all Group I and II airports and major international airports operated by Transport Canada at all times, including peak periods. The only exception is that minimum landing fees were imposed starting in 1991 at Vancouver and Toronto international airports. The use of uniform landing fees at all airports is both inequitable and inefficient because both social marginal cost and price elasticities differ significantly across Canadian airports. Other than charging higher landing fees to international flights, the current landing fees do not distinguish between short and long flights.[27]

Table 4.1
LANDING CHARGES
GROUP I, II AND MAJOR INTERNATIONAL AIRPORTS
CANADA

Domestic Flights		
Charge: $/1000 kg or fraction thereof		
Weight (kg)	**Jet**	**Turboprop**
Less than 21,000	$2.27–2.52	$1.92–2.13
21,000–45,000	$2.87–3.20	$2.33–2.60
More than 45,000	$3.40–3.78	$2.87–3.20
International Flights		
Charge: $/1000 kg or fraction thereof		
Weight (kg)	**Jet**	**Turboprop**
Less than 30,000	$2.94–3.27	$2.39–2.66
30,000–70,000	$3.63–4.03	$3.15–3.51
More than 70,000	$4.01–4.58	$4.38–4.87

Source: Transport Canada (1991).

For major airports such as Toronto, Vancouver, Calgary and Dorval Airport in Montreal, external congestion costs are a major component of the social marginal costs for runway use. Therefore, congestion tolls could become the major source to finance capacity expansion. However, the level of congestion varies substantially by the time of day. This means that peak-period pricing may be an attractive means to approximate social marginal cost pricing for runways and terminals. For example, the British Airports Authority (now privatized as BAA Plc.) has implemented peak-period pricing at all of its seven airports. At major airports such as London's Heathrow, the peak-period landing and terminal fees are about five or six times higher than the off-peak fees. The implementation of peak-period pricing in Canada could lead to a dramatic change in the structure of landing/takeoff and passenger terminal fees at major Canadian airports where congestion problems exist. Peak-period pricing would result in a dramatic increase in peak-period landing and passenger terminal fees, the elimination of the current weight-based landing fees, and the reduction of off-peak landing fees below their current levels.

Table 4.2 illustrates the primary results of a simulation of introducing peak-period pricing at Pearson International Airport (PIA), (Gillen, Oum and Tretheway, 1988). The authors examined the hourly and the daily traffic variations at PIA and found that a three-part pricing schedule made sense; one for weekday evening peak hours (1700 to 2100), one for weekday mornings (0700 to 1000) and weekend evening peak hours (1700 to 2100), and one for all other times, also referred to as off-peak hours. The charges are highest during high peaks, somewhat lower fees are levied in the lower peak, and the smallest fees are assessed in the off-peak periods. Further, an analysis revealed two distinct seasons that dominate aircraft movements: March–October, referred to as the high season and November–February, referred to as the low season. They, therefore, recommended an annual fee structure of six prices, one for each of high-peak, low-peak and off-peak periods for the two (high and low) seasons.

The authors estimated social marginal costs and compared them with the landing fees for selected aircraft in 1985 (See Table 4.2). However, under social marginal cost pricing, general aviation (GA) aircraft (represented in their analysis by light piston aircraft) which previously paid nothing for the use of runways,[28] would pay $31 for landing during off-peak periods, $161 during the summer high-peak, and $43 during the winter low-peak periods.

A heavy aircraft such as a B747-200, which used to pay $769 and $521 for a takeoff/landing of international and domestic flights, respectively, would pay $426 during summer high-peak, $246 during summer low-peak, $271 during winter high-peak, $226 during the winter low-peak and $206 during off-peak periods. The implication is that social marginal cost pricing would reduce the landing and takeoff fees for large aircraft, while substantially raising the fees for small aircraft, especially during peak periods, from their current values.

Table 4.2

Comparisons Between Social Marginal Costs and 1985 Landing Fees for Selected Aircraft, Pearson International Airport, Toronto

(1988 CAN $)

	Aircraft type					
	B747-200	DC10-40	B737-200	Dash 8	Business jet	Piston
1985 fees						
Domestic	$521	$355	$80	$12	N/A	$ 0
International	$769	$524	$85	$15	N/A	$ 0
Social marginal cost						
High months						
High-peak	$426	$376	$269	$213	$211	$161
Low-peak	$246	$196	$154	$ 98	$105	$ 55
Low months						
High-peak	$271	$221	$170	$114	$120	$ 70
Low-peak	$226	$176	$142	$ 86	$ 93	$ 43
Off-peak	$206	$156	$129	$ 73	$ 81	$ 31
Price elasticity	N/A	–0.068	–0.075	–0.086	N/A	–0.58

Source: Gillen, Oum and Tretheway (1988).

The implementation of peak-period pricing based on social marginal costs by type of user is likely to shift demand for airport services to less congested airports. Piston engine aircraft which represent the bulk of the general aviation traffic could be virtually eliminated during high and low peaks in terms of the number of landings. Furthermore, Gillen, Oum and Tretheway's estimates of the price elasticities indicated that peak-period pricing would probably bring some marginal changes to air carrier scheduling. Commuter carriers would make the most changes in their scheduling to avoid the high-peak period fees; thus, the movement of heavier aircraft would increase slightly as the charges decrease.

If social marginal cost pricing had been implemented in 1986, PIA's revenue from runway operations could have doubled to reach $26.6 million. This would allow PIA's airfield operation to generate surpluses, after covering the interest and depreciation expenses. This surplus could be used to expand capacity when economically efficient to do so.

Small Airports

Fees at small airports can also be based on social marginal cost. However, since there is very little congestion at small airports, peak-period user-charges would not be much different from off-peak periods. Therefore, (social) marginal cost pricing would most likely lead to financial deficits since marginal cost would be less than average costs for the reasons discussed in Section 2. The two alternative modifications to efficient prices which allow substitution of a cost-recovery target were discussed in Section 2. First, the Ramsey pricing principle may be applied to achieve a cost-recovery target (including break even). Since the size of the quasi-optimal markup over marginal cost is determined on the basis of willingness to pay (that is, price elasticity), it makes sense to charge higher landing/takeoff fees to larger aircraft, which carry higher payloads than small aircraft. Perhaps this was the economic rationale for the weight-based landing fee structure currently in use in many countries, including Canada. The weight-based fee structure loses its economic rationale when congestion develops. Second, a two-part tariff (usage/access charges) can be applied to cover the system-wide or individual airport shortfall in revenue by charging access (licence) fees while charging site-specific social marginal costs as usage fees.[29]

In principle, the choice between Ramsey pricing and two-part pricing structures depends on the characteristics of demand and the proportions of fixed and variable cost in total cost. It is these elements which effect the welfare gains. Ramsey pricing would be preferred over two-part pricing structures if the elasticity of entry to the system or access to an airport is non-zero. If access is price sensitive, the variable portion of the multi-part price must be increased above marginal cost. In the case of multiple users, this increase would rely, essentially, on a Ramsey concept. On the other hand, usage/access fees would be preferred, from a welfare gain perspective, if the elasticity of demand with respect to access were zero, or equivalently, demand were fixed.

Theoretical second-best and "applied" second-best pricing may lead to the selection of different pricing principles, at least in the short run. An evolutionary rather than revolutionary approach may be required to minimize the complexity and to gain experience with a new system. Both users and suppliers have time to adjust. If a new pricing principle creates uncertainty, it may, in the end, result in a net welfare loss. There is a balance between moving so quickly that markets are disrupted and the potential welfare gains are lost, and moving so slowly that incumbents become entrenched, and institutions are unable to fully implement socially efficient pricing, again resulting in foregone welfare gains.

London airport, a small national airport in southwestern Ontario, is an example of the consequences of introducing new pricing schemes at small airports. The airport depends largely on the presence of the military. It is the home base of a commuter/feeder carrier. In the 1984-85 fiscal year, the airport recovered only about 20 percent of the total airfield costs from landing fees. Gillen, Oum and Tretheway (1986) illustrated the application of both social marginal cost pricing and the Ramsey price principle by type of user for this airport. Table 4.3 is taken from their study. The upper portion of the table summarizes 1985 revenues from the landing fees.

The mid-section of the table shows that application of marginal cost pricing by type of user would eliminate piston aircraft and severely reduce turboprop movements. Jet aircraft movements would also be reduced but only slightly. The marginal airfield costs estimated for this airport are $193, $110 and $78 for jet, turboprop and piston aircraft, respectively. These compare to the current average landing fees of $164, $20 and $0. Under social marginal cost pricing, the airfield revenue was expected to reach approximately $862,000 which is higher than the actual 1985 revenue. Nevertheless, it only represented about 39 percent of the 1985 airfield costs.

To achieve higher cost recovery with minimum loss of economic efficiency, a Ramsey pricing principle can be implemented. Ramsey pricing with a 62.5 percent cost-recovery target, for example, requires charging $361, $166 and $85 to jet, turboprop and piston aircraft, respectively. The results also show that implementing these charges would nearly eliminate piston aircraft movements and reduce jet and turboprop landings to 1,788 and 4,464 from the actual landings of 1,856 and 6,512, respectively. Total revenue could reach $1.4 million, about 62.5 percent of the 1985 actual airfield costs.

Table 4.3
SOCIAL MARGINAL COST PRICING AND RAMSEY PRICING BY TYPE OF USER FOR LONDON (ONTARIO) AIRPORT

	Jet	Turboprop	Piston	Total
1. 1985 situation				
No. of landings	1,856 (7.3%)	6,512 (25.7%)	16,982 (67.0%)	23,350
Average landing fee estimates	$163.66	$ 20.0	$ 0	—
Revenues	$303,760	$130,240	$ 0	$434,000
1985 airfield cost	—	—	—	$2,223,000
2. Marginal cost pricing				
Estimated MC	$192.51	$110.29	$ 77.94	—
Price elasticity	–0.04	–0.065	–0.325	—
No. of landings (forecast)	1,843	4,601	0	6,444
Revenue (forecast)	$354,778	$507,398	0	$862,176
Cost recovery ratio	—	—	—	39%
3. Ramsey pricing with estimated marginal costs				
Ramsey prices	$360.95	$166.41	$ 85.29	—
% Markup	87%	50%	9.4%	—
No. of landings (forecast)	1,788	4,464	0	6,252
Revenue (forecast)	$645,378	$742,854	0	$1,388,232
Cost recovery	—	—	—	62.5%

Source: Gillen, Oum and Tretheway (1986).

Again it should be noted that costs are expected to fall as airports are defederalized. Finally, it is important to point out that marginal social cost pricing may not result in deficits at small airports if there is constant returns to scale and no significant lumpiness in capacity investment.

4.4.2 Summary of Pricing and Airport Cost Recovery

The major conclusions reached and changes advocated with respect to the current user-charge system may be summarized as follows:[30]

- The uniform user-charge system should be repealed in favour of site-specific user-charges.
- The current weight-based landing fees result in fees which are too low for small aircraft (primarily general aviation and corporate aircraft) and too high for large aircraft during congested periods.
- Peak-period user-charges should be implemented in airports with a congestion problem, and the differential user-charges between peak, low-peak, and off-peak periods should reflect their respective social marginal costs, including congestion externality costs. This tends to lead to a large differential between peak and off-peak fees.[31]
- A weight-based landing fee, if appropriately set, may be consistent with the Ramsey pricing principle, and thus may be economically justified at small uncongested airports. This is because the demand for airport services by larger (heavier) aircraft is less price-elastic than by smaller (lighter) aircraft. The rationale for weight-based pricing breaks down, however, when the airport becomes congested and congestion pricing is applied.
- Charging higher landing fees for long-haul international flights at uncongested airports can be economically justified because it is consistent with the spirit of Ramsey pricing; the demand for airport services by long-haul flights is more price-inelastic than short-haul flights.

4.5 ROAD USER-CHARGES

Previous research has led to the conclusion that the best way to economize on maintaining and using an existing road is to set user-charges equal to social marginal costs, that is, the actual cost each user imposes on society, including the effect on the road's condition, noise, pollution and the delay imposed on others. Such a charge would ensure that independent decisions by users reflect the interests of society. On the other hand, the social cost of road use largely depends on the design of the road. Road infrastructure provides capacity in the form of traffic lanes and durability in the form of pavement thickness to facilitate its use by heavy vehicles. Both road capacity and durability investments are expensive and involve scarce resources. The scarcity associated with capacity causes congestion costs while the scarcity associated with durability causes greater costs for road wear and damage (road deterioration) than for the construction of more durable roads.

Road charges should differ by type of user; for example, heavy versus light vehicles, passenger versus freight services or commercial (common carriers) versus private users. The empirical results of two previous studies on this topic are summarized to provide an indication of the direction of adjustment required in road user-charges, particularly by type of road users: a study by Nix (1989) entitled *Road-User Costs* using the most recent Canadian data, and a U.S. study undertaken by Small, Winston and Evans (1989) entitled *Road Work*.

4.5.1 Nix (1989)

The most significant attempt to estimate road-user costs in Canada was by Haritos (1973). He used 1968 data and employed engineering and regression approaches to allocate road costs. Nix (1989) updated Haritos' study with more recent data compiled by the Road Transport Association of Canada (RTAC) and up-to-date information on vehicle characteristics.

Haritos adopted a two-part price structure: maintenance costs (regarded as annual costs) and capital costs (fixed in the short term). The costs considered avoidable within the time frame of one year were considered escapable and, therefore, included in the (short-run) marginal cost. This measure should be used to set the user-charge per vehicle-kilometre. The costs not avoided within a year are considered inescapable, and thus treated as a fixed cost in the short run. Haritos argued that short-run fixed costs should be recovered through annual fees such as vehicle registration and licence fees.

Table 4.4 summarizes the results of Nix's cost allocation study.[32] It compares the 1986 road user-charges and the road costs by vehicle category. The results show that the annual fixed user fees (licence fees) for cars and trucks are too low to cover the short-term fixed cost of roads. In fact, the revenues from the annual fixed user-charges cover only about one third of the fixed road costs. A comparison of the user-charges and road costs imposed by user group (Table 4.4) indicates that charges for passenger cars and other light vehicles are much higher than the financial cost they impose on the road by a factor of 3 to 1 while charges are too low for trucks. In other words, light vehicles, which do not cause much damage to roads, are charged three times the cost per kilometre while heavy vehicles are undercharged. This result may not be interpreted as overcharging light vehicles because neither Nix nor Haritos considered the externality costs of noise, air pollution and congestion delay a driver imposes on others. However, it is safe to interpret the result as undercharging heavy vehicles. Nix also pointed out that, during the past two decades, road user-charges in Canada did not increase as fast as road construction costs.

Table 4.4
COMPARISON OF ROAD-USER-CHARGES AND COSTS[a]
(1986 CAN $)

	User-Charges		User-Costs	
	Range	Mid-point	Range	Mid-point
Annual charges versus annual costs				
Cars	$86		$253–$316	$284.5
3-axle truck (18.1 t)	$250–$675	$462.5	$1881–$2606	$2243.5
3-S2 tractor (32.1 t)	$450–$1600	$1025	$2897–$4014	$3455.5
3-S3 tractor (46-49 t)	$600–$2700	$1650	$4140–$5739	$4939.5
8-axle B-train (62.5 t)	$1450–$4000	$2725	$5394–$7478	$6436
B-train, empty (18.1 t)	$1450–$4000	$2725	$1575–$2182	$1878.5
Trip charges (fuel) versus trip costs, per kilometre[b]				
Cars	$0.012- $0.020	$0.016	$0.004- $0.006	$0.005
3-axle truck	$0.051–$0.084	$0.068	$0.026–$0.056	$0.041
3-S2 tractor	$0.069–$0.097	$0.083	$0.043–$0.090	$0.067
3-S3 tractor	$0.071–$0.117	$0.094	$0.063–$0.130	$0.097
8-axle B-train	$0.086–$0.141	$0.114	$0.084–$0.172	$0.128
B-train, empty	$0.051–$0.084	$0.068	$0.020–$0.046	$0.033

Source: Nix (1989).

a Annual costs include the appropriate annual vehicle charge plus an annual charge for each axle on the vehicle.

b Trip cost includes the appropriate vehicle cost per kilometre given the vehicle weight plus a cost per axle-kilometre for each axle.

The Nix study represents an important conceptual contribution. Clearly, there may be some argument regarding the exact numbers since those reported are based on a particular set of assumptions. What the paper does indicate, however, is that moving to marginal social cost pricing of roads has a relatively greater impact on the structure of prices among road users than on the level of prices. Indeed, with the recent increases in fuel taxes by provincial governments, there is a real possibility that car users are paying their full social costs of road use on uncongested roads. Changing the basis of prices to social marginal costs will affect the distribution of costs between car and truck and most significantly among different types of truck. The following study reinforces this point.

4.5.2 Small, Winston and Evans (1989)

Small, Winston and Evans derived road user-charges under an optimal investment policy (with respect to road capacity and thickness of pavement), taking into account the combined cost to the vehicle owners and infrastructure providers. Inclusion of the cost to vehicle owners of differences in highway capacity and durability was the distinguishing feature of this study as compared to Haritos (1973) and Nix (1989). This allowed Small, Winston and Evans to take into account the effects of a lower standard of road maintenance on the cost of vehicle maintenance and fuel consumption.

The two primary considerations in pricing the use of existing roads are road wear costs and congestion delay costs. The road wear costs include road maintenance costs and the user-costs for operating vehicles, including vehicle repair, vehicle depreciation, fuel expenses and their value of time. The authors computed optimal road durability by minimizing the sum of annualized road maintenance costs, user (passengers and carriers) costs and capital costs for road construction. This procedure is equivalent to determining the optimal level of investment in road quality (thickness) through a series of cost-benefit analyses. The study found that the United States builds roads of lower quality (too little thickness) than economically optimal. This increases both the combined road infrastructure and operator costs, making the road transportation sector less cost efficient than would be the case if roads were built to optimal standards. Computations are made for rigid as well as flexible pavements.

An important conclusion of the study was that a United States road-pricing policy should be targetted to reduce weight per axle as vehicles with heavy loads per axle cause most of the road damage. They emphasized that it is the weight per axle that matters, not total vehicle weight. For example, a 50,000-pound, two-axle truck causes more road wear than a huge twin-trailer rig spreading 100,000 pounds over seven axles. Most road damage is caused by heavy vehicles with a small number of axles.

To measure the congestion costs caused by vehicle traffic, the study used the concept of passenger car equivalents (PCE) per hour. The PCE for each vehicle was determined by the amount of road space it effectively took up, including the space between vehicles required for safety, compared with that of an average car. For example, a typical truck or bus has two-to-five passenger car equivalents. As in Nix's study, the authors relied on an engineering approach to quantify the relationship between traffic volume and speed. This relationship is essential to measure the extra delay caused by adding one passenger car equivalent to the traffic stream. The authors advocated significant changes in road-pricing policy for heavy trucks, shifting from a reliance on fuel taxes and weight-graduated licence fees to one of direct mileage charges steeply graduated with respect to axle loads. They estimated that such a pricing system in conjunction with a modest increase in capital outlays on improving road thickness ("a sensible investment policy") could reduce maintenance costs by about $9.4 billion a year.

Table 4.5 presents the estimates of the effects of their policy on the maintenance costs and "tax revenue" (user-charge revenue) attributable to each truck type, as a share of the total maintenance costs. This table shows that for most truck categories current user- charge revenues fall far short of maintenance costs attributable to them. The user-charge ("tax") revenues for intercity and urban roads cover only 29 percent and 14 percent of pavement maintenance costs, respectively. In contrast, under the proposed policy, the total maintenance costs for both intercity and urban roads would be fully recovered by user-charge revenues. Furthermore, the authors showed that the proposed policy would result in a welfare gain of $8 billion a year over the current pricing and investment policy.

Table 4.5

CONTRIBUTION TO ALLOCABLE MAINTENANCE COSTS BY USER-TYPE[a]

Vehicle type	Current pricing and investment		Optimal pricing and investment	
	Share of allocable maintenance costs (%)	Tax revenue contribution to allocable maintenance costs (%)	Share of allocable maintenance costs (%)	Tax revenue contribution to allocatable maintenance costs (%)
Intercity				
SU2	45.19	9.81	10.07	19.38
SU3	0.74	1.31	1.33	1.35
TT4	0.43	0.11	0.21	0.21
TT5	0.19	0.09	0.33	0.33
CS3	2.11	0.63	1.70	1.71
CS4	0.76	0.33	0.72	0.72
CS5	43.02	13.51	60.32	60.38
CS6	6.34	2.83	14.60	14.62
DS5	1.18	0.42	1.68	1.68
DS6	0.04	0.01	0.04	0.05
Total	**100.0**	**29.05**	**100.0**	**100.43**
Urban				
SU2	88.24	11.88	67.52	67.92
SU3	3.10	1.25	9.61	9.67
TT4	0.61	0.13	0.96	0.96
TT5	0.06	0.01	0.09	0.09
CS3	0.58	0.18	2.49	2.50
CS4	0.25	0.07	0.61	0.61
CS5	5.82	0.40	6.95	6.96
CS6	0.90	0.13	10.51	10.52
DS5	0.38	0.07	0.87	0.87
DS6	0.06	0.01	0.39	0.39
Total	**100.0**	**14.13**	**100.0**	**100.49**

Source: Small, Winston, and Evans (1989).

a Type of vehicles as follows: SU2, Single Unit 2-Axle; SU3, Single Unit 3-axle; TT4, Truck Trailer 4-axle; TT5, Truck Trailer 5-axle; CS3, Conventional Semi 3-axles; CS4, Conventional Semi 4-axle; CS5, Conventional Semi 5-axle; CS6, Conventional Semi 6-axle; DS5, Double 5-axle; DS6, Double 6-axle.

4.5.3 Summary of Road Cost Recovery by User-Group

The studies conducted in Canada and the United States indicate:

- Passenger cars and other light vehicles pay a disproportionately higher share of the total road maintenance costs as compared to trucks and other heavy vehicles.
- Lack of congestion tolls as well as other externality charges in both countries indicate that the current user-charge system undercharges road users on congested roads. It may also undercharge with respect to environmental costs. This is more likely in the United States where fuel tax rates are less than half of those in Canada. With increases in fuel taxes in the last year, it may well be that car users, in Canada, are paying approximately their marginal social costs on uncongested roads.
- Both the Canadian and American studies illustrate that an economically efficient pricing scheme would affect both the average level of road prices *and* the burden across user-groups. If one interprets a fair or equitable pricing scheme as one in which the prices users are paying reflect the costs which they impose, then the efficient pricing scheme which shifts the burden can also be said to be fair.
- Total road costs (the sum of capital and maintenance costs) are sensitive to the level of investment in capacity and durability and to environmental factors. In the United States, the underinvestment in durability (that is, building sub-standard roads) has resulted in an economic loss amounting to billions of dollars annually due to the increased road maintenance costs and damage to the vehicles. The study by Nix, Boucher and Hutchinson (1992) indicated that roads in Canada are built to a higher level of durability due to the harsher climate. Road deterioration costs range between 50 percent and 80 percent of total costs. Therefore, the Small, Winston and Evans (1989) numerical results must be interpreted with some care when placed in a Canadian context.

5. Alternative Methods of Financing Road Infrastructure

5.1 INTRODUCTION

In this section alternative methods of financing road infrastructure proposed by analysts and practitioners are described and evaluated. The evaluation is based on several criteria: the effect on the efficient use of existing capacity, the effect on the efficiency of investment, the implications for equity, administrative feasibility, ease and cost of collection, the practicality of generating revenue and domain of applications (that is, national, local or project basis).

For discussion, the charges related to the road transportation sector are classified into three major groups: vehicle usage, vehicle acquisition and ownership and beneficiaries of road access.[33] The most widely applied user-charges are the motive fuel taxes and annual registration and licence fees. These, along with the general revenue of the governments are the traditional sources of road financing in most countries.[34] This subsection describes each of these categories of payments made by road users. An evaluation of each category as a road financing method is treated in subsection 5.2. And finally, several alternative sets of the charging instruments are discussed and evaluated in subsection 5.3.

5.2 THREE CATEGORIES OF CURRENT USER-PAYMENTS

Charges Related to Vehicle Usage

Fuel taxes form a major portion of the total revenue collected from the road transportation sector. In 1989, they accounted for 77 percent of the total revenue from the road sector in Canada. The use of fuel taxes can be justified in a number of ways including road usage, energy conservation and environmental pollution. The level of fuel tax can have an effect on vehicle usage, as well as the number and type of vehicles owned. Fuel taxes provide a cost-effective and administratively feasible means of allocating variable charges. There are some who argue that they approximate road usage costs for uncongested intercity roads because the total amount of tax paid varies in direct proportion with road usage. Whether fuel taxes are a reasonable proxy for reflecting road costs depends on the type of roadway and the level of traffic.

Charges Related to Vehicle Acquisition and Ownership

This category includes vehicle licence fees, registration fees, vehicle inspection fees, vehicle transfer taxes and excise taxes on the purchase of vehicles, etc. Licence fees can be substantial and are usually used in conjunction with fuel taxes. In some countries, fairly steep taxes are imposed when a car is acquired. The revenues from annual licence fees and payments related to vehicle acquisition usually contribute to those road costs which do not vary, in the short run, with road use. Other payments in this category are generally regarded as fees for the services rendered by government, and thus, are not considered as a source of financing for roads.

Charging Beneficiaries of the Improved Road Access

The construction costs associated with access roads, in Canada and the United States, have usually been financed from general tax revenues (of both the central and the local governments). Local improvement taxes have been used in the past. Now, as their current fiscal position deteriorates, governments are attempting more and more to charge these construction costs against beneficiaries. This approach takes various forms. In the case of new subdivisions, some governments require developers to build the roads and to recover their costs from subsequent purchasers. For major improvements to access roads into a city, some local governments levy land-value increment taxes in the district clearly benefiting from the project (special assessment district). In their recent study, Allen and Floyd (1991) concluded that the tax on special assessment districts along with toll roads were the only promising new sources for funding large-scale expansion projects.

5.3 EVALUATION OF THE ALTERNATIVE FINANCING TOOLS

The pros and cons of each of the charging instruments are evaluated below: motive fuel taxes, a graduated per-kilometre tax based on axle weights, congestion tolls, construction of toll roads, licence fees and taxes associated with acquiring and owning a vehicle of given type, and the methods of charging beneficiaries of the improved road access.

Motive Fuel Taxes

Administratively, fuel taxes are easy to collect and can be used to generate large sums of money for road infrastructure as fuel demand is relatively price-inelastic. At first glance, fuel taxes also appear to be economically efficient because the more extensively road infrastructure is used, the higher the total amount of fuel tax the user ends up paying. Fuel taxes provide a relatively efficient and practical means of pricing road use but only for cars and only for uncongested facilities. Although a high fuel tax (the case in many western European countries) may have some effect in discouraging low-value road users, there are at least two important reasons why fuel taxes are not as efficient as they appear for congested urban and near-urban roads. First, the largest component of the social cost on these facilities is the cost of the congestion externality. Fuel consumption, however, is essentially unrelated to the extent of congestion externality.

Second, as Small, Winston and Evans (1989) pointed out, usage of motive fuel taxes as a major source of road financing encourages operators to use heavy trucks with fewer axles (that is heavier loads per axle). When the load per axle increases, the amount of damage to the pavement also increases exponentially (Winston, 1991). The amount of fuel tax a trucker pays under the current system rises with a vehicle's axles since trucks with more axles require larger engines and obtain lower fuel economy. This provides truckers with an incentive to reduce the number of axles thereby increasing the load per axle. These arguments make it clear that the current fuel tax system is not an economically efficient way to finance road systems. The fuel tax is neither efficient for pricing road damages (maintenance cost), nor is it efficient for pricing congestion externality. Therefore, the mechanism for pricing road use must differ across user-types. To restrict usage fees for all users to the fuel tax may well result in a net welfare loss.

Graduated Per-Kilometre Tax Based on Axle Weights

As a way to price road damages efficiently, Small (1990) argued for revamping the current U.S. system of road taxes, and adopting a "steeply graduated per-mile tax based on axle weights."[35] This pricing system for road damages makes economic sense because as load per axle increases the amount of damage to the pavement increases exponentially. Furthermore, an extensive study by the U.S. Federal Highway Administration found such a graduated tax system to be feasible administratively. It requires slightly more record keeping than the weight-distance charges being used in Iceland, Norway, Sweden and several states in the United States. Small points out that such a graduated per-mile tax, which takes into account both weight and axle configuration, has been in effect, on a systematic basis, in New Zealand for years.

Several recent studies including Nix, Boucher and Hutchinson (1992) and Nix (1989) have shown that a substantial portion of the road maintenance and resurfacing costs in Canada is attributable to "environmental factors" such as elapsed time, weather and other climatic conditions. Therefore, assuming no interaction between deterioration due to environmental factors or due to use, implementation of the steeply graduated per-kilometre tax based on axle weight alone is not going to achieve maximum efficiency. For maximum efficiency, the portion of road maintenance and resurfacing costs attributable to the environmental factors may be recovered by a lump-sum tax such as annual vehicle licence fees, which may not influence usage of roads significantly depending on the value of particular elasticities.

How the fixed or unallocatable costs are allocated between a variable charge and a fixed charge depends on the relative values of three elasticities: the elasticity of access with respect to the fixed fee, the elasticity of access with respect to the variable fee and the elasticity of usage with respect to the fixed fee. If usage and access are completely inelastic with respect to the lump-sum or access fee, this fee should be set equal to average fixed costs. If, however, the just mentioned elasticities are non-zero, economic efficiency dictates that the fixed fee be reduced and the variable fee be increased. Indeed, one may expect that a greater part of the burden of meeting the revenue requirements is placed on the variable charge rather than the fixed charge since it provides more opportunity for those with more elastic demands to avoid the higher charge by consuming less of the good yet still have the opportunity to consume some of the good. A high access fee may preclude a number of people from the market. Clearly, a balance is required since too high a usage charge may also shift users to other markets.

Congestion Tolls

For nearly a century, transport economists have advocated that road authorities implement congestion tolls. The basic analytical framework for peak-period pricing (based on congestion tolls) was pioneered by Pigou (1912) and modelled formally in a short-run framework by Walters (1961). As discussed, Mohring and Harwitz (1962) recast the analysis into a long-run framework and established the relationship between optimal tolls and cost recovery. Other economists who have worked on the subject have recommended the use of peak-period pricing, as a practical means to charge for congestion or variations in demand with fixed capacity. Peak-period pricing has traditionally been advocated in those cases where there is no interaction between the users of a facility. Thus, the use of the facility by one user does not impose a cost on other users such as with congestion. For example, residential users demanding electrical power during the evening do not impose a cost on industrial users who demand it during the day. Similarly, car users of a highway in the evening do not interfere with car users during the day, because in both cases the demands are independent.

Congestion tolls are recommended when the use of a facility by one user imposes a cost on other users. For example, the use of a runway or approach path by one aircraft means that other aircraft cannot use it and must incur a cost by waiting. The demands are interdependent. In the case of congestion there is a difference between private cost and social cost whereas, with peak loads, private and social costs are the same but the issue is one of who should carry the burden of capacity costs. When peak-period pricing is used as a practical alternative, it is being used to proxy the difference between private and social costs.

The chief advantage of using congestion tolls to finance the road system lies with the improved economic efficiency it achieves as it results in users paying the social marginal costs of their transportation choices. This would not only lead to the efficient use of the given capacity in the short run but also act to provide a signal for efficient capacity investment programs in the long run. Despite these advantages, except for the electronic road pricing experiment in Hong Kong (see Hau, 1990),[36] no government uses congestion tolls as a major source of financing for road infrastructure.

In the past, several problems were cited as barriers to the popular acceptance of congestion tolls. First, it is not a straightforward case of calculating socially optimal congestion tolls by road sections and by time of day or day of week. Since congestion consumes enormous amounts of a valuable resource (people's time), a reasonable approximation, which can be done quite easily, is likely to be better than completely dismissing this sound pricing method. The second barrier is that it used to take time (at least to slow down) to collect tolls. This problem is now largely solved since electronic vehicle identification (EVI) technology has become quite reliable, as proven through the Hong Kong experiment. Heggie (1991) indicates that the advent of electronic toll collection can reduce the costs of toll collection on inter-urban toll roads and bridges to under 5 percent of gross revenues. Given the rapidly advancing technology in electronic vehicle identification, toll billing and collection, congestion tolls are likely to become an important new source for road financing.

The third barrier is the alleged unfair distribution of benefits. Opponents of congestion tolls indicate that poor working people and downtown business interests have the most to lose under a congestion charging system. Careful use of the congestion toll revenues may compensate those who are expected to lose under the system. For example, Small (1990) argued that, in the case of major urban areas such as San Francisco, Los Angeles or Toronto, revenues from congestion tolls would be so large that there would be little doubt that it is possible to fully offset the effects of peak user-charges on nearly all groups, including poor working people and downtown business interests. It is also possible to eliminate the need for fuel taxes, registration fees and a large portion of the local property tax and sales tax funds now being used for road maintenance.

In sum, the system of congestion tolls can play an important role as a source of road financing in the future. Even if all of the above problems are solved, the function of congestion tolls as a source of road financing is still limited to some portions of major intercity highways (those which are near-urban), bridges and major urban roads. It would not play a role in low-density roads including rural roads.

Construction of Toll Roads

In an era of fiscal conservatism, reflected by a shift from public- to private-sector provision of services, increasing constraints on governments' abilities to control resources and a general shift to a greater emphasis on efficiency, construction of toll roads is likely to become an important means of dealing with the increasing need for investment in new road systems. Currently, the toll-road networks are growing rapidly with over 5,400 kilometres in France, in excess of 5,100 kilometres in Italy, 4,700 kilometres in Japan and over 7,100 kilometres in the United States (Heggie, 1991). This can be an increasingly important source of financing for new road projects. Numerous states, including California, Colorado, Virginia and Texas, are making use of or are planning to make use of tolls as an alternative funding source (see Allen and Floyd, 1991). The British Columbia government built the Coquihalla toll highway linking Hope and Kamloops and the Okanagan Valley. Many of these toll roads charge differential tolls between peak and off-peak periods and, therefore, their pricing policy is consistent with the spirit of congestion tolls. An appropriately managed and regulated (including the monopoly nature of toll setting) toll road system can promote efficiency in using a given road capacity in the short run, and the attendant market forces will pressure the road authority into putting optimal capacity in place in the long run. Advances in collection technology are expected to encourage the use of toll facilities for building major intercity highways and urban access roads.

Annual Licence Fees and Taxes Involving the Acquisition of Vehicles

Empirical studies which have investigated the impact of higher licensing fees on peak-hour car driving (raising the fixed cost of car driving relative to common-carrier services) have concluded that such fees have failed to produce any substantial improvement in congestion (Allen and Floyd, 1991). They argue that governments are going to have to substantially raise licensing and registration fees in order to have any impact on reducing car use. A substantial increase in both the one-time registration fee (at the time a new or used vehicle is acquired)[37] and annual licence fees from the current level may improve the economic efficiency of using the existing road capacity subject to the condition discussed earlier regarding the elasticity values. If there are substantial fixed road costs which do not vary with traffic volume and the road authority decides to recover these fixed costs from road users (rather than subsidizing them from general tax revenues), it will be more efficient to recover the fixed costs as lump sum annual fees (licence fees) rather than allocating them by marking up the usage-related charges (per-kilometre charge) above the respective marginal costs if the access elasticity is equal to zero. If this is not so, the conditions set out earlier will hold. Specifically, access fees will be reduced and variable charges increased until the net change in welfare with a shift between the two fees is equal to zero; in essence welfare is enhanced by exploiting available gains from trade between the two sources of funds.

Although they are likely to be unpopular, like fuel taxes, these fees are administratively easy to assess and collect. The optimal taxes and charges involving annual ownership and acquisition of vehicles can generate large amounts of revenue for the road sector as these items currently account for 42% of the total charges paid by road users in Japan, 36% in the United Kingdom and 38% in the United States (Heggie, 1991) while accounting for 23% in Canada.

Charging Beneficiaries of the Improved Road Access

For developing local access roads in a suburban area, it is reasonable to require developers to build the roads and to incorporate the costs into the final selling price of the building or service site. This is an economically efficient solution in terms of both the size of the road investment as the developer is expected to make a trade-off between ease of road access and the value of the remaining lands they can sell.

For major improvements in urban road systems, a part of the construction costs may be assessed against the properly identified land-value increments. The concept of special assessment district (under which the geographic areas are identified to levy special assessments), which are being used in several states, is a feasible way to generate funds for large-scale highway reconstruction projects in a reasonably equitable way. This requires close coordination between the government with the taxing authority and the government in charge of road construction. The identification of a special assessment district and actual assessment of special levies can be a controversial process. For example, as soon as the major improvement in the road system is made, the option value of using the road can be realized by the businesses and residents in that area, and, thus, the special levies are economically justified. However, this one-time special levy does not guarantee the efficient use of the road infrastructure, and a combination of one-time special assessment for the construction costs and charging tolls (variable fees) related to road usage would yield a more efficient use of the road and a more efficient level of investment. It may be an interesting option to explore.

5.4 SUMMARY OF ALTERNATIVE ROAD FINANCING METHODS

Several promising sources of road financing are summarized below. Each charging method is evaluated with respect to the promotion of the efficient use of road capacity and durability, the efficiency of (capacity) investment, the potential size of the fund that could be generated, administrative feasibility and collection costs, equity among various groups and the domain of applicability. Several alternative combinations of charging methods are also evaluated and compared.

5.4.1 Summary Evaluation of Financing Methods

Motive Fuel Tax

The evaluation is restricted to its application to car users only. It would be ranked "moderate" in promoting efficient road usage on uncongested roads because the amount of fuel tax one pays is weakly correlated with road damage costs which are the responsibility of the user. It would rank "poor" in promoting the efficient allocation of peak-load capacity since its price is not correlated with road use. It would be regarded as "poor" in promoting investment efficiency on congested roads as the fuel tax does not exert much pressure to build roads with optimal design standards (thickness and capacity). Administrative ease is a strength as large amounts of funds can be collected easily with relatively small administrative and collection costs. Motive fuel tax can be applied for financing nation-wide and/or state-wide road systems. They are regarded as an equitable method of financing uncongested facilities because every user pays according to the user's consumption of fuel. It would be inequitable on a congested facility since it is not highly correlated with congestion costs.

Graduated Per-Mile Tax Based on Axle Weight

It promotes "reasonably high" efficiency in both pricing road damage and inducing an optimal investment in road durability as this tax can be designed to closely reflect the road damage costs each user imposes, and this will generate enough pressure for the road authority to build roads with optimal thickness. However, if a substantial portion of road damage is due to "environmental factors," this tax is likely to achieve maximum efficiency only if it is combined with a lump-sum tax such as vehicle licence fees. This tax is not intended to be a congestion charge or a charge which would optimally allocate scarce road capacity among competing users. It can generate a large amount of funds if it replaces the current fuel taxes, but it will require some administrative work and collection expense. It is feasible to administer such a tax system. Since users pay costs they impose, it can be considered equitable. To be effective, this system should be applied in all parts of a given jurisdiction such as a province or nationally rather than trying to apply it in only a part of a province, for example. If it is applied consistently on all roads within a jurisdiction, it will induce the efficient use of the road system by trucks as well as the efficient choice of trucking technology given the roadway system. If the tax is applied piecemeal, there will be less incentive for the adoption of trucks which minimize road damage. Higher overall road costs may also result if trucks attempt to avoid the charge on some roads by using non-toll roads and generating even more pavement damage.

Congestion Tolls

Tolls promote high efficiency both in the allocation of scarce road capacity and in inducing the optimal investment in road capacity (number of lanes). Congestion tolls are not intended to be a price for road damage but rather to ensure the roads are used and invested in efficiently. Congestion tolls in major urban or near-urban areas could raise a large amount of money. The Hong Kong experiment has provided convincing evidence that such a system is feasible but the cost of electronic vehicle identification and collection is expected to be substantial. Some claim that charging congestion tolls is not equitable because "poor working people" pay high congestion tolls and downtown businesses lose out. It is possible to compensate these losers by subsidizing urban transit services from the congestion toll revenues. Once the way in which the toll revenue is spent is taken into account, the actual and perceived inequity of congestion tolls may disappear. The introduction of congestion tolls or peak-load charges should not be undertaken in isolation with the expectation they will provide the solution to the pricing and investment problems. Pricing is one tool of the transportation planner and policy maker. To be effective, it must be used in conjunction with other strategies, such as substitute mode investment. To gain public support for pricing, it is essential that it not be perceived as simply another way of raising revenue for government nor that it is exploiting a situation in which people have no alternatives. For these reasons it is important that alternatives be available for those facing higher fees.

Toll Roads

If they implement peak-period pricing, toll roads have essentially the same characteristics as implementing congestion tolls on the existing roads. Constructing toll roads allows private-sector participation for large project financing. As collection technology improves, toll collection costs become quite moderate. New toll roads benefit not only those who switch to the toll roads but also those who continue to use the old non-toll roads (due to reduced travel time). This is a Pareto improvement. Construction of toll roads can be considered an equitable method since lower-income users gain by travelling on non-toll facilities. In reality, toll roads can solve financing problems for specific projects only.

Annual Licence Fees

These fees can be used to contribute to road costs which do not vary with traffic volume in the short run. If they are not coupled with a variable charge, as with a two-part tariff, which reflects social costs, annual licence fees will have no direct effect on the efficiency of road investment. As shown in many countries, it can generate large sums of money with relatively low administrative and collection costs. Since some people cannot afford to keep a car when annual licence fees are raised substantially, a high licence fee policy may not be regarded as equitable. One method used in some

U.S. states is to base the fee on the purchase price of the car and have the fee decline each successive year to some floor level where the floor would be the same for everyone. A system of licence fees can be implemented by national, provincial or municipal governments.

Charging Beneficiaries

This one-time charge to beneficiaries (developers, property owners and businesses in a special assessment district) is likely to promote an efficient level of road capacity investment, but has no effect on the efficiency of road usage (use of road capacity or durability) after construction. This is a good source of funds for a well-defined road project but it is difficult to determine the boundaries of a special assessment district, as well as the amount of special levies. Charging beneficiaries is by definition equitable. This method of financing works only on a project basis.

5.4.2 Evaluation of Alternative Sets of Financing Instruments

Government must temper principle with practicality and use a mix of two or more charging methods to improve economic efficiency in the road transportation sector while simultaneously attaining a cost-recovery target. Furthermore, perceived equity cannot be ignored in the selection of the menu of instruments. The following alternative sets of financing instruments are evaluated and compared:

- fuel taxes and licence fees (status quo);
- fuel taxes, licence fees and congestion tolls;
- graduated per-kilometre taxes, congestion tolls and licence fees; and
- social marginal cost pricing (graduated per-kilometre taxes based on axle weights, congestion tolls and other externality taxes).

Fuel Taxes and Licence Fees

Canada, as well as most other countries, currently use fuel taxes and licence fees as the major source of funds for financing road infrastructure. These two instruments together can generate total revenue which is sufficient to finance all of the road infrastructure. They are also administratively easy to collect. Licence fees, if appropriately set, can be used to recover some or all of the short-term fixed cost including the portion of road repair and repayment costs which are unrelated to road usage.

Fuel taxes do not give road users, most importantly heavy trucks and buses, the incentive to choose the optimal axle load which minimizes the combined user and infrastructure costs. This happens because commercial carriers try to reduce fuel consumption which tends to increase with the number of axles, while damage to the pavement increases exponentially with average weight per axle. A second flaw in the use of fuel taxes for pricing urban and near-urban roads is the low correlation between congestion costs and fuel use, hence fuel tax paid. The fuel tax is, therefore, a poor proxy for a congestion tax and would lack any capability for allocating peak-load capacity. Overall, the combination of fuel taxes and licence fees is not an economically efficient method of charging and financing roads.

Fuel Taxes, Licence Fees and Congestion Tolls (Urban and Near-Urban Roads)

The congestion toll component of this package is likely to improve the efficiency of allocating peak-period capacity, and also generate pressure to invest optimally in capacity. However, this package lacks incentives for commercial operators to use optimal axle-load weight.

Graduated Per-Kilometre Taxes, Congestion Tolls and Licence Fees

This combination of fees may be regarded as a transition package en route to a more socially efficient pricing scheme. This study has already addressed the value of fuel taxes and licence fees as providing a practical and effective means of pricing and financing capacity investment for intercity roads on which there is little or no congestion. The weak link is the failure to adequately price truck damage and provide incentives for operators to employ economically efficient axle loads. The fuel tax should therefore be restricted to cars and the scheme supplemented by a charge for trucks.

A graduated per-kilometre axle fee should supplement or replace current truck licence fees and fuel taxes levied on truck use. This will not only efficiently and adequately allocate cost responsibility between cars, trucks and buses but also among the various types of trucks.

Licence fees which are levied will reflect, not necessarily on a one-to-one basis, the magnitude of fixed costs. Licence fees must be allocated both between and within user groups. This means that for fairness and efficiency, trucks, cars and buses should bear a portion of the fixed costs of the roadway system — how much will depend on cost responsibility, the elasticity of access with respect to licence fees and the elasticity of usage with respect to licence fee. User-group allocation would be based on some Ramsey-type approach; for example, car users may have a licence fee based on the purchase price of the vehicle — an approach used in some U.S. states. This study does not advocate the position that the contribution by trucks and buses must be contained in a licence fee. Indeed, it may well be that the graduated per-kilometre usage fee will be the only fee levied and contain both usage and fixed contributions.

For urban and near-urban roads this pricing approach will prove to be neither equitable nor efficient. Some method of congestion pricing must be introduced to efficiently allocate existing capacity and provide for efficient investment in future capacity. Fairness, however, will be determined by the expenditure of the congestion tax revenue and taking a comprehensive approach. That is, it serves no one to focus entirely on one instrument and rely on it to provide a solution. A comprehensive approach is necessary, in which substitutes are provided for those faced with a congestion or peak-load fee. These substitutes can range from flexibility of work hours to providing public transit (of some form).

Social Marginal Cost Pricing

The graduated per-kilometre taxes based on axle weights reflect the cost of road damage users inflict, while congestion tolls and other externality (noise and air pollution) taxes reflect the costs users impose on the community. Together they constitute the social marginal cost of using the road system, and charging the social marginal costs maximizes social welfare. Although the results of Small, Winston and Evans (1989) suggest that this charging scheme would result in an approximate financial break even, based on the assumption of constant returns to scale in joint production of road capacity and durability, other evidence such as Nix (1989), Nix, Boucher and Hutchinson (1992), Kraus (1982) and Oum and Zhang (1990) suggests that break-even performance is not guaranteed. In Canada where the majority of the non-urban roads are of low density (and capacity) and the indivisibility of capacity construction plays some role, the application of social marginal cost pricing is likely to produce a financial deficit in many instances.

As we have also discussed, there are also some practical and administrative difficulties in implementing the graduated per-kilometre charges based on axle weight, and the externality taxes (congestion tolls, noise and air pollution tax). Certainly, it will result in higher costs for administration and collection than the current fuel tax and licence fee combination. However, this is the correct direction in which to move for maximizing economic welfare from the road sector. As the technology of automatic vehicle identification develops and the cost of information processing decreases, the administration and collection is likely to become less burdensome than now.

6. Summary and Conclusions

6.1 PURPOSE AND ORGANIZATION OF THE STUDY

Transportation policy has undergone significant changes over the last 50 years. Canada has moved from a situation of government ownership, management and regulation in all or parts of every mode to a position of deregulation, privatization and a generally greater reliance on market forces. The major transformation has taken place among the carriers while more recent initiatives have been focussing on infrastructure — roads, airports, rail track and terminals.

Much of the stimulus for change has been the result of Canada's need to rethink traditional policies and reposition itself to remain competitive in the North American and world marketplaces. Pricing and investment planning of Canada's transportation infrastructure cannot ignore the forces and pressures developing in the United States and other international markets. Links must be established between investment in infrastructure and the pricing of services delivered by that infrastructure. Indeed, socially optimal modal pricing requires the inclusion of modal air and noise pollution and congestion externalities. Economic welfare will be lower if these factors are not considered in modal prices because demands for massive public investments in infrastructure will continue unabated, infrastructure will deteriorate prematurely, and the distribution of traffic across modes will not reflect the real costs of these modes. Solutions do not necessarily lie with more investment but rather with smarter investment. Smarter investment must start with efficient pricing.

In May 1992, the Congressional Budget Office of the United States released a study, *Paying for Highways, Airways and Waterways: How Can Users be Charged?* The study is in complete congruity with the conclusions stated here. They state in their summary:

> The methods of financing highways, airways, and waterways influence both the amount of revenue that can be raised and the efficient allocation of resources. The concept of revenue adequacy — whether revenues cover costs — is important to the cash-strapped federal government, but it also has implications for efficient allocation of resources in the long run. If the costs of an investment project cannot be recovered from those who use it, the project's feasibility comes into question. But an investment that benefits society is worth making, even though it may not be possible to charge users for it. This often characterizes goods and services provided by the federal government, and it underlies the rationale for government rather than private activity in certain sectors. Revenue adequacy can provide information about the demand by users for public investments, but it alone cannot be the criterion upon which investment decisions are made.
>
> Economic efficiency is the second criterion by which financing mechanisms are evaluated. The standard definition of allocative efficiency is used here: does the price — the value consumers place on the product or service at the margin — equal the marginal cost — that is, the value of resources used in producing the last unit? If the price is less than the marginal cost, consumers tend to overuse the resource; if the price exceeds the marginal cost, they use it too little.
>
> The objectives of revenue adequacy and economic efficiency sometimes conflict. Economic theory offers some ways of minimizing the trade-offs, and these are included in the discussions of alternative pricing mechanisms. (p. xi)

The main purpose of this study is to discuss the principles and methods by which Canada can ensure the optimal use and efficient provision of transportation infrastructure services. Also, to reflect the move to greater fiscal responsibility reflected in a shift from public- to private-sector provision of services, increasing constraints on government's ability to control resources and a general shift to a greater emphasis on efficiency, an emphasis is placed on the cost-recovery issue. The issue is treated in two alternative ways. First, the cost-recovery conditions associated with optimal pricing and investment in infrastructure are studied and compared with the actual cost-recovery situation. Second, the methods of achieving an exogenously given cost-recovery target in such a way as to minimize the efficiency loss are discussed. Although the principles and methods discussed in the study can be applied to all modes of transportation, the emphasis is on roads and airport infrastructure.

6.2 INFRASTRUCTURE PRICING PRINCIPLES

Pricing is a method of allocating resources. There is no such thing as the right price irrespective of issues and objectives. Rather there are optimal prices or pricing strategies given particular objectives to be achieved. Prices can be established to maximize profit, welfare or revenues. They can be used to achieve a particular market share or a desired distribution of demand across products (for example, mode-split in transportation). However, one of the most important goals for pricing goods and services from society's viewpoint is in maximizing economic welfare by optimally allocating scarce resources and goods and services across competing needs in the short run and ensuring optimal investment in capacity in the long run.

It is well known that economic welfare is maximized by pricing infrastructure at marginal social cost which means the externality costs users impose on the system are included. Critics of the current pricing approach agree that governments have been misdirected in their infrastructure policy chiefly in trying to respond to traffic growth by expanding capacity rather than also using demand management through an efficient pricing system. Empirical studies indicate that the net social benefit from effective demand management policy is quite high. The major source of the available efficiency gains for both near-urban roads and airports is the reduction in congestion delays with a consequent savings in operating and time costs. For intercity facilities, efficient pricing leads to a more efficient use of facilities within and across modes.

Efficient user-charges which reflect social costs have three significant impacts. First, economic welfare is increased as the demand for infrastructure is rationed more efficiently in a way which reflects the costs of using the infrastructure. Second, the distortion in traffic allocation across modes is reduced. Third, it contributes to the financing of infrastructure. Furthermore, marginal social cost pricing is not incompatible with fairness or equity. The surplus from congestion tolls may be used to offset losses by some groups who may be adversely affected by increased user-fees. The British Airports Authority (BAA Plc.), for example, has used marginal cost pricing principles to establish runway and terminal charges for the five airports under its control but has redistributed revenues from one airport to another particularly for investment purposes.

Peak-load pricing is an important variant of marginal social cost pricing which formally considers the fact that some users demand and require more capacity than others. This is seen everyday on roadways, runways and in terminals. There is a clear and sound economic efficiency basis for peak-load pricing. Charging higher prices to peak users enhances economic efficiency by inducing them to make rational choice decisions as well as helping to solve financing problems for capacity expansion. Finally, peak pricing is not, in principle, unfair or inequitable. It assigns costs to those who are responsible for them. It makes no economic sense to restrict the use of a facility in off-peak hours because all that results in is the underutilization of an existing facility. In competitive circumstances or under regulation, peak users are not necessarily worse off in terms of what they pay, if there is a lower off-peak price, provided the off-peak users pay at least their variable costs. If off-peak users were eliminated from the market, peak users would pay precisely the same amount as when off-peak users pay their variable costs. The only circumstance when peak users would be better off is if they are cross subsidized from off-peak users. This creates a social loss in terms of both inefficient use of facilities and the loss of consumer surplus.

When the total revenue from the social marginal cost pricing is not sufficient to cover total cost, there are three options open to the infrastructure authority:

- continue to use social marginal cost pricing with subsidization from the general tax revenue;
- use Ramsey pricing which provides a second-best solution with a break-even financial performance; or
- adopt a two-part pricing (access/usage tariffs) scheme.

The question of financing capacity from users versus the general revenue fund (all taxpayers) is an issue of both efficiency and equity. If marginal cost is less than average cost, efficient pricing will result in a deficit. This deficit must be covered. A partial equilibrium approach to this problem would support a policy of funding the deficit out of general revenue while a general equilibrium approach would argue that the relative welfare losses of funding the deficit by users versus the general taxpayer should be considered. Recent evidence, Jorgenson (1992), Jorgenson and Yun (1990) and Ballard, Shoven and Whalley (1985), has shown that the marginal cost of public funds is between $1.33 and $1.45, meaning that between 33¢ and 45¢ per dollar are lost through a private-sector efficiency loss. It is, therefore, not obvious that proposing a policy of marginal social cost pricing and ignoring the costs of public funds to finance the deficit would necessarily improve economic welfare. A second-best pricing scheme in which there is a cost recovery constraint may lead to a higher level of economic efficiency.

Ramsey pricing minimizes the loss of economic efficiency caused by deviating prices from the respective marginal costs in order to achieve financial break even. In effect, it charges higher markup to less price-elastic products or market segments by making the markup inversely proportional to the price elasticity of the demand.

Two-part pricing consists of a flat fee for the right to access a facility (for example, vehicle licence fee to access the road system), and a usage fee (for example, charge per kilometre of road usage). Two-part pricing can lead to a first-best solution if the demand for access is not price sensitive, and if usage charges are set at the marginal costs of usage and the access fee is set at a sufficiently high level to allow the firm to break even. Under these conditions, the regulation of the access fee alone can induce a monopoly firm to charge marginal cost as the usage price. When access demand is sensitive to price, the optimal two-part tariff can be computed by applying the Ramsey pricing rule to the access and the usage demands as if they are two separate products with interrelated demands. This then becomes a second-best pricing approach.

The economics literature is paying increasing attention to the consequences of selective pricing methods for various groups of users and non-users; that is, the income distributional consequences of a pricing method. Problems arise when a pricing method which maximizes economic efficiency does not necessarily yield what is considered a fair or equitable outcome. The most obvious case where concern over income distribution appears critical is the implementation of congestion (or peak-load) pricing for (urban) transportation. The results of a number of empirical and simulation studies vary somewhat, but the basic conclusion is that it is misleading to characterize a congestion or peak-load pricing policy as being regressive.

Marginal social cost pricing is also not incompatible with fairness or equity. Congestion or peak-load pricing in conjunction with a strategy to use the revenues can generate net positive benefits for society. Marginal social cost pricing corrects distortions rather than introduces them. Some groups, however, are made better off and others worse off but this should not justify rejection of this pricing policy. As Hau (1991) notes, it is perhaps asking too much of a pricing mechanism to solve the pricing, investment and income distribution problem. One perspective is to use road and airport funds from efficient pricing to both invest efficiently in capacity (new roads or runways) and to satisfy concerns regarding equity. This study recognizes that while income redistribution issues are better placed in general tax policy, some of the concerns with the perceived inequity of efficient pricing may be allayed if some of the funds from efficient pricing are used to provide substitutes for those most affected by the use of socially efficient pricing. Thus, a transportation fund could potentially invest in public transportation, rural roads or small airports. The magnitude of the fund would depend on scale economies and indivisibilities.

6.3 CARRIER AND INFRASTRUCTURE COSTS

The infrastructure planner must establish user-charges and make capacity investment decisions to maximize the economic welfare of society. Understanding the behaviour of the combined cost of a carrier's service provision and infrastructure provision is essential for the development of a set of socially optimal prices for infrastructure. If short-run costs fall because of increased capacity utilization but long-run costs exhibit constant returns to scale, it is still possible to have marginal social cost pricing and fully cover costs. If, however, long-run costs are characterized by some economies, a second-best pricing approach will be required to have total revenues cover costs and minimize the efficiency loss of deviating from first-best pricing. A prerequisite to understanding the structure of the combined cost is to understand each component, that is, carrier cost structure and infrastructure cost structure.

A number of studies have been directed at determining the behaviour of an airline's cost function with respect to changes in the level and composition of output. The studies have shown that the long-run average cost curve is relatively constant over a wide range of output, that is, there are no economies of scale in the airline industry. This means that the size of a carrier does not generate lower per-unit costs. Studies also concluded that airport capacity construction is as well characterized by constant returns to scale. This implies that the combined cost of carriers and infrastructure is also characterized by constant returns to scale.

Empirical investigations of the truck and intercity bus industries have provided mixed results. The overwhelming evidence is that there are constant returns to scale in trucking, some scope economies and certainly density economies. The limited investigations of the bus industry found weak evidence of non-constant returns. There is no direct evidence of density economies but observing the restructuring of the bus industry in those countries in which there has been deregulation certainly suggests some density economies. Evidence from Small, Winston and Evans (1989) showed constant returns to scale in highway capacity construction. They reported the existence of significant economies of scale with respect to the durability of road and mild returns to scale with respect to traffic volume. They also reported diseconomies of scope from the production of both durability and traffic volume because, as the road is made wider to accommodate more traffic, the cost of any additional thickness rises, since all the lanes must be built to the same standard of thickness.

The final outcome of these three factors at work is that highway capacity construction is characterized by approximately constant returns to scale. In other words, the output-specific economies of scale are offset by the diseconomies of scope for having to produce them jointly. Since they included both infrastructure costs and the costs incurred by road users (individual drivers and transportation carriers) in the total cost of highway modes, their result on the overall constant returns to scale is for the combined cost of highways and users.

6.4 IMPLEMENTATION OF USER-CHARGES

This review of the current charging schemes for roadways and airports and of efficient pricing has led to the following conclusions. In the air sector the uniform user-charge system should be repealed in favour of site-specific user-charges. Furthermore, the current weight-based landing fees are too low for small aircraft (mainly general aviation and corporate aircraft) and too high for large aircraft. This had led to a situation of undercharging the former relative to commercial aircraft (mainly large jet).

Peak-period user-charges should be implemented at airports with congestion problems, and the differential user-charges between peak, low-peak, and off-peak periods can be set to reflect their respective social marginal costs which include congestion externality costs. This leads to a large differential between peak and off-peak fees. Under this system, landing fees for small aircraft become similar to those for large aircraft in the peak periods. It is clear that landing fees for small aircraft at congested airports are even more underpriced. Moving in this direction is highly consistent with the pricing provisions contained in the Transport Canada cost-recovery proposals published in 1990.

A weight-based landing fee, if appropriately set, may be consistent with the Ramsey pricing principle and, thus, have some economic justification for use at small uncongested airports. This is because the demand for airport services by larger (heavier) aircraft is more price-inelastic than smaller (lighter) aircraft. However, the rationale for weight-based pricing breaks down when the airport becomes congested and congestion pricing is applied. Similarly, charging higher landing fees for long-haul international flights at uncongested airports can be justified in an efficient pricing regime since it is consistent with the spirit of Ramsey pricing as the demand for airport services by long-haul flights is less price-elastic than short-haul flights. However, charging higher landing fees for long-haul international flights loses its economic rationale when congestion pricing is applied to congested airports.

In establishing prices for roadways, the empirical studies conducted in Canada and the United States indicate passenger cars and other light vehicles pay a disproportionately higher share of the total road maintenance costs as compared to trucks and other heavy vehicles. The Canadian studies show that the short-run fixed road cost far exceeds the user-charge revenues not related to road usage. This implies that vehicle licence fees may be too low to be optimal, and those who underutilize their vehicles pay less for their option to use the road system than the optimal price for the option to use their vehicle.

Road costs are sensitive to the investment level in capacity and durability. It is therefore important that the road authority make an optimal investment particularly in durability (pavement thickness). In the United States, the underinvestment in durability (that is, building sub-standard roads) has not only substantially increased road maintenance costs more than necessary, but also the costs to the road users.

Currently, the majority of the road user-charges are collected in the form of fuel taxes. Since fuel taxes are not directly related to either congestion or road damage, they are not an efficient means of charging for road usage. Road charges must be related to both road damage and the amount of congestion. Technology now exists for effectively administering congestion tolls either by time of day or by congested road segment. Therefore, the users of urban roads and near-urban highways must be charged both congestion tolls and road damage while users of rural and uncongested intercity roads are charged just for the road damage.

6.5 ALTERNATIVE SOURCES OF FINANCING ROAD INFRASTRUCTURE

In practice, a government must use a mix of two or more charging methods to achieve high efficiency in the road transportation sector while attaining a cost recovery target. The road pricing scheme recommended by this study consists of three parts:

- a system of graduated per-kilometre fees for trucks based on axle weight which will improve efficiency in the use of roads and promote an optimal investment in the durability (thickness) of the road system;
- congestion tolls and environmental externality taxes which will help improve the efficiency of the usage of road capacity for urban and near-urban roads as well as achieve an optimal investment in road capacity (lanes); and
- a fuel tax–licence fee combination for cars on uncongested roads.

The relative mix of fixed and variable fees will depend on the relative values of access and usage price elasticities. It will generally be true that a greater proportion of revenues will be generated from the variable charge.

The first two are essentially the social marginal costs of road usage (that is, charging for road damage and externality costs). Construction of toll roads or charging beneficiaries are essentially local financing solutions for specific road construction or improvement projects. Since empirical evidence suggests that a substantial portion of road costs do not vary with road usage (traffic volume or axle loads), use of substantial licence/registration fees is likely to be an efficient means of recovering the fixed road costs if the number of users is not affected by the magnitude of the licence fee. Methodologically, a two-part pricing system can be applied on a national or provincial scale to charge vehicle licence fees for the right to access the road system and social marginal costs (road damage and externality costs) for the road usage. This two-part pricing system is likely to self-finance the Canadian road system.

This study recognizes that motive fuel taxes have been deeply entrenched in society because it is administratively easy to collect large sums of money from road users but also because it helps achieve other important objectives such as environmental and energy conservation goals. However, when planning for 20 or 30 years into the future, the user-charge plans must include economically rational charging methods such as peak-period and peak-district pricing and graduated per-kilometre charges based on axle loadings if there is to be an efficient system.

ENDNOTES

The authors acknowledge the intellectual contribution of Yimin Zhang, and the research assistance of Eva Lau and Nagy Eltony. We are indebted to Richard Arnott, Sandy Borins and particularly John Sargent for helpful comments and suggestions on substance and exposition. We, however, are responsible for the contents of the document.

1. For the case of a single output with no possibility of effective market segmentation, Ramsey pricing reduces to average cost pricing.

2. Oum and Tretheway (1988) have extended the Ramsey pricing rule to the case where externality costs are present. Their results show that the markup is a weighted average of the inverse elasticity and the ratio of marginal external cost to price.

3. Essentially the same method as Ramsey pricing can be applied to find quasi-optimal prices to achieve a given level of cost recovery (including break even). See Gillen, Oum and Tretheway (1986) for use of the Ramsey framework to compute optimal landing fees subject to various levels of cost recovery for runway services. In the past, some transport economists named it "value-of-service" pricing when firms (airlines and railroads) charged differential (often profit-maximizing) prices for essentially the same services by using differential price elasticities by market segment.

4. The following discussion is in the context of two-part prices but the concepts are not limited to two parts and can apply equally well to multi-part prices.

5. There are a number of examples of two-part prices in the economy: pricing telephone services, golf and other club memberships, amusement parks, roadway access and usage.

6. Train (1991) notes that too high an access fee may result in hardship for lower income groups. If this is perceived as being too inequitable, the efficiency gains may be foregone.

7. See *Aviation Week and Space Technology*, April 27, 1992.

8. Ng and Weisser (1974) have shown that the proportion of the "residual revenue" (the difference between revenue required to break even and the revenue obtained from pure marginal cost pricing) financed by increasing the variable fee above marginal cost, increases with the absolute price elasticity of the number of users and decreases with the absolute price elasticity of the amount of usage.

9. See Ng and Weisser (1974), Spulber (1989) and Train (1991) for further discussions on theoretical and technical issues on multi-part pricing, and its relationship with "block tariffs."

10. Investment to increase slots does not always require a new runway. Construction of a high-speed turnoff, for example, can decrease the amount of time a particular plane occupies a runway.

11. In fact, the landing fee alone is identical in the peak periods for all aircraft types (see Gillen, Oum and Tretheway, 1988).

12. There are numerous studies of urban road pricing and its consequences but they are not reviewed here since the interest is in interurban infrastructure.

13. In all cases the evaluations or comparisons are made against a status quo which is essentially average cost pricing. Thus, the comparison is between average costs and social marginal cost pricing for congested facilities and between average cost and Ramsey pricing for uncongested facilities.

14. Small, Winston and Evans (1989) also developed a set of efficient road prices. These are discussed at greater length later in the section dealing with the question of cost recovery.

15. This was changed in 1991.

16. Since the total disbursements on roads in the United States was $61 billion in 1985 (Small, Winston and Evans, 1989), this means a savings of over 13 percent of the total cost.

17. A recent study in Minneapolis–St. Paul found that the addition of one runway to the international airport would generate such large savings in direct cost to carriers that it (the runway) would have a payback of only three years.

18. Underpricing infrastructure not only leads to excessive use but oversupplying infrastructure leads to long-term structural shifts which almost guarantees today's problems will recur in the future.

19. See Keeler and Small (1977) and Winston (1985) for examples of modelling optimal user charges and investment decisions for infrastructure in such a way as to maximize total benefits for users (social welfare).

20. In the case of a multi-product firm, outputs would be increased in the same proportion or along an output ray.

21. As a corollary, if there are no scale economies, it implies that first-best pricing with no deficit is achievable over the long run.

22. The exception is Air Canada which Gillen et al. (1985) found to have realized most of the available density economies in its network. This may have changed since the empirical evidence is based on data up to 1981 only.

23. Since Canada has a number of small airports in remote communities, indivisibility of capacity at small airports may translate into scale economies.

24. This is one of the significant problems associated with system average costing. It assumes there are not significant differences in costs and demand between facilities.

25. The numbers are also based on costs at the time the study was conducted. The numbers calculated would be as high or lower since lower costs are expected with the move to defederalization (see Hamilton, 1991).

26. Airport planning and construction can in many cases take more than a decade so the characteristics of commercial aircraft soon to be produced are considered in the design.

27. Table 4.1 also shows differential prices by market segment: domestic and international flights. This may be consistent with the spirit of the Ramsey pricing principle considering that most international flights have longer stage lengths and, thus, lower price elasticities of demand for airport services.

28. GA aircraft did pay an airport-specific tax to cover the costs of airport operations but this was not an explicit fee for runway use.

29. As an example of this type of charging structure, New Zealand which has privatized its air traffic control system, charges general aviation aircraft $57 per year (plus sales tax) for the first 50 landings and $3.67 or $4.60 for each subsequent landing depending on airport location (see Paul Proctor, "For Profit New Zealand ATC System Cuts Costs and Increases Efficiency," *Aviation Week and Space Technology*, April 27, 1992).

30. Many of these changes are consistent with a recent Transport Canada pricing policy proposal (Transport Canada, 1990).

31. For example, at Heathrow airport all aircraft, small or large, pay identical landing fees during peak periods. In addition, the combination of landing fees and passenger terminal fees for a large aircraft (for example, Boeing 747) in the peak periods exceeds, by a factor of five, the amount for the same aircraft at off-peak periods. See Gillen, Oum and Tretheway (1988) for the exact fee differentials.

32. Nix allocated costs under two scenarios: with scenario A all capital costs are treated as inescapable (assumed not to vary with vehicle usage). With scenario B one third of the capital costs are assumed to be escapable. Among the capital costs, the pavement costs are allocated to various axle-weight groups, while the road maintenance costs (and one third of the pavement costs in scenario B) were allocated on the basis of vehicle usage. This cost allocation exercise was performed on the following vehicle categories: standard car, three-axle truck, five-axle tractor-semitrailer, six-axle tractor-semitrailer, eight-axle B-train loaded and eight-axle B-train empty.

33. Heggie (1991) categorized the charging instruments into four categories: vehicle usage, vehicle ownership, vehicle acquisition and charging beneficiaries of road system.

34. One exception to this is the case of Japan where substantial revenue from road tolls is generated. Toll revenues accounted for nearly 20 percent of the total revenue collected from road users in 1985. In fact, all three publicly owned road authorities generate financial surpluses after covering capital expansion costs.

35. Small, Winston and Evans (1989) proposed the use of both a steeply graduated per-mile tax based on axle weights and congestion tolls as the main feature of road pricing (pp. 114-19).

36. The Hong Kong experiment is interesting, in part because it is sometimes cited as an example of the failure of congestion charges. Hong Kong did not adopt the congestion toll system after the experiment carried out from 1983 to 1985 primarily because of the complex political factors including people's fears of government intrusion and the desire to exercise the newly won local autonomy from the British governor. However, the test of collection technology in Hong Kong was a resounding success, exceeding by a wide margin the very stringent goals for reliability and ease of use that were established. More than 99.7 percent of vehicles crossing toll sites were correctly identified, and the wrong vehicle was charged less than one time in 10 million.

37. This one-time registration fee may have a small short-run negative effect on the environment since it would discourage the purchase of new vehicles and encourage continued use of older ones. Therefore, it may be desirable to raise annual licence fees substantially without changing the first-time vehicle registration fees.

REFERENCES

Abouchar, A. 1970. "Air Transport Demand Congestion Costs and the Theory of Optimal Airport Use." *Canadian Journal of Economics* 3, pp. 463–75.

Action, J. P. and I. Vogelsang. 1989. "Symposium on Price-Cap Regulation." *Rand Journal of Economics* 20, no. 3, Autumn, pp. 368–472.

Aigner, D. and L. Lillard. 1982. "Measuring Peak-Load Pricing Response for Experimental Data." Working Paper 8224, Social System Research Institute, University of Wisconsin, Madison.

Allen, T. M. and Charles F. Floyd. 1991. "Alternative Financing Techniques in Funding Major Highway Reconstruction Projects." *Transportation Quarterly* 45 (3), July, pp. 357–68.

Anderson, F. J. and N. C. Bonsor. 1974. "Pricing and Valuation of Transportation Facilities in the Presence of Congestion." *Economica,* November, pp. 424–31.

Armstrong-Wright, A. T. 1986. "Road Pricing and User Restraint: Opportunities in Developing Countries." *Transportation Research* 20A, pp. 123–27.

Arnott, R. and J. Stiglitz. 1989. "Congestion Pricing to Improve Air Travel Safety." In *Transportation Safety in an Age of Deregulation.* Edited by Leon Moses and Ian Savage. Oxford University Press.

Arnott, R., Andre de Palma and R. Lindsey. 1990. "Economics of a Bottleneck." *Journal of Urban Economics* 27, pp. 111–30.

Ballard, C., J. Shoven and J. Whalley. 1985. "General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States." *American Economic Review* 75, pp. 128–39.

Baumol, A. W. and D. Bradford. 1970. "Optimal Departures from Marginal Cost Pricing." *American Economic Review* 60, pp. 265–83.

Bisson, B. G. and E. Hildebrand. 1989. "Highway Finance — A Literature Review." Mimeo. Fredericton: The University of New Brunswick Transportation Group.

Boardman, A. E. and Lester Lave. 1977. "Highway Congestion and Congestion Load Pricing." In *Marginal Cost Pricing in Practice.* Edited by James R. Nelson. Englewood Cliffs, N.J.: Prentice Hall.

Boiteux, M. 1960. "Peak Load Pricing," *Journal of Business* 33, April, pp. 157–59.

Bonsor, N. 1984. *Transportation Economics: Theory and Canadian Policy.* Toronto: Butterworths Publishers.

Borins, S. 1978. "Pricing and Investment in a Transportation Network: The Case of Toronto Airport." *Canadian Journal of Economics* 11, no. 4.

— 1982. "The Economic Effects of Non-Optimal Pricing and Investment Policies for Transportation Facilities." *Transportation Research* 16B, no. 1, pp. 17–29.

— 1984. The Economics of Non-Optimal Pricing and Investment Policies for Substitutable Transportation Facilities." *Canadian Journal of Economics* 17.

Braeutigam, R. B. 1979. "Optimal Pricing with Intermodal Competition." *American Economic Review* 69, no. 1, March, pp. 38–49.

— 1985. "Efficient Pricing with Rivalry between a Railway and a Pipeline." In *Analytical Studies in Transport Economics.* Edited by Andrew F. Daughety. New York: Cambridge University Press.

Braid, R. M. 1989. "Uniform versus Peak-Load Pricing of a Bottleneck with Elastic Demand." *Journal of Urban Economics* 26, pp. 320–27.

Bryan, I. A. and Y. Kotowitz. 1978. *Shipping Conferences in Canada.* Consumer and Corporate Affairs Canada, Research Monograph number 2, Research Branch, Bureau of Competition Policy. Ottawa: Supply and Services Canada.

Butcher, E. W. A. 1990. "Inter-State Commission Proposed National Road Pricing and Funding Scheme: A Review." *Australia General Public Service.* July, Canberra.

Button, K. J. 1982. *Transport Economics.* London: Gower Publishing Co.

Canadian Tax Foundation. *Provincial and Municipal Finances.* Various issues, Toronto.

— *The National Finances 1990.* Toronto.

Carlin, A. and R. Park. 1970. "Marginal Cost Pricing of Airport Runway Capacity." *American Economic Review* 60, pp. 310–19.

Carll, R. 1975. "Some Observations on Urban Transport Pricing and its Relationship to Comprehensive Transportation Planning." *Transportation Journal* 14, no. 3 (Spring), pp. 18–30.

Catling, Ian and B. Harbord. 1985. "Electronic Road Pricing in Hong Kong: The Technology." *Traffic Engineering and Control* 26, pp. 608–15.

Caves, D., L. R. Christensen and J. Swanson. 1981. "Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1955–1974." *American Economic Review* 71 (December), pp. 994–1002.

Caves, D., L. R. Christensen and M. W. Tretheway. 1980. "Flexible Cost Functions for Multi-product Firm." *Review of Economics and Statistics,* August, pp. 477–81.

— 1984. "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ." *Rand Journal of Economics* 15, no. 4, Winter, pp. 471–89.

Caves, D., L. R. Christensen, M. W. Tretheway and R. Windle. 1985. "Network Effects and the Measurement of Returns to Scale and Density for U.S. Railroads." In *Analytical Studies in Transport Economics.* Edited by Andrew F. Daughety. New York: Cambridge University Press.

Charney, A., N. Sidhu and J. Due. 1977. "Short Run Cost Functions for Class II Railroads." *Logistics and Transportation Review* 17, pp. 345–59.

Clark, J. M. 1923. *Studies in the Economics of Overhead Costs.* Chicago: University of Chicago Press.

Cohen, Y. 1987. "Commuter Welfare Under Peak-Period Congestion Tolls: Who Gains and Who Loses." *International Journal of Transport Economics* 14, pp. 239–66.

Cox, W. and J. Love. 1991. "International Experience in Competitive Tendering." A Paper Presented at the 3rd International Conference on Privatization and Deregulation in Passenger Transportation. Tampere, Finland, June 16–20, 1991.

Damus, S. 1981. "Two-part Tariffs and Optimal Taxation: the Case of Railroad Rates." *American Economic Review* 71, no. 1 (March), pp. 65–79.

— 1984. "Ramsey Pricing by U.S. Railroads." *Journal of Transport Economics and Policy,* January, pp. 51–61.

Daughety, A. F. 1985. "Transportation Research on Pricing and Regulation: Overview and Suggestions for Future Research." *Transportation Research* 19A, no. 5, pp. 471–87.

Davies, J. E. 1989. "Cost Recovery Arrangements for Interstate Land Transport in Australia: A Review of Recent Policy Initiatives." *Economic Analysis and Policy,* March 19, no. 1, pp. 91–102.

Dawson J. A. L. and F. N. Brown. 1985. "Electronic Road Pricing in Hong Kong." *Traffic Engineering and Control* 26, no. 11, pp. 522–29.

de Borger, B. and W. Nonneman. 1981. "Statistical Cost Functions for Dry Bulk Carriers." *Journal of Transport Economics and Policy* 15, pp. 155–65.

De Palma, A. and Richard Arnott. 1986. "Usage-Dependent Peak-Load Pricing." *Economic Letters* 20, pp. 101–5.

DeVany, A. and T. R. Saving. 1980. "Competition and Highway Pricing for Stochastic Traffic." *Journal of Business* 53, pp. 45–60.

Dewees, D. 1979. "Estimating the Time Costs of Highway Congestion." *Econometrica* 47, pp. 1499–1512.

Dixit, A., J. Mirrlees and N. Stern. 1975. "Optimal Savings with Economies of Scale." *Review of Economic Studies* 42, pp. 131–46.

Douglas, G. and J. Miller. 1974. *Economic Regulation of Domestic Air Transport: Theory and Policy.* Washington, D.C.: Brookings Institution.

D'Ouville, E. L. and John MacDonald. 1990. "Optimal Road Capacity with a Suboptimal Congestion Toll." *Journal of Urban Economics* 28, pp. 34–49.

Dreze, Jacques. 1964. "Some Postwar Contributions of French Economists to Theory and Public Policy, with a Special Emphasis on Resource Allocation." *American Economic Review* 54, (Supplement).

Dupuit, J. 1849. "On Tolls and Transport Charges." Reprinted in *International Economic Papers,* 1962, pp. 7–31.

Foster, C. D. 1974. "The Regressiveness of Road Pricing." *International Journal of Transport Economics* 1, pp. 133–41.

— 1975. "A Note on the Distributional Effects of Road Pricing: A Comment." *Journal of Transport Economics and Policy* 9, pp. 186–88.

Friedlaender, A. F. 1969. *The Dilemma of Freight Transport Regulation.* Washington, D.C.: Brookings Institution.

Friedlaender, A. F. and R. H. Spady. 1981. *Freight Transport Regulation: Equity, Efficiency and Competition in the Rail and Trucking Industries.* Cambridge, Mass.: MIT Press.

Friedlaender, A. F. and S. Mathur. 1982. "Price Distortions, Financing Constraints and Second-Best Investment Rules in the Transportation Industries." *Journal of Public Economics* 18, pp. 195–215.

Friedlaender, A. F. and S. J. W. Chiang. 1984. "Output Aggregation, Network Effects and the Measurement of Trucking Technology." *The Review of Economics and Statistics* 66, no. 2, (May), pp. 267–76.

Friedlaender, A. F. and S. S. Bruce. 1985. "Augmentation Effects and Technical Change in the Regulated Trucking Industry 1974–79." In *Analytical Studies in Transport Economics.* Edited by A. F. Daughety, New York: Cambridge University Press.

Gillen, D. W. and T. H. Oum. 1984. "A Study of Cost Structures of The Canadian Intercity Motor Coach Industry." *Canadian Journal of Economics* 17, no. 2, (May), pp. 369–85.

Gillen, D. W., T. H. Oum, and M. W. Tretheway. 1985. *Airline Cost and Performance: Implications for Public and Industry Policies.* Vancouver, B.C.: University of British Columbia, Centre for Transportation Studies.

— 1986. *Pricing Policies for Canadian Airports with an Emphasis on Airfield Operations,* a report prepared for the Airports Task Force, Transport Canada, Ottawa.

— 1987. *Measurement of Social Costs: An Application to Toronto International Airport.* Vancouver, B.C.: University of British Columbia, Centre for Transportation Studies.

— 1988. *A Study of Peak Period Pricing with an Application to Toronto International Airport.* Report written for Airport Authority Group, Transport Canada, Ottawa.

— 1989. "Airport Pricing Principles: An Application to Canadian Airports." *Journal of the Transportation Research Forum* 29, no. 1, pp. 28–34.

— 1990a. "Airline Cost Structure and Policy Implications: A Multi-Product Approach for Canadian Airlines." *Journal of Transport Economics and Policy*, January, pp. 9–33.

— 1990b. "Airport Pricing and Capacity Expansion: Economic Evaluation of the Alternatives." In *Transportation in Canada.* Transport Canada TP10451-E. Ottawa: Transport Canada, Policy and Coordination, Economic Analysis, Chapter 12, pp. 83–93.

Glaister, Stephen. 1974. "Generalized Consumer Surplus and Public Transport Pricing." *Economic Journal* 84, pp. 849–67.

— 1981. *Fundamentals of Transport Economics.* New York: St. Martins Press.

Glaister, S. and D. Lewis. 1978. "An Integrated Fares Policy for Transport in London." *Journal of Public Economics* 9, pp. 341–55.

Gomez-Ibanez, J. A. and G. R. Fauth. 1980. "Downtown Auto Restraint Policies: The Costs and Benefits for Boston." *Journal of Transport Economics and Policy* 14, no. 2, pp. 133–53.

Goodwin, Phillip. 1987. "Long-term effects of public transport subsidy." In *Transport Subsidy.* Edited by S. Glaister. Bristol, U.K.: Policy Journals, pp. 121–27.

Goss, R. O. and C. Jones. 1977. "The Economies of Size in Dry Bulk Carriers." In *Advances in Maritime Economics.* Edited by R. O. Goss. Cambridge, U.K.: Cambridge University Press.

Government of Canada. *Public Accounts of Canada.* Various issues.

Griliches, Z. 1972. "Cost Allocation in Railroad Regulation." *Bell Journal of Economics and Management Science* 3, pp. 26–41.

Hamilton, Gordon B. 1991. "Cost Competitiveness of Canadian Airports," paper presented to Air Transport Association of Canada 57th Annual General Meeting, Vancouver, Canada, November 10–12.

Haritos, Z. 1973. *Rational Road Pricing Policies in Canada.* Ottawa: Canadian Transport Commission.

— 1975. "Transportation Infrastructure Costs and Revenues in Canada." *Journal of Transport Economics and Policy,* January, pp. 16–33.

Harker, P. T. 1985. "Research Directions in Transportation Regulation and Pricing." *Transportation Research* 19A, no. 5, pp. 489–91.

Harmatuck, D. J. 1979. "A Policy Sensitive Railway Cost Function." *Logistics and Transportation Review* 15 (May), pp. 277–315.

Harris, R. 1977. "Economics of Traffic Density in the Rail Freight Industry." *Bell Journal of Economics and Management Science* 8, pp. 556–64.

Harrison B. 1986. "Electronic Road Pricing in Hong Kong." *Traffic Engineering and Control* 27, no. 1, pp. 13–18.

Hau, Timothy D. 1990. "Developments in Transport Policy: Electronic Road Pricing Development in Hong Kong 1983–1989." *Journal of Transport Economics and Policy,* May, pp. 203–14.

— 1991. *An Economic Analysis of Road Pricing: A Diagrammatic Approach.* Washington D.C.: World Bank. Transportation Division, Infrastructure and Urban Development Department, Policy Research and External Affairs.

Heaver, Trevor. 1970. "Tanker Freight Rates in the Short and Long Run." *Journal of Applied Econometrics* 10, pp. 203–17.

Heaver, Trevor D. and T. Keast. 1974. "Transport subsidies in Canada." In *Transportation Subsidies — Nature and Extent.* Edited by K.M. Ruppenthal. Vancouver, B.C.: University of British Columbia, Centre for Transportation Studies, pp. 21–30.

Heggie, I. G. 1989. *Public Sector Transport Pricing and Allocative Efficiency in Developing Countries.* Washington, D. C.: World Bank. Policy Planning and Research Staff Discussion Paper.

— 1991. *User Charging and Accountability for Roads: An Agenda for Reform.* Washington D. C.: World Bank Infrastructure Division, Discussion Paper.

Hensher, David A. 1991. "Electronic Toll Collection." *Transportation Research* 25A, no. 1, pp. 9–16.

Higgins, T. 1979. "Road Pricing — Should and Might it Happen?" *Transportation* 8, pp. 99–113.

— 1986. "Road-Pricing Attempts in the United States." *Transportation Research* 20A, pp. 145–50.

Hoel, Lester, ed. 1990. "Financing Transportation Infrastructure." Special Issue, *Transportation Research* 24A, no. 4, July.

Holland E. P. and P. L. Watson. 1978. "Traffic Restraint in Singapore." *Traffic Engineering and Control* 19, no. 1, p. 15.

Hurley, W. 1989. *The Efficient Allocation of Runway Capacity.* Final Report. Transportation Development Centre Visiting Experts Program.

International Road Federation. 1990. *World Road Statistics, 1985–1989.* Geneva.

— 1991. *World Road Statistics, 1986–1990.* Geneva.

Interstate Commission. 1986. *An Investigation of Cost Recovery Arrangements For Interstate Land Transport.* Canberra: Australia General Public Service, April 1986.

— 1990. *Road Use Charges and Vehicle Registration: A National Scheme,* vol. 1. Canberra: Australia General Public Service, March 1990.

Jordan, W. J. 1983a. "Heterogeneous Users and the Peak-Load Pricing Model." *Quarterly Journal of Economics* 93, pp. 127–38.

— 1983b. "The Theory of Optimal Highway Pricing and Investment." *Southern Economic Journal* 50, pp. 560–64.

Jorgenson, Dale. 1992. "Fragile Statistical Foundations: The Macroeconomics of Public Infrastructure Investment." Mimeo. Harvard University, Department of Economics.

Jorgenson, D. and Kun-Young Yun. 1990. "The Excess Burden of Taxation in the U.S." Discussion Paper No. 1528. Cambridge, Mass.: Harvard Institute of Economic Research.

Kahn, A. 1970. *The Economics of Regulation,* vol. 1. New York: John Wiley & Sons.

Keeler, T. 1972. "Airline Regulation and Market Performance." *Bell Journal of Economics and Management Science* 3, pp. 399–424.

— 1974. "Railroad Costs: Returns to Scale and Excess Capacity." *Review of Economics and Statistics* 56, pp. 201–08.

Keeler, T. E., K. Small and Associates. 1975. *The Full Costs of Urban Transport, Part III: Automobile Costs and Final Intermodal Cost Comparisons,* Monograph No. 212. Berkeley: University of California, Institute of Urban and Regional Development.

Keeler, T. E. and K. Small. 1977. "Optimal Peak-Load Pricing, Investment and Service Levels on Urban Expressways." *Journal of Political Economy* 85, pp. 1–25.

Keeler, T. E., ed. 1985. *Research in Transportation Economics,* vol. 2. Greenwich, Connecticut: JAI Press Inc.

Knight, F. H. 1924. "Some Fallacies in the Interpretation of Social Cost." *Quarterly Journal of Economics* 38, pp. 582–606.

Koenker, R. 1977. "Optimal Scale and the Size Distribution of American Trucking Firms." *Journal of Transportation and Economic Policy* 11, pp. 54–67.

Kraus, Marvin. 1981. "Indivisibilities, Economies of Scale, and Optimal Subsidy Policy for Freeways." *Land Economics* 57, no. 1, February, pp. 115–21.

— 1982. "Highway Pricing and Capacity Choice Under Uncertain Demand." *Journal of Urban Economics* 12, pp. 122–28.

— 1989. "The Welfare Gains from Pricing Road Congestion Using Automatic Vehicle Identification and On-Vehicle Meters." *Journal of Urban Economics* 25, pp. 261–81.

Kraus, M., H. Mohring and T. Pinfold. 1976. "The Welfare Costs of Nonoptimal Pricing and Investment Policies for Freeway Transportation." *American Economic Review* 66, pp. 532–47.

Kulash, D. 1974. "Income Redistributional Consequences of Roadway Pricing." *Transportation Research Forum Proceedings* 15, pp. 153–59.

Larsen O. I. 1988. "The Toll Ring in Bergen, Norway — The First Year of Operation." *Traffic Engineering and Control* 29, no. 4, pp. 216–22.

Lave L. B. and J. S. DeSalvo. 1968. "Congestion, Tolls, and the Economic Capacity of a Waterway." *Journal of Political Economy* 76, pp. 375–91.

Lee, Douglass B. 1982. "Net Benefits from Efficient Highway User Charges." *Transportation Research Record* 858, pp. 14–20.

Levin, R. 1978. "Allocation in Surface Transportation: Does Rate Regulation Matter?" *Bell Journal of Economics and Management Science* 9, pp. 18–45.

— 1981. "Railroad Rates: Profitability and Welfare Under Regulation." *Bell Journal of Economics and Management Science* 11, pp. 1–26.

Levine, M. 1969. "Landing Fees and the Airport Congestion Problem." *Journal of Law and Economics* 12.

Lewis, W. A. 1949. *Overhead Costs: Some Essays in Economic Analysis.* New York: Rinehart.

Likens, J. D. 1976. "The Welfare Costs of Nonoptimal Airport Utilization." *Journal of Public Economics* 5, pp. 81–102

Little, I. M. D. and K. M. McLeod. 1974. "The New Pricing Policy of the BAA." In *Airport Economic Planning.* Edited by G. P. Howard, Cambridge, Mass.: The Massachusetts Institute of Technology, pp. 445–66. Reprinted from (1972), *Journal of Transport Economics and Policy* 6, pp. 101–15.

Luck, D. P. and I. J. Martin. 1988. "Review of Road Cost Recovery." Occasional Paper 90. Canberra: Australia General Public Service, Bureau of Transport and Communications Economics.

May, A. D. 1979. "Comment: Road Pricing — Should and Might It Happen?" *Transportation* 8, pp. 119–23.

McShan, S. and R. Windle. 1989. "The Implications of Hub-and-Spoke Routing for Airline Costs and Competitiveness." *Logistics and Transportation Review* 35, no. 3 (September), pp. 209–30.

Miyao, T. and P. Shapiro. 1981. "Discrete Choice and Variable Returns to Scale." *International Economic Review* 22, pp. 257–73.

Mohring, Herbert D. 1970. "The Peak Load Problem with Increasing Returns and Pricing Constraints." *American Economic Review* 60, pp. 693–705.

— 1976. *Transportation Economics.* Cambridge, Mass.: Ballinger Publishing Company.

— 1984. "Profit Maximization, Cost Minimization, and Pricing for Congestion-Prone Facilities." *Logistics and Transportation Review,* pp. 27–36.

Mohring, Herbert D. and I. Harwitz. 1962. *Highway Benefits: An Analytical Framework*. Evanston, Illinois: Northwestern University Press.

Morrison, S. 1982. "The Structure of Landing Fees at Uncongested Airports: An Application of Ramsey Pricing." *Journal of Transport Economics and Policy* 16, pp. 151–61.

— 1983. "Estimating the Long Run Prices and Investment Levels for Airport Runways." In *Research in Transportation Economics*, 1. Edited by T. Keeler. Greenwich, Connecticut: JAI Press, pp. 103–31.

— 1984. "Equity, Efficiency and Individual Preference in the Pricing of Transportation Infrastructure." Mimeo. Evanston, Ill.: Northwestern University, Department of Economics.

— 1986. "A Survey of Road Pricing." *Transportation Research* 20A, pp. 89–97.

— 1987a. "The Theory of Optimal Highway Pricing and Investment: A Comment." *Southern Economic Journal* 53, no. 3, pp. 779–82.

— 1987b. "The Equity and Efficiency of Runway Pricing." *Journal of Public Economics* 34, pp. 45–60.

Morrison, S. and C. Winston. 1986. *The Economic Effects of Airline Deregulation*. Washington, D.C.: The Brookings Institution.

— 1989a. "Airline Deregulation and Public Policy." *Science* 245, August, pp. 707–11.

— 1989b. "Enhancing the Performance of the Deregulated Air Transportation System." In *Brookings Papers on Economic Activity: Microeconomics*. Edited by C. Winston. Washington, D.C.: The Brookings Institution, pp. 61–125.

— 1990. "The Dynamics of Airline Pricing and Competition." *American Economic Review* 80, May, pp. 389–93.

Nash, C. 1991. "Rail Policy in the Europe Community." Working paper 324. University of Leeds, Institute for Transport Studies, January.

Nelson, James R. 1964. "Pricing Transportation Services." In *Transportation Investment and Economic Development*. Edited by Gary Fromm. Washington, D.C.: The Brookings Institution.

Neutze, G. M. 1966. "Investment Criteria and Road Pricing." *Manchester School of Economics and Social Studies* 34.

Newbery, D. 1988a. "Road User Charges in Great Britain." *Economic Journal* 98, pp. 161–76.

— 1988b. "Road Damage Externalities and Road User Charges." *Econometrica* 56, no. 2. pp. 295–316.

Newell, G. 1987. "The Morning Commute for Non-identical Travellers." *Transportation Science* 21, pp. 74–88.

Ng, Y. K. and W. Weisser. 1974. "Optimal Pricing with a Budget Constraint: The Case of the Two-Part Tariff." *Review of Economic Studies* 41, pp. 337–45.

Nix, F. P. 1989. *Road-User Costs*, Report for Transport Canada, Economic Research Branch. Ottawa: Transport Canada.

Nix, Fred P., Michel Boucher and Bruce Hutchinson. 1992. "Road Costs." In *Directions: Final Report of the Royal Commission on National Passenger Transportation*, vol. 4. Ottawa: Supply and Services Canada.

Oum, Tae Hoon. 1981. "Efficiency Losses from the Unbalanced Recovery of Costs of Transportation Infrastructure in Canada." *Logistics and Transportation Review* 18, no. 1.

Oum, Tae Hoon and Chunyan Yu. 1990. "An International Comparison of the Economic Efficiency of Passenger Railway Systems." A report prepared for the Royal Commission on National Passenger Transportation, RR-08, December.

Oum, Tae H. and Michael W. Tretheway. 1988. "Ramsey Pricing in the Presence of Externality Costs." *Journal of Transport Economics and Policy* 22, no. 3, pp. 307–17.

Oum, Tae H. and Yimin Zhang. 1990 "Airport Pricing: Congestion Tolls, Lumpy Investment and Cost Recovery." *Journal of Public Economics* 43, pp. 353–74.

Panzar, J. and R. Willig. 1977. "Free Entry and the Subsistability of Natural Monopoly." *Bell Journal of Economics and Management Science* 8, pp. 1–22.

Park, R. E. 1989. *Incremental Costs and Efficient Prices with Lumpy Capacity*. Santa Monica, Calif.: Rand Corporation.

Phillips, A. 1980. "Ramsey Pricing and Sustainability with Interdependent Demands." In *Regulated Industries and Public Enterprise*. Edited by B. M. Mitchell and P. R. Kleindorfer. Lexington, Mass.: D.C. Heath.

Pigou, Arthur C. 1912. *Wealth and Welfare*. London: Macmillan Press.

— 1920. *The Economics of Welfare*. London: Macmillan Press.

Rees, Ray. 1976. *Public Enterprise Economics*. London: Weidenfeld and Nicolson.

Roth, G. J. 1965. "Economic Approach to Road Congestion." *Town Planning Review* 36, pp. 49–61.

Roy, R. 1980. *Economics of Scale in the Airline Industry*. Discussion paper. Ottawa: Canadian Transport Commission.

Royal Commission on National Passenger Transportation. 1991. *Getting There: The Interim Report of The Royal Commission on National Passenger Transportation*. Ottawa: Supply and Services Canada.

Ruggles, Nancy. 1949-50. "The Welfare Basis of the Marginal Cost Pricing Principle." *Review of Economic Studies* 17, no. 1.

— 1949-50. "Recent Developments in the Theory of Marginal Cost Pricing." *Review of Economic Studies* 17, no. 2.

Sarndal, C., T. Oum and W. Statton. 1978. "Further Evidence on Factors Influencing Operating Costs of U.S. Commercial Airlines." *Journal of Transport Economics and Policy* 12, pp. 47–55.

Shah, Anwar M. 1990. "Optimal Pricing of Traffic Externalities; Theory and Measurement." *International Journal of Transport Economics* 17, pp. 3–19.

Sharkey, W. W. 1982. "Suggestions for Game-theoretic Approach to Public Utility Pricing and Cost Allocation." *Bell Journal of Economics and Management Science* 13, pp. 57–68.

— 1985. "Economic and Game Theoretic Issues Associated with Cost Allocation in Telecommunications Networks." In *Cost Allocation: Methods, Principles and Applications.* Edited by P. H. Young. Amsterdam: Elsevier Science Publishers.

Small, K. 1983a. "The Incidence of Congestion Tolls on Urban Highways." *Journal of Urban Economics* 13, pp. 90–111.

— 1983b. "Bus Priority and Congestion Pricing on Urban Expressways." *Research in Transportation Economics* 1, pp. 27–74.

— 1990a. "Highway Infrastructure: Crisis in Finance or Crisis in Management?" *Institute of Transport Studies Review* 13, no. 2 (February), pp. 1–3.

— 1990b. Urban Transport Economics. Draft monograph.

Small, K. and C. Winston. 1986. "Efficient Pricing and Investment Solutions to Highway Infrastructure Needs." *American Economic Review* 76, no. 2 (May), pp. 165–69.

— 1988. "Optimal Highway Durability." *American Economic Review* 78, no. 3, pp. 561–69.

Small, K., C. Winston, and C. Evans. 1989. *Road Work: A New Highway Pricing and Investment Policy.* Washington, D.C.: The Brookings Institution.

Spady, R. and A. Friedlaender. 1978. "Hedonic Cost Functions for the Regulated Trucking Industry." *Bell Journal of Economics* 4, pp. 159–79.

Spulber, D. 1989. *Regulation and Markets.* Cambridge, Mass.: The MIT Press.

Starkie, D. 1982a. "Road Indivisibilities: Some Observations." *Journal of Transport Economics and Policy,* September, pp. 259–66.

— 1982b. "Cost Recovery: An Investment Perspective." In *Pricing and Cost Recovery in Long Distance Transport.* Edited by D. N. M. Starkie et al. The Hague: Martinus Nijhoff.

Starrs, M. M. and David Starkie. 1986. "An Integrated Road Pricing and Investment Model: A South Australian Application." *Australian Road Research* 16, pp. 1–9.

Statistics Canada. *Consolidated Government Finance,* Catalogue No. 68-202, Table 1.

Steiner, P. O. 1967. "Peak Loads and Efficient Pricing." *Quarterly Journal of Economics* 71, pp. 585–610.

Strotz, R. 1965. "Urban Transportation Parables." In *The Public Economy of Urban Communities.* Edited by Julius Margolis. Baltimore, MD.: Johns Hopkins Press, pp. 127–69.

Stubbs, P. W., W. J. Tyson and M. Q. Dalvi. 1980. *Transport Economics.* London: George Allen and Urwin Publications.

Toms, M. 1986. "Towards Cost Related Pricing." Polytechnic of Central London, School of the Environment. Mimeo.

Train, K. 1977. "Optimal Transit Prices Under Increasing Returns to Scale and a Loss Constraint." *Journal of Transport Economics and Policy* 11, no. 1, pp. 185–94.

— 1991. In *Pursuit of Optimal Regulation: The Economic Theory of Natural Monopoly.* Cambridge, Mass.: The MIT Press.

Transport Canada. 1990. *Proposed New Cost Recovery Policy,* Phase II. Discussion paper. No. TP 10041, Ottawa, April.

— 1991. *Air Services Charges Regulations,* Cost Recovery Branch, Finance and Administration, Fifth Edition. No. TP-2590E, Ottawa, April.

Turvey, R. 1968. "Peak-Load Pricing." *Journal of Political Economy* 76, February, pp. 101–13.

U.S. Congressional Budget Office. 1992. *Paying for Highways, Airways, and Waterways: How Can Users be Charged?* Washington, D.C.

U.S. Department of Commerce, Bureau of the Census. 1989. *The Statistical Abstract of the United States.* T.1018, The National Data Bank, Transport section. Washington, D.C.

U.S. Department of Transportation. 1990. *Moving America: New Directions, New Opportunities.* Washington, D.C., February.

— 1991. "Air Services Charges Regulations," 5th ed., April 1.

Vickery, W. 1965. "Pricing as a Tool in the Coordination of Local Transportation." In *Transportation Economics.* New York: Columbia University Press, National Bureau of Economic Research.

— 1968. "Congestion Charges and Transportation Investment." *American Economic Review,* Papers and Proceedings, 59, pp. 107–18.

— 1985. "The Fallacy of Using Long-run Cost for Peak-Load Pricing." *Quarterly Journal of Economics* 100, pp. 1331–34.

Viton, P. A. 1980. "Equilibrium Short-run Marginal Cost Pricing of a Transport Facility: The Case of the San Francisco Bay Bridge." *Journal of Transport Economics and Policy* 14, pp. 185–203.

Walters, Alan A. 1961. "The Theory and Measurement of Private and Social Cost of Highway Congestion." *Econometrica* 29, no. 4 (October), pp. 676–99.

— 1968. *The Economics of Road User Charges.* World Bank Occasional Papers No. 5. International Bank for Reconstruction and Development. Baltimore, MD.: Johns Hopkins University Press.

— 1983. "Airports: An Economic Survey." *Journal of Transport Economics and Policy* 12, pp. 125–60.

Westfield, F. 1966. "Practicing Marginal Cost Pricing — A Review." *Journal of Business* 39, no. 1, part 1, January, pp. 57–74.

Wilson, George W. 1962. *Essays on Some Unsettled Questions in the Economics of Transportation.* Bloomington: Indiana University Press.

— 1980. *Economic Analysis of Intercity Freight Transportation.* Bloomington: Indiana University Press.

Winston, C. 1985. "Conceptual Developments in Economics of Transportation: An Interpretive Survey." *Journal of Economic Literature* 23, pp. 57–94.

— 1991. "Efficient Transportation Infrastructure Policy." *Journal of Economic Perspectives* 5, no. 1, Winter, pp. 113–27.

Wohl, M. and C. Hendrickson. 1984. *Transportation Investment and Pricing Principles.* Toronto: J. Wiley & Sons.

Ying, J. S. 1990. "The Inefficiency of Regulating a Competitive Industry: Productivity Gains in Trucking Following Reform." *The Review of Economics and Statistics* 72, no. 2, (May), pp. 191–201.

Young, Peyton H., ed. 1985. *Cost Allocation: Methods, Principles, Applications.* Amsterdam: Elsevier Science Publishers.

Zajac, E. E. 1974. "Note on an Extension of the Ramsey Inverse Elasticity of Demand Pricing or Taxation Formula." *Journal of Public Economics* 3, pp. 181–84.