13

ENVIRONMENTAL DAMAGE FROM TRANSPORTATION

VHB Research & Consulting Inc.
March 1992

SUMMARY

PURPOSE

The purpose of this document is to assist the Royal Commission in incorporating environmental considerations into its comparison of passenger transport options for Canada.

SCOPE AND METHOD

The document provides a brief review of the major approaches for assessing environmental impacts associated with transport and for valuing these impacts in economic terms. Based on a review of the published literature, estimates of some of these impacts and their economic value are presented. Opportunities for improving these estimates with additional research are identified.

STATUS OF THE DATA

This study relies on secondary sources of data. Reasonable estimates of total emissions are available by type and mode, at the point of emission. In general, other environmental effects (for example, impacts caused by transportation emissions) and the environmental effects of vehicle and

infrastructure manufacture, maintenance and disposal are not readily available. Some important data for comparing modes, such as loading factors, are not directly available.

The best data for estimating the costs associated with emissions are from studies of environmental damage associated with fossil-fired electricity generating stations. There are a number of differences between the types of emissions, their atmospheric dispersion, and patterns of exposure from power stations and transportation systems that necessitate extreme caution in the application of these economic damage estimates to transportation systems.

The data presented are national averages in the late 1980s, based on published information. The conclusions must be tempered with consideration of possible or likely technological, behavioural or regulatory changes (for example, reduced sulphur content of diesel fuel, rail electrification, more efficient vehicles, higher load factors). Similarly, it must be recognized that variations from the national averages may be very significant in some instances, and these may fundamentally change the order of the various modes in those circumstances.

MAIN FINDINGS

There is a range of environmental and economic techniques associated with valuing environmental damage that are applicable to estimating damage from intercity passenger transport. The use of the methods for this application, however, has been limited largely to emission inventories and direct damage methods for assigning economic values. Within the scope of this review, specific estimates of environmental damage associated with transportation modes could only be provided for damage associated with vehicle emissions.

The highest value of environmental damage per passenger-kilometre is associated with rail transportation, the lowest with marine transportation, followed by intercity bus transport. For all modes except air transport, most of the damage is associated with human health effects.

CONCLUSIONS AND RECOMMENDATIONS

The Canadian experience in estimating the social costs of environmental damage is very limited. Evaluations from the United States and Europe should be thoroughly reviewed to assess the applicability of their methods, assumptions and estimates to Canadian conditions. It may be possible to identify studies that could be repeated in Canada, or should be conducted in Canada to fill the most important knowledge gaps left by research elsewhere.

Environmental effects considered in intermodal comparisons are generally limited to vehicle operation. There has been no real attempt to set out the full life-cycle environmental impacts of transportation modes, from raw material extraction through vehicle and infrastructure construction, maintenance and disposal, fuel extraction and processing, and vehicle operation. Broad brush qualitative descriptions have been started for cars, but not for all transportation modes.

Quantitative evaluation of environmental impact assessments of transportation projects or policies is very limited. Greater use of quantitative evaluation could help to indicate some of the local environmental costs, which could be extrapolated for inclusion in the total environmental costs to be considered in transportation policies.

National averages of impacts by mode may obscure significant variations within each mode or significant opportunities that could be pursued. Consideration should be given to evaluation of the environmental impacts of each transportation mode in a multimodal corridor. Windsor to Quebec City or Edmonton to Calgary might be suitable corridors for such a study.

1. Introduction

The Royal Commission on National Passenger Transportation was established to review Canada's intercity passenger transportation system from a comprehensive perspective, looking at all the available modes and their interactions, rather than just one mode at a time.

Environmental considerations are a major concern for future transportation policies. This report is intended to assist the Commission in assessing the relative environmental implications of various modes and to put this assessment in a context that allows the significance of environmental considerations to be compared with other social, technological and policy concerns.

The report consists of six main components:

- a review of methods for assessing the environmental impacts of transportation systems;
- a review of methods for valuing — in economic terms — environmental impacts;
- a summary of published data on environmental impacts of transportation systems;
- a summary of published data on economic values of environmental impacts;
- a consolidation of the impact estimates and their values to estimate the economic value of the environmental impacts of transportation systems; and
- conclusions and recommendations for further research to improve the estimates and their usefulness to the policy-making process.

2. Methods for Estimating Environmental Damage: Theory and Application

2.1 THE ENVIRONMENT

A primary decision in determining the environmental effects of passenger transportation, or any other activities, is to decide what constitutes the "environment." There is no universally accepted definition, and the word conveys different meanings to different persons. The dictionary defines "environment" as the "surrounding; surrounding objects, region, or conditions, especially circumstances of life of person or society" (OCD, 1980).

Even where the "environment" is defined in legislation, the word may take on different meanings within different Acts within the same jurisdiction. For example, in Ontario, the *Environmental Protection Act* s.1(1)m is concerned only with the *natural* environment, which means: "the air, land and water, or any combination or part thereof. . . . "

In contrast, the Ontario *Environmental Assessment Act* s.1 provides a very broad definition of environment:

(a) air, land or water,

(b) plant and animal life, including man,

(c) the social, economic and cultural conditions that influence the life of man or a community,

(d) any building, structure, machine or other device or thing made by man,

(e) any solid, liquid, gas, odour, heat, sound, vibration or radiation resulting directly or indirectly from the activities of man, or

(f) any part or combination of the foregoing and the interrelationships between any two or more of them. . . .

For the purposes of this report, the definition used will fall between these two and is consistent with the definition implicit in the objectives set out in *Canada's Green Plan* (Environment Canada, 1990):[1]

- clean air, water and land;
- sustainable use of renewable resources;
- protection of special spaces and species;
- global environmental security; and
- minimizing the impacts of environmental emergencies.

2.2 SOURCES OF ENVIRONMENTAL DAMAGE

2.2.1 Resource Use

The use of resources is of particular interest from several perspectives. First, the pattern of use may determine whether the resource is being managed in a sustainable manner. Second, resource use is often correlated with other environmental effects, including air or water pollution, or land-use changes. Often, by altering resource-use patterns, it will be possible to reduce environmental damage significantly.

2.2.2 Chemical and Physical Discharges

Some chemicals that are released into the environment as a result of human activities may lead to changes in water, air or soil quality that can result in morbidity or mortality for humans or other species, both plants and animals. These discharges may also damage natural or anthropogenic materials due to their chemical composition.

In addition, discharges may be disruptive even where they are not toxic or corrosive if they lead to physical changes, such as decreased opacity or increased turbidity.

2.2.3 Land-Use Changes

Changes in land use may have an impact on the environment by altering the availability of resources, disrupting habitats or migration routes or affecting resource availability.

2.3 METHODS FOR ASSESSING IMPACTS

2.3.1 Life Cycle Assessment (LCA)

The environmental impacts of an activity may be associated with any stage of the activity, and can only be properly understood — and compared against alternatives — by considering the entire life cycle of the activity, and impacts at each stage of the life cycle. For example, in comparing air emissions associated with electric and diesel trains, it is not sufficient to consider just the emissions from the vehicle, which are higher for the diesel train. One must also consider "upstream" or "indirect" emissions, such as the emissions at the power station where the electricity is generated for the electric train.

In LCAs, the entire life cycle from extraction through material processing, manufacturing, use and disposal or recycling is considered — not just one of these stages. Thus, in the case of an assessment of transportation effects, the effects are not just those associated with operating the vehicle, but also the manufacture of the vehicle, fuel and infrastructure (for example, roads or rail lines), maintenance and disposal. To enable comparisons, these effects are usually considered on a unit basis, such as impact per passenger-kilometre of transportation provided.

Although simple in concept, LCAs are not always easy to implement, and systematic use of quantitative life cycle assessments is limited.[2] Much of the work on life cycle assessment started in the mid-1970s, largely in response to the energy crisis. Some of that early work also considered resource use and environmental discharges. Recently, there has been a renewed interest in LCA, primarily because of concern about waste generation, and many assessments have looked at packaging, diapers and other products associated with waste generation. The methodology for undertaking a life cycle assessment has been described by SETAC (1990), and this approach is becoming widely accepted. It consists of three main stages: inventory, impact analysis and improvement analysis. Much of the focus has been on the inventory stage which quantifies the inputs and outputs from each stage of the life cycle. This is unfortunate because impacts ought to be the basis for comparison, not just resource use and emissions. If one product produces more emissions than another product, but the emissions have no impact because there are no receptors nearby, then the inventory may be misleading as the basis for policy making.

2.3.2 Environmental Risk Assessment

Quantitative environmental risk assessment is one method for assessing the impacts of chemical discharges and may be a part of the impact analysis component of an LCA. The objective of risk assessment is to connect specific environmental releases with specific toxicological responses. Quantitative risk assessment involves the following sequence of steps:

- assessment of the exposure of receptors to the discharge (as discharged or as transformed);
- assessment of toxicological data and likely mortality or morbidity implications of the exposure; and
- assessment of individual or population risk.

Exposure assessment: Exposure assessment will consider the sources of environmental releases, their pathways through the environment to receptors and the magnitude and duration of exposure:

- Emission patterns: emissions, both routine and occasional, are assessed at each stage in the processing sequence. Quantitative information is needed by contaminant because of different toxicities of different contaminants, by quantity over time because impacts are likely to vary with quantity, and by location to determine how the discharge will move through the environment and who is exposed.

- Environmental fate and transport: once released, discharges move through the environment. They may settle in some component or may be transformed. The environmental fate and transport analysis consider such processes as dispersion and dilution, deposition, bio-accumulation and transformation of compounds. The environmental fate and transport assessment consider the pathways of exposure, for example, the air, water, food and soil concentrations of the contaminants.
- Receptor identification and activity patterns: the exposure assessment considers the receptors, and how much of the discharge, or its transformation product, a receptor is exposed to. Receptors considered are typically humans, though they may be domestic or wild plants, animals or structures subject to material damage. Assessment of their exposure requires consideration not only of environmental pathways, but also activity patterns (for example, the duration of time spent in a location where exposure may occur, and how this time is spent). The dose of atmospheric contaminants will be greater if the receptor is engaged in strenuous physical exercise than if it is resting. It may also require consideration of ambient conditions, as well as the specific release being assessed. The measure of the exposure will vary depending on the toxicological response being considered. Carcinogenic responses are usually assessed based on consideration of total annual exposure, and are typically measured in dose per unit body mass per year; non-cancer chronic exposure may be assessed in terms of mass of contaminant per unit body mass per two week period; and acute exposure may be measured in dose or concentration in air or water (g/kg, or mg/m^3).

Toxicological assessment: The toxicological assessment considers dose-response curves. Often these will be based on laboratory studies of indicator species, using toxicological models to extrapolate from high to low doses. For some compounds and toxic end points, actual data for some individuals of the receptor species may be available. A critical issue is often the ambient concentrations, that is, how the specific source under consideration contributes to the total concentration. This is particularly significant for toxic effects which only occur when a threshold is exceeded.

Risk assessment: Finally, the estimated exposure and the dose-response curves may be combined, and the impact of the discharges on morbidity or mortality assessed. Typically the risks are expressed as a probability of morbidity or mortality occurring, and may be given for an individual, or for the entire population exposed.

2.3.3 Ecological Analysis

Resource use is addressed in life-cycle assessments. Air, land and water quality, and environmental emergencies, at least as they affect biological species, may be addressed through environmental risk assessment.[3] However, environmental risk assessment is directed primarily at individuals, and their physiological health as a result of chemical exposures. Ecological analysis considers population, community or ecosystem effects emerging from cumulative effects on individuals, and from a consideration of other ecological factors, such as habitat supply after land-use changes. For example, risk assessment might identify effects of an environmental release on individuals of a certain species, but the ecological analysis would review how these individual effects affect the population and how these population changes might affect other species and the ecosystem as a whole.

For example, the effects on biological communities of filling a marsh to construct a highway would be addressed through an ecological analysis, as would greenhouse gas impacts and ozone depletion.

2.4 EVALUATION OF ALTERNATIVES

Once impacts are identified, it is probable that one alternative (that is, transportation mode) will have lower impacts of some types and higher impacts of other types. Evaluation methods are required to determine which is best overall in a given context. A wide variety of evaluation methods are available (VHB, 1990); the next section considers economic approaches for valuing different types of impacts, thereby enabling evaluation using economic measures.

3. Methods for Evaluating the Social Costs of Environmental Damage: Theory and Applications

3.1 THE ECONOMIC APPROACH TO EVALUATION

The economic approach to evaluation is based on the premise that evaluation takes place whenever a person is confronted by a choice among two or more mutually incompatible alternatives. In such a situation, choosing one of the alternatives requires foregoing whatever could have been attained by choosing any of the other alternatives instead. Economists assume that the alternative chosen by the person in such a situation must be the one the person values most highly. Indeed, such a choice involves an implicit cost of foregoing the next most highly valued alternative, which economists call the "opportunity cost." Thus, if someone chooses to drive to a destination rather than to take a bus, then for that person the net value of driving (the value of the speed, convenience and other valued factors over the cost) must exceed the net value of taking the bus.[4] If taking the bus was the next most preferred alternative to driving, then the net value of taking the bus (perhaps including valued factors such as a more relaxed trip) would be the opportunity cost of driving instead of taking the bus.

This kind of choice occurs whenever we purchase something, or refrain from purchasing something. Purchasing something requires giving up a certain amount of money and thus *foregoing* whatever else could have been purchased with that money. If one refrains from making a purchase, one is implicitly deciding that greater value could be achieved by saving that money to spend on something else. This is the kind of expression of preferences that applies to commodities and services that can be bought and sold in a market. The economic approach to evaluation suggests that, in order to have a unified approach to evaluation *both* for goods and services that are provided through markets *and* for goods and services available outside markets (for example, environmental amenities), the same kind of evaluative procedure should apply to both.

3.1.1 Willingness to Pay and Willingness to Accept

The relationship between value and money provides the economist with a quantitative basis for evaluation: the value of something to a person who does not already have it is the maximum amount of money the person would be willing to pay (WTP) to attain it. Conversely, the value of something already enjoyed by a person is the minimum amount the person would be willing to accept (WTA) to give it up. By either of these different measures of value, the monetary value a person assigns to something will depend on that person's overall assets, including money. Furthermore, people will often assign WTA values several times greater than their WTP values for the same items. Thus, a person's economic evaluation of something will depend on the person's position with regard to ownership of assets in general and of the thing under evaluation in particular.

Given individual monetary evaluations, social value can be most easily computed as the sum of these value assignments. The relative preferences of each person would not count equally in such a social value; rather, each person's preferences would be weighted in proportion to total personal assets or wealth. As a basis for public policy decisions, this social value appears to be at odds with the democratic principle of "one person, one vote." To make social value more in accord with the latter principle, individual evaluations could be weighted to reflect what each person would be willing to pay *if* their incomes were equalized. One crude procedure for this is to apply weights to each person's monetary evaluations inversely proportional to their incomes or wealth (Pearce, 1983). More precisely, it can be shown that this condition is better reflected if the inverse is raised to the power of the person's income elasticity of demand for the good in question (Pearce, 1983). In considering institutional changes that allocate or reallocate rights, it is not only effects by income class that may be important. Bromley (1991, p. 226) observes that "significant progress on air and water pollution turns critically on the distribution (incidence) of different kinds of benefits and costs, not just by income class, but by job category, by location of residence, by education level, and by a number of variables rarely pondered in economic analysis."

These considerations about the appropriate basis for evaluation bear particularly on the evaluation of public goods, including many environmental services (for example, air for breathing and for carrying away exhaust fumes), the allocations of which are public (political) decisions. Who holds rights to which services (to emit exhaust fumes or to breathe free of exhaust fumes[5])? Accordingly, in determining people's values for environmental services, is the correct measure WTP or WTA? Furthermore, are the correct measures of social value based only on people's monetary assets, or on their total assets (including their non-monetary shares of environmental services) or on what they would be if income or assets were equalized?

All of these general considerations with regard to evaluation apply to evaluation of the social costs of environmental damage, including environmental damage from passenger transportation. From an economic perspective, the general problem of environmental damage is that use of one environmental service (for example, air for carrying away exhaust fumes) can alter the environment and reduce the benefits (or impose costs) from use of the same or another environmental service (for example, air for breathing) by someone else or by society at large.[6] There are thus "external costs" from the first use for the second use. Of course, even to refer to "environmental damage" and "external costs" assigns some presumptive rights to the second use.

The most immediate form of environmental damage is damage to our health. In economic terms, health is treated just as any other service. Given an initial endowment of health and other assets, the cost of a loss in health is equivalent to the minimum amount of compensation the person would have required in order to have suffered the loss voluntarily. If other people also suffer a loss from a reduction in health of the person, then the total cost should also include compensation for their losses. There are several types of loss, including direct monetary costs as well as losses in well-being, that bear consideration with regard to compensation: "(1) medical expenses associated with pollution-induced diseases, including the opportunity cost of time spent in obtaining treatment; (2) lost wages; (3) defensive or averting expenditures associated with attempts to prevent pollution-induced disease; (4) disutility associated with the symptoms and lost opportunities for leisure activities; and (5) changes in life expectancy or risk of premature death" (Cropper and Freeman, 1991, p. 166).

Similarly, with environmental damage in the form of damage to natural resources (including crops) and materials, the cost of the loss is equivalent to the minimum amount required to compensate for the loss in the quantity or quality of the products or services derived from these resources or materials.

3.2 TYPES OF VALUE

3.2.1 Direct Use Value

In the case of cultural artifacts and environmental amenities, there are several aspects of value to be considered. This discussion has referred mainly to the value that can be derived from immediate uses, or direct use value, of goods and services (including environmental services). In referring to environmental amenities, it is often necessary to be careful about what constitutes "use." The value of such amenities often derives not only from direct experience, but also indirect experience through literature, photographs, films and television. This type of value is sometimes distinguished as "vicarious use value."

3.2.2 Option Value

Economists also recognize that goods and services can have kinds of value other than their immediate use value. Even if people do not currently use a good or service, if the future supply (availability) of the good or service is uncertain, they might be willing to pay to secure the option to purchase the good or service in the future. Similarly, even if future supply is not an issue, they might be uncertain about whether they want to purchase and use a good or service in the future, and hence might be willing to pay to maintain that option. The types of value implicit in this willingness to pay are referred to by economists as "supply option value" and "demand option value" respectively.

3.2.3 Quasi-Option Value

A somewhat related, but distinct, type of value applies in the case of actions toward the environment which are expected to have irreversible, but uncertain consequences, and where the uncertainty over the consequences could be resolved over time. For example, there might be uncertainty about the benefits that could be derived from preservation of an ecosystem that would be irreversibly harmed by a particular project. Then there is a value, which economists call "quasi-option value," in maintaining options for future use while more complete knowledge is obtained.

3.2.4 Existence Value

Apart from willingness to pay for present and possible future use of environmental amenities, people may be willing to pay just to secure the existence of an environmental amenity (for example, wildlife in a particular area), whether or not they expect to experience it directly or indirectly. Economists refer to this kind of willingness to pay as "existence value." Furthermore, people may be willing to pay for the existence of these amenities not only for themselves, but also for their progeny and future generations. This kind of willingness to pay is sometimes referred to as "bequest value."

In practice, it is often very difficult to know whether people really distinguish among vicarious use value, existence value and bequest value conceptually or in their behaviour. Thus, these three forms of value are often subsumed under a more general concept of existence value; this is the practice that will be followed here.

3.2.5 Summary

Economists have recognized three general types of value that make up people's willingness to pay (or to accept payment) for non-marketed environmental amenities: use value, option value and existence value. Consequently, the social costs of environmental damage can be measured in terms of losses in these values.

It is apparent that the economic approach to evaluation places exclusive weight on the values people attach to things relative to the value they place on their other assets: how much of their other assets they would be willing to forego to obtain a valued item, or how much they would have to be compensated to be willing to give up a valued item. This approach can be challenged in several regards. First, it is clear that it is a strictly anthropocentric basis for evaluation. On the other hand, it is difficult to conceive of a method of evaluation that is not anthropocentric. Nevertheless, a basis of evaluation that is not anthropocentric would seem to be implied by the "biocentric" world views offered by advocates of "deep ecology" (Devall and Sessions, 1985; Naess, 1989). According to such world views, the validity of an evaluation is not derived merely from human volition, but rather is subject to broader ecological criteria. Second and furthermore, the economic approach to evaluation depends substantially on people's knowledge in making evaluations of environmental services, although even scientists are often unsure about the importance of ecosystem components and functions. As remarked by Costanza (1991, p. 336), "[t]he public is most likely far from being fully informed about the ecosystem's true contribution to their own well-being, and they may therefore be unable to directly value the ecosystem's services."

Mindful of these limitations, the strength of the economic approach to evaluation is that it tries to provide an indication of people's "true" preferences in the aggregate (in terms of the maximum they would be willing to give up, and the minimum they would have to be compensated), given the current distribution of assets (and liabilities). One may question whether the status quo is an appropriate basis for social evaluation, and doubt whether people are sufficiently informed or entitled to make evaluations of ecosystems and their services, but nevertheless believe that this measure of value provides useful information to consider in economic and environmental policy making. In any case, it should not be forgotten that it is also incomplete and potentially misleading information, unless it is supplemented by other information, such as economic information on the incidence of costs and benefits, and ecological information on the sustainability of environmental uses. The need for this supplementary information should be borne in mind in interpreting the information on social costs provided here.

3.3 METHODS FOR ASSESSING BENEFITS AND COSTS

The economic benefits and costs of any change in a person's situation fall into two categories:

- gains and losses to their monetary assets (or assets sold for money); and
- increases or decreases in their non-monetary wealth or well-being.

3.3.1 Direct Cost Assessment

The benefit or cost of a change of the first category can be taken to be just the monetary value of the assets, because compensation for the benefit or cost could be made simply by restoring the monetary assets of the person to the initial situation. Some of the costs of environmental damage can be considered as this kind of direct cost.[7] Care must be taken where the damage is sufficiently widespread to affect prices. In this case, costs, and the distribution of costs, will be determined by the market response to the loss incurred by the damage.

This subsection is mainly concerned with methods for evaluating changes of the second category. All of these methods are direct or indirect ways of determining the values people assign to changes in their non-monetary wealth or well-being, usually in terms of WTP for improvements or for avoiding losses, and sometimes in terms of WTA losses. One of the practical problems with these methods is that they can often be applied only to indicate WTP. However, if rights are initially assigned on the basis of prior use or non-destructive use, for example, then the appropriate measure of value is the WTA of the bearers of those rights (Krutilla and Fisher, 1985). Until recently, it was commonly assumed by economists that WTP and WTA should be equivalent, so that a measure of WTP could be substituted for WTA where required. This was based on the assumption that value functions are "smooth," that is, there are no abrupt changes in marginal value according to quantity consumed. As already indicated, there is now substantial evidence that WTA is often several times greater than WTP in common situations (Knetsch and Sinden, 1984; Knetsch, 1984). This suggests that value functions are often "kinked" at some "reference quantity," which is usually the quantity currently consumed.

There are three main methods for evaluating well-being:

- household production function (HPF) methods;
- hedonic methods; and
- contingent valuation methods (CVM).

The first two methods are indirect ways of assessing preferences by inference from market behaviour: these are said to be "revealed preferences" indicated by observable behaviour. The third method determines preferences directly by surveys with interviews.

Two other methods of evaluation for environmental costs are mooted:

- costs necessary to meet national targets for emission reductions; and
- costs implicit in previous decisions or mitigation measures.

Both methods are based on the premise that evaluations already reached through the policy process are a more appropriate guide to further policy development than individual evaluations. There is an apparent circularity in these methods, in that evaluation appears to become immune from any criteria external to the policy process. Nevertheless, it might be maintained that these methods take better account of some of the considerations concerning the incidence of costs and benefits and the sustainability of the environmental uses broached at the end of the previous section. To avoid the circularity of the policy-based methods, these considerations could be taken into account explicitly. For example, if "no net loss" of "natural capital" (Pearce and Turner, 1990) is adopted in policy as a basic condition and commitment of sustainability, a "sustainability premium" (ibid.) could be added to the values determined by the previous methods to ensure this condition.

3.3.2 Household Production Function (HPF) Methods

HPF methods are based on the observation that many goods and services bought by consumers are only intermediate products that households use as inputs into their own production processes to produce goods and services for final consumption. This "household production" also draws on environmental conditions. In particular, there is a strong relation between household expenditures on some goods and services and the use of environmental amenities or avoidance of environmental disamenities. The amounts households are willing to pay for marketed goods and services can thus be used to infer evaluations of environmental amenities and disamenities.

HPF methods usually consider one particular good or service, and assume that prices for that good or service for each household are established already. Given these prices, the HPF method applies to observations of which households purchase the good or service, and in what quantities.

The HPF method used the longest and most extensively is the travel-cost method (TCM) for evaluating recreational sites. In this method, the "prices" that distinguish among households are the costs of travelling to the recreational site, with higher costs generally associated with greater distances from the site.

Recently, the HPF method has also been used for the evaluation of factors affecting health by recording defensive or averting behaviour. For example, frequencies of visits to doctors and purchases of water filters and household detectors or tests for naturally occurring radioactive radon gas have been used to assess knowledge and evaluations of associated health effects (Smith, 1991). In the case of emissions from transportation, if the purchase of air conditioners, especially car air conditioners, could be considered, at least in part, defensive expenditures to filter air, then this method might apply. Of course, there would immediately be a problem of how to attribute expenditures on air conditioning between air filtering and air cooling.

Practical applications of HPF methods generally require restrictive, simplifying assumptions about the factors that determine purchase decisions. The data requirements of these methods grow exponentially with the number of factors under consideration.

3.3.3 Hedonic Methods

Like HPF methods, hedonic methods depend on hypothesized relationships between purchases of marketed goods and services, and evaluations of related environmental amenities or disamenities. In this case the marketed good is often real estate (usually residential property or agricultural land) and sometimes labour.

In hedonic methods (sometimes called hedonic price methods), prices are functions of numerous factors, including environmental factors. For example, it might be hypothesized that residential property prices are related to environmental amenities in the area. If so, this price information can be used to infer values (willingness to pay) for these amenities. For example, houses near an airport affected by noise might sell for less than comparable houses not near the airport, reflecting some of the costs people attribute to the noise.

Similarly, if it is hypothesized that people are willing to work for lower wages in locations with higher environmental amenities, this should be reflected in wage rates, and the value people assign to these amenities can be inferred from wage rates.

Although hedonic methods are straightforward in principle, there are several practical difficulties in applying them. The main difficulties are likely to be in producing statistically significant results, when environmental amenities may only constitute a few minor factors among a large number of other factors which affect prices. In the case of residential properties, in particular, the characteristics of the properties and their neighbourhoods, as well as the accessibility to transportation and work locations, are likely to be factors of greater, or at least equal, importance than environmental amenities. Furthermore, there might not be substantial differences in some environmental amenities *within* the geographic extent that can be identified with a particular residential property market. Differences in other factors could raise difficulties for comparisons between markets.

Another difficulty in applying hedonic methods is making assumptions about the mathematical relationships between prices and the characteristics of relevance in determining price. The more general the starting assumptions, the more data will be required to specify the mathematical relationships. Data requirements are likely to be the greatest difficulty in applying hedonic methods, especially where there is a need to assess demand (willingness to pay) over a wide range of environmental quality. Nevertheless, there have been a number of notable demonstrations of hedonic methods, and progress is being made in refining the methods for easier and more reliable application (Palmquist, 1991; Pearce and Markandya, 1989; Cropper and Oates, 1990).

3.3.4 Contingent Valuation Methods

Contingent valuation methods (CVM) use survey questionnaires to ask people their WTP or WTA for particular benefits or losses. Sometimes, people are asked to indicate a maximum value in dollars, or to select a maximum value from a range of values. More often, they are asked if they would be willing to pay a given amount: if they accept, the amount is raised; if they decline, the amount is lowered, until the maximum amount they are willing to pay is determined. Sometimes, the interviewer indicates verbally or with a chart the amounts paid for similar benefits for comparison.

A study conducted in Berlin in the mid-1980s (Schulz, 1985 as reported in Schulz and Schulz, 1991) is an example of an application of the CVM method to evaluating urban air quality. Residents of Berlin were surveyed to determine their willingness-to-pay for each of four kinds of air quality: Berlin air (where air quality may warrant a smog alarm); big city air (where air quality does not warrant a smog alarm); small town air; and vacation spot or holiday resort air. Further, respondents, through their answers to several questions, were classified according to their knowledge about air pollution and its effects. When the results of this survey were extrapolated for the whole Federal Republic of Germany at that time (population approximately 60 million), the estimated benefit was DM14 billion (approximately CAN$160 per capita) per annum for improvement of air quality to small town air quality, and DM30 billion (approximately CAN$350 per capita) per annum for improvement to the standard of vacation spot/resort air quality. If corrections were made for the "information deficits" of less informed respondents, these willingness-to-pay estimates increased to DM28 billion (CAN$320 per capita) per annum and DM48 billion (CAN$560 per capita) per annum respectively.

Many economists were originally very skeptical whether the CVM method could produce useful information because of the number and magnitude of the biases that can affect results. The different kinds of potential bias have been extensively catalogued and reviewed by Mitchell and Carson (1989). For example, simple applications of the method might be open to a kind of "strategic bias," whereby respondents would give values higher than their actual WTP, in the expectation that this might help to secure the benefit, and believing they would not actually be required to pay in proportion to the value they indicate. If the application of the method includes actual payment (perhaps from an initial fixed "gift"), then respondents might indicate a value lower than their actual WTP, in the belief that their stated value would have little effect on securing the benefit, so that the expected value of a high bid is small compared with the value of limiting their payment.

In any attempt to assess WTA for a loss after the loss has occurred, strategic bias is a problem. It is only somewhat less of a problem before the loss has occurred. Thus, CVM, like the other methods, appears to be most readily applicable to assessing WTP. This is still a problem in using these methods to assess the social costs of environmental damage.

There are many other potential sources of bias in CVM (Pearce and Turner, 1990; Carson, 1991; Mitchell and Carson, 1989). Nevertheless, substantial progress has been made, especially in the last decade, in developing ways of testing for and controlling these biases (Carson, 1991; Mitchell and Carson, 1989). As a result, CVM is now more commonly applied and accepted. It also has the great advantage of often being the only method that can be used to assess all the types of values, not just the use and option values.

In spite of the advances in developing CVM, application design still requires great sophistication to produce reliable results. CVM questionnaires generally require numerous questions to characterize and check the validity of the evaluation responses; sampling and analysis procedures can also be quite complex. Although CVM is gaining acceptance among economists, it is perhaps the method with the furthest to go in winning acceptance outside the profession.

3.4 APPLICATION OF EVALUATION METHODS TO ENVIRONMENTAL DAMAGE

Not all the social costs of environmental damage can be measured by each of the evaluation methods. Table 1 indicates which of the methods might apply to each type of social cost mentioned in subsection 3.1.

The "direct costs" of environmental damage apply where there is a direct monetary ("out-of-pocket") loss from the damage. As shown in the table, this would be the case with medical expenses, lost wages, defensive or averting expenditures, costs of material repair, and loss of the natural resource productivity. Of course, in each case, there will be different degrees of difficulty in establishing the causal relationship between the purported agent of environmental damage and the loss.

Table 1
APPLICABILITY OF EVALUATION METHODS TO THE SOCIAL COSTS OF ENVIRONMENTAL DAMAGE

		Direct cost	Defensive/ averting behaviour method	Travel cost method	Hedonic	Contin- gent valuation method
Health and comfort	medical expenses (incl. opportunity cost of time)	•	•		•	
	lost wages	•	•		•	
	defensive or averting expenditures	•	•		•	
	disutility of symptoms and lost leisure		•		•	•
	changes in life expectancy and risk of death		•		•	•
Material damage	cost of repair/replacement/cleaning	•	•			
	loss of cultural heritage — use value		•	•		•
	loss of cultural heritage — option value		•			•
	loss of cultural heritage — existence value		•			•
Natural environment	loss of natural resource productivity	•	•			
	loss of natural amenities — use value		•	•	•	•
	loss of natural amenities — option value		•		•	•
	loss of natural amenities — existence value		•			•

With regard to HPF methods, the averting behaviour method could, in principle, apply to all the costs, if suitable averting behaviours could be identified for each. In practice, the averting behaviour method has been applied most readily to health effects. In all cases, it should be noted that this method provides only *ex ante* WTP of people's expectations of costs.

It might be argued that it is *ex post* WTA that is likely to be more relevant for the purposes of policy making.

Also among HPF methods, the travel cost method applies to the use values of cultural artifacts and natural amenities. To use the travel cost method for the evaluation of environmental damage, it would be necessary to apply the method before and after the damage occurs, or have multiple "equivalent" alternatives differing only in environmental quality.

The hedonic methods are most likely to be applicable to health damage and reductions in the use and option values of natural amenities in residential areas. In principle, both the wage rate method and the residential property price method could be applied to all these values. In practice, the wage rate method has been applied to the evaluation of workplace health risks, and the residential property price method has been applied to the evaluation of use and option values of environmental amenities in residential areas. The hedonic method could also be applied to evaluate the costs of environmental damage to agricultural land, using agricultural land prices.

In principle, the contingent valuation method could apply to all costs, but it would usually be inappropriate to use it to solicit hypothetical evaluations of direct costs where actual data about these costs can be obtained. If care is taken to avoid "double-counting" (that is, including the same value twice), then the different value or costs can be added to give total values and costs.

Many economists favour research designs in which two or more evaluation methods are applied to the same object of value, in order to compare results among the methods.

4. Environmental Damage

As was indicated in Section 2 there are three aspects of environmental damage to be considered:

- changes in resource requirements, which can be measured using life-cycle analysis techniques;
- changes in physical and chemical discharges, measured by environmental risk assessment techniques; and

- changes in land-use patterns, generally expressed in terms of changes in habitat availability or "ecological analysis."

Though research has focussed on environmental risk assessment because of its direct applicability in determining toxicity to humans and morbidity, all three aspects are necessary if a complete assessment of environmental damage is to be developed. There is a connection between the three types of damage: changes in physical and chemical discharges can often be traced to changes in the processing and use of resources, and physical and chemical discharges alter resource quality which can affect habitat availability. As a result, the techniques cannot be looked at in isolation. They must be examined as a system, with the data coming out of one method of assessment feeding into another method of assessment.

Procuring all the data required for each type of assessment is difficult. Nevertheless, it is possible to draw some conclusions from the available data. Only some of the data will be directly applicable to valuing the damage. This section presents the quantitative and qualitative data which are to be considered in the valuation of the damage attributable to each mode of transport.

4.1 DAMAGE DUE TO CHANGES IN RESOURCE REQUIREMENTS

Resource use is associated with three types of environmental damage: the impact of emissions released during the extraction of resources, the effect of resource extraction on land-use patterns and the reduction in resource capital.

The loss of resource capital is not directly damaging to the environment; rather it is the implication of resource extraction to long term sustainable development which is of concern. Such damage can be expressed in terms of the rate of capital depletion and the potential for resource substitution.

This subsection outlines the resource use associated with the manufacture of transport vehicles and the infrastructure required for their operation as well as the operation of vehicles.

Vehicle manufacture: Resources are required at each stage of manufacturing including:

- the extraction of natural resources;

- the processing of natural resources into final materials;
- parts assembly; and
- vehicle assembly.

A complete inventory of resource use, emissions and waste for each stage cannot be compiled as data are often unavailable, or aggregated to a level where they are no longer representative of individual processes (VHB, 1991c). Assembling a series of comparable data for different transport vehicles is further complicated by inconsistent monitoring and reporting procedures.

Of all the resources under consideration, energy-use data are most readily available. Energy-use data can be assembled for the manufacture of the four basic materials used in transport vehicles — steel, aluminum, rigid plastic and glass.

Where energy data are disaggregated by fuel type, air emissions can be estimated using emission factors.[8] Emissions from vehicle manufacture averaged over the lifetime of the vehicle should be added to emissions from vehicle operation.

Construction and maintenance of infrastructure: Resource use and emissions from the construction and maintenance of roads and depots also contribute to the environmental damage from transportation. Table 2 shows the approximate amount of energy used to make the asphalt and lay one kilometre of a four-lane (two lanes x two directions) highway. Note that the values do not include the energy contained within the tar of the asphalt, the energy required to prepare the ground for paving or the energy required to manufacture and distribute street furniture, such as safety barriers.

The energy used to construct one kilometre of highway is about 32 GJ. However this requirement is negligible when compared to the energy required by the road vehicles using the same length of highway.

Table 2
ENERGY REQUIRED TO PAVE ROADWAYS

Task	Energy use/km	Notes
pavement manufacture	2,000 MJ	4,000 tonnes of asphalt required per km of 4-lane highway (personal communication with Bayne Potapchuk, D. Crupi and Sons, Scarborough) approximately 0.5 MJ/tonne asphalt (Brown 1985)
pavement laying	10,400 to 30,800 MJ	270 to 800 tonnes oil equivalent (OECD, 1988) 1 tonne = 1.17 ML crude (LeBlanc, 1992) 1 ML crude = 38.51 TJ (Statistics Canada, Catalogue No. 57-003)

Fuel manufacture and distribution: The environmental impacts of oil refineries are numerous. Air pollutants released in large quantities include sulphur dioxide, hydrocarbons, nitrogen oxides, carbon monoxide and particulates. Conventional contaminants measured in process effluents include total organic carbon, dissolved organic carbon, ammonia, nitrogen, total and volatile suspended solids, sulphides, phosphorous, oil and grease, phenolics and pH. Benzene, ethylbenzene, toluene, and xylene have been detected in significant amounts in refineries. Chromium and zinc have also been found in high levels in the effluent. In addition, oil refineries produce potentially toxic wastes consisting of sludges from effluent treatment, spent catalysts and tars (VHB and Canviro, 1991).

Gasoline has a high vapour pressure compared to other liquid transportation fuels. This pressure, in combination with the large volume of gasoline used in the transportation sector, means that gasoline distribution and marketing accounts for about 10 percent of the total hydrocarbon emissions in Canada (Table 3). These emissions arise due to:

- the escape of vapour during storage and transfer of fuel between the refinery and retail stations; and
- leakage during vehicle refuelling at the gasoline pump (Transport Canada and Environment Canada, 1989).

Table 3

1985 Emission Inventory in Canada (Kilotonnes)

Source	CO_2[b] (kt)	%	SO_2 (kt)[e]	%	NO_x (kt)	%	HC[a] (kt)	%	PM[c] (kt)	%	CO (kt)	%
Motorcycles	192	<0.01	}		1	0.1	4	0.2			15	0.1
Passenger cars	52,547	12.1	}		392	20.0	471	26.2			4,016	37.2
Light-duty trucks	10,509	2.4	} 47		74	3.8	115	6.4			850	7.9
Medium-duty trucks	7,813	1.8	}		54	2.7	78	4.3	1	0.1	570	5.3
Heavy-duty trucks	27,075	6.2	}	1.2	286	14.6	64	3.5	22	1.0	436	4.0
Air	12,271	2.8	[f]		34	1.7	10	0.6			117	1.1
Marine	5,219	1.2	17	0.4	17	0.9	28	1.6			85	0.8
Rail	3,145	0.7	18	0.5	138	7.0	10	0.6	6	0.3	48	0.4
Transportation total	118,771	27.3	82	2.2	994	50.7	780	43.4	29	1.3	6,135	56.9
Total all sources	**435,180**	**100**	**3,800[e]**	**100**	**1,959**	**100**	**1,798**	**100**	**2,200[d]**	**100**	**10,780**	**100**

Source: Transport Canada, 1989; OECD, 1991; SENES Consultants, 1991.

Notes: Blanks indicate data not available.

- **a** Non-methane
- **b** 1987 fossil fuel only
- **c** Diesel only
- **d** Total particulate from all sources
- **e** 3,800 kt for late 1980s (OECD, 1991); SO_2 distribution from SENES Consultants (1991)
- **f** Included in rail

The development of alternative fuels has long been seen as a desirable measure to reduce the transportation sector's dependence on fossil fuels.

Electric cars are considered by some to be the most promising of alternative vehicles, with prototypes being tested by the major North American car manufacturers. Though these cars have no exhaust emissions, pollutants are released during the generation of electricity, particularly the emissions associated with coal-fired generating plants. The risks associated with nuclear plants should also be considered when determining the total environmental costs of electricity-powered vehicles.

Operation and Maintenance: The disposal of hazardous or potentially hazardous components of transport vehicles may also result in significant environmental damage. Car batteries contain cadmium, mercury and lead which are released as the battery corrodes. Batteries are to be deposited in special hazardous waste landfills which are designed to contain the hazardous materials. However, leaks from such landfills are common.

Coolants used in the air conditioning systems of transport vehicles contain CFC-12 which is known to cause ozone depletion and is thought to contribute to global warming. Although coolants can be refined and recycled on-site, only some service stations offer this service.

Spent motor oil contains organic contaminants such as PCBs, benzene and toluene. It may also contain some lead. Traditionally, waste oil has been used as a dust suppressant on unpaved roads, reused or released to sewer systems or non-hazardous landfills. Because of its potentially hazardous qualities, Ontario recently banned the use of motor oil as a dust suppressant. Waste oil can easily be recycled, and many service stations offer this service.

Used tires when improperly stored are inflammable, as exemplified by the recent used-tire fire in Hagersville. Recycling can only absorb a small fraction of the tires generated. Although tires can be burned for use as an energy source, this has been banned in some jurisdictions due to concerns over the emissions from combustion. Because of the high cost of landfill disposal, tires are often illegally dumped or stored in unsafe piles (Pilorusso, VHB and T.A.G., 1990).

4.2 DAMAGE DUE TO CHANGES IN PHYSICAL AND CHEMICAL DISCHARGES

This section provides a detailed explanation of environmental risk assessment including:

- identification of discharges;
- estimation of pollutant loading;
- a description of the fate of those discharges and some of the secondary effects associated with them;
- an assessment of the receptors which are exposed to those discharges and what the level of exposure might be;
- a review of the toxicological data and implications of exposure; and
- an assessment of the individuals or populations at risk.

As indicated in the last section, discharges include emissions to water and air resources as well as the disposal of solid hazardous and non-hazardous waste. Discharges can be attributed to the manufacture and operation of transport vehicles as well as the construction and maintenance of the transportation infrastructure.

An assessment of the risk associated with physical and chemical discharges should include all these sources of pollution. Research, however, has focussed on air emissions from the operation of transport vehicles, and data on the type and extent of emissions from other aspects of transportation are, in general, not available. As a result, the examples in the discussion focus on atmospheric discharges from vehicles since these data are most readily available.

4.2.1 Identification of Discharges

The emissions of major concern are:

- hydrocarbons (HC)
- oxides of nitrogen (NO_x)

- carbon monoxide (CO)
- carbon dioxide (CO_2)
- particulates, sulphur dioxide (SO_2)
- chlorofluorocarbons (CFCs).

Other more toxic compounds (heavy metals such as cadmium or lead) are of concern in particular instances.

Hydrocarbons: Hydrocarbons (HC) are the volatile portion of unburned hydrocarbons (fuel) emitted from vehicle engines. The main sources of unburned hydrocarbons are exhaust gases and evaporative losses from engines and during refuelling.

Private passenger motor vehicles account for more than 75 percent of HC from transportation sources. The transportation sector as a whole accounts for about 43 percent of HC emissions from all anthropogenic sources (Table 3).

Exposure to hydrocarbons has been shown to cause skin irritation and has also been linked to the increased incidence of leukemia in the vicinity of large point sources (OECD, 1988). The ecological impact of hydrocarbon emissions, however, is most apparent in their contribution to the creation of ozone during reaction with nitrogen oxides and sunlight.

Oxides of nitrogen: The major source of NO_x is the burning of diesel in the internal combustion engine. The transportation sector is the major contributor — about one half of total NO_x emissions from all sources in Canada.

Nitrous oxides result in human respiratory and circulatory ailments, vegetative effects, material damage to textiles, fabrics, plastics and rubber, and impaired visibility. Nitrous oxides are also a precursor of low level ozone, and contribute to ozone layer depletion, global warming and acid deposition (VHB, 1991b).

Carbon monoxide: Carbon monoxide (CO) is a by-product of the combustion of fossil fuels. Private passenger motor vehicles account for about 80 percent of CO emissions from transportation sectors. Just over one half of total CO emissions in Canada originate from transportation sources (Table 3).

Elevated CO levels cause a number of adverse effects on human respiratory and cardiovascular systems. Carbon monoxide is also a contributor to the creation of low level ozone, which affects human health and damage vegetation and materials. It is a radiative gas contributing to global warming (VHB, 1991b).

Carbon dioxide: Carbon dioxide (CO_2) is a by-product of the burning of fossil fuels. The transportation sector accounts for about one quarter of total CO_2 emissions in Canada (Table 3). Carbon dioxide does not pose a risk to humans, materials or ecosystems directly but is a major anthropogenic contribution to global warming which could damage human health, ecosystems and materials.

Particulates: Particulate (aerosol) emissions are a significant portion of diesel exhaust. Only a small amount of total diesel particulate emissions in Canada arise from the transportation sector. Aerosol particulate consists of approximately 75 percent carbon (soot) and 25 percent polycyclic aromatic hydrocarbons (PAHs) (VHB, 1991b). Soot is a major cause of smog. A number of PAHs found in diesel particulate are known or suspected carcinogens and mutagens.

Particulate fibres from the transportation sector also result from the wearing of brake linings and tires containing asbestos (OECD, 1988). Asbestos fibres may cause respiratory and cardiovascular problems in humans.

Visibility may be limited due to haze, and fabrics and surfaces may be discoloured due to the settling of particulates on exposed surfaces.

Sulphur Dioxide: Sulphur dioxide (SO_2) emissions from the transportation sector arise from the burning of fossil fuels and account for only about 2 percent of total SO_2 emissions in Canada. Sulphur is present in diesel fuel. Sulphur dioxide causes adverse respiratory problems and contributes to acid rain which indirectly affects health by increasing the risk of mercury or lead poisoning or intoxication from other elements in the environment. Acid deposition onto certain watersheds also damage the aquatic system through altering the pH level of the water. Sulphur dioxide emissions can also result in materials damage and decreased visibility.

The sulphur content of diesel fuel determines the quantity of particulate emissions emitted by diesel engines. High sulphur content appears to increase the organic soluble fraction of the particulate matter (Pilorusso, 1986). As particulate emissions from diesel engines are reduced, sulphur becomes a larger portion of total particulate emissions. Engine manufacturers and fuel producers in the United States have agreed that engines designed to meet particulate emission standards must have less than 0.05 percent sulphur content in diesel fuel (Transport Canada and Environment Canada, 1989). The result will be a further lessening of sulphur emissions from the transportation sector.

Chlorofluorocarbons: Chlorofluorocarbons, specifically CFC-12, are used as blowing agents and coolants in air conditioning systems and, CFC-11, in the production of foam padding and seats. The transportation sector is a small user of total chlorofluorocarbons. These synthetic chemicals contribute to stratospheric ozone depletion, and may contribute to global warming (IPCC, 1990).

4.2.2 Estimation of Loading

Baseline emissions of these contaminants in Canada are presented in Table 3. The transportation sector accounts for between 25 and 30 percent of all CO_2 emissions in Canada. Transportation is not a major contributor to SO_2 and particulate emissions — about 2 percent and 1 percent respectively of total estimated emissions in Canada. About one half of all NO_x, hydrocarbon and CO emissions in the country originate from the transportation sector. The relative contribution of each transportation mode to the emission of each contaminant of concern is also presented in Table 3. The major source of emissions in the transportation sector is the operation of passenger motor vehicles, except for SO_2, where heavy duty trucks, air, marine and rail modes of transport are the greatest sources of emissions.

Table 4 presents the estimated operating energy use and air emissions of major contaminants by mode of passenger transportation on a unit basis. Energy efficiency data are fleet averages unless otherwise stated. The data reflect the differences in vehicle occupancy between the various modes of passenger transport and provide a standard for comparison.

Table 4

ENERGY USE AND EMISSIONS BY MODE OF TRANSPORTATION[a]

Transport mode	Fuel type	Energy use (MJ/pass-km)		Carbon dioxide (g/pass-km)		Sulphur dioxide (g/pass-km)		Nitrous oxides (g/pass-km)		Hydrocarbons (g/pass-km)		Particulates (g/pass-km)		Carbon monoxide (g/pass-km)	
		Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Intercity bus transport	Diesel	0.52[b]	0.94[h]	36.75	66.70	0.06	0.11	0.50[b]	0.86	0.07	0.10	0.03	0.05	0.21[b]	0.40
Motor vehicles	Gasoline	1.70[b]	1.72	115.55	116.91	0.14	0.14	0.62[c]	0.78[e]	0.78[e]	0.97[e]	0.58	0.58	3.67[e]	5.81[e]
Air transport	Turbine gas	2.00[b]	4.59h	141.66	325.18	0.01	0.02	0.09	0.80[b]	0.06[b]	0.17	0.004	0.01	0.07[b]	0.34
Railroad transport (VIA only)	Diesel	1.73[h]	3.28[c]	122.28	232.03	0.20	0.37	2.47[g]	4.73	0.12	0.64[b]	0.07[b]	0.26	0.10[g]	1.87
Marine transport	Fuel oil	0.11	0.25	8.97	20.40	0.08	0.19	0.01	0.03	0.001	0.003	0.003	0.01	0.04	0.09

Sources: **Environment Canada, 1991. Transportation Systems Division. Khan, 1990. Statistics Canada, 1986. *Railway Transport in Canada, General Statistics, 1985*. Statistics Canada, 1990a. *Canadian Civil Aviation, 1989*. Statistics Canada,1990b. *Passenger Bus and Urban Transit Statistics, 1988*. Statistics Canada, 1990c. *Railway Operating Statistics 69*; 12. Statistics Canada, 1990d. *Shipping in Canada, 1989*. VHB Research and Consulting Inc., 1991. Royal Commission on National Passenger Transportation, 1991.**

Notes:

a Based on 1989 energy use (Statistics Canada, 1986; 1990a; 1990b; 1990c; 1990d and 1989 emissions coefficients VHB, 1991), unless otherwise stated.

b Khan, 1990.

c Statistics Canada (1986; 1990c).

d Based on estimated passenger-kilometres RCNPT (1991), Statistics Canada (1990d) and VHB (1991).

e Environment Canada (1991). Transportation Systems Division, Mobile 4 base case emissions, Ontario. Low estimate is for highway light-duty gasoline vehicle, high estimate for light-duty gasoline truck. Assumes 1.8 passengers per vehicle.

f Intercity bus data are for 1988.

g Based on fuel efficiency and emission factors for California passenger trains (Khan, 1990).

h VHB (1991) estimate of energy use for 1988.

The greatest amount of energy used on a passenger-kilometre basis (pass-km) is from the operation of passenger air service, between 2.00 and 4.59 MJ/pass-km. Rail service is the second largest energy user per passenger-kilometre, 1.73 to 3.28 MJ/pass-km. Marine transport is the most fuel efficient requiring only 0.11 to 0.25 MJ/pass-km. Intercity bus travel is the second most fuel efficient mode of passenger transport demanding 0.52 to 0.94 MJ/pass-km. Passenger motor vehicles required about 1.70 MJ/pass-km, in 1989.[9]

Emissions of each pollutant per passenger-kilometre are also presented in Table 4. The highest emission rate of CO_2 is from air and rail passenger transportation (between 140 and 325 g/pass-km and between 122 and 232 g/pass-km, respectively). The least amount of CO_2 is emitted from marine transport, between 9 and 20 g/pass-km. Passenger motor vehicles emit about 116 g/pass-km of CO_2 and intercity buses between 37 and 67 g/pass-km.

On a passenger-kilometre basis, the predominant source of SO_2 is passenger rail transportation, emitting 0.2 to 0.4 g/pass-km. The other modes emit less SO_2, about 0.02 g/pass-km from air transport, 0.06 to 0.11 g/pass-km from intercity bus, 0.14 g/pass-km from passenger motor vehicles and between 0.1 and 0.2 g/pass-km from marine vessels.

The major source of NO_x is rail transport, 2.5 to 4.7 g/pass-km. Private passenger motor vehicles emit between 0.6 and 0.8 g/pass-km and intercity bus transportation between 0.5 and 0.9 g/pass-km. Nitrous oxides emissions from air range greatly, between 0.1 and 0.8 g/pass-km and emissions from marine transport are small, between 0.01 and 0.03 g/pass-km.

Passenger motor vehicles emit between 0.8 and 1.0 g/pass-km of HC. The remaining modes of transport emit less. Air transport emits between 0.1 and 0.2 g/pass-km and rail transport between 0.1 and 0.6 g/pass-km. Intercity bus transportation emits about 0.1 g/pass-km. Emissions of HC from marine transport are negligible.

The highest emissions of particulates are from passenger motor vehicles, about 0.6 g/pass-km. Rail transport emits between 0.1 and 0.3 g/pass-km and intercity bus transport about 0.04 g/pass-km. Air and marine transport emit negligible amounts of particulates.

Private passenger motor vehicles are the largest contributor of CO emitting between 3.7 and 5.8 g/pass-km. Rail transportation emits between 0.1 and 1.9 g/pass-km. Carbon monoxide emissions from air transport vary between 0.1 and 0.3 g/pass-km, and from intercity bus transport emissions vary between 0.2 and 0.4 g/pass-km. Marine passenger transportation emits between 0.04 and 0.1 g/pass-km.

Detailed data on generation of HC, CO and NO_x for different types of private passenger motor vehicles are presented in Tables 5, 6 and 7. The data indicate the range of pollutant emissions possible. The difference in gasoline emission rates between provinces reflects differences in provincial fuel vapour-pressure standards (personal communication, François Terrillon, Environment Canada). Emissions also reflect the differences in season, climate, type of fuel used and age and maintenance of the vehicle. This does not permit the easy estimation of a national emissions factor. Diesel emissions do not vary significantly between jurisdictions, reflecting the insensitivity of diesel fuel to many of the factors which affect gasoline emissions.

Table 5

Highway Hydrocarbon Emissions by Province, 1989 (g/pass-km)

Province	Light-duty gasoline vehicle	Light-duty gasoline truck	Light-duty diesel vehicle	Light-duty diesel truck	Motorcycles
British Columbia	1.05	0.97	0.13	0.17	1.59
Alberta	0.94	0.96	0.13	0.17	1.72
Saskatchewan	0.97	1.03	0.13	0.17	1.99
Manitoba	1.02	1.03	0.13	0.17	1.97
Ontario	0.69	0.97	0.12	0.17	2.13
Quebec	0.63	0.97	0.12	0.17	2.02
New Brunswick	0.59	0.89	0.11	0.17	1.55
Nova Scotia	0.66	0.93	0.12	0.17	1.78
Newfoundland	0.67	0.93	0.12	0.17	1.68

Source: Environment Canada, 1991. Transportation Systems Division.

Note: Assumes 1.8 passengers per vehicle.

Table 6

HIGHWAY CARBON MONOXIDE EMISSIONS BY PROVINCE, 1989 (G/PASS-KM)

Province	Light-duty gasoline vehicle	Light-duty gasoline truck	Light-duty diesel vehicle	Light-duty diesel truck	Motorcycles
British Columbia	6.25	5.50	0.28	0.31	6.32
Alberta	5.95	6.20	0.28	0.31	6.42
Saskatchewan	5.98	6.56	0.28	0.31	6.15
Manitoba	6.37	6.53	0.28	0.31	6.12
Ontario	3.67	5.81	0.27	0.31	6.00
Quebec	3.37	6.03	0.27	0.31	6.00
New Brunswick	3.33	6.03	0.27	0.31	6.44
Nova Scotia	3.61	5.94	0.27	0.31	6.14
Newfoundland	3.68	6.09	0.27	0.31	6.27

Source: **Environment Canada, 1991. Transportation Systems Division.**

Note: **Assumes 1.8 passengers per vehicle.**

Table 7

HIGHWAY NITROGEN OXIDES EMISSIONS BY PROVINCE, 1989 (G/PASS-KM)

Province	Light-duty gasoline vehicle	Light-duty gasoline truck	Light-duty diesel vehicle	Light-duty diesel truck	Motorcycles
British Columbia	0.90	0.77	0.50	0.57	0.73
Alberta	0.85	0.80	0.50	0.57	0.73
Saskatchewan	0.84	0.81	0.50	0.57	0.72
Manitoba	0.87	0.81	0.50	0.57	0.71
Ontario	0.62	0.78	0.50	0.57	0.71
Quebec	0.58	0.79	0.49	0.57	0.71
New Brunswick	0.59	0.80	0.49	0.57	0.74
Nova Scotia	0.62	0.79	0.49	0.57	0.71
Newfoundland	0.63	0.79	0.50	0.57	0.73

Source: **Environment Canada, 1991, Transportation Systems Division.**

Note: **Assumes 1.8 passengers per vehicle.**

The light-duty gasoline motor vehicle emissions for HC, CO and NO_x given in Table 4 are based on emission factors provided in Tables 5, 6 and 7 for Ontario — the province having the largest proportion of passenger cars registered in Canada.

Limitations of estimates: The emissions data presented in Table 4 are approximations of the expected emissions of each pollutant by transportation mode. It is difficult to estimate a "typical" set of emission factors for any pollutant or mode of transport — age and composition of the vehicle, vehicle fuel efficiency, level of vehicle maintenance, climatic conditions, regulatory standards across jurisdictions, vehicle speed, emissions control technology, vehicle occupancy, fuel vapour pressure and combustion chamber design, all affect emission rates.

Pollutant emissions on a g/pass-km basis also vary based on the occupancy of the transportation vehicle. For example, if the NO_x emissions presented in Table 4 — 4.7 g/pass-km — are for a passenger rail car at 25 percent occupancy then the NO_x emissions for the rail car at 50 percent occupancy would fall to 2.4 g/pass-km. As such, the range of emission factors presented in Table 4 implicitly embody a fixed level of occupancy per vehicle for each mode of transportation.

4.2.3 Estimation of the Environmental Fate of Discharges

After discharge, pollutants may react with other compounds in the environment leading to forms of environmental damage other than those associated with direct deposition. This environmental damage is often referred to as indirect or secondary effects of emissions, though the potential damage caused by those secondary effects may be much more significant than those caused by direct discharge. Secondary damage of concern includes:

- global warming from the emission of greenhouse gases;
- stratospheric ozone depletion;
- the formation of tropospheric ozone or photochemical smog; and
- acid deposition.

Table 8 indicates the individual pollutants from transportation that contribute to these potential causes of environmental damage.

Table 8

SECONDARY EFFECTS OF ATMOSPHERIC POLLUTANTS FROM THE TRANSPORTATION SECTOR

Pollutant	Global warming	Ozone depletion	Photo chemical smog	Acid deposition
Carbon dioxide	•			
Carbon monoxide			•	
Nitrogen oxides	•[a]		•[a]	•
Chlorofluorocarbons	•	•		
Ozone[b]	•		•	
Sulphur oxides				•
Hydrocarbons	•[a]		•[a]	

Source: Barakat and Chamberlin, 1990.

Notes:

a **Ozone is formed in the atmosphere from NO_x, HCs, water vapour (H_2O) and sunshine. NO_x concentrations are believed to be the determinant of the rate of ozone formation.**

b **Ozone is not emitted but the transportation sector is a major contributor of the principal precursors, NO_x and HCs.**

Global warming:[10] The transportation sector produces greenhouse gases through the combustion of fossil fuels. These gases, primarily CO_2, contribute to global warming. In addition, air conditioners and foam padding in transport vehicles may emit CFC-12 which is also a greenhouse gas.

The Earth's atmosphere has a limited capacity to absorb gaseous wastes and the emissions of greenhouse gases from man-made sources may change the world's ecosystems and radically alter the world's climate (Ottinger et al., 1990, p. 127).

Many of the gases emitted by the transportation sector, including CO_2, CH_4, N_2O, O_3 and CFCs,[11] absorb the infrared radiation (heat). Heat is reradiated back onto the Earth's surface, rather than into space, causing the temperature of the Earth's surface to rise. There is scientific consensus that increases in greenhouse gas emission will result in climate change (IPCC, 1990). Although the nature and extent of the change, as well as its ramifications for human well-being, are uncertain.

Stratospheric ozone depletion: The transportation sector emits pollutants which contribute to stratospheric ozone depletion. Of concern are chlorofluorcarbons used in air conditioning systems and foam cushioning. The destruction of the stratospheric ozone layer permits an increased amount of ultraviolet radiation to reach the Earth's surface, which may result in human health, crop, forest and materials degradation.

Acid deposition: Acid rain is caused by emissions of SO_2 and NO_x. Once released into the atmosphere, these substances can be carried long distances by the prevailing winds, and return to Earth in acidic forms as rain, snow, fog or dust. Environmental damage occurs if the acid precipitation cannot be neutralized.

The main sources of SO_2 emissions in Canada are coal-fired power generating stations and industry (non-ferrous ore smelters); the transportation sector contributes a minor amount, about 2 percent of total emissions. The main sources of NO_x emissions, however, are motor vehicles through the combustion of fossil fuels.

Acid rain leads to the acidification of lakes and streams. In some cases aquatic life is depleted. Acid deposition on a watershed causes an increase in the aluminum concentration of freshwater bodies within that watershed; aluminum is more toxic to some aquatic biota than is acidity.

Increases in the acidity of soil, water and shallow ground water are also suspected of causing forest and vegetation damage. Acid rain damages buildings and monuments, and is suspected of causing respiratory illness in humans.

Ozone: Observed for decades in urban areas, ozone is now found in rural areas, especially in the summer (OECD, 1991). Both nitrogen oxides and hydrocarbons contribute to the creation of ozone. The transportation sector, particularly private motor vehicles, is a major source of these pollutants.

The environmental damage caused by ozone includes loss of agricultural crop productivity, human respiratory problems and materials soiling and degrading.

4.2.4 Assessment of the Exposure of Receptors to Discharges

As pollutants move away from the emission source they may be deposited or diluted. They may also break down. Very few data are available on the dispersion patterns and concentrations of pollutants emitted on transportation routes, but impacts can be expected to diminish with distance.

Table 9 shows the dispersion of lead and zinc emitted from road vehicle exhausts. The data in the table indicate that concentration of lead decreases with distance from the road edge.

Dispersion patterns and concentrations of gaseous pollutants depend upon meteorological conditions and can be modelled using computer simulations. Of all possible road situations, rural highways are the easiest to model because the background levels of some pollutants are relatively low when compared to areas where there is a dense road network.

Most models are based on observed data for CO. CO is the easiest gaseous transportation pollutant to model because transportation is the predominant source of this type of emission, and the effect of background concentration is relatively unpronounced.

Table 9

Dispersion of Lead and Zinc Emissions from Road Edge

Distance from road edge (m)	Lead concentration (ppm, d.w.)				Zinc concentration (ppm, d.w.)			
	Road A 50,000	Road B 18,500	Road C 16,000	Road D 3,000	Road A 50,000	Road B 18,500	Road C 16,000	Road D 3,000
0	3,045	858	1,075	465	880	422	272	106
1	2,813	402	457	118	700	198	167	69
5	342	177	136	32	144	122	79	75
15	223	75	163	26	150	55	92	59
30	—	45	63	26	—	63	—	65
50	223	45	95	38	95	69	58	64
100	—	—	60	21	—	—	76	78
Background	14				60			

Source: Freedman (1989) modified from Dale and Freedman (1982).

The rate of fall-off for gaseous pollutants is very rapid. There is little difference in the fall-off rates for each of the gaseous transportation pollutants (personal communication with Rob Bloxam). There are some complications in the modelling of nitrogen oxides due to chemical conversions which take place during transport. Particulate fall-out is particularly difficult to model due to the abundance of sources in the background.

4.2.5 Assessment of Toxicological Data and Likely Mortality or Morbidity Implications of Exposure

Human health: Atmospheric pollutants can cause adverse health effects varying from short- term illnesses (morbidity) to death (mortality). In the transportation sector the pollutants which pose the greatest health risk are NO_x, SO_2, O_3, H_2SO_4 and HNO_3. Table 10 presents the expected health effects for the major pollutants associated with the transportation system.

Table 10
HEALTH EFFECTS BY TYPE OF EMISSION

Pollutant	Health effects
Sulphur dioxide	• respiratory illness for people sensitive to asthma and bronchitis
Carbon monoxide	• interferes with red blood cells and the ability of the body to absorb oxygen • lowers worker productivity • impairs nervous system, coordination, vision, judgement • exacerbates cardiovascular disease symptoms (angina, etc.) • affects the fetus, young children and sickle cell anemics (lower birth rates and increased mortality) • works with other air pollutants to promote morbidity through respiratory and circulatory ailments
Nitrous oxide	• lowers ability of lung to function properly; irritation, emphysema • respiratory problems, coughing, runny noses, sore throat in children • works with other air pollutants to promote morbidity through respiratory and circulatory ailments
Hydrocarbons	• leukemia, mutagens and carcinogens • skin irritation
Particulate matter	• respiratory problems • infant mortality and total mortality in urban areas — bronchitis, asthmatics and cardiovascular patients (already ill) • asbestos (lung cancer, asbestosis, mesothelioma — cancer of lung linings and abdomen)
Ozone	• damage to membranes of the respiratory system • respiratory distress • eye irritation

Sources: OECD, 1988; Freedman, 1989.

Terrestrial Biota: The transportation emissions which are of primary concern for terrestrial ecosystems are nitrogen oxides. Nitrogen oxides affect terrestrial ecosystems through a number of pathways:

- direct fumigation or deposition on the plant;
- the development of photochemical oxidants (specifically ozone) arising from the reaction of NO_x and hydrocarbons with sunlight; and
- through the combination of NO_x with atmospheric water to form acid deposition.

The primary impact pathways of transportation related to airborne emissions on terrestrial systems (and the pollutants corresponding with each of these) are as follows:

- foliar damage due to fumigation (SO_2, O_3, NO_x);
- foliar damage due to deposition (SO_2, H_4SO_4, HNO_3);
- long-term accumulations in soil (H_2SO_4, HNO_3); and
- physiological stress on livestock and wildlife (SO_2, O_3, NO_x).

Although there is a great deal of concern regarding the effects of these pollutants on forest resources, the body of scientific evidence which links specific ambient levels of pollutants to forest damage is meagre. There are a number of reasons for this paucity of information:

- Short duration events are often obscured.

  Pollutant dose is often expressed as a function of the pollutant concentration. This dose-response technique assumes a normal distribution of ambient pollutant concentration over time. Hourly pollutant concentrations (such as, ozone) often vary greatly. Acute effects of periodic short-term high-pollutant concentrations typically differ greatly from chronic effects resulting from long-term exposure (Concord Environmental and VHB, 1991). Standard pollutant exposure estimation techniques tend to underestimate impacts.

- There is a lack of correlation between the findings of laboratory impact assessments and actual changes in yield.

  Many scientific experiments are designed to detect changes to the morphology and physiology of plants. In most cases, however, these changes have not been correlated with changes in yield. This gap severely hampers the use of these results for policy analysis.

Furthermore, observations made in controlled conditions often do not accurately depict what will happen in the natural environment. Opponents to more stringent controls often cite this failure to demonstrate damage under "real world" conditions as a principal argument against tighter regulations (Concord Environmental and VHB, 1991).

- General observations of terrestrial damage have not been validated.

  Validation of the applicability of laboratory results to "real world" conditions can be achieved either through empirical studies or through the development of mechanistic models. The effect of ozone on forest resources is a good example of the difficulty in achieving adequate validation. Although forest decline has been carefully monitored in many locations strict dose-response relationships cannot be developed due to the inherent variability in environmental conditions and the potential for a myriad of factors contributing to the observed response in varying degrees and means (Concord Environmental and VHB, 1991).

These limitations of our scientific understanding pervade the environmental impact literature and are inescapable. Comments do not suggest that the scientific literature should not be used to the greatest extent possible; instead, they suggest caution in their interpretation and the need to be vigilant in recognizing the inherent variability and complexity in the interactions and impacts described.

Forest resources: Although there have been numerous studies on seedling response to nitrogen oxides, ozone, and acid precipitation (Mohnen, 1989; Peterson et al., 1989; Reich et al., 1989; Percy, 1986; Jensen and Patton, 1990; Chappelka and Chevone, 1989), little is known about the effect of these pollutants on mature trees. Most of the dose-response studies use changes in seedling growth as a surrogate for mature trees due to seedling size and manageability. Unfortunately, a quantitative link between the surrogate indicator and the yield has not been developed. Filling in these gaps is crucial if the data provided are to be effectively used in determining the effects of transportation emissions on forest resources. The absence of specific dose-yield response curves for valued resources does not imply that years of research are necessary before these can be produced. Undoubtedly, more research can and should be conducted. Expert judgement and logic can be used to fill in the gaps as an interim measure. This is an essential step for effective policy analysis.

Table 11 summarizes the availability of data relating exposure levels to forest damage.

Table 11

SUMMARY OF THE AVAILABILITY OF DATA RELATING EXPOSURE LEVELS TO FOREST DAMAGE

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available
coniferous trees	fumigation on plant	nitrogen oxides	NAPAP concludes, "NO_x at ambient concentrations is not a direct source of regional scale growth reductions in U.S. forests."[a]	no
		photo-chemical oxidants	no	Both losses and gains of above ground biomass of Pinus taeda in response to ozone exposure are evident and no definitive trend can be discerned.
		dry acid deposition	no	no
	deposition on plant	NO_x	no	Visible foliar injury of red spruce increases with an increase in nitrogen mists. Data is provided in kg/ha/yr; however, such spatial deposition has been determined for only a few sites in Canada.
		acid rain	no	Significant decreases in germination and seedling survival are observed only when the pH of rainfall is below 3.6. This level of acidity is generally not encountered in ambient rainfall. (Significant damage can occur at high elevations (acid fog).)
	deposition on soil		no	no (Most significant impacts: change in Al availability; depletion of Ca and Mg.)
deciduous trees	fumigation	NO_x	NAPAP concludes that "NO_x at ambient concentrations is not a direct source of regional scale growth reductions in U.S. forests."[a]	no

Table 11 (cont'd)
SUMMARY OF THE AVAILABILITY OF DATA RELATING EXPOSURE LEVELS TO FOREST DAMAGE

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available
deciduous trees (cont'd)		photo-chemical oxidants	no	Studies on the effects of ambient ozone on deciduous seedlings have rendered mixed results. (Recent studies looking at the combined effects of ozone and acid rain have rendered mixed results.)
		dry acid deposition	no	no
	deposition on plant	nitrogen oxides	no	no
		acid rain	no	No significant reductions in germination and seedling survival are recorded until pH < 3.6, well below observed ambient rainfall
	deposition on soil		no	no (Most significant impacts: change in Al availability; depletion of Ca and Mg; acidity of soil.)

Source: Concord Environmental and VHB, 1991.

Note:

a U.S. Environmental Protection Agency, National Acid Precipitation Assessment Program ("State of Science & Technology" Series), *Changes in Forest Health and Productivity in the United States of America*, prepared by J.E. Barnard and A.A. Lucier, 1989.

Agricultural Resources: The effect of ozone on yield has been well documented in recent years through the USEPA's National Crop Loss Assessment Network (NCLAN). Figures 1, 2, 3 and 4 show dose-response curves developed for a variety of row and cereal crop species. Losses in yield can occur at well below concentrations observed in Canadian urban areas (see Figure 5). The effect of ozone on fruit species has not been documented.

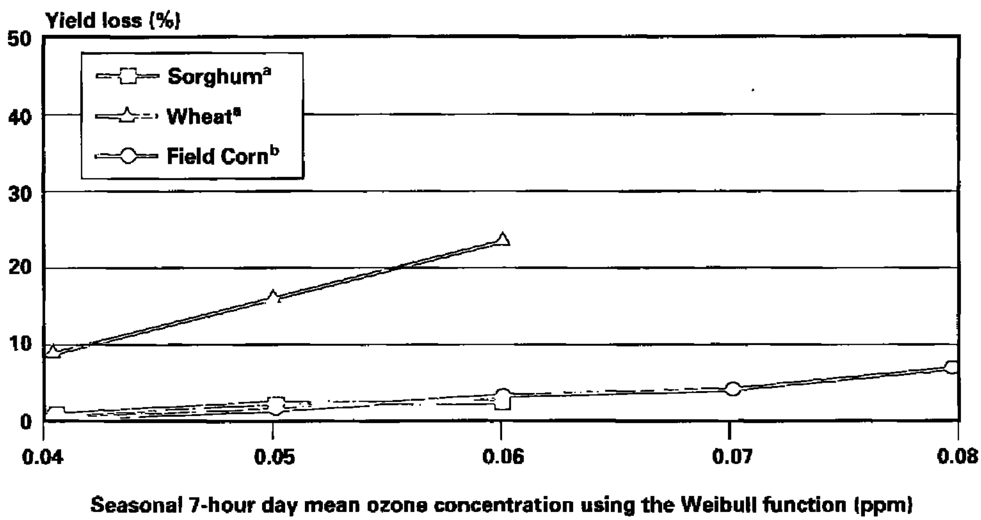
Nitrogen oxides are not thought to cause significant damage to crops. Very high concentrations (>4 ppm of NO_2 for 10 hours) are required for leaf injury to occur. These concentrations are well above concentrations observed in

Canadian urban areas (see Figure 6). In low concentrations, nitrogen dioxides should be accounted for as a source of the necessary nitrogen requirements of plant materials. The fertilizer effect has been observed in a number of studies (Heck, 1989; Lesser et al., 1989; Phytotoxicology Consulting Services, 1989; Victor and Burrell, 1982; DPA, 1987).

Crops exposed to simulated acidic precipitation under controlled conditions showed either no effect on growth or yield, or mixed results at ambient levels. Normal agricultural practices of liming generally prevent nutrient or soil acidification, although the additional costs to ameliorate these effects should be included as part of the impact assessment (Victor and Burrell, 1982). Figures 7 through 14 provide dose-response functions for various cereal, row and fruit crops.

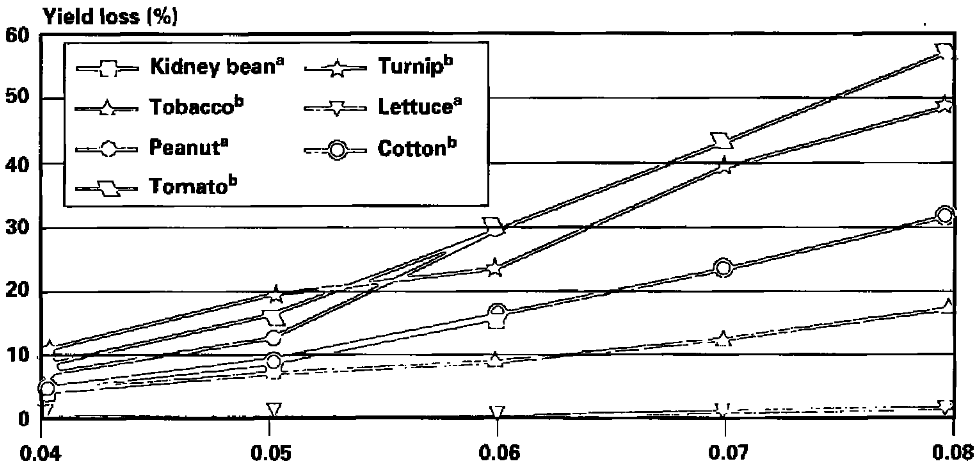
Table 12 summarizes the availability of data relating exposure levels to crop damage.

Figure 1
The Effect of Ozone on Cereal Crop Yield (7-Hour Mean)



a Lesser et al., 1989
b Heck et al., 1988

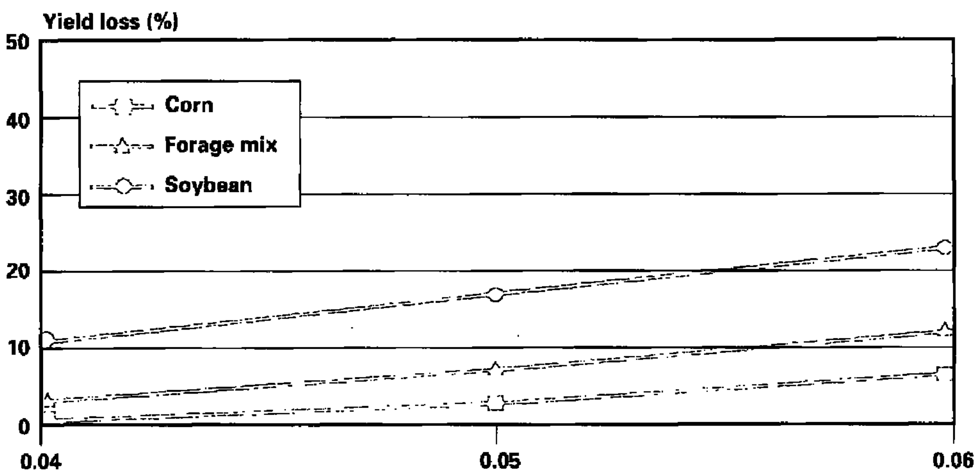
Figure 2
The Effect of Ozone on Row Crop Yield (7-Hour Mean)



Seasonal 7-hour day mean ozone concentration using the Weibull function (ppm)

a Lesser et al., 1989
b Heck et al., 1988

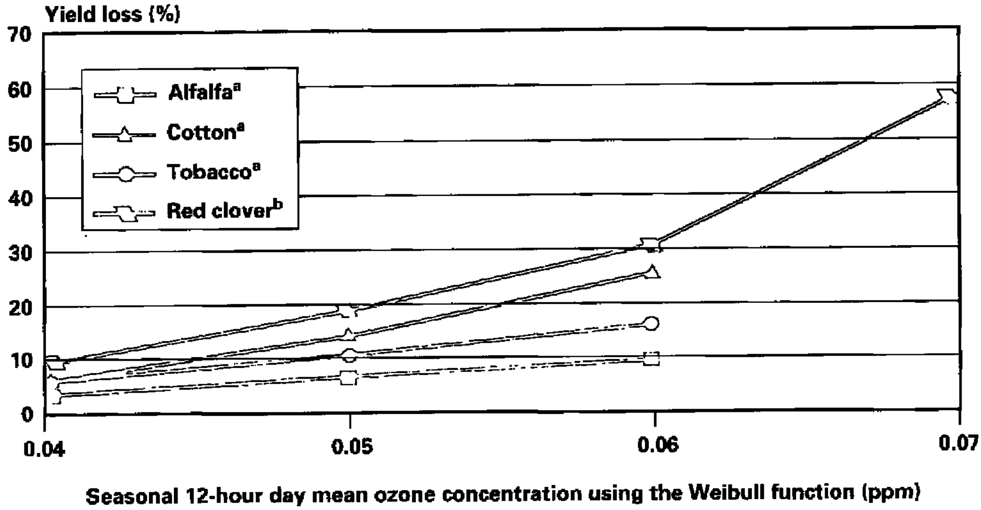
Figure 3
The Effect of Ozone on Cereal Crop Yield (12-Hour Mean)



Seasonal 12-hour day mean ozone concentration using the Weibull function (ppm)

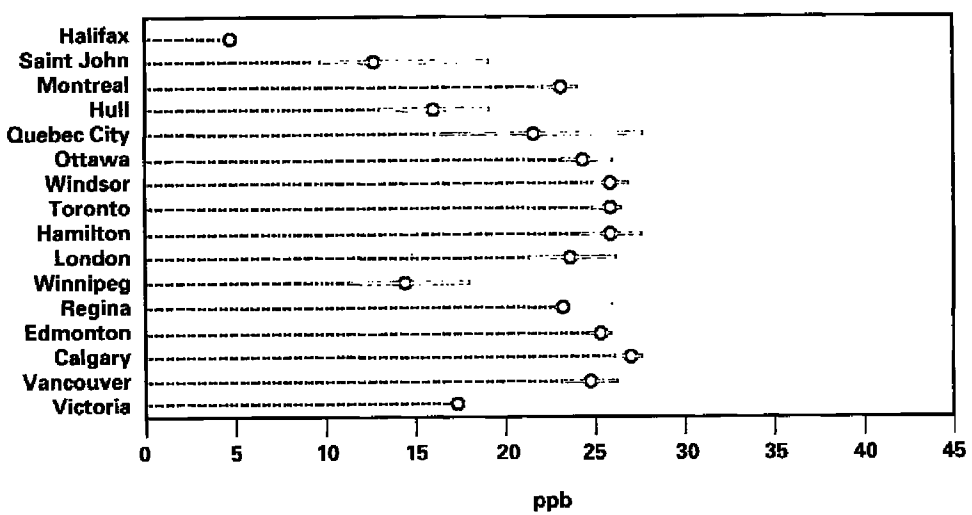
Source: Lesser et al., 1989

Figure 4
THE EFFECT OF OZONE ON ROW CROP YIELD (12-HOUR MEAN)



a Lesser et al., 1989
b Heck et al., 1988

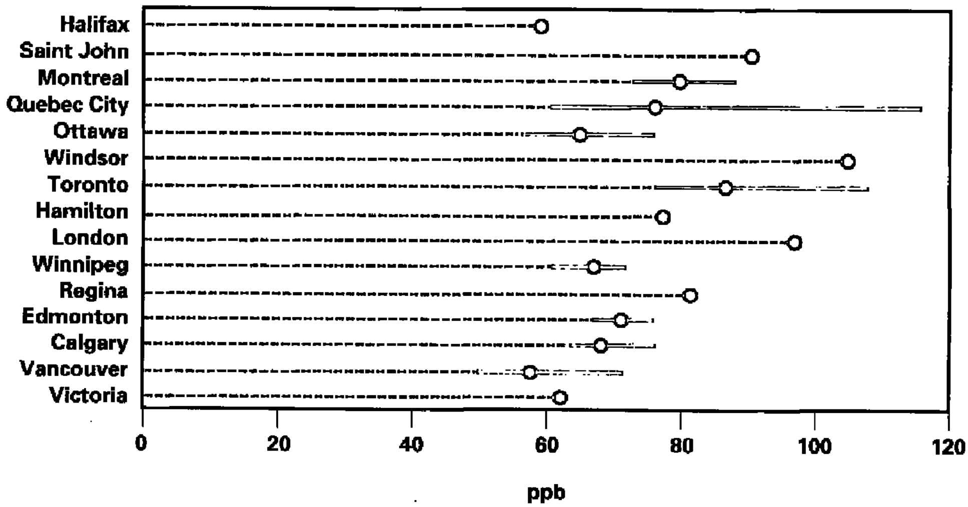
Figure 5
ANNUAL AVERAGE LEVELS OF NO_2 IN SELECTED CANADIAN CITIES



Source: Environment Canada, 1990.

Figure 6

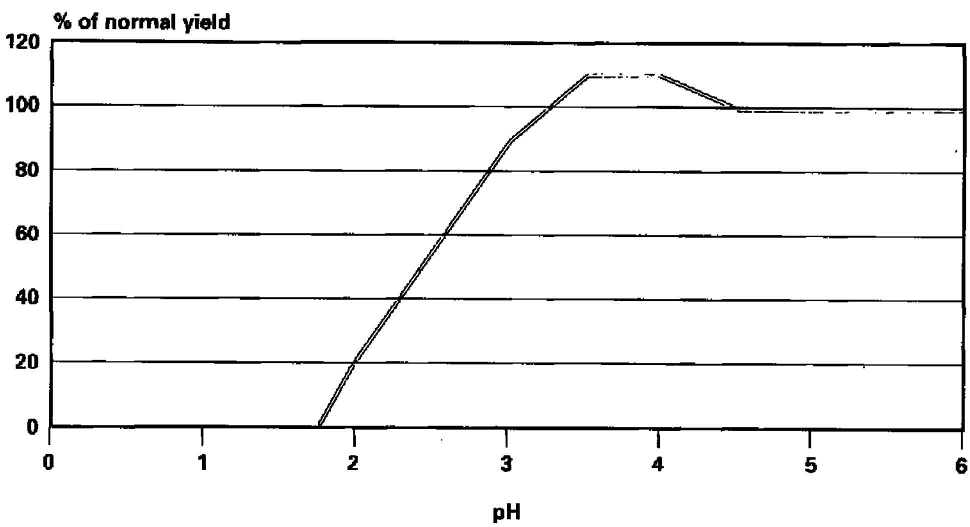
AVERAGE OF PEAK HOURLY OZONE LEVELS IN SELECTED CANADIAN CITIES, 1983–1987



Source: Environment Canada, 1990.

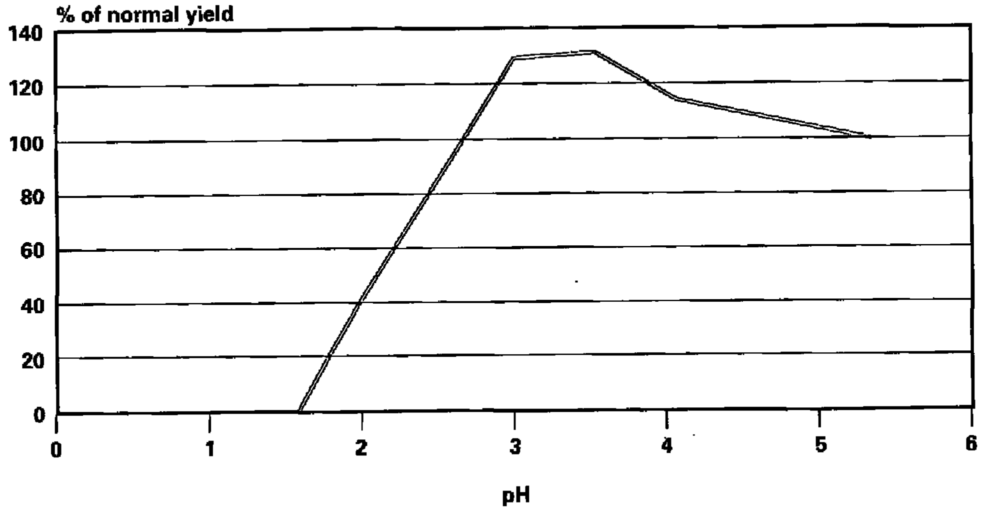
Figure 7

DOSE-YIELD RESPONSE OF COLE CROPS TO PH OF RAINFALL



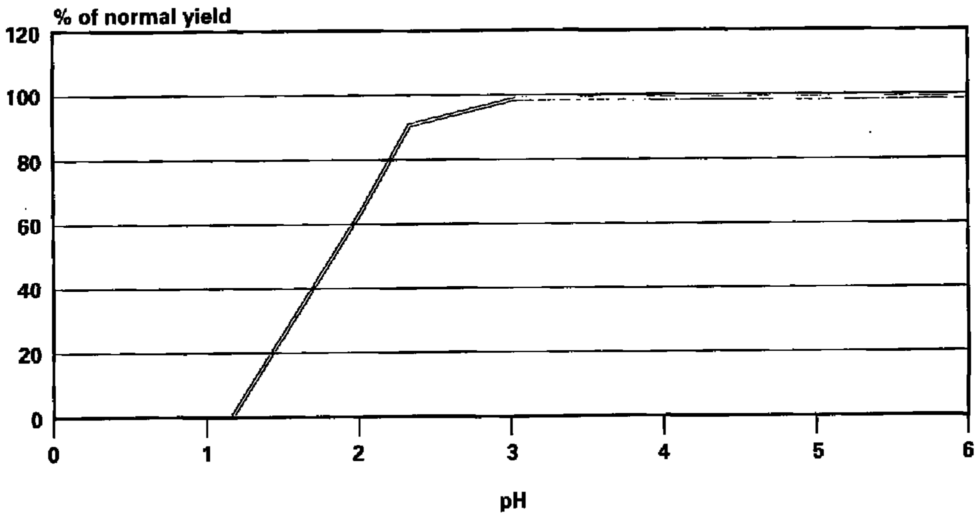
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 8
DOSE-YIELD RESPONSE OF FRUIT CROPS TO pH OF RAINFALL



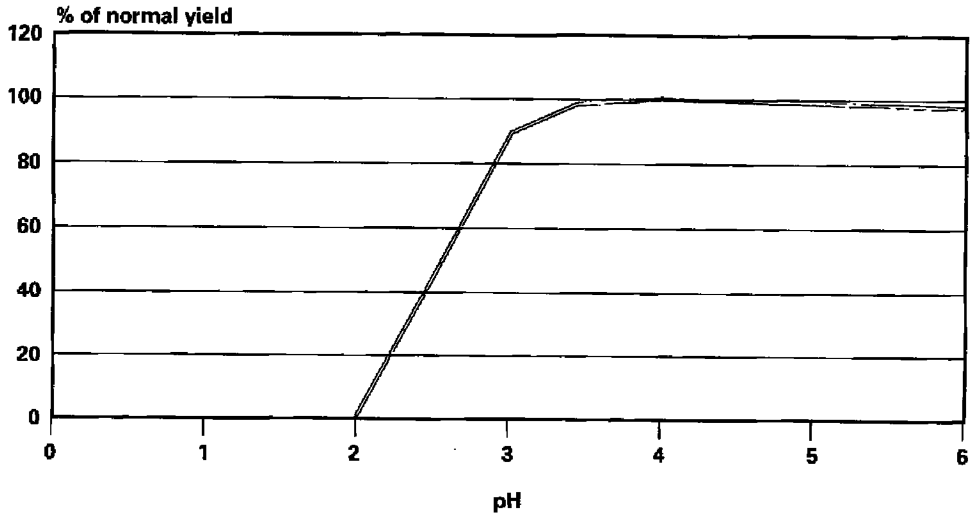
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 9
DOSE-YIELD RESPONSE OF GRAINS AND FORAGE TO pH OF RAINFALL



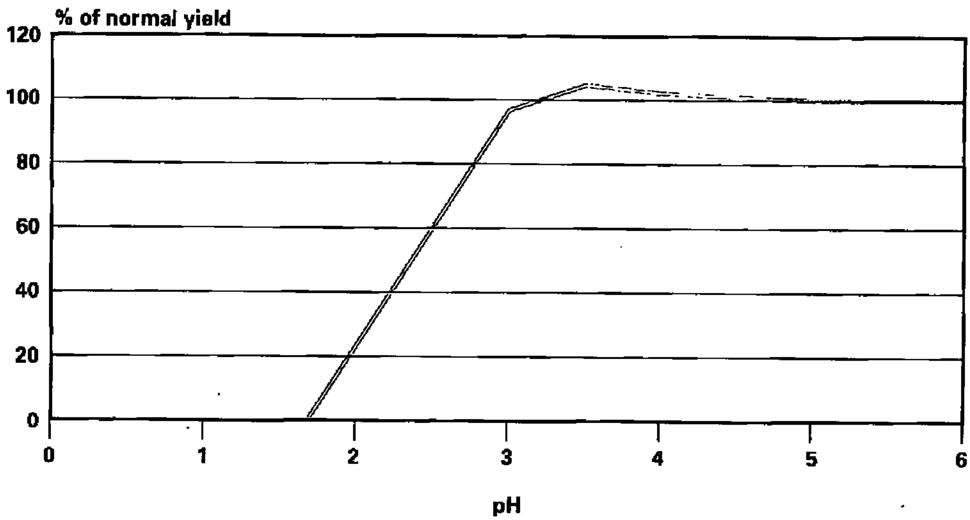
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 10
DOSE-YIELD RESPONSE OF LEAF CROPS TO pH OF RAINFALL



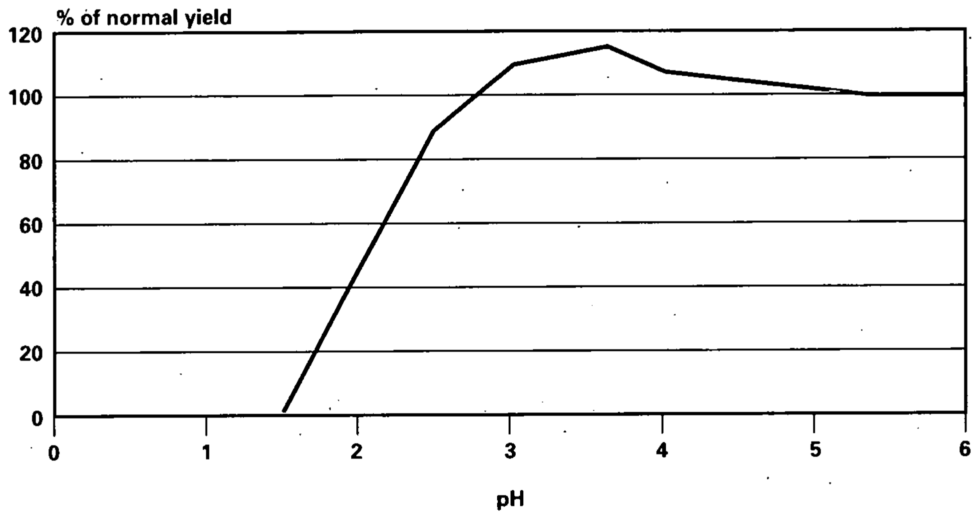
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 11
DOSE-YIELD RESPONSE OF LEGUMES TO pH OF RAINFALL



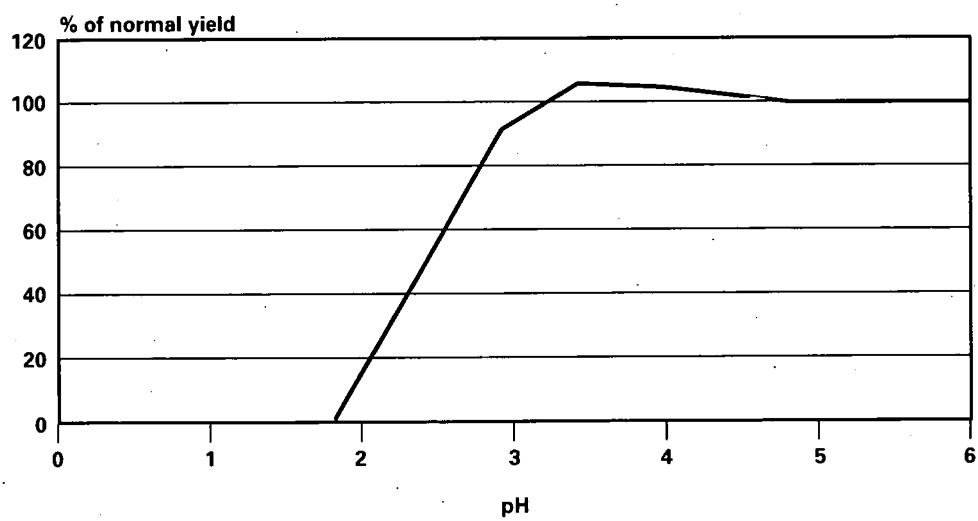
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 12
DOSE-YIELD RESPONSE OF ONIONS TO pH OF RAINFALL



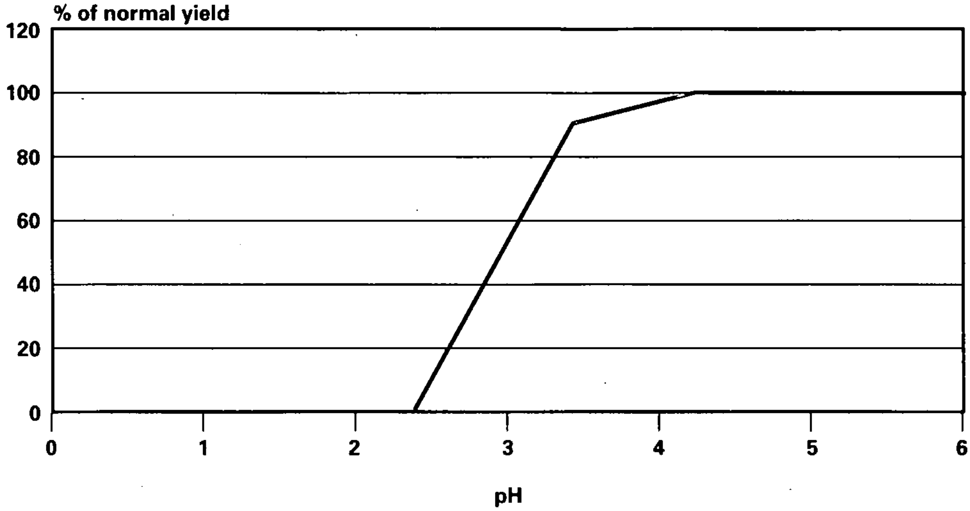
Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 13
DOSE-YIELD RESPONSE OF POTATOES TO pH OF RAINFALL



Sources: Victor and Burrell, 1982; DPA, 1987.

Figure 14
Dose-Yield Response of Root Crops to pH of Rainfall



Sources: Victor and Burrell, 1982; DPA, 1987.

Table 12
Summary of the Availability of Data Relating Exposure Levels to Crop Damage

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available	Notes
cereal crops	fumigation	nitrogen oxides	no	no	deposition causes acidification of soil; however, land management practices generally negate the impact of acidification on soil
		photochemical oxidants	yes	no	variation in response due to exposure length
		dry acid deposition	no	no	
	deposition	nitrogen oxides	no	no	in low concentrations NO_x may cause increased rather than decreased yield due to the fertilizer effect
		acid deposition	yes	no	

Table 12 (cont'd)
SUMMARY OF THE AVAILABILITY OF DATA RELATING EXPOSURE LEVELS TO CROP DAMAGE

Species	Pathway	Pollutant	Dose-yield available	Dose-surrogate response available	Notes
row crops	fumigation	nitrogen oxides	no	no	deposition causes acidification of soil; however, land management practices generally negate the impact of acidification on soil
		photochemical oxidants	yes	no	variation in response due to exposure length
		dry acid deposition	no	no	
	deposition	nitrogen oxides	no	no	in low concentrations NO_x may cause increased rather than decreased yield due to the fertilizer effect
		acid deposition	yes	no	

Source: Concord Environmental and VHB, 1991

Aquatic biota: Of the pollutants being examined, only nitrogen oxides deposited in the form of acid precipitation appear to have direct adverse effects on aquatic ecosystems. Sulphur dioxide from transportation is a minor contributor to acid deposition (OECD, 1988).

Aquatic ecosystems are affected either through direct acidic deposition on the surface waters or through acidification of soils in the watershed which reduces the neutralizing capability of the run-off water. The main effects of increased acid deposition on surface water chemistry include:

- decrease in lake pH;
- decline in lake alkalinity;
- increase in concentrations of trace metals (Hg, Pb, Cd, Zn, Ni, Mn); and
- increase in concentrations of organic and inorganic aluminium.

These changes in surface water chemistry have been shown to cause changes to the production and richness of aquatic communities. Rapid depressions in pH usually occur during spring snowmelt. Acid pulses can hinder reproduction, especially in river spawning species since the largest increases in surface water acidity occur at shallow depths. The degree to which fish species are affected by acid pulses depends on the species and the life stage of the fish.

The biological recovery of lakes after deposition is uncertain. Studies in the Sudbury area suggest that biological recovery does not occur, probably because of the high residual concentrations of trace metals and acidic compounds. The liming of lakes may result in some biological recovery. Whether the recovery will result in the same biological community that existed before remains uncertain (Concord Environmental and VHB, 1991).

The impacts of NO_x emissions on fish populations cannot be described in terms of simple dose-response curves but instead requires extensive modelling to show the contribution of NO_x emissions to surface water acidity and the effects of changes in surface water chemistry to aquatic biota. Response models typically use pH and aluminium toxicity thresholds for individual species as the primary input mechanism.

Table 13 summarizes the availability of data relating acidification to aquatic resources.

Table 13

SUMMARY OF THE AVAILABILITY OF DATA RELATING ACID DEPOSITION TO CHANGES IN AQUATIC RESOURCES

Species	Pollutant and pathway	Dose-yield available	Dose-surrogate available	Link between surrogate and yield known
fish species	acid deposition	Yield models relying on toxicity thresholds of individual species available.	no	no
invertebrates	acid deposition	Existing fish-yield models can be adapted to estimate response of invertebrates.	no	no
aquatic vegetation	acid deposition	no	no	no

Materials: The pollutants of primary concern for materials damage are:

- nitrogen oxides (NO_x) — primarily nitrogen dioxide (NO_2);
- ozone — arising from the reaction of NO_x with sunlight; and
- particulates.

The transportation sector contributes only minor amounts to the total SO_2 concentrations in the atmosphere, however SO_2 and NO_x combined contribute significantly to acid deposition (VHB, 1991a). Acid deposition has long been implicated as a major factor in material degradation. Carbon dioxide is also emitted from motor vehicles and has been linked to the corrosion of metals. The contribution of CO_2 to current ambient concentrations of CO_2 in the atmosphere is minor.

The primary impact pathways for materials damage are:

- corrosion or abrasion due to fumigation and deposition; and
- soiling of surfaces due to fumigation and deposition.

The pollutants corresponding with each of the pathways listed above are as follows:

- NO_x, O_3, HNO_3, H_2SO_4; and
- particulate emissions.

There are few data on the actual effects of these pollutants on materials and most damage is expected to occur in urban areas. Much of the research has concentrated on the effects of sulphur dioxide and acid deposition on materials (Acres, 1991). There are few quantitative data on the effects of other pollutants on materials. Where data are available, they are usually based on lab conditions and are at levels which are orders of magnitude above those existing in the Canadian urban environment. No minimum effect level has been established, and there is no evidence that existing levels in the Canadian environment adversely affect materials.

Table 14 lists the pollutants thought to have some damaging effect on materials.

Table 14

MATERIALS DAMAGE AND THE POLLUTANTS INVOLVED

Material	Effect	Cause	Comments/references
metal materials such as carbon steel, zinc, aluminum and copper	For nickel and zinc there is no significant effect at exposure levels of 1 ppm ozone or 0.5 ppb nitrogen dioxide. The mean annual concentrations of NO_2 and ozone in most urban environments in Canada are significantly lower than this.	NO_2, HNO_3, ozone	Ahuja and Amar (1988) found that exposure of 10 ppm nitrogen dioxide had mild effects on galvanized steel.
paint	Particulates cause the soiling of paints. There is no evidence that NO_2, HNO_3 and ozone affect the rate of paint deterioration at levels found in the Canadian urban environment.	Particulate, NO_2, HNO_3, ozone	Continually changing paint formulations make damage difficult to assess. In addition, it is difficult to distinguish the effects of pollutant deposition and natural weathering processes from ultra-violet radiation, moisture and temperature fluxes.
building stones	Carbonate stone is very susceptible to acidic deposition which causes the formation of a crust on the stone. Limestone exposed to acid rain deteriorates about 10 times faster than when exposed to SO_2 or NO_2 at 90% relative humidity individually (Lindquist et al., 1988). Particulate may catalyze the oxidation of SO_2 and NO_x making them more effective reactants. The contribution of acidic deposition to physical weathering is unknown (NAPAP, 1987).	SO_2, Particulate, NO_2, HNO_3, ozone	Nitrogen dioxide, sulphur dioxide and ozone can increase the rate of calcareous deterioration. There is no evidence, however, that the levels of nitrogen dioxide and ozone found in the Canadian urban environment can significantly affect the deterioration rate of building stone. Haneef (1990) reported a synergistic effect from sulphur dioxide, ozone and nitrogen dioxide compared with exposure to each pollutant alone.
concrete, brick and mortar	Lime mortars are sensitive to acidity due to the acid soluble nature of calcium and magnesium carbonates.	Particulate, NO_x	Concrete is not considered sensitive to acidic deposition due to its thickness.

Table 14 (cont'd)
MATERIALS DAMAGE AND THE POLLUTANTS INVOLVED

Material	Effect	Cause	Comments/references
glass	Soiling due to soil and road dust. Damage to medieval stained or painted glass.	Particulate	Glass is generally quite resistant to the effects of acidic deposition.
wood	Degradation.	NO_x, ozone	For wood treated with a preservative it is the coating that is important in determining the effects.
asphalt	Little is known about the effects of air pollutants on asphalt materials.	ozone	
rubber		ozone	The contribution of ozone to rubber degradation has been significantly reduced by the inclusion of anti-ozonants in the manufacture of tires.
fabrics	Fading of dyes. Particulates may cause the soiling of fabrics.	NO_x, ozone	Much of the research has concentrated on higher levels of nitrogen dioxide and ozone than found in the ambient air.

Source: Concord Environmental and VHB Research and Consulting Inc., 1991.

4.2.6 Assessment of Individuals or Populations at Risk

"Risk is the potential for realization of unwanted, negative consequences of an event" (Rowe, 1977, p. 24). Risk is expressed as the probability of mortality or morbidity for an individual or entire population occurring as a result of a certain level of exposure. Assessing the populations at risk requires knowledge of the following:

- the duration and level of exposure to pollutants at varying distances from the source;[12] and
- dose-response functions for each of the potential receptors to each of the pollutants.

Though some dose-response functions are available (human response to direct deposition of pollutants; response of forest species and crops to varying ozone concentrations; and response of some aquatic species to various levels of lake/stream acidity), ambient conditions attributable to individual modes of transportation are not known and, as a result, risk cannot be calculated.

4.3 ECOLOGICAL ANALYSIS

As indicated in Section 2, ecological analysis refers to population, community or ecosystem effects of emissions or loss of habitat from changes in land use. Table 15 summarizes the expected damage to air, water and land resources by mode of transportation.

Table 15
SUMMARY OF POSSIBLE ENVIRONMENTAL EFFECT BY MODE OF TRANSPORTATION

	Air	Water resources	Land resources	Other impacts
Marine/ ferries	Air pollution (CO, HC NO_x, particulate); global pollution (CFCs released during vehicle manufacture and disposal, CO_2 from fossil fuel combustion)	Discharge of ballast water, oil, spills, etc. Modification of water systems during port construction and canal cutting and dredging	Land taken for infrastructure; dereliction of obsolete port facilities and canals; vessels and craft withdrawn from service	
Railroad	Air pollution (CO, HC, NO_x, particulate and fuel additives such as lead). Global pollution (CO_2, CFCs)		Land taken for rights-of-way and terminals; dereliction of obsolete facilities; abandoned lines, equipment and stock	Partition or destruction of neighbourhoods, farmland and wildlife habitats
Road/ highway	Air pollution (CO, HC, NO_x, particulate and fuel additives such as lead). Global pollution (CO_2, CFCs)	Pollution of surface water and ground water by surface run-off (lubricants, coolants, road salt); modification of water systems by road building	Land taken for infrastructure; extraction of road building materials; abandoned spoil tips and rubble from road works; road vehicles withdrawn from service; waste oil	Partition or destruction of neighbourhoods, farmland and wildlife habitats; congestion

Table 15 (cont'd)
SUMMARY OF POSSIBLE ENVIRONMENTAL EFFECT BY MODE OF TRANSPORTATION

	Air	Water resources	Land resources	Other impacts
Aircraft	(CO, HC, NO_x, particulate); global pollutants (CFCs, CO_2)	Modification of water tables, river courses and field drainage in airport construction	Land taken for infrastructure; dereliction of obsolete facilities; aircraft withdrawn from service	

Source: Adapted from OECD, 1991.

Table 16 shows the land required by mode of transportation. Highways require almost twice as much land per linear kilometre as railroad tracks.

Table 16
LAND USE IN THE TRANSPORTATION SECTOR, 1985

Transportation mode	Width (m)	Land use per km (m^2)
Highway		
Buffer	22.0	22,000
Shoulder	9.5	9,500
Road	14.0	14,000
Total	**45.5**	**45,500**
Air	na	na
Railroad		
Buffer	11.0	11,000
Track	13.9	13,900
Total	**24.9**	**24,900**
Marine	na	na

Source: Statistics Canada, *Human Activity and the Environment*, 1991.

5. OVERVIEW OF ESTIMATES OF ENVIRONMENTAL DAMAGE SOCIAL COSTS

5.1 VALUE OF ENVIRONMENTAL DAMAGE

Table 17 presents the estimated costs of environmental damage by pollutant and type of damage. The damage resulting from the pollutants is the basis for the "starting point" costs caused by each pollutant, except for the costs

Table 17

STARTING POINT COSTS OF ENVIRONMENTAL DAMAGE BY POLLUTANT (1989$/KG)[a]

Damage	Effect	SO_2	NO_x and ozone	Acid deposition	Particulate	Carbon dioxide
Health	Mortality Morbidity Total	4.48 0.13 4.61	0.89 0.76 1.64	na[b] na 0.00[c]	0.86 0.08 0.94	na na na
Materials	Corrosion/ Soiling	0.31	0.03	na	0.00	na
Vegetation	Crops Ornamental forests	0.00	0.03	na	0.00 0.00	na
Visibility		0.36	0.44	na	2.16	na
Ecosystems		na	na	na	0.00	na
Historical monuments		na	na	na	0.00	na
Total		**5.29**	**2.14**	**0.00**	**3.10**	**0.018[d]**

Source: Ottinger et al., 1990.

a **The damage resulting from the pollutants is the basis for the determination of the external costs caused by each pollutant, except for the external costs of CO_2 emissions which reflect the cost of damage control (Ottinger et al., 1990, p. 58). The *starting point* costs are based on a survey of a number of studies estimating the external costs of electricity generation. "The uncertainties associated with these [cost] estimates in their original studies are large; the studies did not have estimates for some potentially important effects; and the existence and magnitude of some of the reported effects are still in dispute. These estimates should be used with great caution, as they indicate only the magnitude of damages." (Ottinger et al., 1990, p. 228)**

b **na — not available. Value of environmental damages not estimated, though they may exist.**

c **Value of environmental damage estimated and found equal to zero.**

d **The control cost of sequestering the CO_2 emitted into the atmosphere by planting trees or other vegetation (Ottinger et al., 1990, p. 138).**

of CO_2 emissions which reflect the cost of damage control (Ottinger et al., 1990). The "starting point" costs are based on a number of surveys of external costs from electricity generating stations and estimates of the external costs of typical coal generating stations. Caution is advised in their use. The estimates are generic in nature for stationary sources in or near urban areas. Intercity transportation environmental damage will often arise in more isolated regions of the country.

The estimates of the values of environmental damage found in Ottinger et al. (1990) were the result of a two-year study reviewing evaluation methods and environmental cost estimates for power generation utilities in the United States. The analytical framework used can be described as follows:

- identify pollution sources and the quantity of emissions;
- estimate the dispersion of these emissions;
- identify the groups (humans, ecosystems, materials, etc.) exposed to the pollutants;
- estimate the responses between the groups exposed and the pollutants; and
- estimate the value of these responses (replacement costs, contingent valuation, hedonic prices, travel costs, market prices).

This study utilizes the estimates of the value of the dose-response relationships reviewed in Ottinger et al. (1990) as the best estimate of the potential value of environmental damage from different modes of intercity passenger transportation.

Alternative costs of environmental damage from power generation exist. Ontario Hydro developed estimates of damage from NO_x and SO_2 and particulate emissions arising from the export of electricity (Concord Environmental and VHB, 1991; Acres, 1991; Phytotoxicology Consultant Services, 1990; VHB, 1991a). The estimates for SO_2 environmental damage are presented in Table 18. The estimates of damage from other air emissions in the Ontario Hydro studies are not available for comparison with Ottinger et al. (1990). The estimates of environmental damage from SO_2 in the Ontario Hydro study are lower than those suggested in Ottinger et al. (1990).

The estimates of damage developed by Ontario Hydro were not used in this study since they have not yet been subject to peer review.

Table 18

COMPARISON OF ONTARIO HYDRO AND OTTINGER ET AL. (1990) ENVIRONMENTAL DAMAGE FROM SULPHUR DIOXIDE EMISSIONS

Location	Mortality[a]		Morbidity[a]		Materials damage[b]
	Fatalities/kt	(1989$/kg)	Illnesses/kt	(1989$/kg)	(1989$/kg)
Lambton	0.03	0.12	15	0.01	0.04
Nanticoke	0.07	0.28	17.5	0.01	0.05
Lakeview	0.5	2.00	162.7	0.08	0.09
Ottinger, 1991	0.95	4.48	54	0.13	0.31

a Concord Environmental and VHB Research and Consulting Inc., 1991.
b Acres, 1991.

Note: Assumes $4 million per fatality and $485 per illness for the Canadian generating stations.

The major source of environmental damage is from SO_2, particularly health and materials damage. These costs (1989$) total an estimated $5.29/kg of pollutant emitted. The estimated cost of impaired visibility due to SO_2 is $0.36/kg. The transportation sector is not considered to be a major source of SO_2 emissions. Only about 2 percent of all emissions originate from the transportation sector.

Particulates are the major contributor to impaired visibility, resulting in an estimated cost of $2.16/kg of pollutant. The estimated health cost of particulate is $0.94/kg. NO_x and ozone result in estimated health costs of $1.64/kg, materials damage of $0.44/kg, crop (vegetation) damage of $0.03/kg and visibility damage of $0.03/kg.

An estimate of the control cost of global warming (CO_2) is also presented in Table 17. The control cost presented is for establishing and maintaining a forested area that is capable of absorbing CO_2 through photosynthesis (that is, carbon sequestration). This forest must be maintained in perpetuity to continue to store the carbon released once through fossil fuel combustion. The cost is about $0.02/kg of CO_2.

The direct costs of environmental damage is greatest from SO_2, estimated at $5.29/kg. The cost of environmental damage arising from particulate emissions is estimated at $3.10/kg and the costs of NO_x and ozone emissions are about $2.14/kg.

5.2 APPLICATION AND LIMITATIONS OF ESTIMATES

The estimated values presented in this study are for the power generation sector or other stationary sources. The power generation sector is a significant contributor, through the burning of fossil fuels, to atmospheric emissions of the same pollutants emitted by passenger transport vehicles. The costs presented in Table 17 are those that seem to most reasonably represent the range of values in the studies reviewed (Ottinger et al., 1990), considering the location at which the studies were performed, their documentation and their thoroughness. Too many relevant costs are overlooked by the studies and the studies included vary in quality and level of documentation. As a result the figures presented are not actual cost estimates but serve as a "starting point" to provide an order-of-magnitude of the studies reviewed and a starting point for further research (Ottinger et al., 1990).

The direct cost estimates of environmental damage presented in Table 17 leave out many potential effects and are likely to be conservative. More complete estimates should include the environmental damage arising from air emissions not included in the present estimates, such as the greenhouse gases, methane, N_2O and toxics, heavy metals, ozone precursors and hydrocarbons, and other environmental damage such as water and land use and solid waste disposal (Ottinger et al., 1990).

Cost estimates of the environmental damage of the construction and maintenance of infrastructure, the manufacture and disposal of vehicles and the manufacture and distribution of fuel do not exist and must be developed.

6. Costs of Environmental Damage Between Different Modes of Passenger Transportation

Costs of each transport mode by contaminant and damage are presented in Table 19. These cost estimates are based on the direct costs of environmental damage and the emissions of pollutant per passenger-kilometre (pass-km) developed previously. The range of environmental damage presented reflects the range of emissions of each pollutant by mode of transport presented previously. The environmental damage shown in Table 19 varies by mode of transport and type of damage.

Table 19

TOTAL ENVIRONMENTAL DAMAGE

1989$/1,000 PASS-KM

	Health mortality		Morbidity		Total health		Materials corrosion		Crops		Visibility		Global warming		Total	
Transport mode	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
Intercity bus transport	0.73	1.28	0.39	0.67	1.12	1.95	0.03	0.06	0.01	0.02	0.30	0.52	0.66	1.20	2.12	3.75
Motor vehicles	1.66	1.81	0.53	0.65	2.19	2.46	0.06	0.06	0.02	0.02	1.57	1.66	2.08	2.10	5.92	6.31
Air transport	0.11	0.79	0.07	0.61	0.18	1,40	0.00	0.03	0.00	0.02	0.05	0.38	2.55	5.85	2.79	7.68
Railroad transport (VIA only)	3.13	6.09	1.90	3.64	5.03	9.73	0.13	0.24	0.06	0.12	1.32	2.80	2.20	4.18	8.74	17.07
Marine transport	0.38	0.87	0.02	0.05	0.40	0.92	0.03	0.06	0.00	0.00	0.04	0.10	0.16	0.37	0.63	1.44

Note: **Values are rounded.**

The health costs per 1,000 passenger-kilometres for rail passenger transportation are estimated at between $5 and $10 (1989$)/1,000 pass-km. Private passenger motor vehicle transportation health costs are estimated at between $2.2 and $2.5/1,000 pass-km. The health costs for intercity bus transportation are valued between $1.1 and $1.9/1,000 pass-km. Marine transportation health damage is valued between $0.4 and $0.9/1,000 pass-km. Air passenger transport health costs are valued at an estimated $0.3 to $1.4/1,000 pass-km.

Rail passenger transportation causes the highest estimated materials damage, $0.1 to $0.2/1,000 pass-km. Private motor vehicles cause an estimated $0.05 to $0.06/1,000 pass-km of materials damage. Materials damage from marine passenger and intercity bus transport are estimated at between $0.03 and $0.06/1,000 pass-km. The cost of materials damage arising from air transport are negligible, less than $0.03/1,000 pass-km.

Rail passenger transport results in an estimated $0.06 to $0.12/1,000 pass-km in crop damage. Private motor vehicles cause about $0.02/1,000 pass-km in crop damage. Crop damage from intercity passenger bus transport is estimated to cost between $0.01 and $0.02/1,000 pass-km. Crop damage from air and marine passenger transport is negligible, less than $0.02/1,000 pass-km from air transport and less than $0.001/1,000 pass-km from marine transport.

Rail passenger transport results in an estimated $1.3 to $2.8/1,000 pass-km and private motor vehicles between $1.6 and $1.7/1,000 pass-km of visibility damage. Visibility damage from intercity passenger bus transport is estimated at between $0.3 and $0.5/1,000 pass-km. Air passenger visibility damage is estimated to cost between $0.05 and $0.4/1,000 pass-km. Visibility damage arising from marine passenger transportation is less than $0.1/1,000 pass-km.

The control cost of global warming resulting from CO_2 emissions is greatest from air transport, between $2.6 and $5.9/1,000 pass-km. Rail passenger transport results in a control cost estimated at $2.2 to $4.2/1,000 pass-km. Global warming control costs from private motor vehicles are estimated at about $2.1/1,000 pass-km. Global warming control costs from intercity passenger bus transport are estimated at between $0.7 and $1.2/1,000 pass-km. Marine global warming control costs are estimated at between $0.2 and $0.4/1,000 pass-km.

Table 20 presents a comparison of the environmental costs of transportation modes. At present only preliminary cost estimates of the operation and maintenance of transportation systems are available. On a passenger-kilometre basis, railroad transportation results in the highest environmental damage, up to $17/1,000 pass-km, and marine passenger service the least damage at between $0.6 and $1.4/1,000 pass-km. Private passenger motor vehicles result in the second highest damage-between $5.9 and $6.3/1,000 pass-km, followed by air and intercity bus transport at $2.8 to $7.7 and $2.1 to $3.8/1,000 pass-km respectively.

Table 20
DAMAGE BY MODE OF TRANSPORTATION AND ORIGIN

Transportation mode	Construction and maintenance of infrastructure	Vehicle manufacture and disposal	Fuel manufacture and distribution	Vehicle operation and maintenance (1989$/ 1,000 pass-km)
Intercity bus	na	na	na	2.1–3.8
Motor vehicle	na	na	na	5.9–6.3
Air	na	na	na	2.8–7.7
Railroad	na	na	na	8.7–17.1
Marine	na	na	na	0.6–1.4

Note: na — not available.

6.1 APPLICATION AND LIMITATIONS OF ESTIMATES

The estimated values of environmental damage presented in Table 19 are based on estimates of the value of environmental damage in the power generation sector. The valuing of environmental damage undertaken for electricity generating stations is not completely applicable to the transportation sector. Of particular concern is the costs arising from SO_2. Power generating utilities have placed a large emphasis on the costs arising from SO_2 emissions due to regulatory and public concerns about acid rain. The transportation sector is not a major contributor to SO_2. As a result the values attributed to SO_2 emissions from power generation may not be totally appropriate for the transportation sector.

Also, the health damage used in the electricity generation sector studies relate to power stations in or near urban areas. Although intercity travel involves transportation in and through urban areas, the actual exposure of humans to emissions from passenger transportation would be less than those stated in Ottinger et al. (1990).

The estimates of the value of environmental damage may leave out many potential effects and, as a result, damage may be undervalued. More complete estimates should include the environmental damage arising from air emissions not included in the present estimates, such as the greenhouse gases, methane and N_2O, toxics, heavy metals, ozone precursors and HC, and other environmental damage such as water, land use and solid waste disposal (Ottinger et al., 1990).

Finally, estimates of environmental damage on a passenger-kilometre basis can fluctuate widely depending on the occupancy of the passenger vehicle. As the occupancy of the transport mode increases, the emissions and environmental damage on a passenger-kilometre basis will decrease.

7. Major Obstacles to Estimating Social Costs of Environmental Damage

It is apparent from the previous sections that both the theoretical and empirical bases of evaluating the social costs of environmental damage from transportation continue to have major gaps and deficiencies. Nevertheless, substantial progress has been made, especially during the last decade, in developing a more robust theoretical framework for evaluation. There has also been a growing body of increasingly sophisticated empirical studies with principal contributions coming originally from the United States and now more from Europe. Unfortunately, empirical studies directed to producing definitive results are still demanding and costly undertakings. Furthermore the quantitative results of these studies cannot usually be generalized, and may only be suggestive in other contexts.

There are several theoretical and practical issues which recur in efforts to evaluate the social costs of environmental damage. Some of these issues are discussed briefly in the following subsections, to provide an assessment (in the next section) of the potential for research to produce more precise and reliable estimates.

7.1 THEORETICAL ISSUES

In conducting studies to estimate the social costs of environmental damage, or in interpreting these studies, it is important to keep in mind several fundamental theoretical issues. These issues are inherent in the evaluative framework and in any attempts to apply it. The issues that are discussed here include only those that are general and are not specific to particular methods of evaluation.

7.1.1 Scoping

In assessing environmental damage, especially from a "non-point source" such as transportation, it is often difficult to define the geographic and temporal extent (that is, the scope) of the effects. For example, in an environmental impact assessment for a transportation infrastructure project, such as an airport or highway, scoping may involve distinguishing the distances over which impacts will be assessed. Similarly, there will be explicit or implicit distinctions of the time horizons over which impacts and effects occur.

Scoping is based on assessment of the nature and magnitudes of possible impacts, and on consideration of the scientific (ecological) *and* socio-economic significance of these impacts. Thus, socio-economic significance is recognized as one of the constitutive requirements for an environmental impact or effect.

This accords with the perspective of economic theory, whereby a social cost is incurred only if there is *both* a physical effect *and* a loss in someone's welfare as a consequence of this effect. As previously suggested, restriction of concern to the social costs of environmental damage can be criticized as being anthropocentric. This criticism can be met, at least partially, by taking a broad view of the environmental services that contribute to human well-being (for example, including the existence values that humans attribute to other species).

The economic perspective of equating effects with social costs has at least one compelling advantage in that it provides criteria for assessing the importance of effects, and thus for devoting attention to them accordingly. It also accords salience to environmental effects in economic and policy arenas.

Although the economic perspective helps to focus the problems of spatial and temporal scoping, it does not eliminate them. It remains difficult to determine, for example, how far to go in following cause-effect linkages in assessing the effects of emissions. The cause-effect linkages quickly ramify and become more diffuse. Although more diffuse, the effects can be more widespread and subtle, and thus still have greater consequences overall. Clearly, there are practical limits to how far it is possible to explore cause-effect linkages, but we should not confuse these limits with the limits of the effects themselves.

In the case of intercity passenger transportation, the spatial extent of known effects ranges from local impacts, such as particulate emissions, to global implications from carbon dioxide emissions contributing to global warming. In the temporal dimension, the effects may range from immediate irritation and discomfort to long-term implications for health, and possible modifications of ecosystem function and structure. These latter effects may extend not only over current generations, but also over future generations.

Economic theory provides very uncertain guidance for addressing the problem of temporal scoping. The usual approach in economic evaluation has been to discount future costs and benefits, that is, to reduce the present value equivalence of future costs and benefits according to their distance in the future. This usually has the effect of making costs and benefits, more than about 10 years in the future, almost insignificant. Concerns about the environmental sustainability of economic development have raised questions about this practice.

There is limited scope for addressing these concerns within current economic theory, but three avenues hold some promise:

- Arguments can be made that the value of natural amenities will increase, relative to marketed goods and services, as the relative (or absolute) abundance of these natural amenities decreases and as the demand for them grows with increasing discretionary time and income. This tends to offset the effects of discounting in diminishing the costs of losing environmental amenities (Krutilla and Fisher, 1985).

- It may be possible to find ways of incorporating intergenerational equity as a more explicit consideration in economic evaluation, beyond the notion of bequest value, introduced previously, which considers the value of making bequests only to those who make them.
- Several further arguments can be invoked (Pearce and Turner, 1990) to support a general principle that economic activities should be guided to ensure that "natural capital" is maintained or enhanced.

The need for new approaches, such as those mentioned above, is made more compelling when the following issues are considered.

If human values are considered to apply, (in a sufficiently broad fashion) over ecosystem components and functions, and over time, then the problem of scoping is constrained more by available knowledge of biophysical causes and effects than by economic criteria.

7.1.2 Valuation

Although it is possible in theory to take a broad view of human values relating to the environment, in practice the methods for determining these values are far from simple or comprehensive. This means that it is usually necessary to be selective about which environmental values are evaluated in a particular case. There is justifiable concern that in such a process, the "soft" and diffuse environmental values will be undervalued compared with the usual "hard" economic values.

Whether or not there is such a selection, it is always necessary to discriminate among the values that are being assessed, so there can be clarity about what one is evaluating. In practice, it is often very difficult to disentangle values or motives, and to compare value measures produced by different methods.

This difficulty may be present not only for the economist attempting to assess values, but also for the people themselves who are doing the evaluating. This applies particularly to existence values, but existence values may themselves be confounded with other values.

Sagoff (1988) has questioned whether the "self-regarding values" of standard economic evaluation are commensurable with the "group-regarding" values attributed to the natural and cultural heritage, such as unique or special natural areas or historic monuments. The attempt to make such environmental values in dollars may implicitly contribute to the "down-valuation" of the former (Kelman, 1981).

Furthermore, as previously noted, evaluation presupposes knowledge of those things that contribute to our well-being. Costanza (1991), as quoted previously, questions to what extent this knowledge can be presupposed on the part of the general public with regard to ecosystem functions, when scientists are only just beginning to discover the contributions of these functions to our life support and well-being.

When such values are at stake in public policy, evaluations by individual members of the public must be circumscribed by a broader process of public deliberation which includes contributions from scientists. These deliberations must also recognize the limited knowledge that scientists have about ecosystem processes and functions, and the unknown risks of actions and practices that interfere with these functions. With improvements in knowledge, it is presumed that people's most basic evaluations would converge with the imperative to protect the ecosystem functions on which life depends.

7.1.3 Equity

The economic measures of value discussed here are based fundamentally on people's willingness to pay (WTP) or willingness to accept (WTA). These measures raise considerations about equity within generations (intragenerational equity) and equity across generations (intergenerational equity).

With regard to intragenerational equity, it is clear that people's willingness to pay depends on their ability to pay, and that ability to pay differs greatly among individuals in society. Similarly, compensation demanded is evaluated relative to one's overall assets, and accordingly is greater for those with greater assets.

These disparities mean that individual evaluations carry different weights in the overall social evaluation according to each individual's ability to pay. Disparities may also be manifest in the different incidence of environmental harm for individuals in society, with those in less affluent areas subject to greater harm (such as higher levels of pollution).

Although there are many types of environmental harm which individuals in society have a common interest to prevent, regardless of their level of affluence, the disparity in weightings of individual evaluations according to affluence may be reflected in relative evaluations of environmental problems. These problems tend to be in the form of amenity losses for the more affluent, and in the form of health implications for the less affluent.

These potential biases should be borne in mind at the scoping stage. Reference has been made earlier (Section 3) to possible methods of compensating or correcting for these disparities in evaluation.

Future generations, by definition, are not present to express their views, or have their WTP or WTA count in current evaluations. This leads to the problem of intergenerational equity. Again, impacts for future generations should be given special attention at the scoping stage, and methods should be developed to account for these impacts in overall evaluation.

7.1.4 Uncertainty

Scientific uncertainty about environmental impacts is pervasive in evaluation. Uncertainty (in the sense of only probabilistic knowledge of the parameters' bearing on cause-effect relationships) can be addressed through expected utility theory if risk profiles and attitudes can be determined. A more considerable kind of "uncertainty" is a complete lack of knowledge of potentially important causal relationships.

Much of the uncertainty arises from the difficulties of establishing thresholds of effects in complex physiological and ecological systems. Scientists are continually finding effects at lower levels of environmental stress than were noted previously. For example, the history of research on health effects of lead exposure follows this general pattern; this case is particularly relevant to the present study because the research eventually contributed to the phasing out of lead in gasoline. Research and policy on acid rain have a similar history.

Uncertainty may be compounded where there are cumulative effects from many sources over time. Thus, the environmental effects of emissions from intercity passenger transportation are part of the overall effects of transportation emissions generally, and of the effects of emissions from all sources. Furthermore, these emissions may be only one factor among several which determine physiological and ecological stress. In this case, intercity passenger transportation will be one among many contributors to environmental damage, and the appropriateness of policies to address environmental effects from this contribution must be judged with regard to the efficacy of policies for all contributions.

In the case of intercity passenger transportation, the major uncertainties are the effects of emissions on materials, biota and human health and comfort, and the effects of land uses on wildlife habitat and space available for other uses. Even if dose-response relationships or other measures of impact could be determined in experimental settings, there would still be great uncertainty about their implications on the magnitudes and geographical extents of impacts in the real world. It would be very difficult to distinguish the effects of different modes of transportation, especially where they make use of practically the same corridors.

The foregoing discussion of uncertainty has focussed on biophysical impacts. Apart from biophysical uncertainty, there is great uncertainty about future evaluations of environmental amenities. This is especially true over the course of generations, through which technologies may change, and the relative demands for resources and amenities may shift in one direction or another. It is thus very difficult to anticipate what combinations of natural capital and artificial capital will be most valued by future generations. As previously indicated, some economists (Krutilla and Fisher, 1985; Pearce and Turner, 1990) have offered reasonable arguments suggesting that the relative value of natural capital, especially life-support functions and amenities, is likely to increase in the future.

Because determining the social cost of environmental damage is of little practical value in itself, a third kind of uncertainty arises. In practice, the ultimate concern is the costs and potential benefits of policies to reduce and control environmental damage. There can often be considerable uncertainty about the efficacy of policies, and about the costs of implementing them. The importance of uncertainty about costs is indicated again by the policy processes with regard to acid rain and the phase-out of leaded gasoline. In both cases, the costs and efficacies of policies were one of the major areas of uncertainty impeding legislative or regulatory action. The same might be said about the current debate over greenhouse gas emissions and global warming.

7.1.5 Irreversibility

Evaluation becomes especially difficult (and many of the previous issues are compounded) in cases where there are environmental irreversibilities. An irreversibility is a permanent loss in an ecosystem component or in an ecosystem's capacity to maintain a particular function. For example, expansion of a highway or an airport might affect the habitat of a rare species. Emissions of greenhouse gases could also contribute to changes in global climate, which would be practically irreversible. Concerns about irreversibility arise mainly with regard to the former kind of discrete loss. Gradual losses of ecosystem functions, such as stratospheric ozone depletion and climate change, raise broader issues of environmental sustainability.

Clearly, if the value of preserving an ecosystem component or function is known, then the value of a development or activity must exceed this preservation value, if there is to be a net benefit from the development or activity when it incurs a loss of the preservation value. Often, however, preservation values are neglected, and there is undue optimism or over-estimation of the net benefits of development.

The problem of irreversibilities increases the salience of the question of intergenerational equity. Irreversible losses are incurred not only by the present generation, but also by all following generations. Often, however, only the present generation, or those in the near future, receive the benefits of the development or activity that imposes the environmental loss. The question therefore arises whether it is fair for the present generation to benefit at the expense of all succeeding generations. At the least, consideration needs to be given to the total economic and environmental legacy of the present generation for future generations.

The problem of irreversibilities is compounded by uncertainties about the values of ecosystem components or functions that are subject to irreversible loss. It is often difficult or impossible to know what the values of ecosystem components or functions are or will be. An often-cited example of this is a rare species or stock that could be the only source of a life-saving drug. A particular species could also have an unrecognized function in maintaining the balance of a whole ecosystem.

In the face of such uncertainties, the most risk-averse strategy would be to minimize the maximum losses that could occur. In this case, development would be a preferable option only if the costs of preservation, that is, the foregone development benefits, exceed these maximum possible losses from development. This approach would tend to correct any bias toward development and against preservation.

It may be difficult to keep track of all the ecosystem components and functions that human activities put at risk and the uncertainties about their values and probabilities of loss. Accordingly, it has been suggested (Ciriacy-Wantrup, 1952) that safe minimum standards should be developed so that irreversible losses of important ecosystem components and functions can be avoided. This approach has been elaborated in particular with regard to endangered species (Bishop, 1978).

8. The Potential for Research to Improve the Precision of Estimates

This report has provided a general overview and estimates of the environmental costs of intercity transportation. A number of difficulties and deficiencies in such estimation have been noted in previous sections. Here, some of these deficiencies are discussed with a view to the potential for research to improve the precision and reliability of estimates.

Attempts to improve the precision of estimates must address several layers of uncertainty. For example, with regard to emissions, there is uncertainty about each step from sources to fates and effects: quantities of emissions, transport and disposal, exposure, response and evaluation. Priorities for addressing uncertainties from emissions to response are discussed in general terms in the following subsection on biophysical estimates, and priorities for addressing uncertainties in evaluation are discussed in the subsection on economic estimates. These discussions lead to general recommendations for research priorities.

Two important limitations in these cost estimates have been mentioned previously, but bear prominent reiteration here. First, the cost estimates are all either average costs or marginal costs (that is, increments in costs for unit increments in emissions or resource use), given the current levels of emissions and environmental quality. For example, the costs of carbon dioxide emissions are based on the current costs of purchasing and afforesting land to take up the carbon dioxide through photosynthesis. Once the land currently available for this purpose is used up, the cost of more land to absorb additional carbon dioxide emissions is likely to be higher. Accordingly, the current average cost or marginal cost could be misleading.

Similarly, the marginal costs of environmental impacts can increase rapidly as ecological or physiological thresholds are approached. For example, some lakes may be little affected by acid precipitation until the buffering capacity of the surrounding watershed is depleted; then, acidification may be rapid. Again, current marginal costs could be misleading in such a case. Clearly, it is necessary to consider not only the costs of small, marginal changes, but also take account of biophysical limits and thresholds that might be breached in the longer term.

The second important limitation of the cost estimates presented here is that data were sparse for considering the complete life-cycle costs for each mode of transportation, including construction and maintenance of infrastructure, vehicle manufacture and disposal, fuel manufacture and distribution, and vehicle operation and maintenance. These front-end, or up-stream, costs could also include many subsidized and external costs, including environmental costs.

8.1 BIOPHYSICAL

8.1.1 Emissions

The emissions estimates in this report are based on existing regulations, typical fuel characteristics and average fuel efficiencies for each mode. For some modes — notably cars — emissions are determined largely by environmental regulations which are expressed in mass per unit distance (g/km); for other modes, emissions are estimated using the fuel efficiency (MJ/km) and unit emission rates (g/MJ) that reflect fuel characteristics. In the long term, choice of mode could also take into consideration the prospects for more stringent regulations, changes in fuel types or fuel characteristics (for example, alternative fuels or reduced sulphur content in diesel fuel) and improving efficiencies.

The comparative results for emissions, and hence costs, per passenger-kilometre are also very dependent on vehicle occupancy assumptions. Therefore, the prospects for future increases or decreases in vehicle occupancies also bear consideration.

8.1.2 Transport and Disposal

The cost estimates presented in this report have been derived mainly from cost studies for electric power generating facilities (Ottinger et al., 1990). Nevertheless, emissions from transportation have different chemical compositions and characteristics than emissions from power plants.

Sulphur dioxide and particulates have been the power plant emissions of greatest concern and have consequently received the greatest attention and rigour in evaluation. These emissions are of less concern for transportation compared to nitrogen oxides, hydrocarbons and carbon monoxide. These latter emissions have received less attention in impact assessment and evaluation.

Transportation emissions are generally released close to ground level, and may therefore have more concentrated local impacts. These would not be reflected in the estimates, which assume that transportation emissions undergo the same transport and disposal as electric power plant emissions. In fact, the impacts of air travel are likely to be concentrated near airports, of bus and car travel near roadways, and of train travel near rail lines. This is true not only for emissions, but also for land use impacts (such as the disturbance of wildlife habitat, modification of hydrological regime, and aesthetic considerations). The cumulative impacts of all these effects should be considered.

Finally, the altitude at which the pollutants are released may affect damage estimates. A recent study shows that the radiative forcing of surface temperature is most sensitive to changes in tropospheric ozone at a height of greater than 12 kilometres — the altitude at which the maximum emissions of nitrogen oxides from aircraft occur. The study concludes that the environmental damage of NO_x emissions from aircrafts is about 30 times greater than from surface vehicles on a unit basis (Johnson, Henshaw and McInnes, 1992). The impact of release altitude on damage should also be considered.

8.1.3 Exposure

The concentration of transportation effects near transportation infrastructures means that exposure to impacts of transportation may be incurred, not only by nearby people, materials and ecosystems, but also by travellers, vehicles and transportation infrastructures themselves. Nevertheless, there is little research that directly measures exposures from transportation. Instead, estimates of exposure are based on ambient conditions. This may give an incomplete perspective on total exposure through multiple pathways.

Research is also lacking on behaviour to avert exposure to transportation externalities. It is therefore difficult or impossible to assess the costs and foregone benefits implied by this averting behaviour.

8.1.4 Response

Most of the estimates presented here depend critically on the direct cost method of damage estimation, and therefore on dose-response functions or damage functions relating total exposure to the incidence of particular effects and costs (that is, the effect for which a dose-response function or damage function must be clearly identifiable and measurable). This approach may neglect effects which are more subtle and less easily measurable, or for which the long-term costs are less obvious.

To establish that a response arises as a consequence of transportation or transportation emissions, it is necessary to substantiate each of the previous steps. At each step, the effects from transportation may be confounded with other effects. This may also limit the environmental costs that can be attributed to transportation.

The estimates quoted here indicate that human health is the category in which the greatest costs of emissions can be attributed. Visibility is the category with the next highest costs. Costs for several categories, notably forests and ecosystems, are not indicated. As pointed out by Ottinger et al. (1990, p. 553): "For all pollutants, valuation of ecosystem effects, wildlife, endangered species, and historical or cultural assets has been very limited. This is an important limitation that needs more attention." These costs are evident with regard to traffic in urban settings, but may be more difficult to discern for intercity transportation. Numerous studies have investigated roadside contamination by heavy metals from traffic, especially lead contamination when leaded gasoline was used (Freedman, 1989). The other metals are presumed to derive from vehicle wear. These metals have also been found to be accumulating in the floors of forests remote from roadways, but the ecological implications of this increasing contamination is not yet understood (ibid.).

Overall, these considerations suggest that efforts to improve estimates of the environmental costs of intercity transportation might be directed most usefully to assessment and evaluation of local impacts to supplement the estimates of the more widespread costs of emissions reported here. These local impacts could generally include the cumulative effects of emissions, land-use changes and aesthetic considerations. Although these impacts might be noted in environmental impact assessments (for example, for road improvements or airport expansions), there is usually no systematic attempt to evaluate their costs. Nevertheless, these impact assessments (or better, post-audits of impact assessments) might serve as starting points in determining some of the local environmental costs of transportation to supplement the estimates of the more widespread costs from emissions.

With regard to both local and distant impacts, greater attention needs to be devoted to irreversible changes, such as the accumulation of toxic contaminants and greenhouse gases; which have uncertain or unknown ecological or physiological consequences.

8.2 ECONOMIC

The estimates of environmental costs of transportation reported here are based mainly on studies, performed in the United States, of the direct costs of environmental damage from emissions. These estimates, and the assumptions underlying them, have not been thoroughly checked for their applicability to Canadian conditions. Such checking would be necessary before any reliance is placed on these estimates for Canadian policy.

When environmental or health damage occurs, the costs include any direct costs of averting or mitigating the damage *and* the residual welfare loss after the averting or mitigating behaviour. Although some degree of consistency has been achieved across evaluation studies for direct costs, such consistency has proven elusive for the costs of residual welfare loss. In the case of changes in risks of morbidity and mortality, in particular, there appear to be large discrepancies in how people evaluate risks, as determined by hedonic price methods, household production function methods and contingent valuation methods, especially between evaluations produced by different methods. Fundamental questions remain about applications of these methods, and the inferences that can be drawn legitimately from them, especially with regard to the evaluation of morbidity and mortality. Nevertheless, the costs of morbidity and mortality are a major component of the costs attributed to transportation emissions.

There are obvious gaps in the estimates with respect to some damage, especially potential damage to forests and ecosystems. This possible damage must be subject to further ecological research before economic assessment can provide quantitative estimates of costs. In qualitative terms, economic theory can only provide the guidance indicated earlier about the caution that is especially justified given both uncertainties and irreversibilities.

Nevertheless, the policy implications of the environmental impacts of transportation might be made clearer by a judicious selection of evaluation studies. These studies would apply some of the evaluation methods indicated here to at least some of the environmental costs that can reasonably be attributed to transportation. The attempts to estimate these costs would also help in identifying the gaps or uncertainties in biophysical knowledge that are most important for policy.

8.3 RECOMMENDATIONS

Although there is now a growing body of work evaluating the social costs of environmental damage in the United States and Europe, there has been as yet almost no original evaluation research conducted in Canada. This is the situation with regard to evaluation of environmental damage generally, and perforce with regard to evaluation of environmental damage from transportation.

Given this situation and the preceding discussion, there are three areas of support for evaluation research which appear to offer the greatest prospects of producing results that will be useful in the near term, while also promoting a general body of research and capabilities which can provide the basis for future advances.

First, even without full quantitative monetary evaluation, it would be very useful to set out the full range of possible environmental impacts from each transportation mode, using the complete life-cycle analysis advocated here.

Second, greater use could be made of quantitative evaluation in connection with environmental impact assessments of transportation projects or policies. Evaluation in environmental impact assessments of projects could help to indicate some of the local environmental costs, which could be extrapolated for inclusion in the total environmental costs to be considered in transportation policies.

Third, it might be particularly revealing, and at the same time provide results that people could most easily relate to, to conduct an evaluation of the environmental impacts of each transportation mode in a multimodal corridor. Windsor to Quebec City or Edmonton to Calgary might be suitable corridors for such a study.

Finally, outside the scope of evaluation of the environmental impacts of transportation, but relevant to its long-term prospects, consideration might be given to the sources of demand for transportation, and whether the social and environmental costs of transportation might be reduced (or perhaps even augmented) by the introduction of advanced telecommunications technologies.

ENDNOTES

The study team thanks John Lawson of the Royal Commission for his assistance throughout the project in providing direction and in identifying relevant reference documents. In addition, he provided insightful and helpful comments on a draft of the report. Useful comments were also provided by other Commission staff, in particular John Sargent, Director of Research; these are appreciated.

Ed Hanna and Peter Victor of VHB also pointed to relevant references and served as sounding boards for nascent ideas.

The study team for the project at VHB consisted of David Heeney, Peter Stokoe, Murray Trott and Evelyn Nepom. David Heeney was the overall project manager. Peter Stokoe prepared the review of economic methodologies, Murray Trott prepared the assessment of emissions and their economic value, and Evelyn Nepom prepared sections on environmental effects and non-emission impacts.

1. *Canada's Green Plan* also identifies preserving the integrity of the North and environmentally responsible decision making at all levels of society as primary objectives.

2. The comparison of alternatives is fundamental to environmental assessment practice (and the rational planning model), and in principle considers the entire life cycle of each alternative. In practice, however, these have rarely been quantitative or comprehensive assessments of impacts.

3. Conceptually, material damage may be assessed in an analogous manner to health, based on exposure and dose-response considerations.

4. Note, however, that the net values of driving and going by bus might be different for society as a whole than for the person who makes the choice, because the social costs of the two choices include not only the private costs to the person but also the external costs of traffic congestion, environmental damage and other such costs that are borne by other members of society. It is the external costs component of social costs that are our main topic here.

5. Note that even breathing involves emitting "exhaust fumes" of carbon dioxide, a gas which contributes to global warming.

6. Such effects can occur not only through the natural environment, but also exclusively within the built environment, as with traffic congestion and accidents, but effects through the natural environment are the main subject of this report.

7. Note that although such costs are direct, they must often be inferred indirectly by means of dose-response functions or damage functions that relate environmental parameters to damage (for example, concentrations of sulphur dioxide and materials or crop damage). Some references (for example, Pearce and Markandya, 1989) distinguish evaluation methods primarily according to this interpretation of indirect and direct (that is, whether or not dose-response or damage functions are applied).

8. The actual emissions from plants vary according to the regulations in individual jurisdictions. Tougher regulations can effectively limit the choices of materials available to the manufacturer. For example, regulations limiting the emission of solvents have forced some car makers to abandon solvent-based paints in favour of waterborne paints.

9. Emissions per passenger-kilometre change proportionately with passenger occupancy. If vehicle occupancy for passenger rail transport doubled, the emissions per passenger-kilometre would decrease by one half. Estimates of vehicle occupancy are not explicit in the calculations made in this report and are not stated.

10. Parts of this section follow the discussion in Ottinger et al., 1990, p. 127.

11. A recent study suggests that CFC emissions may not contribute to global warming as previously thought. The holes in the stratospheric ozone layer resulting from CFC emissions may in fact cause a reduction in radiative forcing of the Earth in the middle to high latitudes. This reduction in radiative forcing may more than offset the warming effect of the CFC emissions. (World Meteorological Organization/United Nations Environmental Program, 1991).

12. Emissions from the source and background levels of pollutants both contribute to total exposure.

REFERENCES

Acres International Limited. 1991. *The Effects and Social Costs of Fossil-fired Generating Station Emissions on Structural Materials.* Update. Toronto: Ontario Hydro.

Ahuja, M. and P. Amar. 1988. *California's Program to Assess Economic Impact of Acidic Deposition on Materials.* Eighty-first annual meeting, American Pollution Control Association, Dallas, Texas.

Badr, O., S. D. Probert and P. W. O'Callahan. 1990. "Chlorofluorocarbons and the Environment: Scientific, Economic, Social and Political Issues." *Applied Energy* 37, pp. 247–327.

Barakat & Chamberlin. 1990. *"Contribution of the Transportation Sector to Atmospheric Pollution Problems."* A working paper prepared for the Royal Commission on National Passenger Transportation.

Beanlands, G. and P. Duinker. 1983. *An Ecological Framework for Environmental Impact Assessment in Canada.* Halifax: Institute for Resource and Environmental Studies, Dalhousie University. Published in cooperation with the Federal Environmental Assessment Review Office.

Bishop, R. 1978. "Endangered Species and Uncertainty: the Economics of a Safe Minimum Standard." *American Journal of Agricultural Economics* 60, pp. 10–18.

Bloxam, Rob. (Ministry of the Environment, Air Resources Branch). Telephone conversation with author. November 13, 1991.

Braden, J. B. and C. D. Kolstad, eds. 1991. *Measuring the Demand for Environmental Quality*. Amsterdam: Elsevier.

Bromley, Daniel W. 1991. *Environment and Economy: Property Rights and Public Policy*. Oxford: Blackwell.

Brown, H. L. 1985. *Energy Analysis of 108 Industrial Processes*. New York: Fairmont Press.

Carson, Richard T. 1991. "Constructed Markets." In *Measuring the Demand for Environmental Quality*. Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Chappelka, A. H. and B. I. Chevone. 1989. "White Ash Growth Response to Ozone and Simulated Acid Rain." *Canadian Journal of Forestry Resources* 16, pp. 786–90.

Ciriacy-Wantrup, S. 1952. *Resource Conservation: Economics and Policies*. Berkeley: University of California Press.

Concord Environmental and VHB Research and Consulting Inc. 1991. *Update of Information on Health and Environmental Effects of Air Pollutants from the Transportation Sector*. Draft report. Ottawa: Environment Canada.

—. 1991a. *The Effects and Social Costs of Fossil-fired Generating Station Emissions due to Electrical Exports on Human Health*. Update. Toronto: Ontario Hydro.

Costanza, Robert. 1991. "Assuring Sustainability of Ecological Economic Systems." In *Ecological Economics: The Science and Management of Sustainability*. Edited by R. Costanza. New York: Columbia University Press.

Cropper, M. L. and A. M. Freeman III. 1991. "Environmental Health Effects." In *Measuring the Demand for Environmental Quality*. Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Cropper, M. and W. Oates. 1990. "Environmental Economics: A Survey." *Discussion Paper QE 90-12.* Washington, D.C.: Resources for the Future.

Dale, J. M. and B. Freedman. 1982. "Lead and Zinc Contamination of Roadside Soil and Vegetation in Halifax, N.S." *Proceedings of the Nova Scotia Institute of Science,* vol. 32, pp. 327–36.

Devall, Bill and George Sessions. 1985. *Deep ecology: Living as if Nature Mattered.* Salt Lake City: Peregrine Smith Books.

DPA Group Inc. 1987. *Review and Update of Bio-economic Models of Acid Deposition.* Toronto: Ontario Ministry of the Environment.

Environment Canada. 1990. *Canada's Green Plan.* Ottawa: Supply and Services Canada.

Freedman, B. 1989. *Environmental Ecology: The Impacts of Pollution and Other Stresses on Ecosystem Structure and Function.* San Diego: Academic Press Inc.

Haneef, S. J., C. Dickenson, J. B. Johnson, G. E. Thompson and G. C. Wood. 1990. "Effects of Air Pollution on Historic Buildings and Monuments and the Scientific Basis for Conservation." *European Cultural Heritage* 4(3), pp. 9–16.

Heck, W. W. 1989. "Assessment of Crop Losses from Air Pollutants in the United States." In *Air Pollution's Toll on Forests and Crops.* Edited by J. J. MacKenzie and M. T. El-Ashry. New Haven: Yale University Press.

Intergovernmental Panel on Climate Change (IPCC). 1990. *Climate Change.* New York: Cambridge University Press.

Jensen, K. F. and R. L. Patton. 1990. "Response of Yellow-Poplar (*Lirondendron tulipifera* L.) Seedlings to Simulated Acid Rain and Ozone — Growth Modifications." *Environmental and Experimental Botany* 30, pp. 59–66.

Johnson, C., J. Henshaw and G. McInnes. 1992. "Impact of Aircraft and Surface Emissions of Nitrogen Oxides on Tropospheric Ozone and Global Warming." *Nature* 355, no. 6355.

Kelman, S. 1981. *What Price Incentives?* Boston: Auburn House.

Khan, A. M. 1990. *Energy and Environmental Factors in Intercity Passenger Transportation.* Ottawa: Transport Canada.

Knetsch, J. 1984. "Legal Rules and the Basis for Evaluating Economic Losses." *International Review of Law and Economics* 4, pp. 5–13.

Knetsch, J. and J. Sinden. 1984. "Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value." *Quarterly Journal of Economics* 1984 (August), pp. 507–21.

Kolomeychuk, R. J., K. L. Yeager and J. M. Spiegel, The Environmental Applications Group Limited and A. Yassi. 1984. *Effects of Automotive Emissions.* Ottawa: Environment Canada.

Krutilla, J. and A. Fisher. 1985. *The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources.* Revised. Baltimore, Md.: The Johns Hopkins University Press.

LeBlanc, N. (National Energy Board). Telephone conversation with author. January 20, 1992.

Lesser, V. M., J. O. Rawlings, S. E. Sprull and M. C. Somerville. 1989. "Effects of Ozone on Agricultural Crops: Statistical Methodologies and Estimated Dose-response Relationships." *Crop Science.*

Linquist, O. R., R. E. Mangeo, L. E. Olsson and J. Rosvall. 1988. "Case Study on the Deterioration of Stone. The Cathedral Well (Domkyrkobrunnen) in Goteburg." *Durability of Building Materials* 5, pp. 581–611.

Mitchell, R. C. and R. T. Carson. 1989. Using Surveys to Value Public Goods: *The Contingent Valuation Method.* Washington, D.C.: Resources for the Future.

Mohnen, V. A. 1989. "Elevational gradients/local chemistry." *Biologic Markers of Air-pollution Stress and Damage in Forests.* Washington, D.C.: National Academy Press.

Naess, Arne. 1989. *Ecology, Community and Lifestyle.* Translated and edited by David Rothenberg. Cambridge: Cambridge University Press.

National Acid Precipitation Assessment Program (NAPAP). 1987. *Interim Assessment: The Causes and Effects of Acidic Deposition,* Vol. 1, Executive Summary. Washington, D.C.: National Acid Precipitation Assessment Program.

Ontario Round Table on the Environment and the Economy (ORTEE). 1991. *Transportation Sector.* Toronto: ORTEE.

Organisation for Economic Co-operation and Development (OECD). 1986. *Environmental Effects of Automotive Transport.* Paris: OECD.

—. 1988. *Transport and the Environment.* Paris: OECD.

—. 1990. *Transport Policy and the Environment.* Paris: OECD.

—. 1991. *The State of the Environment.* Paris: OECD.

Ottinger, R. L., D. R. Wooley, N. A. Robinson, D. R. Hodas and S. E. Babb. 1990. *Environmental Costs of Electricity.* New York: Oceana Publications, Inc.

Oxford Concise Dictionary (OCD). 1980. Sixth edition. Edited by J. B. Sykes. Oxford: Oxford University Press.

Palmquist, Raymond B. 1991. "Hedonic Methods." In *Measuring the Demand for Environmental Quality.* Edited by J. B. Braden and C. O. Kolstad. Amsterdam: North-Holland.

Pearce, D. and A. Markandya. 1989. *Environmental Policy Benefits: Monetary Valuation.* Paris: OECD.

Pearce, David. 1983. *Cost-Benefit Analysis.* 2nd edition. New York: St. Martin's Press.

Pearce, David W. and R. Kerry Turner. 1990. *Economics of Natural Resources and the Environment.* New York: Harvester Wheatsheaf.

Percy, K. 1986. "The Effects of Simulated Acid Rain on Germinative Capacity, Growth and Morphology of Forest Tree Seedlings." *New Phytology* 104, pp. 473–84.

Peterson, C. E., K. G. Mattson and R. A. Mickler. 1989. *Seedling Response to Sulphur, Nitrogen and Associated Pollutants.* U.S. EPA Environmental Research Laboratory. Corvallis, Oregon, EPA 600 3-89/081.

Phytotoxicology Consultant Services. 1990. *Study of the Effects of Fossil-fired Generating Station Emissions on Terrestrial Ecosystems.* Toronto: Ontario Hydro.

Pilorusso Research Associates Inc. 1986. *Control of Heavy Duty Vehicle Emissions.* Ottawa: Transport Canada.

Pilorusso Research Associates Inc., VHB Research and Consulting Inc. and T.A.G. Resource Recovery. 1990. *Scrap Tire Management in Ontario.* Toronto: Ontario Ministry of the Environment.

Reich, P. B., A. W. Schoettle and R. G. Amundson. 1989. "Effects of O3 and Acidic Rain on Photosynthesis and Growth in Sugar Maple and Northern Red Oak Seedlings." *Environmental Pollution,* series A 40, pp. 1–15.

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.

Rowe, William D. 1977. *An Anatomy of Risk.* New York: John Wiley and Sons.

Sagoff, M. 1988. *The Economy of the Earth: Philosophy, Law and the Environment.* Cambridge: Cambridge University Press.

Schulz, W. 1985. "Besser Luft, was ist sie uns wert?" *Eine gesellschaftliche Bedarfsanalyse auf der Basis individueller Zahlungsbereitschaften.* Texte. 25/85 des Umweltbundesamtes. Berlin.

Schulz, W. and E. Schulz. 1991. "The Federal Republic of Germany." *Valuing the Environment: Six Case Studies.* Edited by J. P. Barde and D. Pearce. London: Earthscan.

SENES Consultants Limited. 1991. *Countdown Acid Rain Future Abatement Strategies.* Toronto: Ontario Ministry of the Environment.

Smith, V. Kerry. 1991. "Household Production Functions." In *Measuring the Demand for Environmental Quality.* Edited by J. B. Braden and C. D. Kolstad. Amsterdam: North-Holland.

Society of Environmental Toxicologists and Chemists (SETAC). 1990. *A Technical Framework for Life-Cycle Assessment.* Washington, D.C.: SETAC and SETAC Foundation for Environmental Education.

Statistics Canada. 1986. *Railway Transport in Canada, General Statistics,* 1985. Ottawa: Statistics Canada.

—. 1990. *Canadian Civil Aviation, 1989.* Ottawa: Statistics Canada.

—. 1990. *Passenger Bus and Urban Transit Statistics, 1988.* Ottawa: Statistics Canada.

—. 1990. *Railway Operating Statistics.* Ottawa: Statistics Canada.

—. 1990. *Shipping in Canada, 1989.* Ottawa: Statistics Canada.

—. 1991. *Human Activity and the Environment, 1991.* Ottawa: Statistics Canada.

Transport Canada and Environment Canada. 1989. *A Plan to Identify and Assess Emission Reduction and Opportunities from Transportation, Industrial Engines and Motor Fuels.* Ottawa: Transport Canada.

VHB Research and Consulting Inc. 1990. *Evaluation Methods in Environmental Assessment.* Toronto: Ontario Ministry of the Environment.

—. 1991. *Effects of Transportation Emissions on Ecosystems and Materials.* Background material prepared for the Royal Commission on National Passenger Transportation.

—. 1991a. *A Bio-economic Evaluation of the Effects of Fossil-fired Generating Station Emissions on Water Quality.* Update. Toronto: Ontario Hydro.

—. 1991b. *Reduction of Energy Use and Emissions in Ontario's Transportation Sector.* Toronto: Ontario Ministry of the Environment.

—. 1991c. *Review and Evaluation of Data Considerations in the Conduct of Life-Cycle Assessments.* Unpublished manuscript. Ottawa: Environment Canada.

VHB Research and Consulting Inc. and Canviro Consultants. 1991. *Water Pollution Abatement Technology and Cost Study.* Toronto: Ontario Ministry of the Environment.

Victor and Burrell. 1982. *Estimated Impacts of Acid Deposition in Ontario anc Their Economic Value.* Toronto: Ontario Ministry of the Environment.

World Meteorological Organization/United Nations Environmental Prograrr 1991. *Scientific Assessment of Stratospheric Ozone, 1991.* New York: United Nations Environmental Program.

World Resources Institute. 1990. *World Resources 1990-91.* New York: Oxford University Press.

14

DEREGULATION AND COMPETITION IN THE CANADIAN AIRLINE INDUSTRY

Steven A. Morrison*
March 1992

I. INTRODUCTION

Any change in regulatory regimes is bound to affect the relevant industry. Airline deregulation is no exception. This report details the results of this study's measurement of the effects of airline deregulation in Canada. In particular the assessment looks at how industry concentration, load factors (percentage of seats filled) and air fares have changed under deregulation.

Over the course of a decade, Canada relaxed regulation on its airlines until official deregulation began on January 1, 1988. Three data sets provide the basis for much of this assessment. First, the Airport Activity Data Base contains data on the passenger (and cargo) traffic flows into and out of Canadian airports. Second, the Revenue Passenger Origin and Destination Data Base is a 10-percent sample of passenger flight coupons. Finally, the Fare Basis Survey Data Base contains fare data from a 100-percent survey of passenger flight coupons conducted on 56 days of the year.

The intended methodology for this study was to begin with a time period before Canadian (and U.S.) deregulation to serve as a basis of comparison for future developments in Canadian aviation. Since the Fare Basis Survey

* Department of Economics, Northeastern University

did not begin until 1983 however, this was the first year used in this analysis. The year 1987 was selected as an intermediate year, with 1990 being the last year selected. The analysis pertains to Level I carriers.

This report uses what could be called a *factual* methodology. That is, it chronicles the changes in the industry from 1983 to 1990. Some — perhaps most — of these changes may be due to regulatory reform. But observed changes could have nothing to do with airline deregulation and would have occurred anyway. The alternative methodological approach is a *counterfactual* one, where a model of how the regulated airline industry would have evolved (during the years of actual deregulation) is compared with the (actual) performance of the deregulated industry. Because this counterfactual model is a control, the observed differences between actual deregulation and counterfactual regulation can be attributed to deregulation. Unfortunately, developing a counterfactual methodology is beyond the scope of this study. Thus, care should be taken in interpreting the results presented below.

II. Industry Concentration

When most people think about the effects of airline deregulation they think about its effect on fares. However, given the structure-conduct-performance paradigm that has been part of industrial organization for many years, industry structure (that is, the degree of concentration) is thought to affect fares. Thus, other things being equal, the fewer the number of competitors (that is, the more concentrated the industry), the higher the fares. (Fares and the effect of concentration on fares are examined in subsequent sections of this report.)

There are many ways to measure concentration in a network industry such as air transportation. First, there is concentration at the national level. This is an oft-cited figure when discussing the effects of airline deregulation. Typically one calculates a "concentration ratio," that is, the percentage of the market controlled by the *n* largest firms, where *n* is a number such as two, four or eight. These measures in Canada (and in the U.S.) show that the deregulated industry is more concentrated than it was before, in the sense that fewer firms control a large share of the national market. Note that there is no such thing as a national market. The markets for air transportation are the thousands of city-pair markets between which transportation takes place.[1] Thus, another measure of airline industry concentration is to measure concentration at the route level. Finally, because of the concern about a few airlines dominating hub airports, concentration can be measured at the airport level. In this case, concentration can be based on the percentage of passengers or flights controlled by each carrier at an airport. Alternatively, the extent of route-based concentration at an airport can be assessed by aggregating the route concentration of all routes originating at that airport.

In assessing the extent of industry concentration at the route level or airport level, the concept of "number of effective competitors" is used. The number of effective competitors is the inverse of the Herfindahl index. Rather than a simple count of the number of carriers in a market, the Herfindahl index adjusts for unequal market shares by summing the square of each airline's market share. Thus, if two airlines each had a 50 percent market share on a route, the Herfindahl index would equal $0.50^2 + 0.50^2 = 1/2$. Inverting the Herfindahl of 1/2 gives 2 (effective competitors). The inverted figure has a more intuitive interpretation than the Herfindahl index.[2]

Figure 1 shows industry concentration measured at the route level. Industry observers in Canada (and in the U.S.) point out that there are fewer airlines serving passengers today. Figure 1 indicates, however, that the number of effective competitors at the route level has increased from about 1.3 in the first quarter of 1983 (1983:1) to about 1.6 in the fourth quarter of 1990 (1990:4) and reached a peak of nearly 1.7 in the second quarter of 1987 (1987:2).[3]

Figure 2 presents the data on route concentration in a more intuitive way. It shows the percentage of passengers who fly on carriers with 90% or more route market share (that is, near-monopoly carriers). This fell from about 40% in 1983:1 to about 14% in 1990:4. Figure 3 shows the percentage of passengers who fly on carriers with 100% route market share (literally monopolies). The percentage of travellers captive to one carrier fell from about 26% in 1983:1 to about 7% in 1990:4. Similarly, Figure 4 shows the percentage of passengers on competitive carriers (those with route market shares less than 20%). This figure fell from around 6% percent in 1983:1 to about 3% in 1990:4.

Figure 1
INDUSTRY CONCENTRATION: ROUTE LEVEL

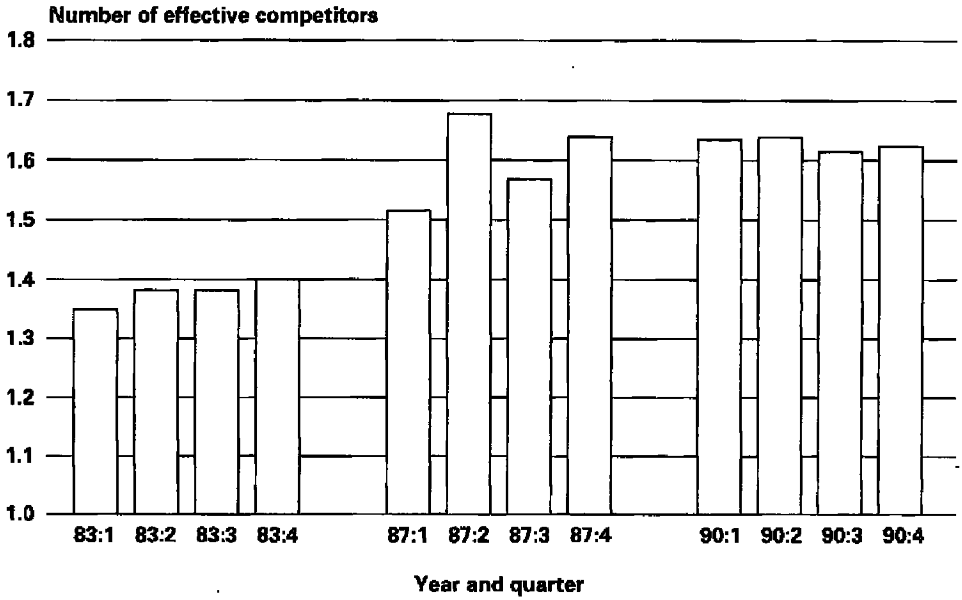


Figure 2
PERCENTAGE OF PASSENGERS FLYING ON CARRIERS WITH >90% ROUTE MARKET SHARE

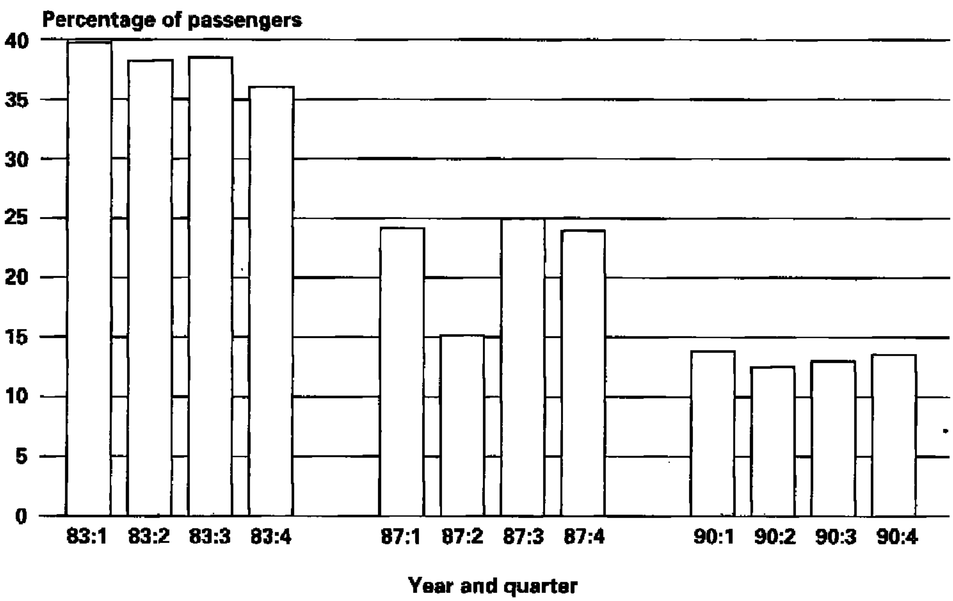


Figure 3

PERCENTAGE OF PASSENGERS FLYING ON CARRIERS WITH 100% ROUTE MARKET SHARE

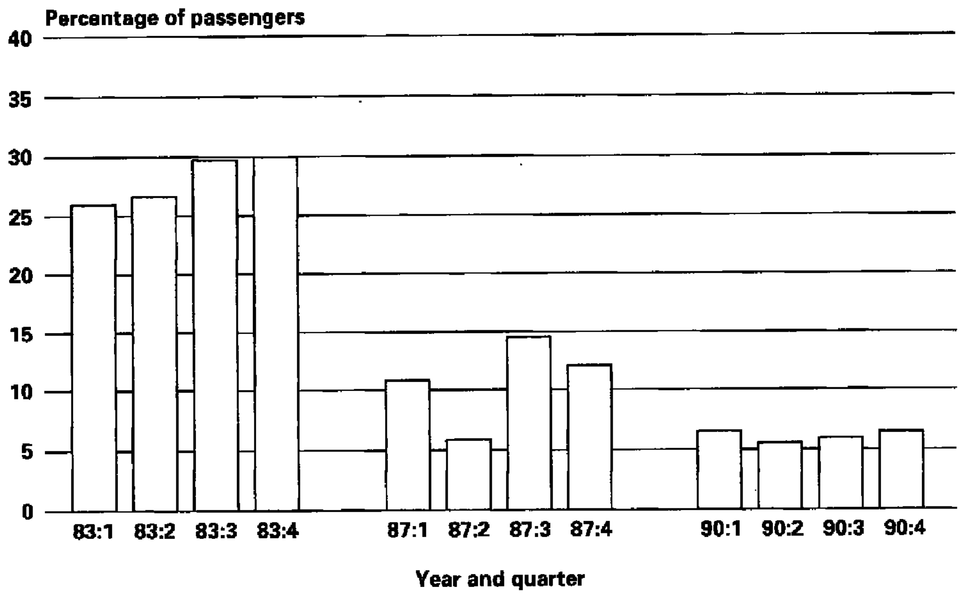
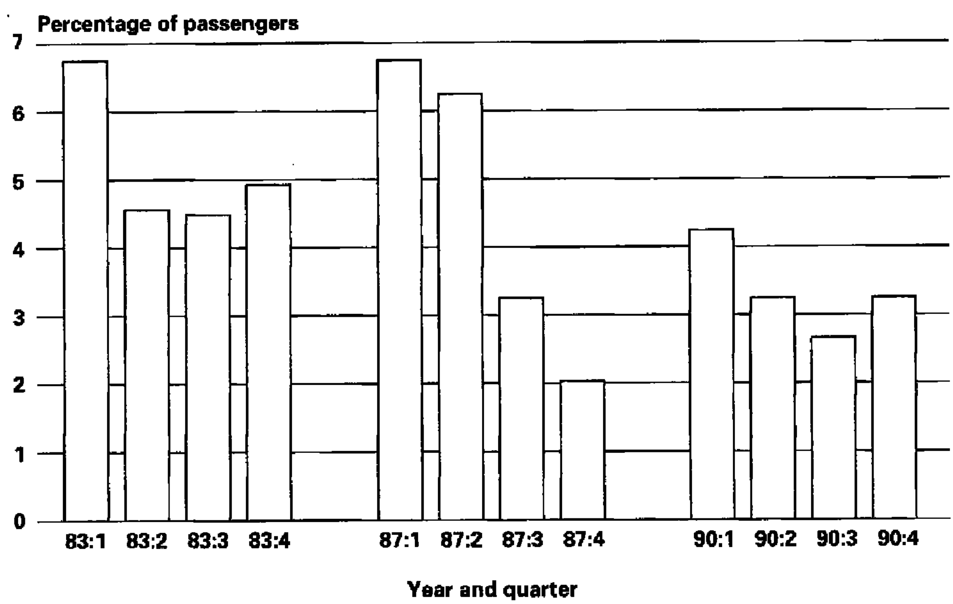


Figure 4

PERCENTAGE OF PASSENGERS FLYING ON CARRIERS WITH <20% ROUTE MARKET SHARE

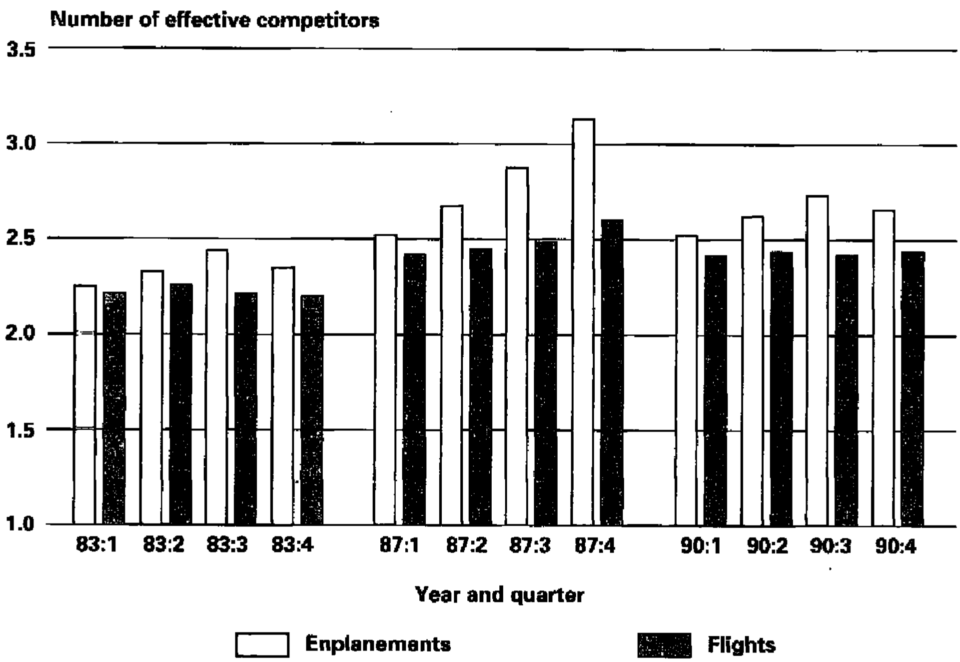


Although there were fewer airlines serving passengers in 1990 than in 1983, airlines competed more frequently at the city-pair level. The number of passengers with a wide choice of carriers declined by three percentage points, while the percentage of travellers captive to one airline declined by about 19 percentage points.

Another measure of industry concentration of interest is airport concentration. This measure of concentration may be relevant because if airports become more concentrated, it may be more difficult for other airlines to enter routes serving those airports. Figure 5 shows airport concentration based on enplanements and flights.[4] Using the enplanement measure, airport concentration has decreased, with the number of effective competitors rising from 2.2 in 1983:1 to nearly 2.7 in 1990:4. Using the flight-based measure, the number of effective competitors has increased from about 2.2 to about 2.4 during the same period. (These average figures may mask the diversity of changes at the individual airport level.)

Figure 5

Airport Concentration Based on Enplanements and Flights



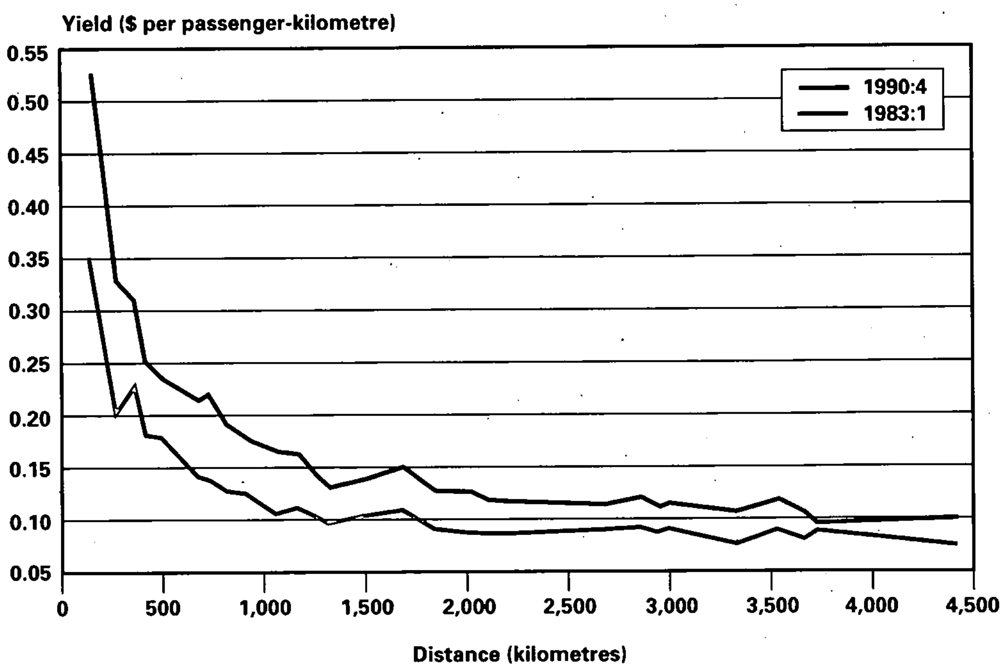
Overall, other than at the national level, the Canadian airline industry was, on average, less concentrated in 1990 than in 1983.[5] Of course, different routes and different airports have not shared equally. Some are less concentrated, while others are more concentrated.

III. Fares[6]

The concentration information outlined in Section II suggests that fares under deregulation should be lower (in real terms) than previously. This section examines the relationship of fares in Canada over time as well as compared with the United States.

Figure 6 shows nominal yield (revenue per passenger-kilometre) as a function of distance in 1983:1 and 1990:4. The first aspect of this graph to notice is the fare "taper." Yield declines with distance, initially at a very rapid rate and then at a slower rate. This reflects (among other possible factors) the

Figure 6
Yield vs. Distance
1983:1 and 1990:4 (Southern)



fixed costs of take off and landing, which, when amortized over longer distances, result in lower yields.

When comparing the yield curves for 1983:1 and 1990:4 it is difficult to see any trend, other than that nominal yields have risen. To get a better idea of fare changes, Figures 7 and 8 show the nominal and real percentage change in fares for several routes. Although a negative trend appears to be present, there is a lot of variation in fare changes across routes. The negative trend is more pronounced in Figures 9 (nominal) and 10 (real), which present average fare changes (averaged across various distance bands). As shown in Figure 10, real fares (using the CPI as the deflator) have risen for routes less than about 1,300 kilometres and have fallen for routes longer than about 2,100 kilometres.

Figure 7
AIR FARE CHANGES BY ROUTE FROM 1983:1 TO 1990:4

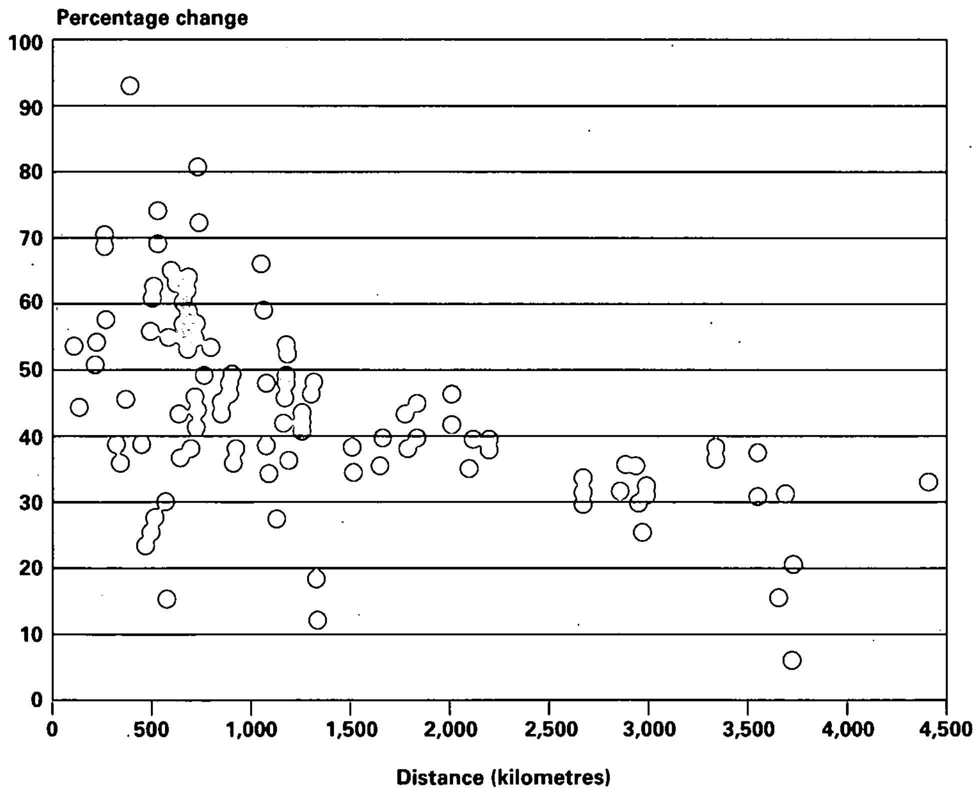
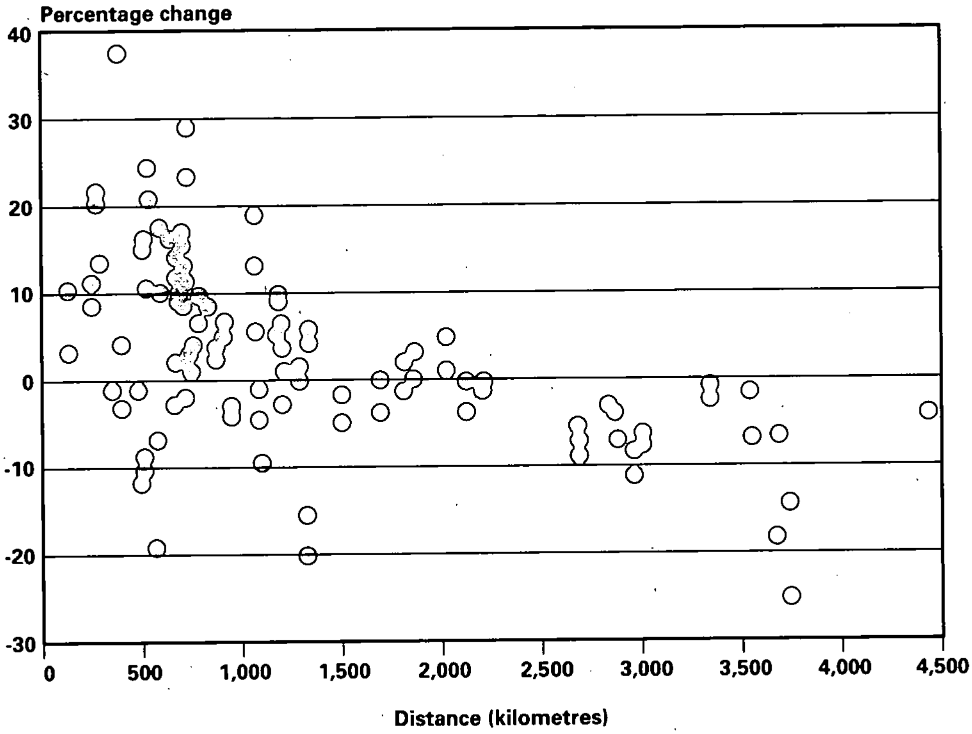


Figure 8

Real Air Fare Changes by Route from 1983:1 to 1990:4



Thus, independent of the overall effect of deregulation on fares, there has been a significant change in the fare taper. There is now more of a taper with distance. Relative to each other and relative to inflation, short-haul fares have risen and long-haul fares have fallen. This pattern is quite similar to that observed in the United States in the wake of airline deregulation there.

Figures 11 and 12 show the relationship of fares in Canada to fares in the United States as a function of distance for 1983:1 and 1990:4.[7] In both years, Canadian air fares were lower than U.S. air fares for short-distance flights (less than about 1,500 kilometres) and greater than U.S. fares for long flights. Since many factors other than distance affect fares (for example, route density, load factor) one cannot conclude from these figures that Canadian long-haul fares are too high (or that U.S. short-haul fares are too high). In fact, the result that short-haul fares in Canada were, and remain, lower than U.S. air fares for the same distances may only reflect an anomaly

Figure 9
PERCENTAGE CHANGE IN AIR FARES BETWEEN 1983:1 AND 1990:4 (SOUTHERN)

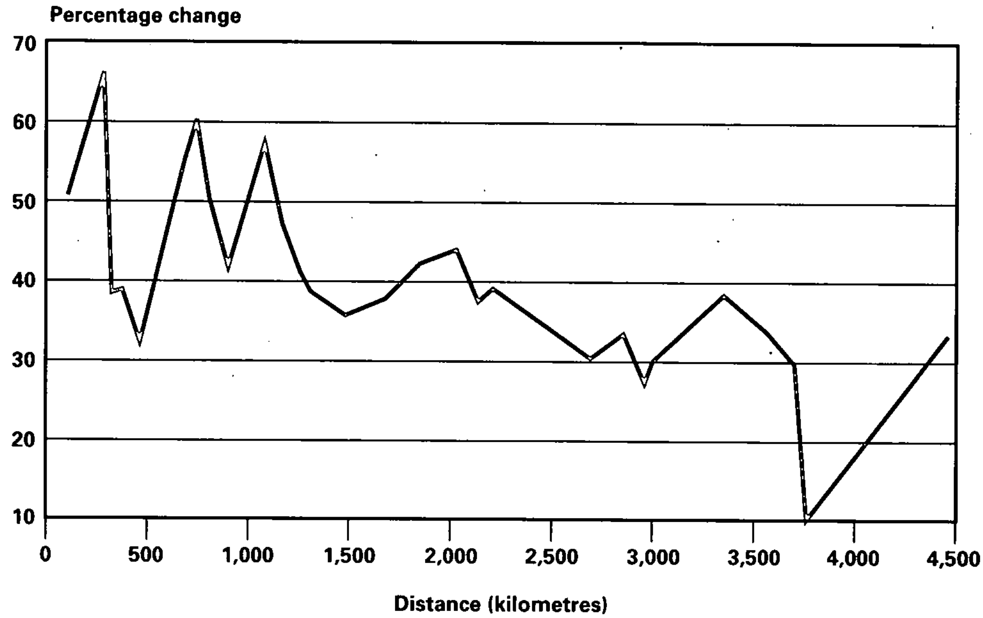


Figure 10
PERCENTAGE CHANGE IN REAL AIR FARES BETWEEN 1983:1 AND 1990:4 (SOUTHERN)

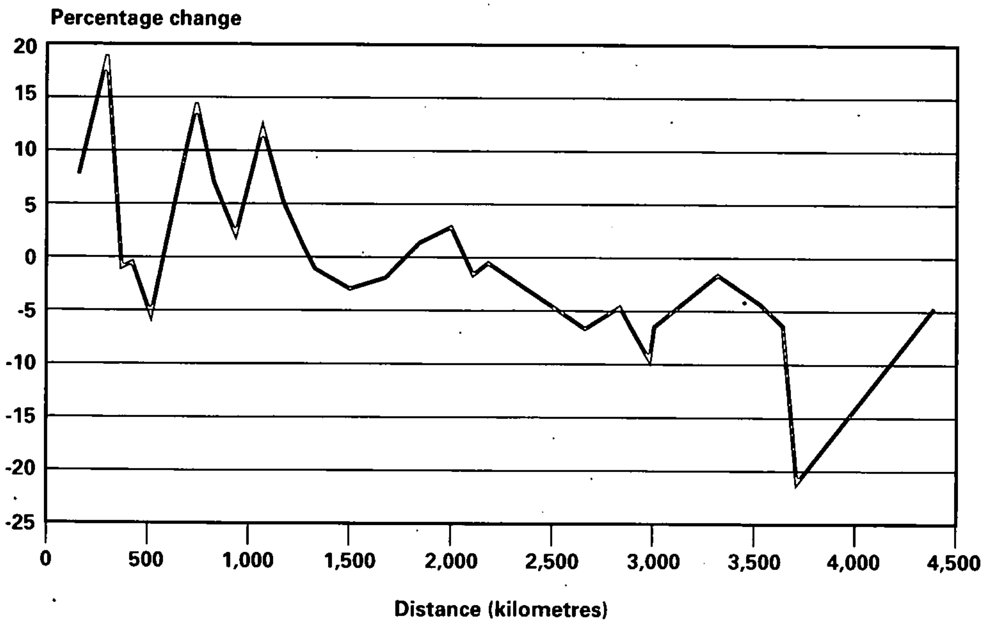


Figure 11
RELATIONSHIP OF CANADIAN AIR FARES TO U.S. AIR FARES IN 1983:1

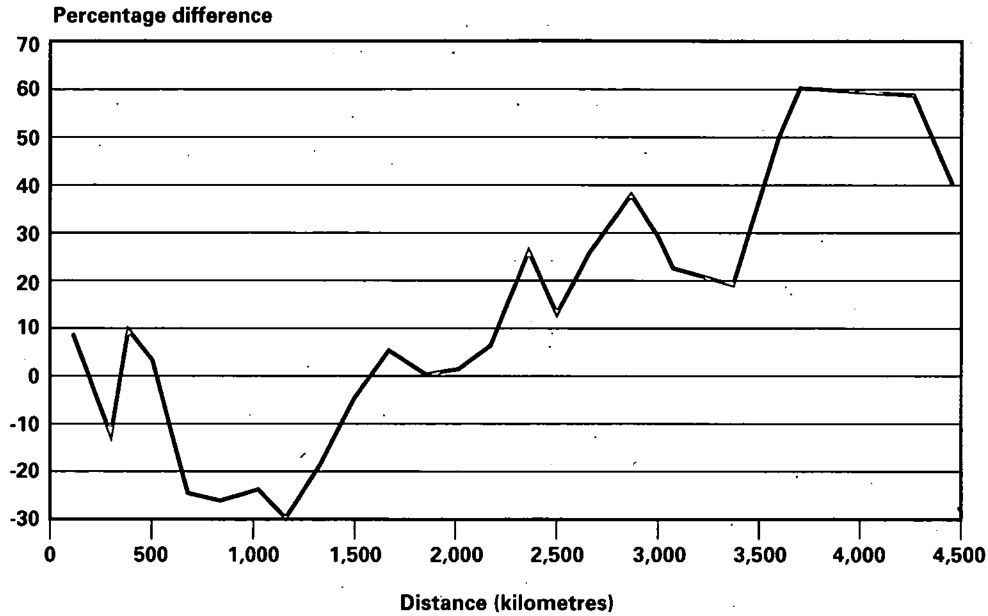
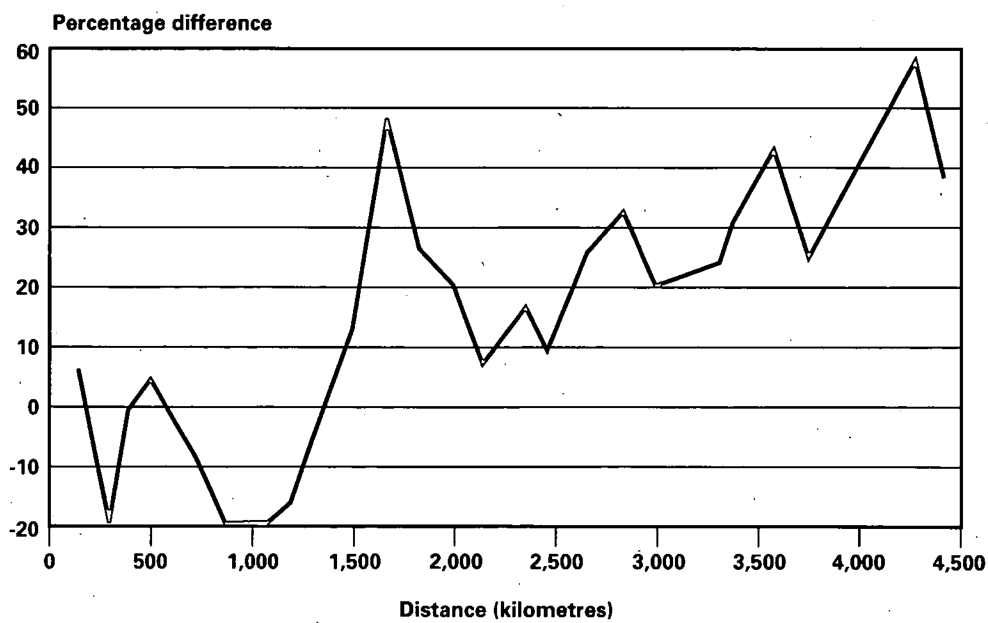


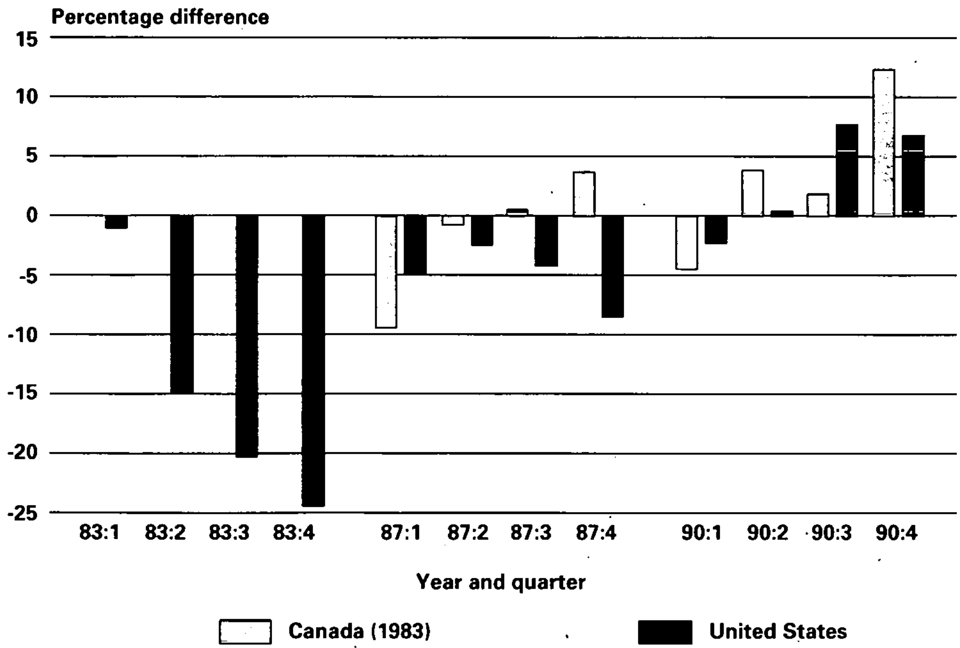
Figure 12
RELATIONSHIP OF CANADIAN AIR FARES TO U.S. AIR FARES IN 1990:4



in the Canadian fare data. If a passenger's journey from A to B involves a connection at point C, the Fare Basis Survey allocates the A–B fare to the segments A–C and C–B. For example, if a passenger flew from Vancouver to Toronto and made a connection at Winnipeg, the Fare Basis Survey would (effectively) show the short-haul Vancouver–Winnipeg and Winnipeg–Toronto legs as having a low yield that should more appropriately be assigned to the long-haul Vancouver–Toronto segment. How much of the observed pattern is due to this anomaly is difficult to say.

Figure 13 presents an attempt to estimate how fares in Canada have changed over time, relative to the fares that prevailed in Canada in 1983 and relative to those in the U.S. in the corresponding year and quarter. In particular, each coupon in the Fare Basis Survey was used as the basis for figuring out what was actually paid for a given trip (that is, segment). This was compared with what would have been paid had the real fares prevailing in Canada in 1983 (for the same distance) been charged. Also, actual fares were compared with those that would have been paid if passengers paid the fares charged in the U.S. (adjusted to reflect the exchange rate) for the same distance during the same year and quarter.

Figure 13
Air Fares in Canada Relative to Comparison Groups



In 1987, fares in Canada were nearly 10% lower in 1987:1 than they were in 1983:1. Over the course of 1987, relative fares increased; by 1987:4 fares were nearly 4% higher than they were in 1983:4. In 1990:1 fares were nearly 5% lower than fares in 1983:1. By 1990:4 relative fares increased to nearly 13% higher than in 1983:4.

The figure also compares fares in Canada with those in the U. S. Due to the anomaly in the Canadian data discussed above, however, *these figures must be interpreted with caution.* Based on these results, fares in Canada were nearly 25% lower than U.S. air fares in 1983:4. By 1990:4 they were about 7% higher than U.S. air fares.

Because of the caveat above, it is difficult to put much credence in the Canada-U.S. comparison. In Canada-Canada inter-temporal comparisons, any bias should cancel out. If the fourth quarter of 1990 is treated as an anomaly due to the Persian Gulf Crisis, it can be concluded that, when compared with 1983, Canadian airline deregulation did not change the level of fares very much, but significantly changed the structure (with distance). Presumably this changing structure makes the variation of fares with distance more in line with the variation in costs.

These results are at odds with those of Oum et al. (1991)[8] who reported that Canadian air fares declined by 18% in real terms between 1983 and 1989 (the last year they were studied). What accounts for the discrepancy? First, it appears that the figures in Oum et al. were for all operations of Canadian air carriers. In particular, it appears that they included northern and southern domestic Canadian traffic, transborder traffic and international traffic. The figures in this study relate only to southern (deregulated) domestic Canadian traffic. Second, because they used aggregate data and were thus forced to focus on (average) yield, Oum et al. picked up some of the effect of the changing composition of routes in the sample. For example, as short routes become (relatively) more expensive and long routes become (relatively) cheaper in the wake of deregulation, one would expect the deregulated era to contain a larger proportion of long trips than the regulated era. Indeed, even holding constant the routes in question, as in Figure 7, the average length of haul (passenger-kilometres per passenger) increased to 1,381 kilometres in 1990:4 from 1,193 kilometres in 1983:1. Since the coefficient of the log of distance in the fare regressions (see below) is approximately 0.5, the change in the distribution of trips alone would lower (average) yield by about 7% even if fares on each route remained constant. In particular, using the routes represented in Figure 7, comparing the yield in 1983:1 with the yield in 1990:4 reveals a nominal increase of 31.5% and a real decrease of 6.3%. When the weight attached to each route was held constant, however, the average yield increased by 38.8% (with 1990:4 weights) and 40.1% (with 1983:1 weights). Real fares decreased by 1.1% with 1990:4 weights and by 0.2% with 1983:1 weights.

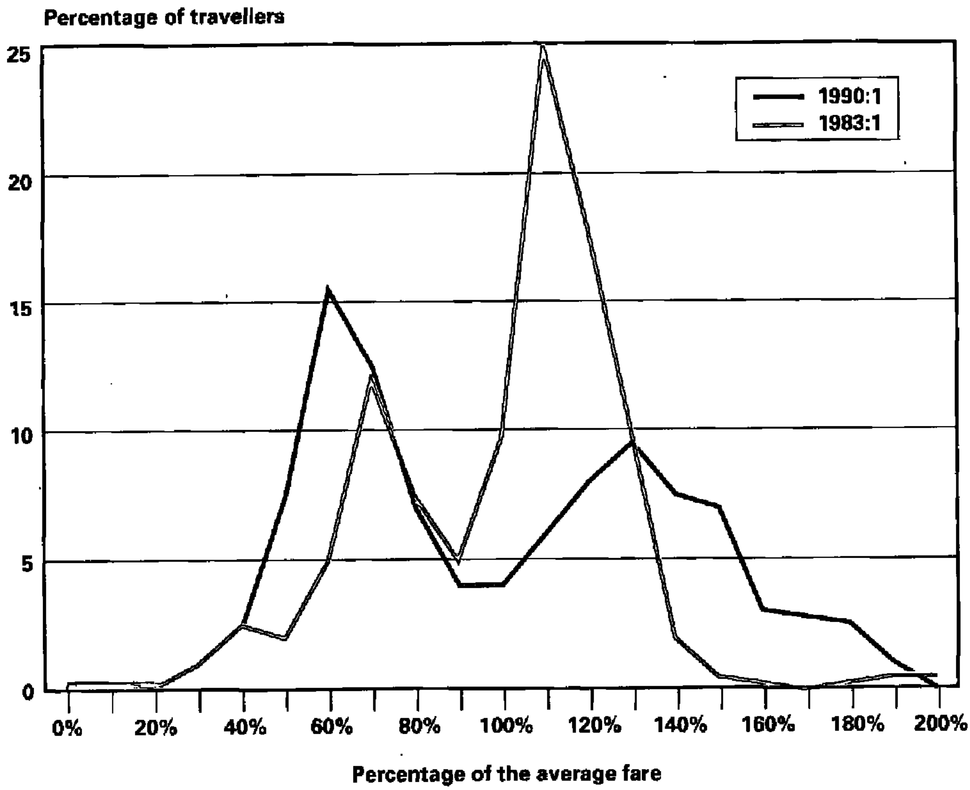
Another issue relating to fares is their distribution on each route. Table 1 shows the percentage of passengers in each of eight fare classes in 1983:1 and 1990:4.[9] The most dramatic change in the table is in the changes in the proportions of regular economy and advanced purchase fares. Over 50% of travellers in 1983:1 flew on a regular economy fare. By 1990:4 this had fallen to less than 25%. Advanced purchase fares show the opposite pattern. About one quarter of travellers flew on advanced purchase fares in 1983:1. By 1990:4 over one half of passengers availed themselves of the discounts available by making reservations and purchasing tickets in advance. During this time the depth of the discount of advanced purchase fares relative to regular economy fares increased to 37% from 24%.

Table 1
PERCENTAGE OF PASSENGERS BY FARE CLASS

Fare Class	1983:1	1990:4
First class	0.7	0.3
Regular economy	56.4	23.6
Advanced purchase	28.8	52.0
Non-advanced purchase	7.6	13.1
Other	0.1	0.6
Industry and agency discount	1.0	1.5
Unknown	5.3	1.5
Business class	0.0	7.6

The proliferation of advanced purchase discount fares raises the question of the distribution of fares on each route. Figure 14 looks at the change in fare distribution. It plots the distribution of fares around each route's average for 1983:1 and 1990:1. The fare structure was bimodal in 1983:1 and remained bimodal in 1990:1. However, the fare structure was more dispersed in 1990:1 than in 1983:1. This could be because airlines erected "fences" and priced based on willingness to pay (that is, demand elasticity). Alternatively, prices could be based on costs, in which peak-period flights and on-demand service are more costly to provide. Most likely, both factors are influencing the increased spread of fares.

Figure 14
DISTRIBUTION OF AIR FARES



IV. LOAD FACTORS

Load factor is the percentage of seats occupied by paying passengers. From the point of view of passenger convenience, low load factors are good because the lower the load factor, the more likely a passenger can get a seat on the flight that he or she desires. On the other hand, since airlines must recover their costs from the passengers they carry, the higher the load factor, the lower the fare. The optimal load factor reflects a trade-off between these two conflicting components. The optimal load factor rises with distance (other factors being constant). Because long-distance flights are more expensive than short-distance flights, it is optimal to have fewer excess seats.

Figures 15 through 22 show load factor comparisons as a function of distance for each quarter in 1983 with the corresponding quarter in 1990. The first graph of each pair shows the load factor for each of the years. The second graph shows the percentage change in the load factor. In all cases, load factors rise with distance. In comparing the first three quarters of 1983 with the corresponding quarters in 1990, load factors fell for short-distance flights and rose for long-haul flights. This is consistent with the fare changes: up for short-haul flights and down for long-haul flights. This pattern disappears in comparisons of the last quarter in 1983 with the corresponding quarter in 1990. Whether this reflects the evolution of deregulation or anomalies of the Persian Gulf Crisis remains to be seen. Thus, at least through the first three quarters of 1990, passengers paid higher fares for short-haul flights but found it easier to get a seat on the flight of their choice. Long-haul fares fell but it was harder to find a seat.

Figure 15

Relationship of Load Factor to Distance in 1983:1 and 1990:1

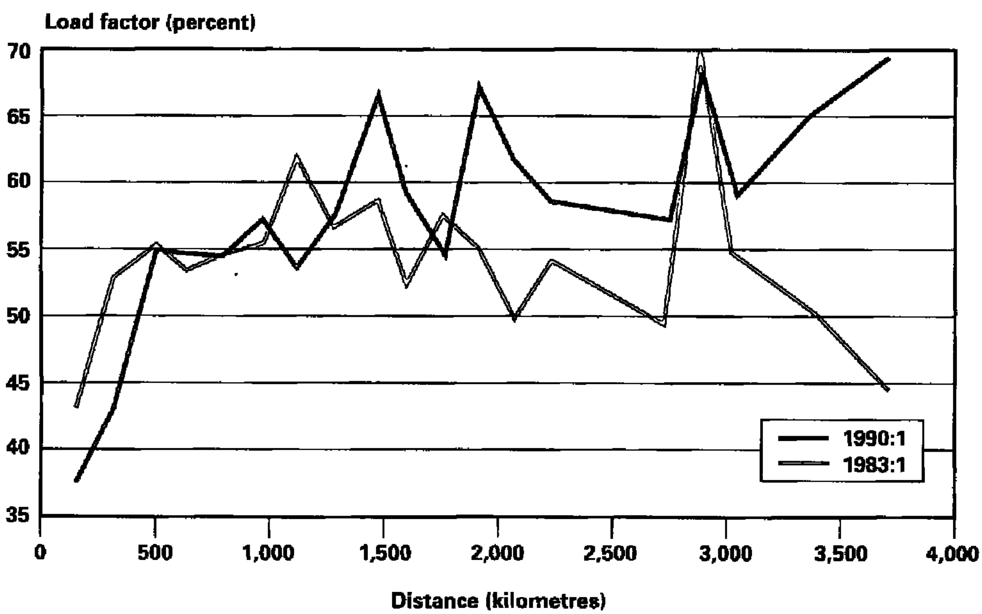


Figure 16

PERCENTAGE CHANGE IN LOAD FACTOR BETWEEN 1983:1 AND 1990:1

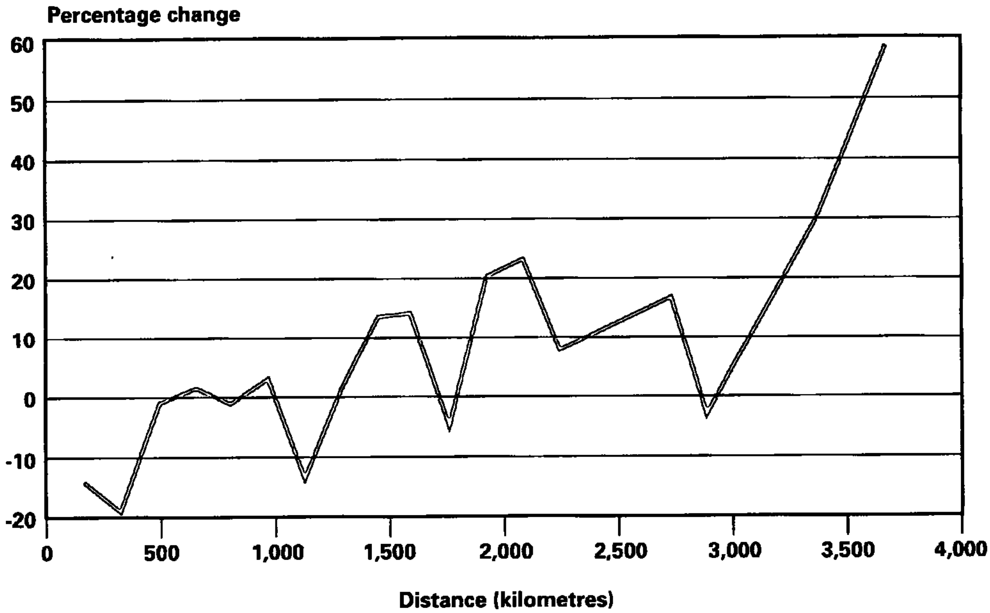


Figure 17

RELATIONSHIP OF LOAD FACTOR TO DISTANCE IN 1983:2 AND 1990:2

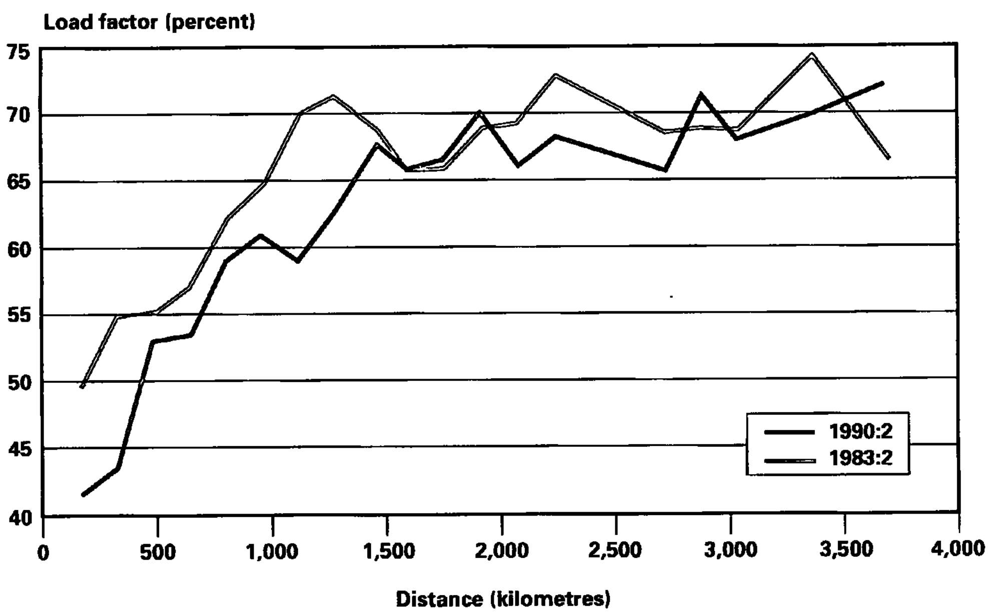


Figure 18
PERCENTAGE CHANGE IN LOAD FACTOR BETWEEN 1983:2 AND 1990:2

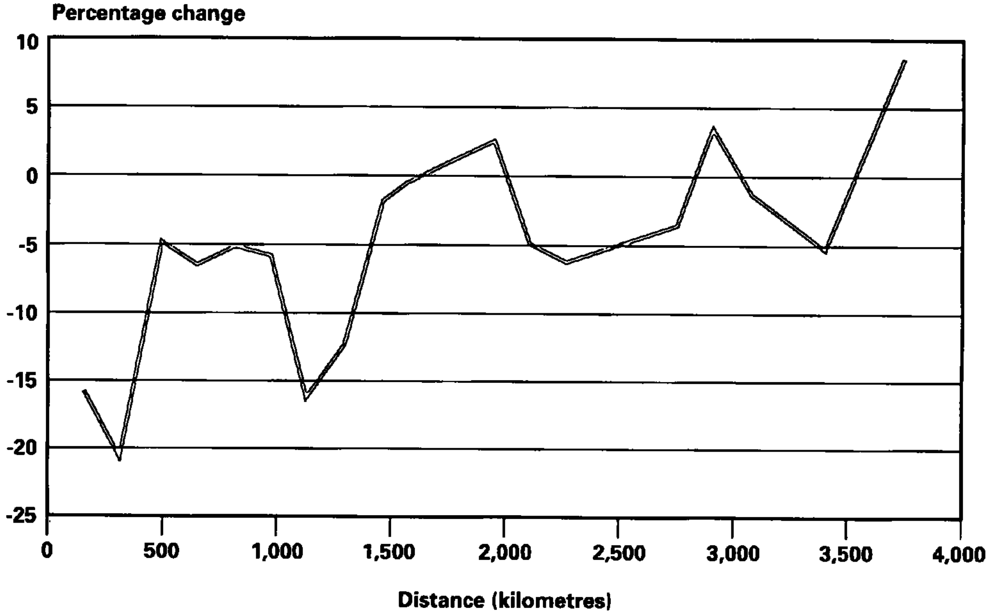


Figure 19
RELATIONSHIP OF LOAD FACTOR TO DISTANCE IN 1983:3 AND 1990:3

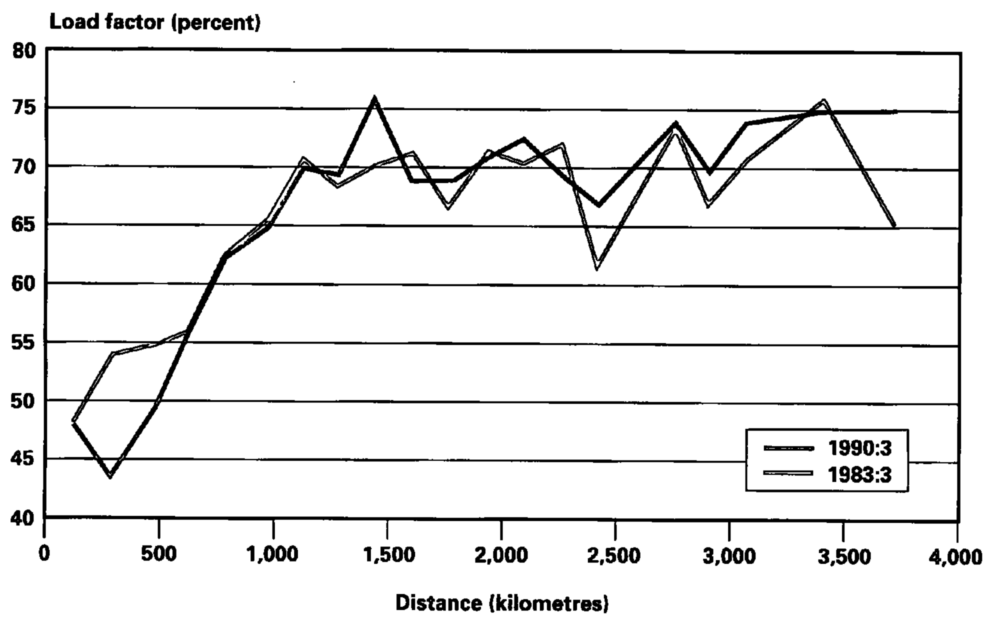


Figure 20
PERCENTAGE CHANGE IN LOAD FACTOR BETWEEN 1983:3 AND 1990:3

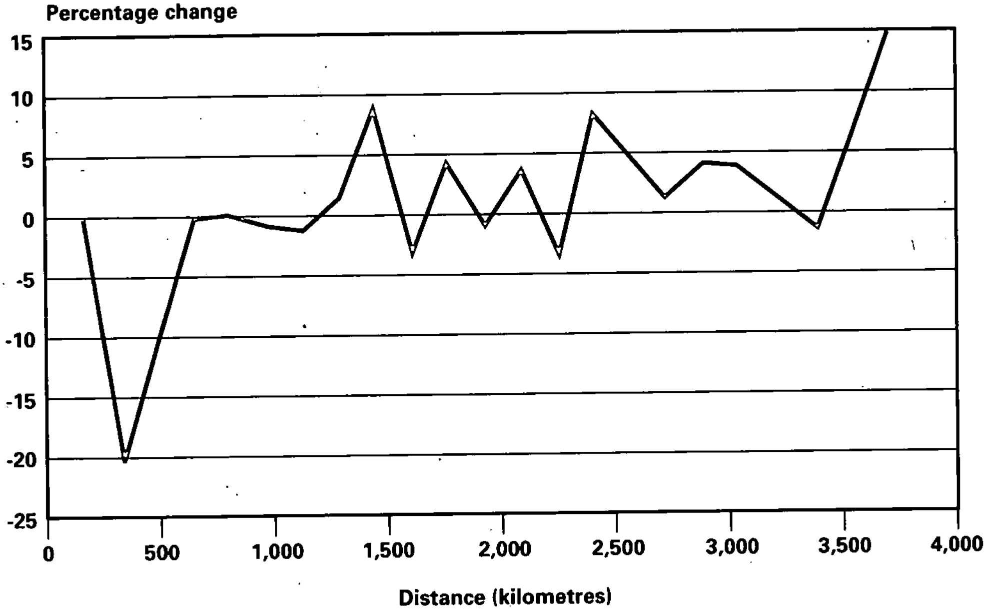


Figure 21
RELATIONSHIP OF LOAD FACTOR TO DISTANCE IN 1983:4 AND 1990:4

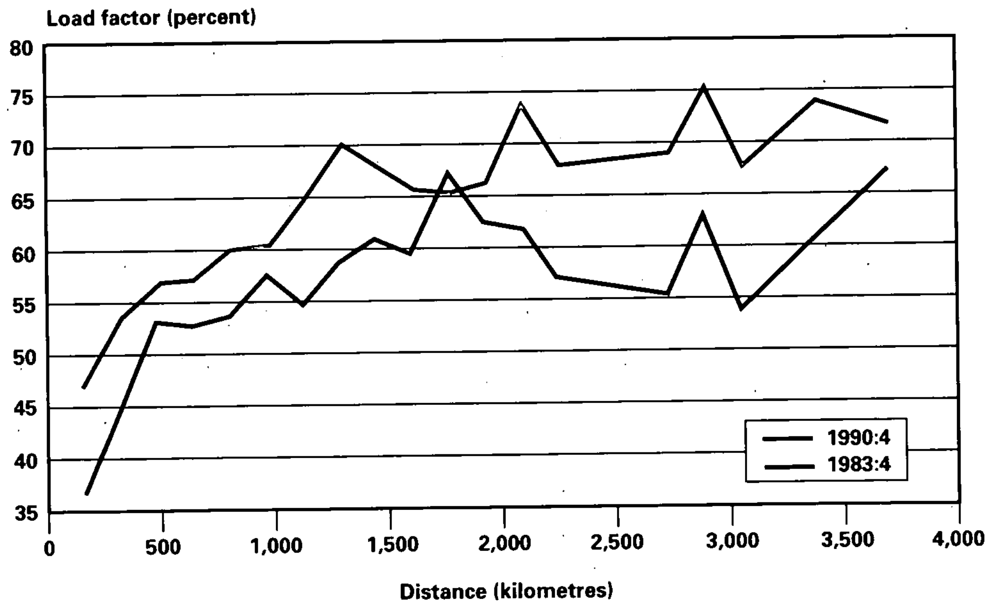
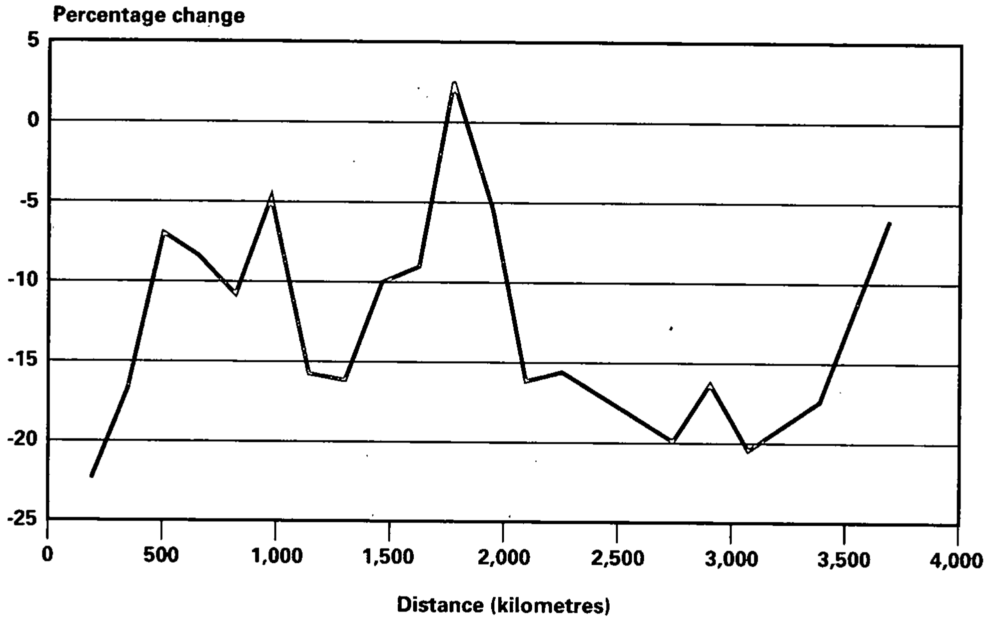


Figure 22
PERCENTAGE CHANGE IN LOAD FACTOR BETWEEN 1983:4 AND 1990:4



V. FARE REGRESSIONS[10]

The previous sections presented data on what has happened to route concentration, airport concentration and air fares since 1983. This section amalgamates those pieces and attempts to explain air fares on each route as a function of the concentration variables, a route density variable and distance, for 1983, 1987 and 1990.

The data sources for these regressions are the same as those used already in this study: Fare Basis Survey Data Base (for fares), Airport Activity Data Base (for airport concentration) and Revenue Passenger Origin and Destination Survey (for route concentration). The Fare Basis Survey is completed and filed by Canadian Level I air carriers operating scheduled passenger service, both domestic and international. In 1983, Air Canada, CP Air, Eastern Provincial Airways, Nordair and Pacific Western Airlines participated in the Fare Basis Survey. In 1987, Air Canada, Canadian Airlines International and Wardair participated. In the 1990 survey, Air Canada and Canadian Airlines International participated. The routes included in the analysis were all domestic routes (for which data were available) in the deregulated southern sector. For a city-pair route to be included in the reported regressions, appropriate data for that route and for its end points had to be available in all three data bases. Given that two of the data bases are samples from a larger population, there were some cases where routes had to be dropped because of lack of data.[11]

Table 2 reports the results of a cross-section regression of fare in 1983 on distance (LDIST), a measure of route concentration (LACTRTE),[12] a measure of airport concentration at the origin (LCOMPO) and destination (LCOMPD),[13] and a measure of route density (LPOPPOP).[14] A measure of route density is needed because, other things being equal, as route density increases, fares are expected to fall and more competitors are expected on high-density routes. Failure to control for route density may result in route density being captured by the coefficient of the competition variable, resulting in a route competition coefficient that is larger in magnitude than the true effect of competition on fares. The variable used here, while not a direct measure of route density, does not suffer from being endogenous, as would a direct measure of route density, for example, the number of passengers travelling on that route.

The first equation shows that fares increase with distance, although less than proportionally (reflecting the fare taper). Also, as the number of effective competitors on a route increases (LACTRTE), fares decrease. Since the equation is in logarithms, the -0.0507 coefficient indicates that a 1% increase in the number of effective competitors on a route reduces fares by 0.0507%. This coefficient is statistically significant at conventional significance levels. Neither of the airport concentration variables (LCOMPO, LCOMPD) is statistically significant. The route density variable (LPOPPOP) is negative and statistically significant indicating that as route density (that is, the product of the populations of the origin and destination) increases by 1%, fares decline by 0.0177%. Collectively, these variables explain 86.1% of the variation in (the log of) fares. Distance alone explains 85.4%.

Table 2
FARE REGRESSION FOR 1983

LS // Dependent Variable is LFARE
SMPL range: 1–1,955
Number of observations: 1,955

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.5148964	0.0311805	48.584731	0.0000
LDIST	0.5086341	0.0050752	100.22025	0.0000
LACTRTE	-0.0507408	0.0194973	-2.6024466	0.0093
LCOMPO	-0.0010721	0.0104145	-0.1029449	0.9180
LCOMPD	0.0066027	0.0103955	0.6351446	0.5254
LPOPPOP	0.0177170	0.0025000	-7.0867148	0.0000

R-squared	0.860949	Mean of dependent var.	4.680293
Adjusted R-squared	0.860592	S.D. of dependent var.	0.465855
S.E. of regression	0.173938	Sum of squared resid.	58.96593
Log likelihood	648.3842	F-statistic	2413.482
Durbin-Watson stat	1.964760	Prob(F-statistic)	0.000000

Table 3 shows the results for the same specification for 1987. No variable, except distance, is statistically significant, and all coefficients, except route density and distance, have the "wrong" sign. Collectively, these variables explain 61.5% of the variation in (the log of) fares. Distance alone explains 61.4%.

Finally, Table 4 presents regression results for 1990. Once again, only distance and the route density variable are statistically significant. Collectively, these explanatory variables account for 79.1% of the variation in (the log of) fares. Distance alone explains 78.2%.

Because it is possible that more observations would lead to more precisely estimated coefficients, the data for 1987 and 1990 were pooled and a single regression was estimated with dummy variables for each year serving as "intercepts." These results are reported in Table 5. Unfortunately, they differ very little from the separate results for the two years.

Table 3
FARE REGRESSION FOR 1987

LS // Dependent Variable is LFARE
SMPL range: 1–1,546
Number of observations: 1,546

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.7161854	0.0722403	23.756617	0.0000
LDIST	0.4608607	0.0103730	44.428721	0.0000
LACTRTE	0.0128799	0.0335448	0.3839607	0.7011
LCOMPO	0.0307646	0.0193679	1.5884307	0.1124
LCOMPD	0.0305367	0.0193790	1.5757625	0.1153
LPOPPOP	-0.0048699	0.0057713	-0.8438266	0.3989

R-squared	0.615462	Mean of dependent var.	4.818028
Adjusted R-squared	0.614213	S.D. of dependent var.	0.521164
S.E. of regression	0.323704	Sum of squared resid.	161.3677
Log likelihood	-446.8992	F-statistic	492.9605
Durbin-Watson stat	1.973949	Prob(F-statistic)	0.000000

Table 4
FARE REGRESSION FOR 1990

LS // Dependent Variable is LFARE
SMPL range: 1–754
Number of observations: 754

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	1.8991473	0.0885634	21.443923	0.0000
LDIST	0.5124343	0.0103094	49.705524	0.0000
LACTRTE	0.0408411	0.0416958	0.9795019	0.3276
LCOMPO	0.0369906	0.0271501	1.3624462	0.1735
LCOMPD	-0.0215412	0.0271244	-0.7941638	0.4274
LPOPPOP	-0.0313062	0.0078153	-4.0057451	0.0001

R-squared	0.790645	Mean of dependent var.	5.108579
Adjusted R-squared	0.789245	S.D. of dependent var.	0.517822
S.E. of regression	0.237722	Sum of squared resid.	42.27066
Log likelihood	16.37009	F-statistic	564.9753
Durbin-Watson stat	1.890526	Prob(F-statistic)	0.000000

Table 5
FARE REGRESSION FOR 1987 AND 1990 (POOLED)

LS // Dependent Variable is LFARE
SMPL range: 1–2,300
Number of observations: 2,300

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
D87	1.6794516	0.0554100	30.309522	0.0000
D90	1.9107242	0.0585146	32.653812	0.0000
LDIST	0.4792344	0.0076954	62.275472	0.0000
LACTRTE	0.0313808	0.0262754	1.1943023	0.2325
LCOMPO	0.0291124	0.0157870	1.8440745	0.0653
LCOMPD	0.0140270	0.0157869	0.8885218	0.3744
LPOPPOP	-0.0129064	0.0046296	-2.7877905	0.0054

R-squared	0.690728	Mean of dependent var.	4.913278
Adjusted R-squared	0.689919	S.D. of dependent var.	0.537556
S.E. of regression	0.299338	Sum of squared resid.	205.4600
Log likelihood	-485.8338	F-statistic	853.5311
Durbin-Watson stat	1.950103	Prob(F-statistic)	0.000000

What are we to make of these results for the functioning of deregulated Canadian airline markets? Perhaps the most interesting result is that route competition does not have a statistically significant effect on fares. Unfortunately, the interpretation of this result is ambiguous. If these markets were (perfectly) contestable, route level competition would not have an effect on fares (because the threat of entry would force incumbents to keep their prices down). The same result would be expected, however, if the industry were a cartel. But profit data show no hint of monopoly rents. Another possibility is that the industry is still in transition (1990 was a recession year, and fuel prices increased during the Persian Gulf Crisis).

VI. CONCLUSIONS

It is clear that, on average, deregulation has increased the extent of competition at the route level and at the airport level (although it has declined at the national level). It is also clear that the structure of fares (with respect to distance) has changed, presumably to one more in line with the costs of production. Load factors have changed accordingly, falling on short routes and rising on long routes. These changes are consistent with market behaviour. It appears, however, that deregulation has had a neutral effect on the overall level of fares (when compared with Canadian air fares in 1983).

Route-level competition does not affect fares. Although this is consistent with a cartel, profit data show no hint of monopoly rents. The results are also consistent with a contestable or competitive market. The poor profitability of the industry suggests that this interpretation should be seriously considered.

ENDNOTES

1. Because of this, in calculating the percent of the national market controlled by each firm, analysts must aggregate the city-pair level outputs into a single national output. Typically, this is done by calculating the number of passenger-kilometres (or passengers) transported on each route and summing across routes.

2. Likewise, if one firm had two thirds of the market and two other firms each had one sixth of the market, the Herfindahl index would also equal 1/2. In other words, even though there are three firms in the industry, the relative size of the largest firm makes the industry behave as if it had only two firms.

3. In this case, effective competitors is the inverse of a passenger-weighted average of each route's Herfindahl index. Figures 1 to 5 were calculated from data in the Revenue Passenger Origin Destination Data Base (ticket origin destination).

4. The data for these figures are from the Airport Activity Data Base. Airport-based Herfindahl indices were calculated using share of enplanements and share of flights. These measures for each airport were then weighted by percent of total enplanements or flights, as appropriate. The number of effective competitors is the inverse of this figure.

5. Although competition takes place at the route level and not the national level, there is a concern that, as the number of carriers declines nationally, the possibility of collusion (at the route level) increases.

6. All fare data in this section relate to domestic Canadian services in the southern sector.

7. The exchange rate prevailing at the time of the comparison (that is, 1983:1 and 1990:4) was used to convert U.S. dollars to Canadian dollars.

8. Tae Oum, William Stanbury and Michael Tretheway, "Airline Deregulation in Canada," in *Airline Deregulation: International Experiences*, edited by Kenneth Button, (London: David Fulton, 1991).

9. These results are from the Fare Basis Survey and include only domestic Canadian routes in the southern sector flown by Level I carriers. Only those routes represented in the sample in all quarters of 1983, 1987 and 1990 were included.

10. The variables used in the regressions reported in this section are the natural logarithms of those discussed in the text.

11. The fare regression for 1990 reported in Table 4 has about one half of the observations (routes) as the 1987 regression. These were all the routes for which complete data were available. Presumably the routes formerly served by Level I carriers and now served by their (commuter) affiliates are not part of the Fare Basis Survey.

12. This is the "number of effective competitors" (that is, the inverted Herfindahl index based on passenger shares) for each route in question.

13. Each of these airport measures represents the number of effective competitors at the origin and destination airports, respectively. In particular, these measures are the inverse of the origin and destination airports' Herfindahl indices based on passenger enplanement (both domestic and international) shares.

14. This is the product of the populations of each end point on the route.

15

THE EFFECTS OF U.S. AIRLINE DEREGULATION: A REVIEW OF THE LITERATURE

Ron Hirshhorn*
December 1990

1. BACKGROUND

Prior to the mid-1970s, interstate airlines in the U.S. were subject to a comprehensive framework of controls.[1] The Civil Aeronautics Board (CAB) controlled entry and exit, regulated fares, administered subsidies and oversaw mergers and intercarrier agreements. Under its own initiative the CAB liberalized fare regulation beginning in 1976.[2] Two years later, in October 1978, Congress passed the *Airline Deregulation Act.* This formally acknowledged the need to place "maximum reliance on competitive market forces" and set in place a process for the gradual dismantling of regulatory controls. Since 1982, air carriers have had virtually complete freedom of market entry and exit. The CAB lost control over rates, mergers and acquisitions on January 1, 1983, and it ceased operation entirely at the end of 1984. The Department of Transportation assumed responsibility for approving mergers and for administering a subsidy program (the Essential Air Service Program) to guarantee service to small communities.[3] Safety regulation was not affected and was retained at the Federal Aviation Administration (FAA).

* Research Division, Royal Commission on National Passenger Transportation.

2. General Effects: Overview

The consensus in the literature is that, based on the most important criterion, the effect on consumer welfare, U.S. deregulation has been highly beneficial. The prevailing view is summarized in a recent paper by Alfred Kahn (1988a, p. 321) which acknowledged a number of "unpleasant surprises," but then came to the following conclusion:

> The last ten years have fully vindicated our expectations that deregulation would bring lower fares, a structure of fares on average in closer conformity with the structure of costs, an increased range of price-quality options, and great improvements in efficiency . . . all this along with a 35 percent or so decline in accident rates.

In what is, probably, the most widely cited study of U.S. experience, Steven Morrison and Clifford Winston (1986) estimated that, taking account of both fare and quality of service changes, deregulation has resulted in a $5.7 billion (in 1977 dollars) annual improvement in the welfare of consumers. This welfare gain, which amounted to 35 percent of actual airline revenues in 1977, translated to a benefit of $10.62 per traveller per round trip. In addition, deregulation was estimated to have led to at least a $2.5 billion (1977 dollars) annual increase in industry profits. Although the financial performance of U.S. carriers has been poor, the implication is that it would have been significantly worse in the absence of deregulation.

Enthusiasm for deregulation has waned, but only somewhat, in face of the recent spate of mergers and acquisitions, along with evidence that entry into the industry is not as easy as had been presumed. Few economists would view these developments as testimony to the failure of deregulation. While competition may be imperfect, the attendant costs are not seen to approximate nearly the costs of what was a highly distortionary regime of price and entry regulation.

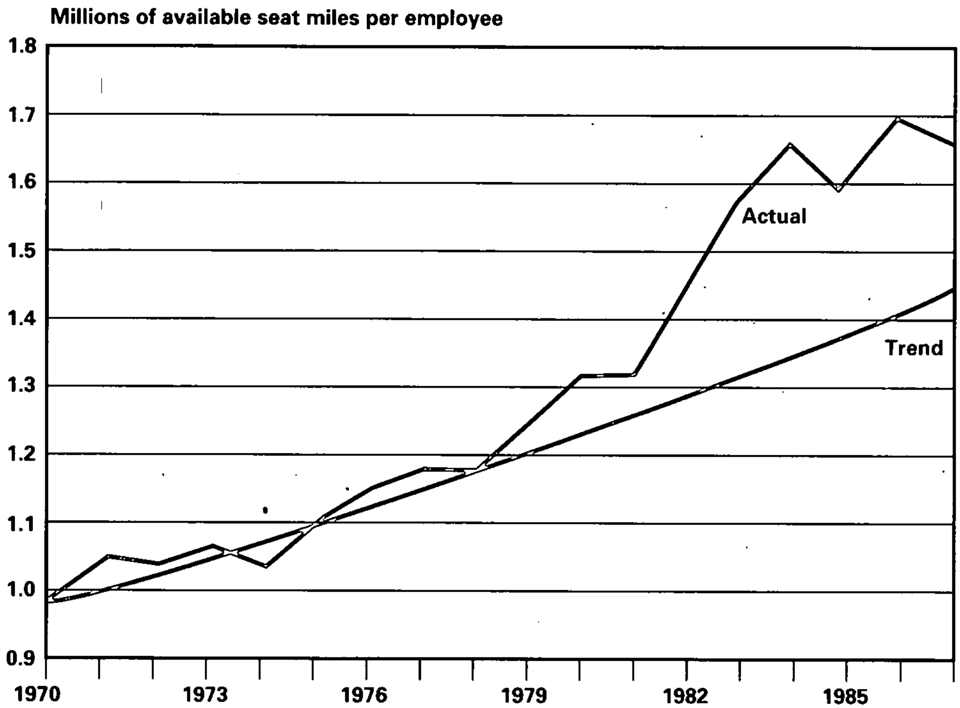
Recent developments, however, do underscore the need to understand the nature and effect of structural changes in the U.S. airline industry. High concentration and impediments to entry will undermine the benefits of deregulation if they reduce the pressure for carriers to improve efficiency or result in monopolistic pricing practices. There is, therefore, a continuing need for some form of monitoring of the conduct and performance of carriers. It is also important to attempt to understand the extent to which recent structural developments are the result of technological characteristics of air service as defined from certain features of the policy environment.

3. Main Sources of Gain from Deregulation

Several studies, in addition to Morrison and Winston (1986), have examined the impact of deregulation on airline efficiency. Bailey, Graham and Kaplan (1985), reported that competition from new low-cost entrants had led to a substantial increase in employee productivity among the formerly regulated carriers. More recent data gathered by the Congressional Budget Office (1988) substantiate the improvement; labour productivity, defined as seat-miles per full-time employee, increased at an annual average rate of 3.8% in the decade after 1978, compared with 2.3% between 1970 and 1978 (see Figure 1). At the same time, average load factor, another indicator of efficiency, increased from around 55% prior to deregulation to between 60% and 65% in recent years.[4] Caves et al. (1987) assessed the performance of U.S. carriers before and after deregulation against a control group of some 27 large carriers operating in countries where there was little or no deregulation. The evidence on foreign experience provided the authors with an alternative basis to consider what would have occurred in the absence of deregulation in the U.S.[5] They found that the total factor productivity of U.S. airlines, which accelerated after deregulation, would have instead fallen very substantially had the performance of U.S. carriers after 1975 changed similarly to that of foreign carriers.[6]

David Sawers (1987) examined the operating costs of U.S. carriers and attempted to distinguish cost reductions that are attributable to new aircraft from those that are due to managerial effort. It is the latter that is the test of airline efficiency and reflects the contribution of deregulation. His results, which are shown in Table 1, reinforce the finding that the efficiency gains achieved by U.S. carriers accelerated after deregulation.[7]

Figure 1
LABOR PRODUCTIVITY
ACTUAL VS. TREND UNDER REGULATION



Source: Congressional Budget Office (1988).

Note: Includes both domestic and international operations. In computing the trend, labor productivity is assumed to have grown after 1978 at the same rate it had grown between 1970 and 1978. Employment data are for December. Part-time employees are counted as one-half full-time workers.

Table 1
COST REDUCTIONS IN U.S. DOMESTIC TRUNK AIRLINES, 1970–1984

	1970	1978	1984
Cost per available ton-mile (ATM) at 1978 input costs (cents)	47.37	38.7	34.03
		percent	
Annual rate of reduction in costs (since 1970 or 1978)	—	2.1	1.9
Annual rate of reduction in costs attributable to new aircraft	—	0.9	0.4
Annual rate of reduction in costs attributable to management	—	1.2	1.5
Annual increase in capacity of fleet in ATMs	—	3.2	1.6
Annual increase in traffic in revenue passenger-miles (RPMs)	—	6.9	1.4

Source: Sawers (1987)

Although questions can be raised about various aspects of these studies,[8] the cumulative evidence provides persuasive documentation of the favourable influence of deregulation on airline efficiency and costs. Airline performance improved significantly in the post-1978 environment when carriers enjoyed greater operating freedom and were subject to more intense competitive pressures. Airline costs declined partly because of the elimination of high rents enjoyed by airline workers under a regulatory regime that permitted high labour costs to be largely passed on to consumers. For example, Bailey, Graham and Kaplan (1985) found that, after deregulation, airlines were instituting two-tier wage structures (under which new employees are paid substantially less than employees already on the payroll), and introducing much more competitive work rules.[9] In addition, gains were achieved because airlines responded to the new opportunities and pressures under deregulation by taking advantage of available economies, and by relating their fares and quality of service more closely to costs and the nature of consumer preferences. These latter developments merit elaboration.

In general, the costs of air service per passenger are positively related to distance, and negatively related to the number of passengers. But while costs are lowered when larger aircraft are run at higher load factors, the resulting decline in flight frequency will inconvenience passengers. To maximize profits, carriers must balance cost savings from economies of scale and utilization against the revenue implications of the associated decline in quality of service.

Bailey, Graham and Kaplan (1985) have shown that the trade-offs are such that profit maximization will lead to the use of larger aircraft at higher load factors as distance and market density increase. In high-volume markets, the economies from more efficient use of resources exceed the decline in demand from reduced flight frequency. The same is true in longer-distance markets where consumer preferences are strongly influenced by the difficulty of substituting other modes of transport for air travel.

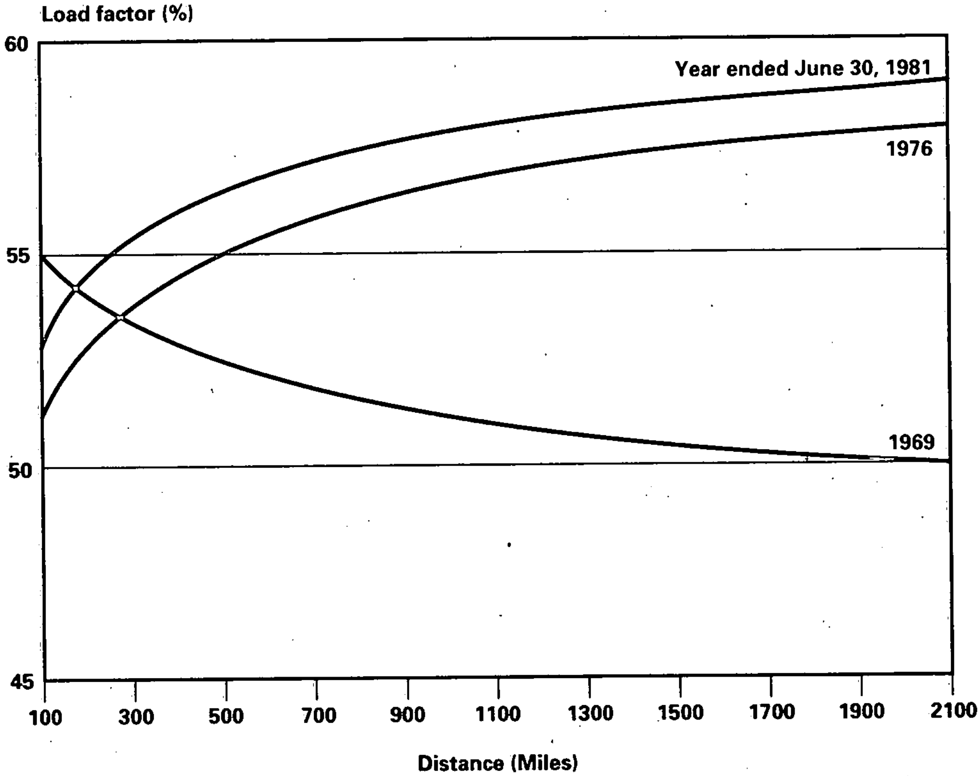
Under CAB regulation, carriers could not respond to these market realities. Part of the problem was a system of regulated prices which did not reflect costs. In short-haul markets, prices were set below costs, leading to an insufficient allocation of resources. By contrast, in long-haul markets, airlines engaged in non-price competition by increasing flight frequency and lowering load factors. The result was exactly the reverse of what we should expect from the economics of air service; load factors were lower on long-haul than on short-haul routes (see Appendix A).

In a study predating deregulation, Douglas and Miller (1974) examined the costs of the service competition which took place under CAB regulation as a substitute for price competition. They found that the benefits to travellers from more frequent flights and improved in-flight amenities were more than offset by the attendant higher fares. The resulting welfare loss to travellers was estimated to be greater than $1 billion per year during the early 1970s.

Deregulation eliminated the misleading signals that were a product of CAB price regulation. Figure 2, which comes from Graham, Kaplan and Sibley (1983), shows that the perverse relationship between load factor and distance, which Douglas and Miller had identified for 1969, had been reversed with the relaxation of regulatory controls and the replacement of service with price competition.[10] Load factor increased at most distances because most consumers preferred a lower combination of price and service quality than was available under regulation. At the same time, however, deregulated carriers continued to offer a more expensive product to meet the special needs of those passengers who place a high value on service convenience.[11]

A second important consequence of deregulation was that it allowed carriers to alter their market coverage to exploit available economies more fully. This has resulted in the increased use of hub-and-spoke route structures, under which passengers from various cities are fed (by spoke routes) into a centralized airport (the hub), from which they take connecting flights to their destinations. Hub-and-spoke systems take advantage of economies of scope that exist because it is often less costly to produce airline outputs jointly than independently. Passengers from various origins who are going to a common destination are collected at a hub airport and put onto a single large aircraft, thereby producing a joint output. Underlying the advantages of joint production are the economies of larger aircraft and higher load factors that were discussed earlier. When these savings exceed the costs of rerouting traffic, economies of scope exist.

Figure 2
LOAD FACTORS AND DISTANCE[a]



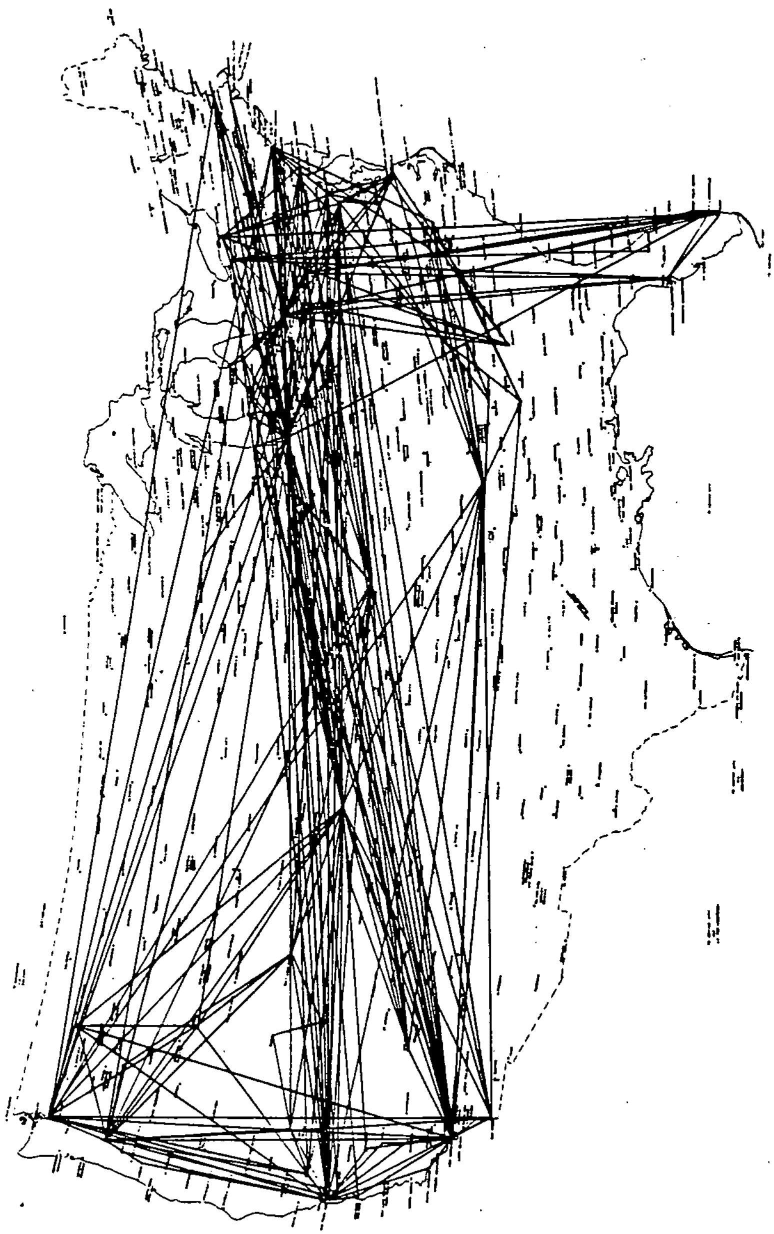
Source: Graham, Kaplan and Sibley (1983)

a For each year, curve is drawn for the mean values of passengers and number of airlines or Herfindahl Index.

Hub-and-spoke systems were in operation before deregulation. The elimination of entry controls accelerated their use and allowed for a more rational structuring of hubs. The experience of United Airlines, which is illustrated in Figures 3A and 3B, typifies the dramatic changes in route patterns that occurred after deregulation.

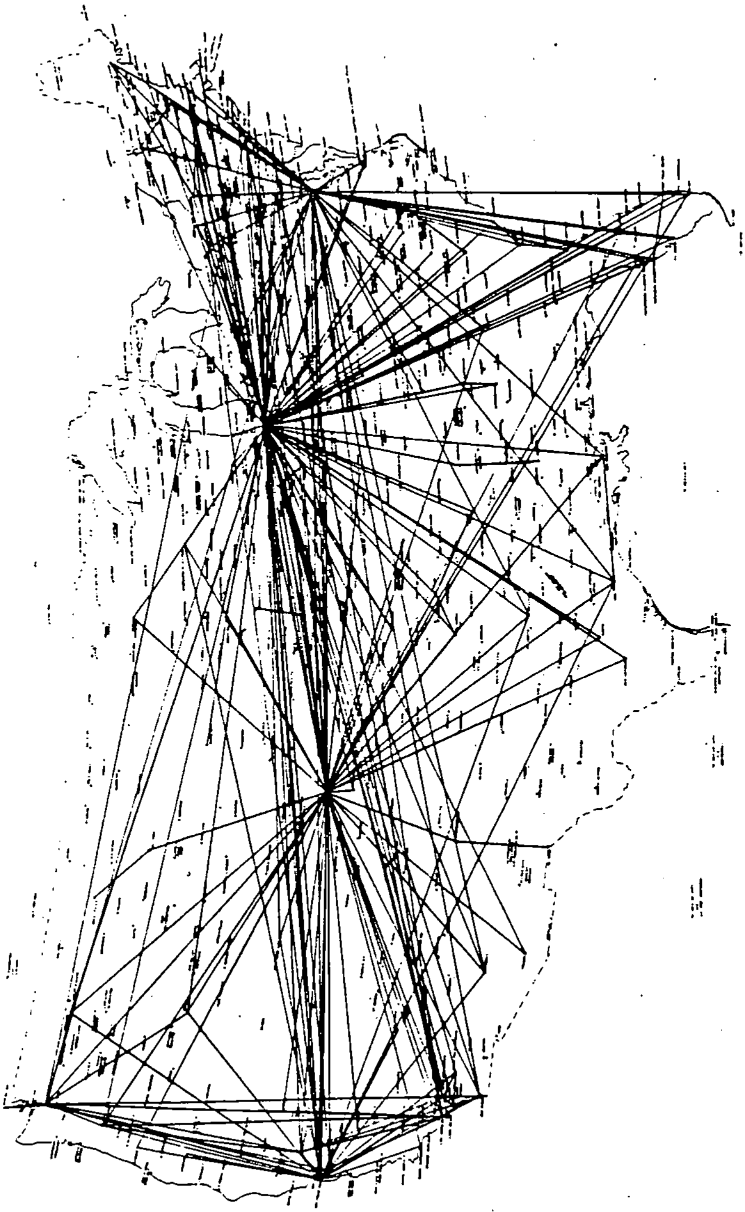
The hubbing process requires carriers to synchronize flights to and from the various points in their system. Timing is important; to cross-connect passengers, carriers must move waves (or banks) of aircraft through connecting points at about the same time. A successful hubbing system also depends

Figure 3A
United Airlines, Inc.
Unduplicated Flight Pattern, July 1979



Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990d)

Figure 3B
UNITED AIRLINES, INC.
UNDUPLICATED FLIGHT PATTERN, MARCH 1989



Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990d)

on the establishment of an extensive network, involving cross-connections to a large number of points. An airline with several spokes emanating from its hub will be well placed to offer passengers service to their ultimate destination. The high "feed" from incoming flights will also help the airline achieve a high load factor on flights departing from its hub. While the hub-and-spoke system has increased efficiency, it has also had important implications for service quality and for airline concentration aspects (which we discuss in later sections).

4. AIRLINE FARES

It is generally agreed that deregulation brought fares more in line with resource costs of providing alternative services. Fares have risen on short-haul routes (where they were below costs under regulation) and fallen on long-haul routes. But the price disparities that have emerged do not simply reflect variations in cost; there has also been an intensification of price discrimination. The airlines' use of deep discount fares to attract marginal customers has been of great benefit to many discretionary travellers — although as Borenstein (1990) noted, this tends to be overlooked because the term "price discrimination" carries such negative connotations. Of more concern are those price disparities which reflect differences in the intensity of competition between different markets (for discussion see Section 8).

Several studies document the decline in average fares, adjusted for inflation, but results are sensitive to the time period selected, and the adjustments (if any) made to remove the effects of those factors unrelated to deregulation. Early assessments, such as the study by Bailey, Graham and Kaplan which focussed on the 1978–1982 period, were complicated by a number of important exogenous events: the 1981–1982 recession, the second oil shock and the U.S. flight controllers' strike. Assessments of the 1985–1987 period reflect the effect of the severe price wars which led to industry-wide losses. Real airline fares fell sharply, but the low rates were not sustainable.

Since 1987, real yield (that is, revenue per passenger-mile adjusted for inflation) has increased somewhat, but in 1988 it was still almost 22 percent below its level of a decade earlier. In itself, however, this does not substantiate the benefits of deregulation. Studies, such as that by the Secretary's Task Force on Competition in the U.S. Airline Industry (1990), which emphasize the decline in inflation-adjusted fares are open to the sorts of arguments illustrated in Table 2. This comes from a study by Paul Dempsey (1990) which attempted to document the failure of deregulation. What the table shows is that real yields were declining prior to deregulation and, indeed, when adjusted for changes in real fuel prices, real yields declined more rapidly in the decade before, rather than in the decade after, deregulation.

These results are significantly influenced by the choice of 1978 as the starting year of deregulation. The CAB had begun to allow discount fares and to relax substantially its price controls in early 1977. If, following Kahn (1988b), we choose 1976 as the critical dividing year, the growth rate differences indicated in the last column of Table 2 virtually disappear.

Table 2
Yield and Fuel Price Indices
(1978=100)

	Real yield (revenue per passenger mile)	Real fuel prices	Fuel adjusted real yields
1967	129.2	55.9	143.8
1968	123.5	54.1	138.0
1969	121.7	50.9	137.1
1970	117.3	47.0	133.4
1971	117.7	46.2	134.2
1972	114.3	46.3	130.3
1973	112.3	47.7	127.6
1974	116.2	82.1	120.9
1975	111.0	90.2	113.3
1976	110.1	92.5	111.9
1977	109.3	99.3	109.5
1978	100.0	100.0	100.0
1979	94.2	131.7	88.0
1980	104.9	180.1	90.0
1981	106.9	189.8	90.3
1982	95.9	168.6	84.0
1983	91.9	148.0	83.3
1984	91.8	135.7	84.9
1985	85.4	124.0	80.7
1986	77.5	140.7	79.1
1987	76.5	81.6	78.4
1988	78.4	74.9	81.4
Growth rates:			
1967–77	-1.7	5.9	-2.7
1978–88	-2.4	-2.8	-2.0

Source: Dempsey (1990)

Moreover, fuel-adjusted real yields represent a crude and misleading attempt to adjust for the effects of inflation. A methodologically more acceptable approach was adopted by Morrison and Winston (1986, Chapter 3), who constructed a fare deflator by estimating the relationship between fares, input prices and output characteristics. The deflator allowed the authors to predict what 1983 (their post-deregulation benchmark) fares would have been if flight characteristics, fuel prices and wages were the same as in 1977 (their pre-deregulation reference year).

However, Morrison and Winston's methodology is, itself, open to criticism. The treatment of wages as an exogenous variable omits what some would claim to be a main impact of U.S. deregulation: the elimination of the rents that had accrued to airline employees under regulation. On the other hand, Morrison and Winston may have attributed too much of the decline in air fares to deregulation by failing to adjust for the cost reduction attributable to the introduction of new, more efficient aircraft. The latter has been an important factor underlying the long-term decline in real yields as indicated in Table 1. Sawers (1987) showed that the introduction of new aircraft did not significantly lower costs between 1978 and 1984, roughly the period examined by Morrison and Winston (see Table 1).

Despite the difficulty of estimating the precise impact of deregulation on prices, there is agreement that expectations created in earlier years by the disparity in prices between regulated interstate and unregulated intrastate carriers have, in general terms, been satisfied.[12] The gains in efficiency following deregulation have largely benefited consumers rather than investors; after deregulation, the rate of return on investment in the airline industry has remained less than half the average rate for all U.S. industries. In addition, consumers have benefited from an elimination of the biases toward higher fares that were incorporated in the CAB pricing formula. Updated calculations of the Standard Industry Fare Level (SIFL) formula undertaken by the Department of Transportation indicate that, as of the third quarter of 1989, CAB regulation (applied to the cost levels of deregulated carriers) would have resulted in prices, on average, 8 percent higher.[13]

5. Service Quality

Based on the evidence that price controls caused wasteful service competition, it was expected that deregulation would lower the average quality of service — albeit providing an overall price-service package that was more reflective of consumer preferences. To some extent this has occurred. Responding to market demands, airlines have traded off, for example, wide seating and various in-flight amenities, for lower prices. There has also been a strong growth in the discount fare portion of the market (from 48 percent in 1979 to 91 percent in 1988) and, because of their various restrictions, discount fares represent a somewhat lower quality of service than coach fares. But the overall impact of deregulation on service quality is quite different from expectations, in large part because of the consequences of the hub-and-spoke system.

The hubbing process itself involves an additional connection, and thus increases travel time for some passengers. This is more than offset, however, by the benefits of hub-and-spoke operations: more on-line service (rather than interline service in which passengers must switch both planes and carriers to complete their flight) and more frequent flights to most destinations. The latter has been a particularly important development. The airlines have greatly improved the service available on low-density routes by providing more frequent flights using smaller aircraft between non-hubs and hubs. According to the Secretary's Task Force (1990b), most small points now receive three or more flights a day to a connecting hub that provides access to most large cities throughout the country. Under the previous linear system, small cities received only one or two flights to particular destinations, with these generally including one or more intermediate stops.[14]

Morrison and Winston found that business travellers, who place a high value on finding flights that correspond to their desired departure time, were the main beneficiaries of the increase in flight frequency. Indeed, this was a main source of the almost $6 billion annual gain in consumer welfare identified in their study.

Morrison and Winston's estimates did not capture what many travellers would consider to be two of the most important and disturbing recent developments: the greatly increased congestion and delay at major U.S.

airports. These have been caused by the increased traffic concentration at hubs, but more importantly, by the recent growth in air travel, which is due to deregulation as well as other factors such as the buoyant U.S. economy. Most economists, however, would agree with Kahn (1988a, p. 321) that airport congestion is not a necessary and inherent cost of deregulation; rather, it is due to the failure of government authorities "to expand airport and air traffic control capacity and . . . to price those scarce facilities at their marginal opportunity costs." In a more recent Brookings study, Morrison and Winston (1989) estimated the potential net benefits from combining more economically appropriate landing fees with optimal investments in runway capacity to be at least $11.0 billion annually.

6. Service to Small Communities

The fear that carriers would take advantage of the freedom of entry or exit under deregulation to abandon service on low-density routes has turned out to be unfounded. The findings of the Secretary's Task Force (discussed above) support an earlier General Accounting Office (GAO) study (1985) which found that between October 1977 and October 1984, weekly departures increased by 31.3 percent at small hubs and by 20 percent at non-hubs. Moreover, because of the nature of the hub-and-spoke system, each of these departures offered connections to destinations throughout the country — a major improvement over the access available under the previous system.

In addition to the benefits from adoption of the hub-and-spoke system and the replacement of jets with smaller aircraft that could economically serve these thinner markets, air service to small communities was protected through specific provisions in the *Airline Deregulation Act.* Section 419 of the Act, dubbed the Essential Air Service Program (EAP), guaranteed the provision of service to points that were certified and thereby protected under earlier legislation, as well as to some points that had been decertified over the previous decade. As of February 1, 1986, the CAB, which administers the program, was providing $24.1 million in annual subsidies to preserve service to 105 points. The EAP was a significant improvement over its predecessor, which had inadvertently encouraged the use of inappropriate aircraft and the provision of a lower quality of service.[15] It was also less costly: Meyer and Oster (1984) estimated that the EAP provided savings of around 30 percent, compared to the per passenger subsidy provided under the previous program.

The EAP has not prevented carriers from terminating service to uncertified points. The GAO (1985) reported that 114 communities lost all scheduled air service between October 1978 and October 1984. This was, however, a continuation of a trend that had begun well before deregulation. Morrison and Winston (1986) attempted to isolate the impact of deregulation from other factors, for example, GDP, fuel prices and interest rates, that influence the number of points in the scheduled air service network. Their finding, that economic conditions rather than deregulation, caused the observed loss in service to small communities, reinforces the conclusion that deregulation has, on balance, had a highly favourable impact on passengers in low-density markets.

7. Air Safety

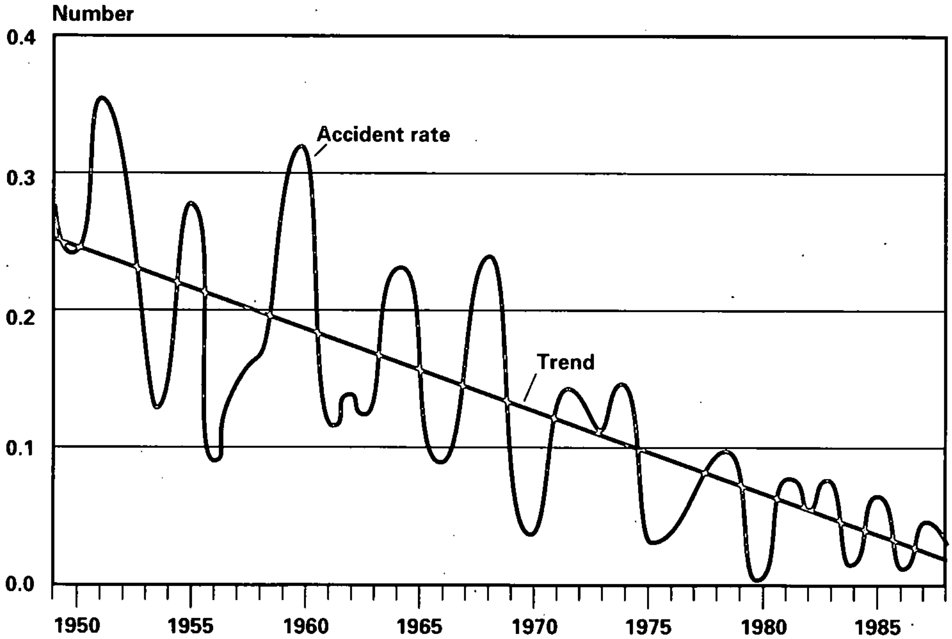
In a perfect market, carriers will treat safety in the same fashion as they treat other aspects of service quality and will make an optimal investment reflecting travellers' willingness to pay for increased safety precautions. Morrison and Winston (1989) showed that travellers attach significant importance to a carrier's safety record. But while the market encourages carriers to operate safely, the incentive structure will not be optimal because of the inadequacy of the information available to consumers. The costs of this market failure could conceivably increase under deregulation where carriers are under strong pressure to gain a competitive advantage.

Empirically, it is very difficult to determine if this has been the case. Airline accident rates — that is, fatal accidents per 100,000 departures — have declined secularly in the U.S. due to various technological developments and training improvements (see Figure 4). To determine if the decline in the recent period would have been greater in the absence of deregulation, it is necessary to control for all other possible influences (besides regulation) on accidents, some of which are very difficult to identify and measure.

Studies that have examined the causes of airline accidents do not suggest that deregulation has been a significant factor. Oster and Zorn (1987), for example, found that accident rates based on equipment failure were lower in the post-deregulation period, indicating there has been no slackening in maintenance practices. They also found that the growth of the hub-and-spoke system has not led to an increase in accidents related to air traffic control. Kanafani and Keeler (1987) found that concerns relating to the safety practices of new entrants were unjustified; the overall safety record of new entrants was no worse than that of established carriers.

Critics of deregulation maintain that accident data do not reflect the true situation. Dempsey (1990) pointed to a number of potentially troubling indicators including the increase in the age of the fleet; the decline in maintenance expenses as a proportion of total operating expenses; and the decrease in the average level of experience of pilots. He referred to a survey of airline pilots in which 97 percent were reported to concur with the view that deregulation has adversely affected safety.

Figure 4
Fatal Accidents per 100,000 Departures, U.S. Certificated Air Carriers 1949–1988



Source: Morrison and Winston (1989)

While most economists are sympathetic to the need for further research and improved understanding in this area, no economist would regard safety concerns as an indictment of price and entry deregulation. The *Airline Deregulation Act of 1978* did not affect the authority of the Federal Aviation Administration (FAA) over airline safety. To the extent the evidence on airline safety raises questions, these pertain not to the merits of deregulation, but to the adequacy of administration and enforcement by the FAA. Even researchers, such as Morrison and Winston (1989), who believe that deregulation has not reduced long-run safety, have suggested that there is considerable scope to improve institutional arrangements so that the government is better able to manage and promote air safety.

8. Industry Structure and Competition

The long-run consequences of deregulation depend crucially on the existence of effective competition. It is competition which generates the pressure to increase efficiency and to pass the resulting gains on to consumers. The relevant issues in this area can usefully be discussed by addressing three basic questions:

- What has occurred?
- What are the implications of the structural changes that have occurred?
- To what extent do these changes reflect underlying production technology, as opposed to the influence of government policy?

WHAT HAS OCCURRED?

One perspective on what has occurred comes from an examination of recent mergers and acquisitions which have greatly increased the degree of industry concentration. Between 1978 to 1984, the industry expanded but then went through a period of consolidation (see Figure 5) that, by 1989, resulted in 92 percent of the industry's revenue passenger-miles being concentrated in the eight largest airlines.

Organization and route restructuring has also given rise to increased concentration at large- and medium-sized hub airports.[16] Figure 6 shows the increase in concentration at many of the most important U.S. airports between 1979 and 1988. At six large hubs, a single carrier accounts for 75 percent or more of all enplanements. This concentration is partly a logical consequence of the pressures within the hub-and-spoke system for carriers to establish a high degree of control over traffic to and from their connecting hubs.

Figure 5

MAJOR AIR CARRIER MERGERS, ACQUISITIONS, PURCHASES AND CONSOLIDATIONS SINCE PROMULGATION OF THE AIRLINE DEREGULATION ACT OF 1978

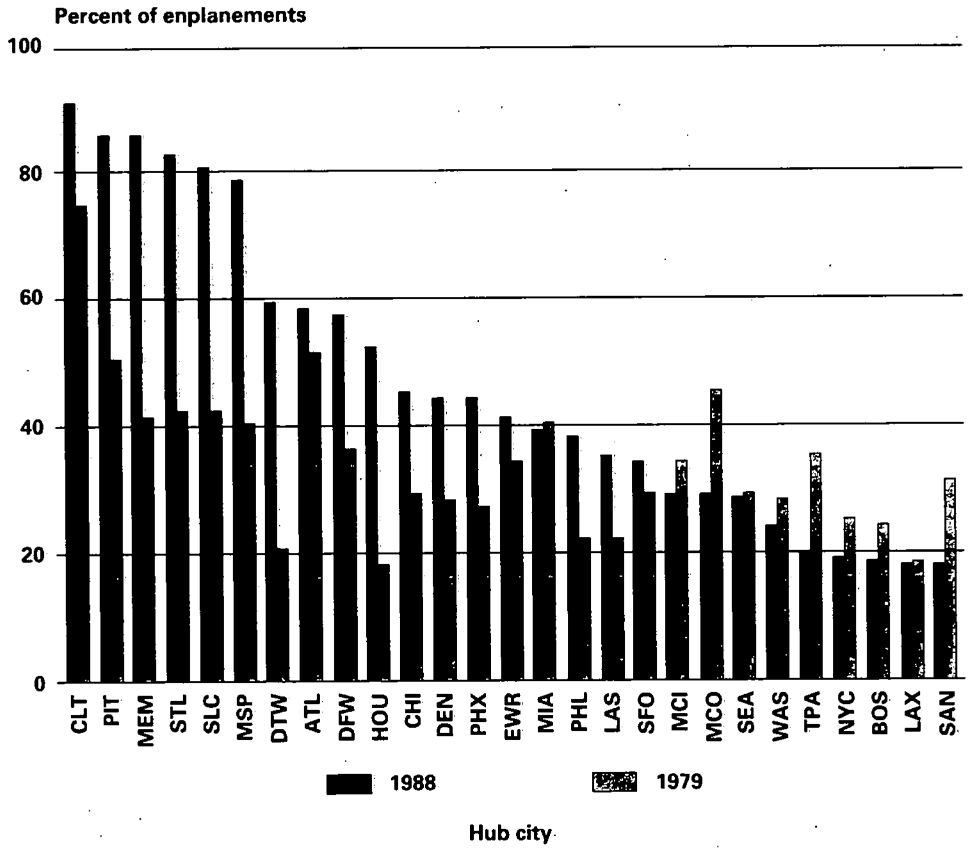
			Market Share		
			1987	1988	1989
American		AMERICAN	13.8	15.2	16.6
Air Cal					
United		UNITED	16.9	16.4	16.2
Pan Am (transpacific routes)					
Texas International		TEXAS AIR	19.0	19.3	15.9
Continental					
New York Air					
Frontier	People Express				
Britt					
PBA					
Braniff (Latin America)	Eastern				
Rocky Mountain					
Delta		DELTA	12.2	12.0	13.3
Western					
Northwest		NORTHWEST	10.3	8.9	9.6
North Central	Republic				
Southern					
Hughes Airwest					
TWA		TWA	8.2	7.4	7.2
Ozark					
USAIR		USAIR	7.1	7.2	7.2
PSA					
Empire	Piedmont				
Henson					
Pan Am		PAN AM	6.3	7.1	5.9
National					
Ransome					

Source: **Dempsey (1990)**

Note: **Since 1989, the industry has become more concentrated. Among the more important recent developments: Braniff and Eastern have closed down, and Pan Am and Continental are operating under the protection of U.S. bankruptcy laws.**

Figure 6

Dominant Carrier Share at Large Hubs Based on Enplanements and 1989 FAA Hub Classifications



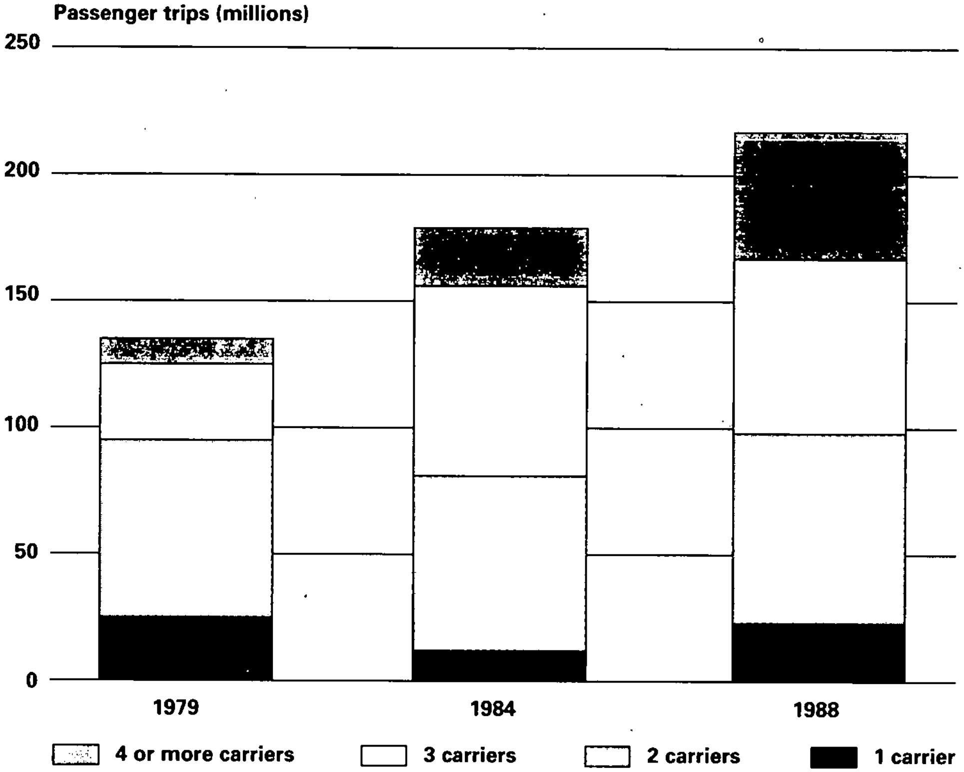
Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990b)

A different and more encouraging perspective emerges from an examination of individual routes which are generally considered to be more relevant indicators of what is happening. Here, competition has increased. The situation, as portrayed in Figure 7 and described by the Secretary's Task Force (1990a, p. 4) is as follows:

> Whether measured by the volume of passenger trips (as shown), revenue passenger miles, or number of travel markets, competition in city-pair markets grew between 1979 and 1988. In 1988, more than 55 percent of the passengers travelled in city-pair markets served by three or more air carriers, up from 28 percent in 1979.

Figure 7

PASSENGER TRIPS BY COMPETITIVE STATUS



Source: Secretary's Task Force on Competition in the U.S. Domestic Airline Industry (1990a)

Of the top 1,000 passenger markets in the U.S., 72 percent were less concentrated in 1988 than in 1979. This increase in route competition can be reconciled with the increased degree of industry concentration at the national level by the fact that the major carriers greatly extended their networks in the decade after deregulation.

In other words, while there are now fewer carriers, each is a participant in a much larger number of markets. The ten major carriers and their regional partners served 1,361 points in 1988, as compared to only 531 points in 1979.[17]

The main exception to this positive trend in the structure of city-pair markets is the reduced degree of competition on short-haul trips — (under 1,000 miles) — to or from a concentrated hub. Non-hubbing carriers have stopped competing in these markets, so competition is essentially limited to carriers that hub at either end point. But it is important to distinguish between "local" traffic moving between a spoke city and a hub, and "flow" traffic for which the hub is likely to represent only one of a number of possible connecting points to the ultimate destination.[18] Concentration is an issue only with respect to local traffic, while most movement from smaller centres to hubs is with respect to flow traffic.

WHAT ARE THE IMPLICATIONS OF THESE STRUCTURAL CHANGES?

Several studies have investigated the relationship between market structure and airline fares. The results indicate that fares fall as the number of competitors on a route increase, and that an additional carrier has greater impact on a market where there are fewer existing competitors. For example, Morrison and Winston (1989) found that losing a competitor on a route raised fares by 2 to 32 percent, with the high end of the range represented by situations in which there was only one carrier left serving a market. At the mean distance in their sample of 112 routes, the round-trip fare increased by $89 when the number of competitors went from two to one; the increase was only $6 when two or more carriers remained after one carrier had exited. These results are similar to the findings of Hurdle et al. (1987).

The effect of market concentration on fares tends to be greater where entry barriers exist. One potential barrier is the existence of a hub at the point of origin or destination on a route. Levine (1987, p. 412) elaborated on this:

> At a "strong" hub — one where one or two airlines have large connecting complexes which account for a large proportion of departures — any new entry that does occur ordinarily seems to be limited to service to and from other airlines' strong hubs. The new entrants do not appear to compete on the hubbing airline's spoke segments to other non-hub cities. Infrequently, an attempt will be made to establish a competing hub structure at another airline's strong hub or a struggle will develop among several airlines attempting to establish a hub where none existed before. But the contests that occur are usually treated by contestants and observers alike as battles for the survival of only one or two carriers, who are expected to earn rents at the hub once the smoke clears and the dead and wounded are carted away.

Hurdle et al. (1989), Morrison and Winston (1989) and Borenstein (1989) all found that the effect of a merger which reduces the number of airlines on a route is more serious when the origin or destination is a hub.[19] In Morrison and Winston's study, loss of a competitor increased fares as high as 55 percent when hub effects were taken into account. Borenstein (1989) found that a carrier that had a dominant position at an airport charged higher prices than it did elsewhere, although per passenger costs tended to be lower. He also observed that the dominant carriers were able to charge higher prices than their competitors; in other words there was no "umbrella effect" from which carriers with smaller operations may benefit.

The evidence on pricing indicates that U.S. airline markets are not perfectly contestable.[20] Under the contestability hypothesis, the threat of entry should suffice to keep fares at competitive levels; high levels of market concentration should not result in higher prices. Contestability depends on ease of entry and exit, but there are a number of impediments to entry into U.S. airline markets. These include carrier reputation (including its safety record), hub-and-spoke route systems, computer reservation systems, frequent flyer programs, travel agent commission override programs, and the limited availability of airport slots and gates. Many of these factors are mutually reinforcing. For example, at a major hub, the airline that controls the converging network and can provide single-line service to a wide range of destinations, has a strong advantage.[21] This is reinforced by frequent flyer programs, which create strong brand loyalty among passengers, and by travel agent commission override programs, which increase the commissions payable to travel agents who sell a certain proportion of their tickets on a particular airline.

While airline markets do not satisfy the conditions for perfect contestability, this does not mean that the competitive threat posed by potential entrants is inconsequential. The most detailed examination of this issue, by Hurdle et al. (1989), indicates that potential (as distinct from actual) competition is a significant factor in *some* markets. The relevant city pairs are short-distance routes (that is, under 800 miles), where there are no barriers to entry due to low-density traffic or because the city pair is primarily providing feed to a connecting hub which is part of a well developed hub-and-spoke system. Indeed, markets in which potential entry is deterred because of economies of scale (arising from low-density traffic)[22] or economies of scope (associated with a hub-and-spoke system) account for a small proportion of domestic passenger traffic. The Secretary's Task Force (1990) corroborated the results of other studies with respect to the higher fares on short-haul trips (defined, here, as under 1,000 miles) originating or terminating at the nation's most concentrated hubs,[23] but it found that the relevant markets represented just over 4 percent of domestic revenue passenger-miles.

Still, the number of passengers affected by the combination of high concentration and high barriers to entry is not insignificant. And there is concern that continuing airline consolidation and route restructuring could result in substantial growth in the number of city-pair markets where competition is weak or absent. These concerns must be set against the favourable general finding that competition on most city-pair markets has increased since deregulation.

In addition, some consideration should be given to the threat that the small number of major U.S. airlines will engage in collusive behaviour. Collusion on a nation-wide basis could only be successful if there are substantial barriers to new entry in all city-pair markets, or if entry is successful only on a large scale.[24] The U.S. government is currently investigating possible collusion among the airlines through the use of the computer reservation system. Anecdotal evidence suggests that airlines use the computer system to signal intended price increases which can then be abandoned if other carriers do not go along.

STRUCTURAL CHANGE AND GOVERNMENT POLICY

There is widespread agreement that the benefits of U.S. airline deregulation have been reduced as a result of inadequacies in various other (that is, non-regulatory) aspects of public policy. It is very difficult, however, to determine to what extent more appropriate policies and better management of the transition to deregulation could have improved the competitive environment.

Contrary to earlier expectations, size does confer an advantage on carriers. While the strength of the major airlines partly reflects sunk costs and long-term contracts which provide them with a high degree of control over strategic airport facilities, it is also a result of their ability to operate an elaborate network and maintain high flight frequencies. The latter aspects can be partly attributed to certain benefits of scale.[25] Larger firms are better able to operate an efficient hub-and-spoke network. They can more easily finance the lumpy investment associated with the use of price-optimizing technologies. They are in a better position to induce brand loyalty through the use of frequent flyer programs, and to establish incentive schemes to influence travel agent recommendations. And only larger carriers are able to take advantage of the benefits of owning computer reservation systems. Bailey and Williams (1988) have noted that American's Sabre system, and United's Apollo dominate the travel agencies in the hubs where these carriers operate.[26] Despite efforts to eliminate the "display bias" in the way flights are listed, travel agents still tend to favour those carriers who own the reservation system.[27]

The public policy problem arises from the fact that these factors are not simply sources of market power; they also convey some real economic benefits. Earlier, we discussed the economies of scope associated with a hub-and-spoke routing system. Levine (1987) also credited the system with providing savings in consumer search costs. In particular, large network operations reduce the costs of acquiring reputation and service information "which is subtle and difficult for individuals to accumulate and assimilate." Bailey and Williams (1988, p. 189) suggested that the ownership of a computer reservation system can also be a source of efficiency gains: "Carriers with these systems can fill more seats by more efficient matching of customer preferences with available discount fares, and they can more effectively schedule capacity, resulting in higher load factors." It is similarly argued that frequent flyer programs enhance airline efficiency. By appropriately funnelling frequent flyer awards, the airlines are able to fill seats that would otherwise be empty.

One area where the need for reform is apparent is airport policy. Governments can much more efficiently allocate scarce facilities at the major commercial airports. Currently capacity at the busiest airports is rationed through the use of take-off and landing rights or "slots."[28] Although slots can be purchased, it may be very difficult for a new carrier to acquire the necessary rights to enter a slot-constrained airport. Entry is also complicated by the shortage of ground facilities such as gates and check-in areas at large hub airports. Part of the solution, as Morrison and Winston (1989), Kahn (1988a) and others have pointed out, is to replace current landing fees with those that reflect marginal opportunity costs and thus fully take account of an aircraft's contribution to congestion during peak periods.

Airport authorities could also attempt to buy back under-used gates and to extricate themselves from arrangements with carriers which limit their ability to expand groundside facilities.[29]

As well, substantial benefits may be realized through an efficient expansion in airport capacity. This would require local authorities to increase capacity up to the point where the extra costs are equivalent to the benefits in terms of reduced congestion and delay. However, as the Congressional Budget Office (1988) has noted, many large airports do not have the space for new runways, and noise and land-use concerns are a major obstacle to airport expansion.

Airline marketing practices have been the subject of a number of inquiries. Frequent flyer programs, computer reservation systems and travel agent override commission programs impede competition. They can also cause serious principal-agent problems; frequent flyer programs create a possible divergence in interests between business travellers and their employers, while commission override programs, which affect the self-interest of travel agents, may adversely affect travel agents' clients. Since these marketing arrangements appear, at the same time, to have some beneficial consequences, the appropriate policy response is far from certain. Recently, ownership of three of the five computer reservation systems in the industry has become more diversified. This lessens the concern that computer systems will be used to thwart competition.

To those who believe that the high level of airline concentration reflects the failure of government policy, one of the most significant aspects of this failure was the Department of Transportation's inability to control the merger wave that occurred between 1984 and 1987. The Congressional Budget Office (1988) found that there is basis to criticize the decisions of the Department of Transportation (which mistakenly assumed airline markets were perfectly contestable), but these decisions "did not play a large role in the consolidation of the industry." On the other hand, Morrison and Winston analyzed six mergers approved over 1986–1987, and found that all resulted in decreased competition and significantly higher fares. It is only when they took into account the non-price benefits of mergers, and particularly the effect of mergers in raising the value of travellers' frequent flyer benefits, that some of the mergers seemed, on balance, to be positive. As Morrison and Winston themselves acknowledged, however, the evaluation of frequent flyer benefits is problematic. The authors interpret their results not so much as a criticism of previous merger decisions, but as an indication of the limitations of merger policy in terms of its ability to take account of the long-run opportunity cost of network consolidation.

Conclusion

There is general agreement that airline deregulation in the U.S. has been a success. The changes in route patterns and organizational structures which have occurred in response to market forces have resulted in major improvements in efficiency. Wasteful service competition has been eliminated, and the price-quality offerings of U.S. carriers now more closely reflect costs and traveller preferences. Some questionable distributive aspects of the previous regime have also been eliminated; in particular, airline employees no longer enjoy rents in the form of high wages and inefficient work rules. Deregulation has not reduced the service available to small communities, nor has it disrupted the long-term trend toward improved airline safety.

These positive developments have been only partly offset by some unfavourable structural trends. While there has been an increase in the number of competitors in most city-pair markets since deregulation, the airline industry as a whole has become quite concentrated. Moreover, the expectation, that deregulated airline markets would be perfectly contestable and that the threat of entry would thereby limit the opportunity for the exercise of market power, has proven incorrect. Airline fares are higher where routes are highly concentrated and/or where one or two carriers are in a dominant position at an airport. Along with concerns pertaining to the power of carriers in particular markets, the high degree of industrial concentration and the current use of computer reservation systems are seen, by some, to pose a threat of collusion at a national level.

The competitive advantage of large carriers is partly a natural consequence of their efforts to develop an efficient network and to build a reputation as reliable providers of air service. It is also attributable to airline marketing practices involving frequent flyer programs, computer reservation systems and travel agent commission override programs. And it is a result of entry barriers arising from the lack of slots and ground facilities at some airports. To some extent, these reflect certain inadequacies in U.S. policy. As Elizabeth

Bailey, an ex-member of the CAB, has herself acknowledged (1989, p. 113), "the Civil Aeronautics Board [failed] to treat civil aviation policy as a complete system of interrelated elements. [It] did a partial equilibrium analysis, deregulating the rate-and-route authority that [it] had within the CAB, without fully addressing such policy issues as airport capacity and air traffic control."

At the same time, however, some of these issues pose very difficult trade-offs between increased efficiency and reduced competition. For example, a merger may eliminate competition on a route, but traffic density may be such that only one carrier can efficiently serve the market. An acquisition may substantially strengthen the position of a carrier at a given hub, but the threat from this increase in concentration must be balanced against the real efficiency gains from a better developed network which generates increased feed to and from all points. Moreover, the threat to market forces from such developments may be quite subtle. The anti-competitive effect of a merger, for example, may come from the *potential, long-term* impediment to entry from network consolidation.

Similarly, the incremental contribution of particular marketing practices to the market power of large carriers will often be very difficult to isolate. At issue is not simply the administration of competition policy, but the more fundamental question as to whether competition policy can provide a sufficient safeguard against the exercise of market power. While the evidence argues strongly against the reregulation of the airline industry, it does not allow one to entirely dismiss the possibility that some protection may be required for passengers in particular airline markets.

APPENDIX A

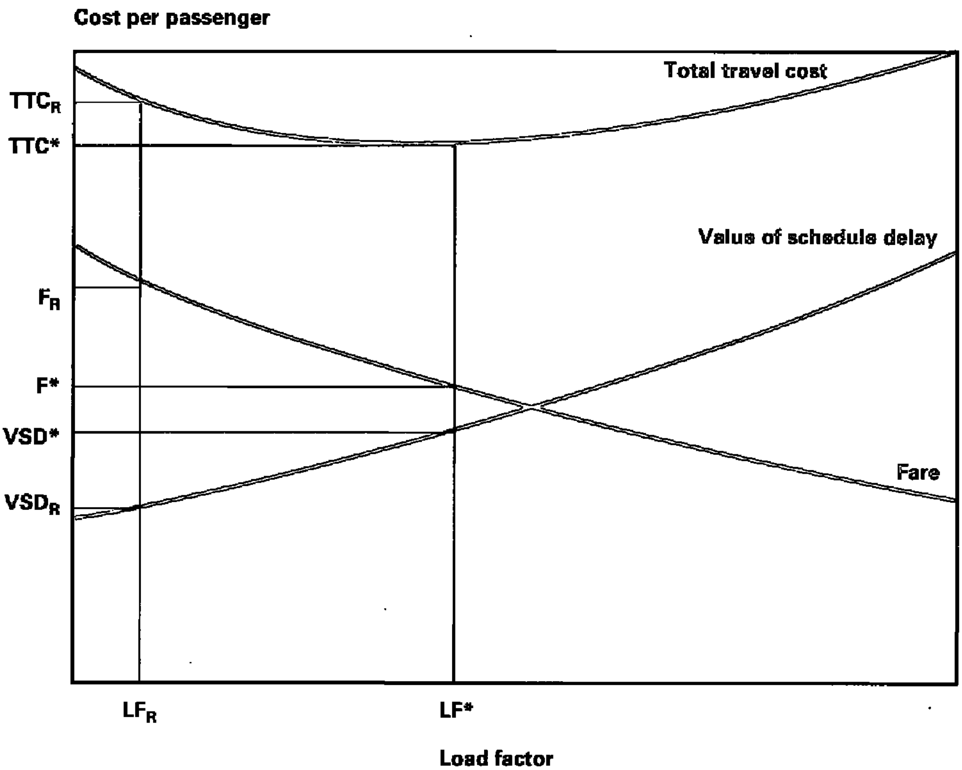
THE OPTIMAL LOAD FACTOR

Following Morrison (1989), the effect of regulation on load factor and quality of air service can be illustrated using Figure A-1. The line labelled "Fare" represents the combinations of fares and load factors that result in normal profits. At higher load factors, lower fares are consistent with normal profits. The line labelled "Value of Schedule Delay" represents the loss to passengers arising from the gap between their actual and their desired departure time. This slopes upward because higher load factors are associated with fewer flights and this, in turn, is likely to result in increased passenger delay. "Total travel cost" represents the cost to passengers, taking account of the impact of increasing load factor on both fares and passenger delay.

In a competitive market, carriers will be under pressure to minimize the total cost of travel, consistent with their own need to earn normal profits. This will result in the establishment of fare F* at load factor LF*. Under CAB regulation, fares were set at a level such as F_R, and carriers competed by increasing flight frequency until all excess profits were eliminated. In the diagram this occurs when the load factor has been reduced to LF_R. Total cost to the passenger, TTC_R, is above the minimal point on the travel cost curve, and the traveller is presented with a price-quality package that is much less desirable than that which would result under competition.

This distortion was particularly marked in long-distance markets. In the figure, the fare curve for a long-distance flight would be above that for a short-haul flight, and the "Total Travel Cost" curve would move up and to the right. As a result, minimum total travel cost occurs at a higher load factor in long-distance markets. But, because CAB fares were set at a level that initially provided substantial excess profits in long-distance markets, the ensuing competition led to abnormally high flight frequency and extraordinarily low load factors on these routes. Hence, contrary to the optimal relationship, CAB regulation resulted in load factors that declined with distance.

Figure A-1
Optimal Load Factor Determination



Source: Morrison (1989)

The figure can also be extended to take account of the differing preferences of business and vacation travellers. Since business travellers tend to attach a higher cost to schedule delay, the minimum point on their "Total Travel Cost Curve" will be to the left of that for vacation travellers. Business travellers would be best served through a lower-than-average load factor, and a correspondingly higher fare. The reverse is required to meet the needs of vacation travellers. This distinction was not recognized under CAB regulation, where fares were based mainly on distance. In consequence, regulated fares were above the optimal level on vacation routes, as well as on long-distance routes.

ENDNOTES

1. Federal regulation did not apply to *intra-state* carriers. Evidence of the superior performance of unregulated intra-state carriers in California and Texas influenced the public policy debate leading up to deregulation.

2. The liberalization process accelerated after Alfred Kahn became Chairman of the CAB in 1977. "Supersaver" fares were introduced in the spring of 1977. In the fall of 1978 a "zone of reasonableness" was established which allowed carriers to set fares as much as 10 percent above or 50 percent below a CAB standard fare. In May 1980, the upper boundary of this zone was extended and the lower boundary entirely removed.

3. This Essential Air Service Program is still in operation although it was initially scheduled to expire in 1988.

4. Average load factor had dropped below 50 percent in the early 1970s as the opportunity for higher profits resulting from the development of increasingly efficient jet aircraft led to increased service competition; flight frequency increased and load factor declined until excess profits were eliminated. After 1972, the CAB tried to take account of this effect in setting rates which resulted in some improvement in the load factors of the major U.S. airlines just before deregulation.

5. In most studies, the counterfactual case on how the industry would perform were it still regulated is developed from an extrapolation of trends observed in the period before deregulation.

6. For U.S. carriers, the growth of total factor productivity accelerated from a 2.9% average rate over the period 1970 to 1975 to a 3.1% rate over the period 1975 to 1983. Had the performance of U.S. carriers changed in the same way as non-U.S. carriers, total factor productivity would have grown by only 1.8% per year over 1975 to 1983. The results from accumulating these differences over the eight years suggest that, in the absence of deregulation, 1983 airline costs would have been 10% (or $4 billion) higher. Caves et al. dated deregulation from the relaxation of CAB price control in the mid-1970s. However, they found that using 1978, the year the *Airline Deregulation Act* was passed, as their transitional year did not change their general conclusions regarding the positive effect of deregulation on the growth of total factor productivity.

7. Sawers pointed out that this improvement was all the more impressive in the context of the much slower growth in air traffic (due in part to the oil shock and recession) after 1978. In this environment, as opposed to that prevailing between 1970 and 1978, efficiency gains were more difficult to achieve.

8. For example, it has been argued that the industry was not in equilibrium during the period observed by some of the studies. With the exception of Morrison and Winston, the studies have not attempted to adjust for differences in the mix and quality of output before and after deregulation. The studies' assumptions about what would have occurred in the absence of deregulation are also open to challenge.

9. New entrants created strong pressures for incumbents to achieve less restrictive work rules. Graham, Kaplan and Sibley cited a number of pertinent examples. Pilots at Southwest, a former intra-state carrier, averaged more than 50 percent more flying hours per month than pilots of the formerly regulated carriers. Southwest's cost of operating a Boeing-737 was initially much less than that of United, a formerly regulated carrier, in large part because United's labour contract required it to use a three-person cockpit crew to fly a two-person

aircraft. United subsequently renegotiated its labour contract so it could use a two-person crew. In addition, Southwest and other new entrants had more flexibility in allocating labour; for example, baggage handlers could be called up to load in-flight meals.

10. The load factor curves in Figure 2 were derived from equations estimating the relationship between load factor, distance, concentration and traffic volume. While the 1976 results predate deregulation, the CAB had, by this time, recognized the effect of its rate controls on load factor, and it had adjusted its pricing formula to reduce airlines' incentives to engage in non-price competition.

11. Airlines have met the need of time-sensitive passengers by setting aside a number of seats on each flight for full-fare passengers. They have also introduced peak-load pricing which reserves seats on peak flights for those who place a high cost on delay time.

12. A number of studies have highlighted the lower fares charged by unregulated intra-state carriers. For example, Jordan (1970) found that, in California, the fare level of intra-state carriers was 32 to 47 percent lower than the comparable rates charged by CAB regulated trunk carriers. From a model based on the costs of California carriers, Keeler (1972) estimated that the disparity between regulated and unregulated fares was between 20 and 95 percent, with the gap increasing with distance.

13. This estimate does not take account of the contribution of deregulation to reducing carrier operating costs because the SIFL formula is partly based on costs. Fares were set by a fixed charge plus a per-mile addition, calculated so as to allow the industry to earn a 12 percent rate of return on its base, assuming a 55 percent load factor.

14. The ability of a hub-and-spoke system to provide increased service to small centres can b e seen by looking at a simple network of five cities. As Morrison (1989) showed, using a linear system, 20 non-stop flights per day are required to provide one non-stop flight between each city. With a hub-and-spoke arrangement, those 20 flights will provide two connecting flights per day between each of the cities.

15. The previous subsidy program was under the authority of section 406(a) of the *Civil Aeronautics Act of 1938*. The individual carrier subsidies introduced by the CAB in 1954 were replaced by a "subsidy class rate" system in 1961 which remained in effect until the termination of the 406 program in 1983. Meyer and Oster (1984) noted that the "class rate" formula was generous in identifying the portion of the carrier's investment base relating to subsidized operations. (This portion is the basis for calculating the amount of subsidies required to provide a reasonable return.) The effect was to encourage carriers to invest in jet aircraft that were more suitable to long-haul, high-density markets and, hence, to hasten the abandonment of small communities by local service carriers. Also, since subsidy rates were based on the volume, and not the quality, of service, there was an incentive for carriers to provide a lower quality of service to small communities; carriers could maximize their revenue by flying subsidized routes during off-peak periods (when they could earn less revenue on their nonsubsidized routes), and by flying a multi-stop route among subsidy-eligible points.

16. The Federal Aviation Administration defines large hubs as those that account for 1% or more of all domestic enplaned passengers. Medium hubs account for 0.25 to 0.999%; small hubs, 0.05 to 0.249%; and non-hubs, less than 0.05%.

17. These data come from the Secretary's Task Force on Competition in the U.S. Airline Industry, (1990d, I).

18. The following example, from the Secretary's Task Force (1990b, p. 16) illustrates this point:

> A small hub like Akron, Ohio, is a monopoly spoke that contributed to Piedmont's concentration at Dayton. It is also a monopoly spoke for six other connecting hubs, contributing to the dominance of one carrier at each of those hubs. Nevertheless, Akron has available hundreds of connections to many points as a result of being a spoke city to each of these hubs and, obviously, when Akron passengers move beyond one of the connecting hubs, various competitive alternatives are typically available. Most Akron passengers, in fact, do move beyond these connecting hubs. This is another example of how the hubbing process simultaneously increases concentration at a connecting hub and creates competitive alternatives for passengers moving beyond the connecting hubs.

19. Dresner and Windle (1990) suggested that the effect of hub concentration on airline fares could be explained by the gains to the consumer in the form of frequent flyer benefits. However, their focus is on the difference in fares charged by airline A, flying from X to Y as compared to flying from Y to X. They found that higher fares would be charged on X to Y if the airline was the dominant carrier at airport X but that the fare disparity could be accounted for by the value of frequent flyer benefits. Morrison and Winston (1989) and Borenstein (1989) were concerned with the effect of hub concentration at either the point of origin or destination on the fares charged by airline A relative to its smaller competitors. Their finding that, in this context, hub effects are significant is not challenged by the results of the Dresner and Windle study. The latter, though, does highlight the difficulty of controlling for all the factors besides market share that may explain higher airline yields.

20. The airline industry, in fact, violates all the theoretical conditions for perfect contestability. Contrary to the contestability hypothesis, new entrants to airline markets may be at a significant cost disadvantage because of the existence of economies of scale and scope. For contestability, incumbent firms must be unable to respond quickly enough with a price reduction to foreclose a profit opportunity to entrants; but route fares in the airline industry can be adjusted almost instantly. Contestability also depends on the absence of sunk costs, but a carrier's service and safety reputation is an important aspect of air service, and an investment that is unrecoverable.

21. This advantage is in terms of both higher service quality and lower production costs. The former relates to the carrier's ability to provide more frequent service to a wide variety of destinations. Lower costs are a consequence of the carrier's ability to generate the feed required to run larger planes and to achieve high load factors.

22. Caves et al. (1984) distinguished between economies of firm size and economies of traffic density. The latter focusses on the behaviour of costs, as output is expanded, by increasing service within a given network (number of points served remains constant). The former focusses on costs, as output is expanded, by adding points to a network with a given traffic density. They found there were no significant economies of firm size (unit costs are roughly constant within a broad range of network size), but there were sizeable economies of traffic density (unit costs decline as service within a given network increases). Within this framework, the advantages of larger airlines is explained by their greater ability to increase demand and thereby take advantage of economies of density. If output was measured differently, however, so as to capture differences in the characteristics of the output produced by small and large carriers, it is possible that the marketing advantages of large firms would be reflected in increasing returns to firm size.

23. After adjusting for differences in trip distance and market size, the Secretary's Task Force (1990a) found that fares on trips originating or terminating at the nation's most concentrated hubs averaged 18.7 percent higher. The hubs were: Minneapolis, Cincinnati, Dayton, St. Louis, Pittsburgh, Memphis, Salt Lake City and Charlotte.

24. *De novo* entrants are in a much more difficult position than established carriers who already have built a reputation, and are linked into a frequent flyer program and a computer reservation system. New carriers must deal with these hurdles, in addition to those which confront all aspiring entrants (new carriers and incumbents) to those city-pair markets characterized by a concentrated hub or a scarcity of airport slots and gates.

25. As noted in note 22, these advantages were not reflected in some of the earlier studies of economies of scale which ignored the effect of size on airlines' ability to influence the demand for their services. These studies are the source for the traditional view that the airline industry is not characterized by economies of large-scale production.

26. Besides Sabre and Apollo, the other computer reservation systems are: System One, Pars and Datas 11. In 1988, the five systems generated $1.2 billion in revenue, approximately 70% of which was from booking fees paid by carriers, with the remainder from fees paid by subscribers. Sabre accounted for 38.8% of industry revenue, Apollo 27.6%, System One 16.6%, Pars 11.1% and Datas 11 5.9%.

27. The reasons for this, according to the Secretary's Task Force (1990c, p. 5) are as follows:

> Some CRSs may exhibit superior functionality for booking on the vendor airline, which facilitates more bookings on that airline. In addition, travel agents often have more confidence in the accuracy and timeliness of the seat availability information on the vendor airline and the reliability of having their bookings recorded, which again results in additional bookings. Some agents also maintain that it is easier and faster to make and change reservations, issue tickets and boarding passes, select seats, clear wait lists, and perform several other routine tasks for the vendor airline's flights. Moreover, the ongoing business relationship between vendors and subscribing agents — including the ability to receive preferential treatment — is a major factor in explaining agents' loyalty to their vendor airline.

28. There are four airports in which the Federal Aviation Administration has imposed restrictions on take-off and landing slots: National Airport in Washington, La Guardia and J.F.K. airports in New York, and O'Hare Airport in Chicago.

29. Where carriers provide the financial underpinning for the sale of airport bonds by agreeing to cover any revenue shortfall, these tenant carriers will have a large voice in airport operating and investment decisions. Bailey and Williams (1988) observed that some airports were responding to the resulting problems by attempting to obtain financing on the basis of the underlying strength of air traffic markets rather than on the basis of carrier guarantees.

REFERENCES

Bailey, E. 1989. "Comments" on S. A. Morrison and C. Winston, "Enhancing the Performance of the Deregulated Airline Industry." *Brookings Papers: Microeconomics.*

Bailey, E., D. Graham and D. Kaplan. 1985. *Deregulating the Airlines.* Cambridge, Mass.: The MIT Press.

Bailey, E., and J. R. Williams. 1988. "Sources of Economic Rent in the Deregulated Airline Industry." *Journal of Law and Economics* 31 (April).

Borenstein, S. 1989. "Hubs and High Fares: Airport Dominance and Market Power in the U.S. Airline Industry." *Rand Journal of Economics* 20 (Autumn).

______. 1990. "Transportation Deregulation in the United States." Presented at Pacific Economic Cooperation Conference, Kyoto, October 11 and 12.

Caves, D. W., L. Christenson and M. Tretheway. 1982. "Airline Productivity under Deregulation." *Regulation* (November/December).

______. 1984. "Economics of Density versus Economics of Scale: Why Trunk and Local Airline Costs Differ." *Rand Journal of Economics* 15 (Winter).

Caves, D. W., L. Christenson, M. Tretheway and R. J. Windle. 1987. "An Assessment of the Efficiency Effects of U.S. Airline Deregulation via an International Comparison." In *Public Regulation: New Perspectives on Institutions and Policies.* Edited by E. E. Bailey. Cambridge, Mass.: The MIT Press.

Congressional Budget Office. 1988. *Policies for the Deregulated Airline Industry.* Washington, D.C.: U.S. Government Printing Office.

Dempsey, P. 1990. *Flying Blind: The Failure of Airline Deregulation.* Washington, D.C.: Economic Policy Institute.

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

Dresner, M. and R. Windle. 1990. "Airport Dominance and Pricing in the U.S. Airline Industry." Paper presented at Transportation Research Forum, Long Beach, Calif., October.

General Accounting Office. 1985. *Deregulation: Increased Competition is Making Airlines More Efficient and Responsive to Consumers.* GAO/RCEO-86-26. Washington, D.C., November 6.

Graham, D. R., D. P. Kaplan and D. S. Sibley. 1983. "Efficiency and Competition in the Airline Industry." *Bell Journal of Economics* 14 (April). (Figure 2 — Copyright © 1983, American Telephone and Telegraph Company, reprinted by permission.)

Hurdle, G., R. Johnson, A. Joskow, G. Werden and M. Williams. 1989. "Concentration, Potential Entry and Performance in the Airline Industry." *The Journal of Industrial Economics* 38 (December).

Jordan, W. A. 1970. Airline Regulation in America: Effects and Imperfections. Baltimore, Md.: Johns Hopkins University Press.

Kahn, A. 1988a. "Surprises of Deregulation." *American Economic Review* 78, no. 2 (May).

______. 1988b. "Airline Deregulation — A Mixed Bag, But a Clear Success Nevertheless." *Transportation Law Journal* 16.

Kanafani, A. and T. E. Keeler. 1987. "New Entrants and Safety: Some Statistical Evidence on the Effects of Airline Deregulation." Presented at Transportation Deregulation and Safety Conference, Northwestern University, Evanston, Illinois, June.

Keeler, T. E. 1972. "Airline Deregulation and Market Performances." *Bell Journal of Economics* 3 (Autumn).

Levine, Michael E. 1987. "Airline Competition in Deregulated Markets: Theory, Firm, Strategy, and Public Policy." *Yale Journal on Regulation* 4 (Spring).

Meyer, J. R. and C. V. Oster. 1984. *Deregulation and the New Airline Entrepreneurs.* Cambridge, Mass.: The MIT Press.

Morrison, S. A. 1989. "U.S. Domestic Aviation." In *The Age of Regulatory Reform*. Edited by K. Button and D. Swann. Oxford, England: Oxford University Press. (Figure A-1 by permission of Oxford University Press.)

Morrison, S. A. and C. Winston. 1989. "Enhancing the Performance of the Deregulated Air Transportation System." *Brookings Papers: Microeconomics.*

______. 1986. *The Economic Effects of Airline Deregulation.* Washington, D.C.: The Brookings Institution.

OECD. 1988. *Deregulation and Airline Competition.* Paris: Organisation for Economic Co-operation and Development.

Oster, C. V. Jr. and C. K. Zorn. 1987. "Airline Deregulation: Is It Still Safe to Fly?" Presented at Transportation Deregulation and Safety Conference, Northwestern University, Evanston, Illinois, June.

Sawers, D. 1987. *Competition in the Air: What Europe Can Learn from the U.S.A.* London, England: The Institute of Economic Affairs.

Secretary's Task Force on Competition in the U.S. Domestic Airline Industry. 1990a. *Executive Summary.* Washington, D.C.: U.S. Department of Transportation.

______. 1990b. *Industry and Route Structure: Executive Summary.* Washington, D.C.: U.S. Department of Transportation.

______. 1990c. *Airline Marketing Practices.* Washington, D.C.: U.S. Department of Transportation.

______. 1990d. *Industry and Route Structure,* I and II. Washington, D.C.: U.S. Department of Transportation.