TRANSPORTATION RESEARCH RECORD

No. 1395

Planning and Administration

Finance, Taxation, Pricing, Economic Analysis, Socioeconomics, Education, and Management

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1993

Transportation Research Record 1395
ISSN 0361-1981
ISBN 0-309-05465-6
Price: $41.00

Subscriber Category
IA planning and administration

TRB Publications Staff
Director of Reports and Editorial Services: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Norman Solomon, Susan E. G. Brown
Graphics Specialist: Terri Wayne
Office Manager: Phyllis D. Barber
Senior Production Assistant: Betty L. Hawkins

Printed in the United States of America

National Research Council. Transportation Research Board.

Sponsorship of Transportation Research Record 1395

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Chairman: Sally Hill Cooper, Virginia Department of Transportation

Management and Administration Section
Chairman: Thomas F. Humphrey, Massachusetts Institute of Technology

Committee on Taxation, Finance and Pricing
Chairman: Jenifer Wishart, World Bank
Secretary: Thomas W. Cooper, Federal Highway Administration
Thomas W. Bradshaw, Jr., Thomas M. Corsi, Ralph C. Erickson, Mark A. Euritt, David J. Forkenbrock, Thomas D. Hopkins, Arthur C. Jacoby, Anthony R. Kane, Colin Ian MacGillivray, Roger D. Mingo, Shiva K. Pant, Elizabeth Pinkston, Clyde E. Pyers, Arlee T. Reno, Steven M. Rock, Suzanne H. Sale, Roger L. Schrantz, Frank J. Smith, Norman H. Wuestefeld

Committee on Management and Productivity
Chairman: Kathleen E. Stein-Hudson, Howard/Stein-Hudson
Secretary: Barbara Martin, Montana Highway Department
Luz B. Allende, Walter E. Bortree, Steven L. Butts, John F. DiRenzo, Henry Droughter, Donald O. Eisele, Judy Guerrero, Barbara Thomas Harder, Jerry A. Hawkins, Laura A. Jibben, Robert C. Johns, P. A. Koushki, George T. Lathrop, Carol LeFevre, Arthur J. Morse, Jane L. Shiatte, Roy Sparrow, William G. Stringfellow, Cindy M. Turvy

Committee on Transportation Education and Training
Chairman: Louis J. Pignataro, New Jersey Institute of Technology
Edward A. Beimborn, Michael S. Bronzini, Katherine L. Davis, Roger L. Dean, Marvin C. Gersten, Edwin W. Hauser, Joyce H. Johnson, Elaine E. Joost, C. Jotin Khisty, Patricia L. Lees, John M. McCullough, Charles T. Morison, Jr., Neville A. Parker, Calvin Roberts, John Ross, Kumares C. Sinha, Louis B. Stephens, Jr., C. Michael Walton, Eugene M. Wilson

Committee on Local Transportation Finance
Chairman: David J. Forkenbrock, University of Iowa
Chris P. Bertram, Donald J. Emerson, Norman S. J. Foster, Genevieve Giuliano, Terry L. Gotts, Sharon M. Greene, Randall Halvorson, Douglass B. Lee, Michael A. Lipsman, Deb Miller, Robert D. Owens, Wayne B. Placide, Marilyn Ragusa, Richard J. Reeder, Suzanne H. Sale, Frank J. Smith, Barbara W. Weiss, Bruce B. Wilson, Norman H. Wuestefeld, Robert J. Zuelsdorf

Transportation Forecasting, Data, and Economics Section
Chairman: Mary Lynn Tischer, Virginia Department of Transportation

Committee on Transportation Economics
Chairman: Ralph C. Erickson, Federal Highway Administration
Gary R. Allen, David A. Aschauer, Kazem Attaran, Fredric S. Berger, Susan J. Binder, Stewart E. Butler, Peter D. Cook, Asif Faiz, Z. Andrew Farkas, Edwin W. Hauser, Loyd R. Henion, R. Ian Kingham, William F. McFarland, Marcelo J. Minc, Bahar B. Norris, Maximus H. Perera, Farideh Ramjerdi, Christian Reynaud, Gabriel J. Roth, Aad Ruhl, Roger L. Schrantz, Guillaume Shearin, Yorgos J. Stephanedes, Wayne K. Talley, C. Michael Walton

Committee on Social and Economic Factors of Transportation
Chairman: Thomas N. Harvey, Cambridge Systematics Inc.
Secretary: Mary R. Kihl, Iowa State University
Susan J. Binder, Jesse L. Buffington, Burton Clifton, John W. Fuller, Gary L. Groat, William A. Hyman, Snehamay Khasnabis, Brenda C. Kragh, Carl-Heinz Mumme, Glen Weisbrod, Katie N. Womack, Norbert Y. Zucker

Kenneth E. Cook, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1992.

Transportation Research Record 1395

Contents

Part 3—Socioeconomic Impacts

Part 4—Management and Productivity

Part 5—Education and Training

Foreword

The papers in this Record relate to the following transportation issues: finance, taxation, and pricing; economic analysis and socioeconomic impacts; management and productivity; and education and training.

The first three papers deal with matters of transportation finance. Topical issues discussed include public-private partnerships, the valuation of time to be used in the pricing of transportation services, and treatment of roads as a public utility, charging a fee for their use in the same way that a fee is charged for the use of other public utilities.

The second group of papers describes a variety of economic analysis techniques, including the use of toll route diversion behavior as a means of measuring drivers' value of time and the derivation of value of time curves for Bangkok. Other papers evaluate road user cost comparisons between signalized intersections and interchanges under rural expressway conditions, welfare maximization with financial constraints for bus systems, and alternative tunnels to handle international double-stacked container trains.

A number of papers deal with economic analysis techniques and consider issues of cost analysis for paved shoulders, cost methodology for maintenance management, costs of network preservation strategies, and vehicle operating cost models. Other papers of economic interest describe a different perspective on highway investment and economic growth, social costs of peak-period road pricing, external factors that will influence transportation logistics in the future, and the estimation of fuel costs and tail pipe emissions. An international paper addresses the economic theory of travel decisions.

The socioeconomic impact papers address the consequences of highway rehabilitation on businesses, the effects of highway bypasses, and travel patterns of extremely poor people.

Several papers discuss performance management, customer satisfaction, fare policy as part of transit strategic management, and response consistency of management questionnaire surveys.

The final paper discusses enhancement of the future pool of civil engineers for transportation.

PART 1

Finance, Taxation, and Pricing

Economic Aspects of Public-Private Partnerships for the Provision of Roadway Services

FRANK P. STAFFORD AND KAN CHEN

Although road pricing is an old and, theoretically, effective approach to traffic congestion relief, it has problems of public acceptability. The basic concept and recent developments in road pricing were reviewed, and a theoretical framework was developed for the broader issue of public-private partnerships for the provision of roadway services. Within this framework the basic concept of road pricing may be implemented in an innovative bundling of private intelligent vehicle-highway systems services with economic incentives for traffic diversion. Future research is suggested for building basic economic models of excludable public goods that would include congestibility. An operational field test is suggested to try out the idea of bundling private services to trucks: public authorities would offer economic incentives to divert trucks from congested routes.

The formation of partnerships between the public and private sectors appears to offer promise in the delivery of a menu of intelligent vehicle-highway systems (IVHS) functions from which users of the roadways choose, toward the improvement of travel efficiency and in the provision of needed financial sources for the IVHS infrastructure (*1*). Key elements underlying the viability of these partnerships are user fees for newly developed services, ranging from in-vehicle delivery of timely and relevant traffic information to nonstop toll collection for access to bridges, special lanes, or use of roads during peak traffic periods.

The idea of any type of fees for the use of free roads or freeways is likely to be met with public skepticism. Yet two motivating factors for a closer examination of options in this area are that (a) existing traffic patterns are clearly ineffective and (b) new highway technologies may allow the public collection of road-use fees in conjunction with private services—services that may be purchased voluntarily from private suppliers. A key illustration of the latter is what has been referred to as the bundling of public and private road services.

The need for some new approach is highlighted by the fact that, during peak traffic times, Los Angeles freeways handle a far smaller volume of cars per hour than in off-peak hours (*2*)! Evidently, free access to roads can and does create a result that is far below what can be regarded as system optimal. Even simple access limits, such as entry ramp metering, could improve overall traffic flow.

F. P. Stafford, Department of Economics, University of Michigan, 312 Lorch Hall, Ann Arbor, Mich. 48104-1220. K. Chen, Department of EECS, University of Michigan, 4112 EECS Building, Ann Arbor, Mich. 48109-2122.

ROAD PRICING: CONCEPT AND RECENT DEVELOPMENTS

Basically, road pricing works through the provision of economic incentive or disincentive to influence drivers' behavior, that is, demand management. The concept is not new because it dates back to the 1920s (*3*), but it has been reassessed and improved at various times (*4,5*). The concept has been attractive to economists who argue that excessive congestion is a phenomenon of inefficient allocation of scarce resources. An efficient way to reduce congestion is thus to introduce a market mechanism to road transport. Without road pricing, increasing highway capacity through road building or automation would simply attract more traffic to the new roads, and the previous level of congestion would return as the system finds a new equilibrium. In the long run, the only way to reduce congestion is by charging the less urgent users—some critics would say the less affluent users—sufficiently to keep them off the congested routes. Although this concept does not require implementation of IVHS, electronic toll collection technology has made road pricing practical and has given the concept a new life (*5*).

Compared with the incremental approach to congestion relief through traveler information and route guidance, road pricing may be considered a radical approach (*6*). Its impact on urban traffic congestion in Singapore has been dramatic. Interestingly, IVHS was not used in Singapore to set up its current road pricing scheme, although some form of electronic toll collection is expected to be installed there soon. The toll collection system will facilitate the future expansion and management of that country's road-pricing scheme. In Singapore, a manually operated road pricing system (an area licensing scheme) to keep most of the motor traffic from its central business district has been in operation since the mid-1980s. The scheme was dramatically successful in reducing traffic congestion in the central business district. In fact, it was overly successful to the extent that the roads became highly underused in the district, and the price for any vehicle to enter the restricted zone during peak hours was reduced from $3 to less than $2 (*7*). Those who used to drive to the central district now either ride the subway or drive to the periphery of the central district and walk or take a taxi in.

Although road pricing has been successful in Singapore, it has not been accepted in other congested cities. In fact, the first attempt at electronic road pricing was actually made by Hong Kong in the mid-1980s, when motor-traffic congestion and pollution in the central business district became intolerable. However, even after money and effort had been spent to install it, the system was never put to use because of political unacceptability. In a recent interview, the Hong Kong authorities attributed the public rejection to the unfortunate timing in the road pricing installation. The Hong Kong authorities did not anticipate that, shortly after the installation, the United Kingdom and China would sign the treaty that reverts Hong Kong to China in 1997. The Hong Kong populace became highly suspicious that the road pricing system might be the beginning of Big Brother watching the residents' movements. Thus, although road pricing is still an official policy in Hong Kong, the authorities resorted to an increase in car ownership taxes as the more practical means to achieve a marked, although perhaps temporary, traffic reduction in Hong Kong.

In Europe, there is a joint manual and automatic toll cordon for Oslo, Norway, and similar plans are under consideration for Stockholm, Sweden. The Dutch government initiated the now-tabled Rekening Rijden (traveling accounting) project, which was due to implement the first part of a road pricing system by 1992, with complete coverage of the Randstat (Rotterdam, Amsterdam, and the Hague) by 1996. In the United Kingdom, serious consideration of road pricing has been coupled with innovative ideas for its implementation. For example, a Timezone concept has been proposed for London, which would be ringed with roughly concentric circles representing progressively more expensive tolls as one approached the center (*8*). This approach would prevent traffic diversion at zone boundaries as has happened around the central business district of Singapore, causing congestion around its boundaries. It was reported that GEC, a U.K. firm, would begin a pilot test of this concept in early 1992 in the southwest London borough of Richmond upon Thames, using a radio frequency communications link that activates an in-vehicle meter (*9*). An even more radical concept, known as congestion metering, has been under consideration by the city of Cambridge (*10*). Unlike the usual road pricing scheme (as in Hong Kong), in which a congested zone is predetermined and a fixed fee for entry is charged whether or not the zone turns out to be congested, congestion metering will levy a charge only when a vehicle experiences actual congestion (defined by a threshold of vehicle speed or numbers of stops per unit distance or both). It is believed that such a scheme will induce a more economically rational behavior from the driver and will result in more effective relief of congestion. Because of the unpopularity of road pricing, the Cambridge term, congestion metering, apparently has been adopted in place of road pricing to represent the generic concept of demand management through economic incentives.

The rejection of, or hesitancy in, adopting the radical solution of road pricing has led to a number of analyses of its political unacceptability. Road pricing has many opponents. Besides the impression that road pricing favors the rich, the strongest public sentiment against road pricing is its appearance as another tax. The general public feels that it already has paid too many taxes. Moreover, the gasoline taxes at both the national and the state levels have not been used entirely for road construction and maintenance. Why not use some of those taxes for roads instead of charging more for road use? Most of the highway users are against road pricing, which is considered a deterrent to automobile travel and another potential imposition that favors public transit versus car use. As has happened in Hong Kong, the privacy issue also has been raised elsewhere as a negative factor by the opponents of road pricing when it is implemented with automatic vehicle identification technology.

On a rational basis, the proponents of road pricing seem to have answers to all the objections that have been mentioned (*8*): for example, reduced rates may be charged to the poor and privacy may be protected by the use of anonymously prepaid smart cards. Depending on the economic assumptions made, no net increase in taxes or costs would result from road pricing; families would be induced to own multiple vehicles; and therefore the automotive industry might even get a 13 percent increase in the market (*11*). Perhaps the best conclusion to the political controversy of road pricing is that although the net social benefit is maximized by the introduction of road pricing, the realistic distribution of this benefit will leave some of the interested parties (including those who cannot afford to pay) worse off than the status quo, and strong opposition from these parties usually has succeeded in blocking the implementation of road pricing (*12*). Any realistic introduction of road pricing must consider some innovative compensation arrangement so that all the major interested parties will be better off than the status quo. While this debate continues, resolution of the key issues and consensus forming will be difficult without field tests of the basic concept of congestion pricing. Interestingly, the recent U.S. legislation (*13*) has provided $25 million per year for six years to support such field tests.

ROAD PRICING—THEORETICAL ISSUES

This section contains an assessment of issues in the theory of road pricing and suggestions of how these issues may be dealt with in the context of emerging IVHS technologies and the potential for public-private partnerships. There are six features of such partnerships that shape a new perspective on road pricing:

1. IVHS technologies offer a large potential set of new services that can be offered on a fee-for-service basis through electronic pricing. A precondition is a significant willingness to pay for these services. The rapidly spreading electronic toll and traffic management applications in the United States and the privately operated TrafficMaster system in the United Kingdom, which provide traffic conditions to fee-paying subscribers (*14*), are encouraging indications of the existence of this precondition.

2. Delivery of these services creates a potential for new relationships between public agencies and private firms. Traditionally, the public sector has contracted on a one-time basis with private firms for delivery of new roadways built according to a design and specifications of the highway department. The new partnership is more likely to be established on a continuing basis with a need for revisions in the relationship as new ideas flow from the learning experience. Contracting must take place with a new emphasis on functional outcome rather than on the basis of predetermined design features.

3. The private sector is assumed to be better equipped to develop pricing and compensation relationships with more flexibility across user groups and more flexibility through time. In other words, for IVHS functions to be fully deployed, there should be the flexibility of unbundling any package of services (*15*) as well as the possibility of bundling public and private services, as suggested in the section on bundling of public and private services.

4. More options will be needed. The fee-for-service approach might start with a fee for basic services and then offer options for those who are interested and willing to pay (*16*).

5. The baseline of fees for users opens up new public-sector opportunities. As an important possibility, public-sector user fees can be offered in the form of discounts or negative prices (subsidies and discounts) to system users (*17*).

6. In developing such systems it is important to offer the user the status quo as an option. In this way citizens will not be forced to accept new technologies, which they may see as experimental. Only in the distant future will it be necessary to create mandatory participation in some elements, in which a consensus may be important for safety or to provide benefits for the traffic system as a whole, such as in the mandatory installation and use of the seat belt.

We now turn to application of these features to fee-for-service partnerships. Recent studies have shown that, from the public perspective, the idea of road pricing or fee for service has a negative connotation. As stated previously, although the basic idea of road pricing dates back to the 1920s and has been reassessed and improved at various times, proposals to implement pricing systems have met with resistance. To counter public resistance, added efforts will be required to present a clear and well-reasoned plan if there is to be any hope of achieving a consensus of support.

One could argue that the implementation of the theory has been naive, issues of redistribution among various groups have been ignored, explanations to the public about the purposes have been inadequate, and cumbersome or unreliable procedures have been used in collecting fees. We now turn to the discussion of six topics in the area of fee for service.

FEE-FOR-SERVICE SYSTEMS

Why Fees Fail To Gain Public Support

The road pricing controversy has been assessed recently in a working paper (*12*). It is easy to see from the basic theory of road congestion where opposition can arise. As shown in Figure 1, the net benefits to road users are reduced by congestion, but the fee used to "correct" the problem makes them still worse off unless they see the revenues from the fee going to their benefit in some other way.

Some users, notably those who value their time highly, will be better off even if the revenues are not returned through lower taxes or in-kind travel-related benefits. Because these users are likely to be in the minority, there is widespread resistance to user fees by road user associations, leading to the political demise of naively formulated proposals to charge for the use of what are regarded as "free" roads. Moreover, some of the losers in a road pricing scheme may feel strongly about the loss and organize highly vocal opposition to user fees.

As will be seen in the section on bundling of public and private services, if road users are paying a fee for service to a private firm (Feature 4), then instead of a fee for using the road at peak congestion times, they may be offered a credit to their monthly statement for diverting off the congested

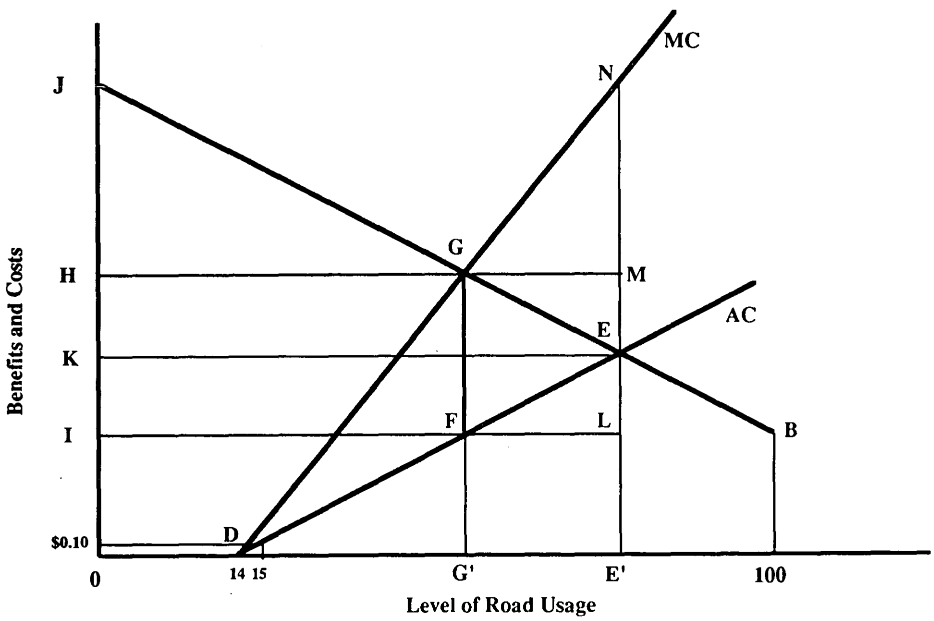


FIGURE 1 Road congestion fees and payment to diverters.

roadway (Feature 5) or for the condition that the roadway was not used at all at peak times during the billing period. Although in theory, paying a fee to travel a congested road should have identical incentives to receiving a subsidy for not traveling a congested road, the user reaction to the latter may be far more favorable (see the appendix). This is analogous to the traveling public reacting negatively to designating an existing lane for HOV while being neutral to designating a newly constructed lane for the same purpose. Moreover, it is the public partner that should have a predominant interest in system benefits from the diversion of travelers rather than the market response of a minority of individuals seeking a specific service option, and this division of responsibility creates a motivation for a public-private partnership. Using the concept of the status quo as an option (Feature 6) means that no one has to divert or pay to continue on the planned route. This necessarily means that they will not be worse off.

IVHS as a Congestion or Loss-Reducing Strategy

The fee for service provision of IVHS technology, such as traveler information and route guidance, can be thought of as shifting out the point (Point D in Figure 1) at which the added flow of traffic begins to congest a roadway or roadway network or reduces the congestion functions. Another dimension to IVHS benefits can be in the area of predictability of trip times instead of reductions in average duration of trip. Some of the congestion on a given road can be predicted by users, but an unexpectedly high level of congestion, discovered only after the trip is in progress, may be unavoidable because the commitment to the trip already has been made: there may be no turning back.

Congestion with long delays has been argued to have a particularly acute cost if, as seems highly plausible, there is a rising marginal cost of added delay in trip time or loss of function convexity (*18,19*). A classic and extreme example of this would be the speed-flow curve, wherein at high levels of traffic flow it is possible to achieve either a high speed and a low trip time or near gridlock with lengthy trip time. If travelers are averse to the risk of lengthy delays, then some type of public-private partnership seems essential for offering users an optional fee-for-service (Feature 4) advisory information system, which would allow the users to avoid unexpectedly long delays from congestion. Again those not interested can have the status quo (Feature 6) with the added plus that some of the diversion of participants can facilitate their trips. Here too is a partnership role. If diverters provide benefits to others, but little or none to themselves, it is in the interest of the system to find a way to compensate them.

If participation in the advisory system is on a voluntary subscription basis, then those who are advised to divert or postpone a trip over a congested segment will presumably save time for themselves or avoid the risk of near gridlock. By not adding to or reducing the trip time of others not diverting, subscribers will provide benefits to nonsubscribers as well. These system benefits, particularly to the nonsubscribers, should motivate a continuing (Feature 2) public-sector commitment to traveler advisory systems offered through private contractors. These public benefits can be seen as a rationale for some public cooperation or financial support for the provision of infrastructure.

Multiple Routes: Existing or Newly Created?

In the modeling of multiroute congestion (*4*), the approach is to represent congestion of a road network instead of a single road. In the simple case that was examined, there are two routes connecting Point Y to Point Z, with different functions for congestibility on the two routes. In this case the user fee on one route needs to be set, taking into account the effect of diversion on the traffic pattern and congestion on the other route. There are two messages from this network approach to public-private partnerships. (a) In setting a user fee on a given (private) route in a road network, there is a need to factor in the possible cost of excessive diversion onto the (public) alternative routes. This supports the continuing relationship concept (Feature 2). (b) The newly constructed private tollways or bridges create benefits on the publicly held routes by relieving congestion there. Here the status quo user (Feature 6) of public routes is better off. In this sense new private roadways partially solve the problem of public road congestion just as public transportation systems ease commuting delays, thereby eliciting an endorsement by resolute automobile commuters.

Excludable Public Goods Offered by a Sole Seller

There is previous literature on private provision of excludable public goods (*10*). Excludable public goods are those to which a potential user can be denied access, such as a museum, a road, an airport, or a park. Excludable public goods are distinct from weather forecasts or other pure public goods (from which users cannot be easily excluded). This excludable public good literature indicates a potential conflict between the private interest of the supplier and user benefits: the supplier offers too little of the public good at too high a price—a situation parallel to the main result of monopolies.

These existing models have not incorporated overuse and consequent congestion. If there is congestibility, the supplier will factor in the interest of users, insofar as congestion will diminish the user community's willingness to pay. There is an element of fortuitous circumstance in that the incentive of the sole seller to undersupply coincides with a social benefit from restricting use below the free access equilibrium. The formal model for this has not yet been worked out in the literature but seems to be a useful project given the range of circumstances that coincide with these conditions.

From the perspective of public-private partnerships, one could think of a case in which it could be effective to charge a relatively high user fee on the toll road to limit excessive congestion. This fee could be combined with a sharing formula for revenues (Feature 2). The public-sector share of revenues could be used to fund a diversity of transportation activities.

Highways as Public Utilities?

A review of legal dialogue (K. D. Syverud, personal communication, 1992) on the recent innovations providing authorization for private toll roads indicates wide differences across jurisdictions in the extent to which private toll roads are regarded as being subject to public utility (cost plus normal profit) rate regulation. In some cases it is as if it were a foregone conclusion that the public utility rate commission is applicable, and in others it is as if this were not even a question.

This lack of clarity concerning rate regulation could create a situation parallel to that with cable television, in which suppliers were granted an exemption from public rate regulation as a condition for investment in the systems. Now that the systems are in place and subscription rates have been rising, various groups have asked for some type of limitation or review of cable rates through mechanisms similar to those used in public utility rate regulation. These issues are bound to arise in road pricing.

Bundling of Public and Private Services

An asserted advantage of public-private partnerships has been the possible augmentation of resources for transportation outside the traditional tax revenue sources (*20*). Here the idea is a bit different: the partnership revenue could give the public sector not only added revenue but an opportunity to set incentives for users (Feature 5). This idea arose in the context of a particular application but seems to have a wider applicability. The specific context was the issue of a privately provided service to truck fleets on I-75 in Michigan (*21*). For a subscription fee and a per-use-of-service fee, trucks would have the benefit of electronic weigh-in, messages from the private fleet controller, travel advisory messages, and other new services.

The public-sector interest would be the longer-term (Feature 2) savings as a result of fewer weigh-station personnel, a reduction in pollution from stop-and-start traffic, and the possible use of economic incentives as a means of diverting traffic. Specifically, at certain points along the route that are subject to periods of congestion, the public partner could offer financial incentives for trucks to divert (Feature 5) or possibly postpone trips to a less congested time. Note that truck drivers so inclined could stick with the status quo (Feature 6). The assumption here is that there are alternate routes that have below-capacity traffic or that trip timing could be set for nonpeak times (Feature 1).

The incentive could be in the form of a reduced monthly charge for each diversion, or frequent diverter credits, for those with no trips at peak times. This idea, of course, assumes some longer-term sharing (Feature 2) of the subscription revenues between the public and private partners. The important point here is that the private partner would not be expected to have a long-term interest in creating incentives out of its revenue share for trucks to divert because a good part of the benefits would accrue to vehicles outside the system (private passenger cars in this case). Here we can see a division of interest and responsibility within the partnership that creates real possibilities for a complementary relationship between the public and the private partners.

SUGGESTIONS FOR FUTURE RESEARCH AND FIELD TEST

The key ideas in the area of public-private partnerships from the perspective of economics, as developed and summarized in this paper, constitute a framework for future research, which can range from basic investigations to operational field tests.

At the fundamental level, we would suggest the development of formal models for studying the concept of congestibility in excludable public goods. Such models can be used as a basis for legal and economic policy analysts to debate and develop rate regulations to resolve conflicts between, and to protect the interests of, users and suppliers of excludable public goods, such as toll roads and IVHS infrastructures. These models also can provide a more solid ground for estimating the private versus external benefits/costs in a total system, so that the optimal share of public versus private financing (for IVHS infrastructures) can be determined on a more rational basis.

The institutional issue of public-private financing is intertwined with technical issues and requires a sociotechnological approach to consider both types of issues simultaneously (*21*). For example, for IVHS infrastructures, a key prior issue is what type of standard should be used regardless of how it is financed. The whole issue of standards is a subject unto itself and highlights questions such as early commitment to what turns out to be a poor standard (consider the issue of analog or digital standards in HDTV) on the one hand, and procrastination, which prohibits anything from starting, on the other (*22*). Suppose the question of infrastructure type is settled. One could argue that the public could pay for infrastructure and charge a user fee to suppliers, who would, in turn, pass this cost on to users. This has a parallel in the payment of landing fees by airlines to a publicly financed airport authority. If an infrastructure were privately financed and the private firm charged a user fee on the basis of use of services, the issue of public regulation of the fee structure could arise here too.

One of the exciting ideas that has emerged from our work on this paper is the innovative implementation of congestion pricing through the bundling of public and private services. This idea has been discussed in the context of a service privately provided to truck fleets on I-75 in Michigan in conjunction with an economic incentive for the trucks to divert from congested routes. Although the validity of the basic concept has been proven by our economic analysis, the practical problems in implementation will need further consideration. We would suggest a concerted effort among the interested parties to identify and resolve these practical problems. For example, given the current traffic information, how do we provide economic incentives according to the true intentions (and not false reports) of the truck drivers about their preferred routes? Can we design the road network, including the access and egress control from freeways, to compel the truck drivers to reveal their true intentions? What about the conflicts between truck drivers and their fleet operators? How do we avoid some of the problems such as the potential oscillations between alternative routes as have been revealed by computer simulation (*23*)? What is the ultimate benefit to the public authority to bringing user optimum to system optimum in traffic assignment so that we can determine the maximum justifiable economic incentives for diversion?

Some of these problems probably cannot be solved, and other related problems cannot be identified, unless the idea of bundling private services with congestion pricing is tested in the real traffic environment. Given the encouragement and substantial funding for congestion pricing field tests in the recent legislation (*13*), we suggest an exploration of the feasibility of establishing a congestion pricing project on I-75 in Michigan with federal funding, augmented by private resources. Such a project will help to bring the interested parties together for a serious and concerted effort to test the exciting ideas of simultaneous bundling and congestion pricing.

APPENDIX
Model of Road Pricing

To understand this model of congestion, break the analysis into two parts. In the first part we have users of a roadway segment over a given interval. The potential users are arrayed in terms of benefits as indexed by their willingness to pay. They may or may not need to pay. This is summarized in Curve *JB* (see Figure 1), which ranks potential users from highest value of the trip to lowest value.

Now turn to the congestion side. We suppose that over some range of usage, there is no problem of road congestion. This range is from 0 to *D* users per hour. Beyond that point congestion sets in. Here the key distinction is between marginal congestion (MC) and average congestion (AC). Average congestion shapes individual behavior. Marginal congestion describes system costs more accurately. Marginal congestion rises above average congestion.

The idea may be illustrated numerically. Suppose that as the number of users rises from 14 to 15 per unit time, the trip is slowed by a few seconds for everyone. That is, average trip time rises by a few (3) seconds. The 15th traveler could possibly notice the slowdown. Here we assume, perhaps unrealistically, that the slowdown is noticed. This is not so critical to the argument. The critical point is that the 3-sec slowdown applies not to just Driver 15, but to all (14) preexisting drivers as well. This 45-sec (15 drivers × 3 sec) slowdown, translated into congestion cost, is what defines marginal congestion cost (to the entire set of users). But individual behavior is shaped by average congestion costs. Added travelers are discouraged from driving only at Point *E*, where average congestion equals the congestion-free benefit to the user from the *JB* schedule. Socially efficient congestion is back at G' travelers.

A commonly proposed remedy is a user fee of *GF*. The added user at G' would then face the cost of average congestion plus the user fee, and use would equilibrate to the level of G' travelers. A problem with this remedy is that the fee revenue of *HGFI* comes out of user pockets. This makes them worse off, and possibly more so than the original problem of congestion. In theory, this revenue should be used to reduce taxes or provide some offsetting benefit, but motorists, like other taxpayers, are skeptical of government. A simple restriction on quantity of drivers to G' would seem to solve this question of the government getting the revenue. A drawback is that all drivers along the *GB* segment of the benefit curve would be interested and there would be a type of non-price rationing or roadway lottery. Some who value the trip very little would end up traveling at peak times and would displace those who value the peak trip time more highly.

A different approach is negative pricing or to pay for not using the road at peak times over some billing period. (This is paid to those who do not travel at peak times!) This has usually been only a theoretical possibility. With the idea of diverter discounts (administered through in-motion metering) in a joint public-private venture, as discussed in the section on bundling of public and private services, such an arrangement might be a practical possibility as well.

REFERENCES

1. F. P. Stafford and K. Chen. *Self-Financing HELP and Rerouting Incentives*. Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, 1992.
2. M. Cameron. *Transportation Efficiency: Tracking Southern California's Air Pollution and Congestion*. Environmental Defense Fund and Regional Institute of Southern California, March 1991.
3. A. C. Pigou. *Wealth and Welfare*, 1st ed. London, 1920.
4. A. A. Walters. The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica*, Oct. 1961, pp. 676–699.
5. K. A. Small, C. Winston, and C. A. Evans. *Road Work*, Brookings Institution, Washington, D.C., 1990.
6. K. Chen. Radical Approaches to IVHS Goals. Presented at 2nd Annual Meeting of IVHS America, May 17–20, 1992.
7. B. G. Field. *From Area Licensing to Electronic Road Pricing: A Case Study in Vehicle Restraint*. SAE Paper 911677. Society of Automotive Engineers, 1991.
8. N. Green. London "Timezone." Presented at National Economic Development Organization Conference on Information Technology and Traffic Management, London, Nov. 1990.
9. London Borough Sets Road Pricing Trial of GEC Built System. *Inside IVHS*, Vol. 1, No. 20. Oct. 14, 1991, pp. 4–5.
10. B. Oldridge. Electronic Pricing: An Answer to Congestion. Presented at National Economic Development Organization Conference on Information Technology and Traffic Management, London, Nov. 1990.
11. T. Karlsson. *Road-Pricing*. DRIVE Program, March 1990.
12. T. Nemoto and J. O. Jansson. *Road Pricing—Political Controversy and Its Modeling*. Discussion Paper 16, DRIVE/SECFO. Feb. 1991.
13. *Intermodal Surface Transportation Efficiency Act of 1991*.
14. D. Martell. Navigation Systems. Presented at National Economic Development Organization Conference on Information Technology and Traffic Management, London, Nov. 1990.
15. G. A. Robertson, and M. A. Roberts. *Intelligent Vehicle Highway Systems: Private-Sector Investment Capital and Regulatory Issues*. Research paper. Alex. Brown & Sons, Inc., April 9, 1992.
16. M. P. Ristenbatt. A Communications Architecture Concept for ATIS. *Proc., 2nd Vehicle Navigation and Information Systems*, IEEE/SAE, 1991.
17. K. Chen and R. D. Ervin. *American Policy Issues in Intelligent Vehicle-Highway Systems*. SAE Paper 901504. Society of Automotive Engineers, 1990.
18. K. Chen. *Nonlinear Disutility of Travel Delays: Implications in Diversion Benefits and Strategies*. IVHS Technical Report 90-3. University of Michigan, Ann Arbor, 1990.
19. F. P. Stafford. *Social Benefits of IVHS*. SAE Publication SP-833. Society of Automotive Engineers, 1990.

20. J. A. Gomez-Ibanez, J. R. Meyer, and D. E. Luberoff. What Are the Prospects for Privatizing Infrastructure? Lessons from U.S. Roads and Solid Waste. In *Is There a Shortfall in Public Capital Investment?* (A. H. Munnell, ed.), Federal Reserve Bank of Boston, 1990, pp. 143–174.
21. K. Chen and F. P. Stafford. A Sociotechnological Perspective on Public-Private Partnership for IVHS Infrastructure. Presented at the Workshop on Public/Private Sector IVHS Deployment Issues, Rockville, Md., April 8–9, 1992.
22. J. L. Gifford. *Standards for Intelligent Vehicle-Highway Systems*. SAE Paper 920938. Society of Automotive Engineers, 1992.
23. P. D. Halas. Resource Allocation Mechanisms in IVHS: Information and Incentives. *IVHS Research Topics*. University of Michigan, Ann Arbor, April 27, 1992.

Publication of this paper sponsored by Committee on Taxation, Finance, and Pricing.

Time in Transport: A Perverted Problem? Arguments for a Fresh Look at Time Utility Research and Its Application

SVERRE STRAND

The way in which time is converted into money is becoming increasingly important in transportation planning. In particular, the pricing of time is becoming more and more decisive for the calculated profitability and realization of road projects. The general trend is toward more and more traffic but smaller and smaller individual time savings, although, as always, such savings are discrete in time and space. These and other circumstances create very real problems of pricing, especially in view of the aggregation problem. It is argued that the problems of aggregation are ignored in the application of, and maybe in, economic theory itself. Whether this assertion is wrong or right, it raises the all-important question of how robust a theory is with respect to deviations from the underlying constraints before the application of that theory collapses in terms of validity of the results. The impression is that in the application of the time utilization theory the effects of such deviations have been ignored, which makes much of the present use of time in transport both meaningless and misleading. Pertinent questions about research problems and bottlenecks for a credible practical application of time utilization theories are raised.

How important is the issue of time in transport? It has been well documented that for the average road project, 70 to 80 percent or more of the total benefits are attributed to the time savings of the project—that is, benefits are defined and delimited in standard cost-benefit analyses. In other words, the issue is all important. By the same token, it is all important how we treat time in transport.

The point of departure was curiosity about the behavioral justifications (guidelines for how to aggregate and price accordingly) for time savings in time and space. What started as curiosity became skepticism, not because of what was said, but mainly because of what was not said: the discussion of these problems was found to be practically nonexistent. Rather, the process of obtaining the kind of time values we really want should be regarded as a three-stage rocket that can function only if the stages are released in the correct order:

1. Measurement of time and time savings;
2. Aggregation, that is, adding together time and time savings; and
3. Valuation of time in monetary terms.

It is imperative to approach Problems 1 and 2 critically because the answers to these determine how Problem 3 can be confronted.

Institute of Transport Economics, PO Box 6110 Etterstad, 0602 Oslo, Norway.

The important question in relation to Problem 1, the measurement of time savings, is as follows: How large must a single savings be to be judged as significantly larger than 0 so that a person will make alternative use of it? In other words, should all savings below a certain limit (1 sec, 1 min, 10 min, 30 min, and 1 hr) be considered equal to 0 in practice, or should they be retained; what is the threshold value in different situations? (The problem of errors of measurement, which is important in itself, is disregarded, particularly when small time savings are concerned.)

Problem 2 comprises yet another dimension:

- What can be aggregated? Which assumptions are necessary to allow time savings, or rather their effects, to be aggregated in time and space under what rules of aggregation?
- Which threshold values are appropriate when the individual savings are to be aggregated into hours and days of work and person-labor hours, days, and years?

The perception is that Problems 1 and especially 2 generally are treated as nonproblems in economic theory and practice. It should not be necessary to separate the different phases as rigidly as is suggested here. On the contrary, problems relating to the functional measurement of aggregation of time cannot be seen independently of the problem of valuation in monetary terms. That is because ideally we would like to be able to put a shadow price on every individual, nonaggregated time savings, depending on such factors as the person involved, the amount of time saved, and the circumstances under which the time was saved. But again, the crucial question is the relevance of the willingness to pay that is possible to measure or that we are able to measure. Therefore, this three-stage rocket should be seen as just as much an educational and practical aid as a fundamentally new way of looking at things.

THE EYE OPENER

Geographers—even though they have a subject called time geography—have always stayed in the background and left the problem of time and time valuations to economists, planners, and engineers.

To me, as a geographer, the case of the Sollihøgda Road changed all this. There I happened to discover that the realization of a major transportation project was being decided by the arguments related to time and savings of time. This fact in itself was hardly a shock, but the discovery of the reasons behind the conclusions certainly was a shock, professionally speaking.

In fairness to the consultants and the decision makers involved, it should be stressed that the road project in question is not unique; it just happened to be the one that caught my attention. Such projects are becoming more and more typical, both in character and in analytical approach. In addition the project is not especially negative in the sense that the consultant did a poor job. Moreover, the road project is not necessarily economically unprofitable.

Until 1991, about NOK 150 million (approximately U.S. $20 million) will have been invested in various improvements to the existing Sollihøgda-Vik road some 30 mi from Oslo, Norway. Operated as a toll road, the length of the highway will be reduced by 2.4 km.

The consultant prepared an impact analysis that took into consideration costs related to construction and maintenance; vehicles; time; and accidents and noise. Only the time component shall be discussed here. In any case there is no conceptual confusion involved in the calculations of the others.

The three subprojects represent a total shortening of about 2.4 km. As a result of the improved standard, the average speed may be increased from about 50 km/hr to about 60 km/hr. This results in a total time savings per automobile trip of some 5 min, including a stipulated loss of time of 20 sec for toll collection. The following quote from the consultant's report may serve as an example of how mechanistic assessments of the sensitivity of people's behavior to changes are (*3*, p.48, translated): "Estimations for the section Sønsterud-Rørvik have been made which show that income-maximizing toll money rated will be close to . . . , on the assumption that the old road is subject to restrictions corresponding to 2 minutes addition in traveling time."

As an example, one car (containing one person) commuting between Hønefoss and Oslo 250 days a year will represent an aggregate annual time savings of about 40 hr. For such journeys, the conversion rate for time is conventionally NOK 25.20/hr. On this road, the car would then, equally conventionally, represent an economic savings of about NOK 4/day and NOK 1,000/year. If this were a car at work, the price per hour would be NOK 95.50. In this case the savings would be about NOK 3,900/year.

In this project, the actual volume and structure of the traffic forecast gave an aggregated savings of about NOK 18.4 million/year, in which the aggregation process in essence was nothing, as usual, but a series of straightforward additions.

The main point of the impact analysis is that the time savings constitutes the major and decisive factor. Without the monetary value of time savings, the project would be defined as unprofitable. With the time savings, the project suddenly becomes very profitable, since the ratio between benefits and costs jumps from being less than the necessary value to indicate profitability to twice that value:

Item	*Value*
Construction costs (NOK, millions)	138.0
Net benefit per annum with time valuation (NOK, millions)	29.2
Net benefit per annum without time valuation (NOK, millions)	10.7
Benefit/costs ratio with time	0.21
Benefit/cost ratio without time	0.08
Minimum benefit/cost ratio for profitability	0.10

What kind of time savings are we in fact talking about in this case, and what rationale permits the aggregation procedure? Despite considerable uncertainty and variability, the answer may be on the order of 5 min per vehicle per trip, assuming that the trip is long enough to take advantage of the entire savings provided by the system. The monetary valuation of these savings follows the specifications of the *Norwegian Driving Cost Manual*. It is assumed that all individual savings can be added to hours and person-labor years and then converted into money using time-price criteria. These aggregates become large because there are many cars (assumed to be "cooperating" with a view to obtaining something productive from the "unproductive" individual savings). And the aggregates become so big that they change the projects from nonprofitable to profitable and therefore allow their implementation—a particularly important point in a situation involving planning and decision making.

Should this be considered an acceptable procedure, representing public interest? In other words, does the benefit side of the calculation represent money that is as "real" as that on the cost side undoubtedly is? Are we comparing comparable units?

MANUALS: THE NEWEST TESTAMENT?

All Norwegian discussions on road investments refer to the *Driving Cost Manual*, as is presumably the case in other countries (*4*). Earlier editions of this manual dating back to 1959 also seem to have held an important position, although perhaps not to the same extent as now. The reason for that is that time was not as crucial to investment decisions as it has become in later years. The influence of the manual today appears to be so great that a reference made to the driving cost manual in itself legitimates without further discussion the use of time in transport in one very specific manner. This makes it clear that it, and presumably equivalent manuals, has a very strong influence on cost/benefit analysis and subsequent decisions on road investment and "competing" transport investments. What is written there should therefore be assessed equally seriously.

If we consider the manual in relation to the three-stage rocket, that is, measurement, aggregation, and valuation in monetary terms, we find nothing about Problems 1 and 2. These problems are implicitly "solved" by not defining them as problems.

The Swedish "Effect Catalogue" (*5*) may be considered equivalent to the Norwegian manual. We find no discussion of Problems 1 and 2 in this book, but we find the same arguments and standards with respect to the pricing of time.

A Manual of User Benefit Analysis of Highways and Bus-Transit Improvements (*6*) is less absolute in its presentation. Although the aggregation problem is not explicitly discussed here either, small and large savings are at least evaluated against each other in fixing prices. Because a savings of 0 to 5 min is considered to be small, 5 to 15 min average, and over 15 min large, and the curves climb steeply after only 5 min, leveling off again at 15 min, we also run into Problem 1—the problem of errors of measurement—although unfortunately only implicitly. This handbook was critically evaluated a few years after it was published (*7*). This evaluation was studied with one question in mind: Do the critics see as a problem the way the manual deals with time and time costs? The answer is a definite no.

If biblical texts are used axiomatically, it is the users who must be held accountable. But who should be blamed here? The users, the authors, or both? Both, of course. On the other hand, it is hard to blame a user, usually a consultant, too hard for misuse. An author must take the main responsibility for misuse by users of the manual. Misuse by the users should be considered as mainly de facto rather than intentional. An important reason for this point of view lies in the structure of the manual, which provides answers but virtually no assumptions or interpretative reservations concerning the application of these answers. However, in all fairness one must also presume that use is based on professional acceptance of the manuals, not on blind faith. Therefore, any blame should be shared, the more so as the years go by, without anyone challenging the underlying assumptions of the procedures and thus the procedures themselves.

VALIDITY OF CORE ARGUMENTS

How time in transport is applied in cost-benefit analyses is based on the economic welfare theory, that is, on the individual's maximization of benefits: the marginal value of time is the one the consumer is willing to pay for a marginal reduction in travel time (*8–11*).

On the basis of this theory there is a vast amount of literature on the subject of pricing time. This literature will not be discussed, except for the observation that there seems to be little willingness to consider alternatives to the classic assumptions about marginal utility, momentary consumption, free individual adjustment of working hours, and others. Beyond this, examples relating only to the initial point on the validity of the core of a theory, its central adjustment, in view of deviations from the assumptions that underpin this theory, will be referred to.

Fridstrøm (*12*, p.3) illustrates very well one of the behavioral inconsistencies of the economic time utilization theory, as he describes its dependence on growth for the theory to hold and the relationship between time, goods, and welfare in economic theory. Referring to this theory, he states that the individual's welfare consists of two components: (material) goods and time. Only the first component can grow in volume, the number of hours in a day of course being constant. Thus, the only way to increase welfare is to increase the consumption of goods, implying for instance that the only way to increase the welfare by *x* percent is by increasing the consumption by *x* percent. This obviously perverts the concept of welfare if we want welfare to imply well being or even happiness. And still, this is one of the fundamental prerequisites of the economic time utilization theory.

Heggie (*13*) supports in this example the view that there is an imbalance between theoretical refining and empirical calibration, that is, a disagreement between planning models and research models:

> The reasons for these disagreements are various. Important amongst them are that much recent research has concentrated on theoretical and statistical issues. Little substantive work has been done on the empirical ones. Indeed, theoretical developments have tended to outstrip the practitioner's ability to estimate empirical coefficients. The main dispute over the appropriate value of travel time savings may thus be an empirical, rather than theoretical one.

The particular issue of different valuation of small versus greater time savings—without explicitly coupling it to any aggregation problem—has attracted considerable attention.

Gårder (*14*) has made an in-depth study of the valuation of short time periods. In that study he also conducted an interesting inquiry among several outstanding people in transportation research to determine whether a small savings involved a different valuation.

Expert	*Yes*	*No*
D. Solomon	X	
F. A. Haight		(X)
W. F. McFarland		X
T. Miller		X
M. Luger	X	
E. Hauer	X	
A. Timar	X	
J. Lawson	X	X
S. R. Jara-Diaz	X	
L. Needleman	X	

The results speak for themselves: confusion; this is in a context in which implications of the one choice or the other are staggering, considering the size of time savings that we most often must speak of nowadays.

The pioneers among time economists were aware of the aggregation problem, as the following quote from Harrison and Quarmby (*15*) shows. One can argue whether their suggestion for tackling the problem is the right one and whether, for instance, the behavioral link to their "marginal consumer" is solid enough. This is not in itself a problem, but a natural, necessary, and continual challenge. It is, however, a real problem that the discussion of aggregation and behavior never, as far as can be seen, came to fruition, but was in reality rejected. This is even more regrettable in view of the fact that Harrison and Quarmby (*15*, pp.183–185) acknowledged as much as 25 years ago that

> this problem (of size of time saving) is put in two basic forms: first, is one saving of ten minutes worth ten savings of one minute; and second, do savings under some given amount have any value at all. Before these questions can be tackled, some theory must be developed as to the way in which time savings are of value to people . . . It is not difficult to construct simple arithmetic examples which show the equivalence of the numerous small savings with the single large saving, but it is apparent that any argument for equivalence must depend heavily on the validity of the probability distribution assumed.

And it will soon be as long since Burenstam Linder (*16*, pp.115–116) wrote:

> It must not be overlooked that what we have stated here as to savings and time allocation are criteria of efficient behaviour and not necessarily propositions about actual behaviour. It may be advantageous to combine consumption goods and enjoyment time in certain proportions over the years. But from our discussion of decision-making by households, it should be clear that such an efficiency criterion is built on an assumption of perfect knowledge and zero information costs. In reality the situation is different. We must expect that people will not systematically try to estimate their future earnings. There is quite a widespread use of current earnings as an index of the future situation. In such a case, actual behaviour will differ from what would be efficient behaviour under perfect knowledge.

Carlstein et al. (*17*, p.4) put it like this in their well-known treatise on time geography: ". . . shows how the rope of continuity can become a hangman's noose if basic facts of life such as indivisibility are not taken note of."

Although the following quotation appears only to deal with the Level 3 problem, monetary valuation, it is equally related to Level 2 (*18*, p.250, translated)

> As a measure of the value of the time spent on a specific activity, it is reasonable to use the value of time used in the best alternative application. Here it is reasonable to use the person's salary as a measure. Economically, this will also be the correct measure of the value of time, under normal conditions.

When it comes to the rationale of adding together time savings, small or great, which are also discrete in time and space, I have yet to find a discussion that reflects that behaviorally defined marginals are not or do not need to be the same as mathematical marginals. It is definitely pertinent that discussions of the conditions for identity between such marginals should be on the agenda. After all, we are talking about "normal conditions" such as the following:

- Individuals act independently of one another, taking their environment for granted;
- Individual preferences are rational (mutually compatible) and constant; and
- The employee is free to decide his own working hours.

These conditions are nothing less than conditions that must be fulfilled for the theory to be applicable in the sense we are talking about here. Therefore we should not avoid a discussion of the behavioral realism of such assumptions and to which of these, and to what extent, the models are sensitive. For instance, how important is the quoted condition of freedom with regard to the conversion of time to money?

The behavioral content represents the bottleneck criterion in a situation in which money valuation and profitability are what the theory must justify. Therefore, how far has the theory of time utility been calibrated against empirical observation?

Why then is it so hard to find such discussions? One can only speculate that positive causes would be those that assume that problems of aggregation are being solved indirectly through proper measurements of the individual's willingness to pay. A negative cause would be, for instance, whether the analysis technique required linear aggregation, and this was contrary to one's own conviction of how they should be added together, and whether this conviction was allowed to yield to the formal demands of the technique in question.

The following quotation from Bates (*19*, p.15) is probably a very plausible explanation, unfortunately, for why methods and techniques should not become straightjackets for thought, but tools for thought: "The problem is that their application would lead to distortions of the CBA calculus . . .".

In today's standard procedures for aggregating time savings and transforming them into real money—to be weighted against the real money costs of the project in question—it is assumed that these two products are equivalent: 1 year × 1 person and 1/360 year × 360 persons.

My point of departure would be that in general they may be equal from the point of view of individuals, but not from the point of view of society. In other words the individual is selfish, and may not, will not, or cannot act in the same way as a private person as he does as a citizen. Thus, the aggregation of individual willingness to pay may not—maybe it usually does not—add up to what society as a whole should be willing or should prefer to pay for a given time savings, in view of the alternative use of that same money. Surely, this sort of schizophrenia is a well-recognized phenomenon. But it is also as surely not dealt with accordingly in practical application. And that is what matters here.

CONCLUDING REMARKS

This paper has been an attempt to point out some methodological problems regarding measurement, aggregation, and pricing of time savings, without regard to whether we are in a situation of strong traffic growth. Until now, the present and, it is argued, sometimes sloppy way of applying economic theory has with the greatest of ease made feasible almost any road project calculation. Because growth has been what it has been, there has been little need to check out the forecasts as such. But this situation is changing. Because of higher mobility levels in the before situation and because marginal traffic generated from the same project will be less than it used to be, the marginal willingness to pay must be expected to decrease (Figure 1). This is definitely adding to the pressure on the relevance of CBA procedures in general and on the relevance of time valuation procedures in particular.

Finally, the question is not whether time is important, but where, when, and how time in transport is important. It is not the critical use of time that is criticized, but its uncritical use. What is attempted is to bring back to the agenda a problem that has been defined as a nonproblem for too long. In terms of the three-stage structuring of the problem of time in transport, the neglect of Stages 1 and 2—the problems of measuring and aggregating—is particularly stressed.

Still, one has heard and will continue to hear that all the problems touched on here are well known. That may be so.

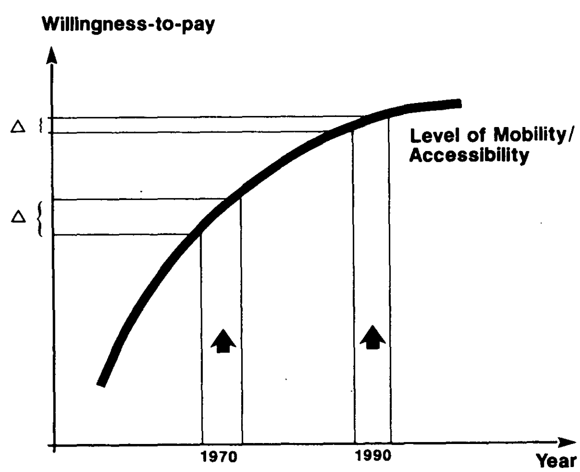


FIGURE 1 Variations over time in view of different preimpulse levels of mobility/accessibility: identical impulse–nonidentical effects.

However, this is not a very productive argument in the present situation, one characterized by an increasing lack of credibility: the feeling is that neither the individual user, as reflected in the failure of traffic forecasts, nor governmental decision makers, as reflected in their rejection of the recommendations of the cost-benefit analyses, believe any longer in what we are doing. Neither the individual nor the government behaves according to our findings on the value of time. In other words, we are not very good at unearthing the behavioral content of time savings. This general lack of credibility may be the best indication that something is wrong and that something must be done. In many ways, the utilization theory and the possibilities of applying it are found to be valuable but unused. The fact that there are no ready recipes for how to use this theory would be a matter to worry about only if one could not agree on the credibility diagnosis. Because if one could agree, it would be possible to find that receipe in only a matter of time and of that a very short time.

REFERENCES

1. D. A. Hensher. Behavioural and Resource Values of Travel Time Savings: A Bicentennial Update. *Australian Road Research*, Vol. 19, No. 3, 1989, pp. 223–229.
2. *The Value of Travel Time Savings*. Policy Journals. MVA Consultancy, Newbury, 1989.
3. *Utredning om bompenger på E68 i Hole* [The Sollihøgda Toll Road (E68)]. Sandvika, Asplan H 2411/P10313, 1985.
4. N. Skarra, and K. O. Gabestad. *Kjørekostnadshåndboken (1983), med reviderte tabellhefter og eksempelsamlinger 1985 og 1987* [*A Driving Cost Manual (1983), with Revised Tables and Examples, 1985 and 1987*]. Transportøkonomisk institutt og Vegdirektoratet, Oslo, 1983.
5. *Effektkatalog: Väg- og gatuinnvesteringar (Effect Catalogue: Road and Street Investments)*. Vägverket, Knivsta, 1986.
6. *A Manual on User Benefit Analysis of Highway and Bus-Transit Improvements*. AASHTO, Washington, D.C., 1977.
7. A. Kimboko, L. Henion, et al. Critical Evaluation of AASHTO's Manual on User Benefit Analyses of Highway and Bus-Transit Improvements, 1977. In *Transportation Research Record 826*, TRB, National Research Council, Washington, D.C., 1981, pp. 22–28.
8. De Serpa. A Theory of the Economics of Time. *Economic Journal*, Vol. 81, 1971, pp. 828–845.
9. G. Becker. A Theory of the Allocation of Time. *Economic Journal*, Vol. 9, 1965, pp. 493–517.
10. N. Bruzelius. *The Value of Travel Time Theory and Measurement*. Croom Helm, London, 1978.
11. S. Algers. *Samordnade tidsvärdsstudier i Norden (Co-ordinated Time Value Studies in Norden)*. Transek, Stockholm, 1991.
12. I. Fridstrøm, *Tidsknapphetens paradoks. (The Paradox of Time Scarcity)*. TØI 0167, Oslo, 1971.
13. I. G. Heggie. Value Savings in Non-Working Travel Time: The Empirical Dilemma. *Transportation Research 17A*, Vol. 1, 1983, pp. 13–23.
14. P. Gårder. *The Value of Short Time Periods*. Stockholm, 1989.
15. A. J. Harrison, and D. Quarmby. *The Value of Time in Transport Planning: A Review*. ECMT, Paris, 1969.
16. S. Burenstam Linder. *The Harried Leisure Class*. Colombia University Press, London, 1970.
17. T. Carlstein et al. (eds.). *Time and Regional Dynamics*. Edward Arnold, London, 1978.
18. *Arbeidstidsreformer (Working Hour Reforms)*, Vol. 9A and 9B. Norges Offentlige Utredninger, Oslo, 1987.
19. J. Bates. *The British Value of Time Study*. John Bates Services, London, 1965.

Publication of this paper sponsored by Committee on Taxation, Finance, and Pricing.

Transportation Utility Fees

REID EWING

In June 1992, Port Orange, Florida, became the 10th U.S. city (and the first east of the Mississippi River) to adopt a transportation utility fee (TUF). Initially, TUF funds will replace a 0.287-mil subsidy from the city's general fund and eliminate a shortfall in the city's road maintenance budget. Eventually, funds will be used to pave dirt roads, construct bike paths, and reconstruct and widen deficient city streets. In a TUF ordinance, roads are treated as a public utility, and developed properties are charged a fee for service in much the same way they are charged for water, sewer, trash collection, and, increasingly, storm water utility services. Like other utility fees, TUFs are imposed on a jurisdictionwide basis and continue in perpetuity, financing ongoing operations. A nationwide search uncovered nine localities outside Florida with TUF ordinances as of late 1991. From a review of their experiences and an analysis of legal cases relating to TUFs and other user fees, a TUF ordinance was drafted and a fee structure was developed for Port Orange. The rationale for Port Orange's fee structure, focusing on legal considerations, is outlined.

In June 1992, Port Orange, Florida, became the 10th U.S. city (and the first east of the Mississippi River) to adopt a transportation utility fee (TUF). Initially, TUF funds will replace a 0.287-mil subsidy from the city's general fund and eliminate a shortfall in the city's road maintenance budget. Eventually, funds will be used to pave dirt roads, construct bike paths, and reconstruct and widen deficient city streets.

In a TUF ordinance, roads are treated as a public utility, and developed properties are charged a fee for service in much the same way they are charged for water, sewer, trash collection, and, increasingly, storm water utility services. Like other utility fees, TUFs are imposed on a jurisdictionwide basis and continue in perpetuity, financing ongoing operations.

A nationwide search uncovered nine localities outside Florida with TUF ordinances as of late 1991 (Table 1). From a review of their experiences and an analysis of legal cases relating to TUFs and other user fees, a TUF ordinance was drafted and a fee structure was developed for Port Orange. This paper outlines the rationale for Port Orange's fee, focusing on legal considerations.

IT BEATS A PROPERTY TAX

A TUF is not a user fee in the classic sense. By the textbook definition, user fees are "payments for voluntarily purchased, publicly provided services that benefit specific individuals" (*1*). A TUF is not comparable with fees for, say, the use of public swimming pools. It is not voluntarily paid and does not fund a service that benefits "specific individuals" to the exclusion of nonfeepayers. In this sense it is more akin to a tax than a fee.

Popular wisdom would say that roads are not suitable for funding with user fees. "User charges can function only when activities financed meet two necessary conditions: benefits separability and chargeability. These are features absent from pure public goods. . . . The farther a good or service departs from publicness and the closer it approximates a private good, the more feasible are user charges" (*2*, p.360). Although roads are not a pure public good, they certainly fall near the public end of the public-private spectrum. Provision of such goods historically has been funded with taxes rather than user fees.

Then why contemplate a fee for road use? User fees are the fastest-growing source of revenue for local governments. They are popular for three reasons.

New Revenue

Historically, cities and counties paid for roads with taxes and special assessments. Since the mid-1980s, localities in certain states have also imposed transportation impact fees. But even tapping all these sources, many localities find their road needs outpacing their revenues. Port Orange, for example, is facing a $245,000 shortfall in its road maintenance program.

Property tax hikes are death politically, even where (as in Port Orange) a locality has the ability to raise them under state law. Transportation impact fees are inherently volatile and can be used only to meet the needs of new development (under applicable case law). Local governments must come up with the funds to operate and maintain the new roads built with impact fees and to eliminate congested conditions on existing roads. These are precisely the uses of funds for which TUFs are intended (Table 2).

Equity

With a property tax, a significant percentage of trip generators pay nothing because of their tax-exempt status. In contrast, with a TUF, every local trip generator pays to support the local road system.

Of course, some inequity creeps into a transportation utility fee schedule because road use is estimated rather than measured, and because estimates are based on averages for entire classes of property. Still, this shortcoming may be less problematic than the exemption of entire classes of developed property from any payment of property taxes. This is clearly the case in Port Orange, where the amount of tax-exempt property is a sore point with local officials.

Joint Center for Environmental and Urban Problems, Florida Atlantic University/Florida International University, 220 Southeast 2nd Avenue, Suite 709, Fort Lauderdale, Fla. 33301.

TABLE 1 National Experience with Transportation Utility Fees

Fee Characteristics	Fort Collins Colorado	La Grande Oregon	Pocatello Idaho	Ashland Oregon	Tualatin Oregon
Name	Transportation Utility Fee	Street User Fee	Street Restoration and Maintenance Fee	Transportation Utility Fee	Road Utility Fee
Status	Implemented 1984 Discontinued 1987	Implemented 1985	Implemented 1986 Discontinued 1987	Implemented 1986	Implemented 1990
Fee Basis	Front footage Trip generation	Flat fee	Trip generation	Flat fee for residential uses Required parking spaces for non-residential uses	Trip generation
Use of Funds	Road maintenance	Road construction Road maintenance	Road maintenance	Road maintenance Bikeways Pedestrian improvements	Road maintenance Street lighting
Land-Use Categories	Single-family Multifamily 1 Non-residential category	1 Land use category	Single-family Multifamily Mobile homes Retirement homes 79 Non-residential categories	Single-family Multifamily 10 Non-residential categories	Single-family Multifamily 7 Non-residential categories
Exemptions	Undeveloped land	Properties without water service	Undeveloped land Unoccupied structures (must be unoccupied at least 90 days)	Undeveloped land Churches Nursing homes	Undeveloped land Unoccupied structures without water service (after 30 day vacancy) Park-and-ride lots Public park land without parking Railroad right-of-way
Fee per Single Family Unit	$.75 - $1.50 per month ($1.00 target)	$2.50 per month	$2.65 per month	$1.20 per month (20 cents for bikeways)	$1.42 per month
Billing Mechanism	Monthly utility bill	Monthly utility bill	Monthly utility bill	Monthly utility bill	Monthly utility bill
Individual Billed	Owner or occupant (owner is liable)	Owner or occupant (owner is liable)	Owner or occupant (jointly liable)	Owner or occupant	Owner or occupant (owner is liable)
Enforcement Mechanism	Discontinuance of utilities Creation of lien	Creation of lien Legal action	Discontinuance of utilities Creation of lien	Discontinuance of utilities Legal action	Discontinuance of utilities Creation of lien
Discounts (upon application)	Low-income elderly	Low-income elderly	Low-income elderly	Low-income elderly	Vacant structures with water service

Fee Characteristics	Beaumont Texas	Soap Lake Washington	Austin Texas	Medford Oregon
Name	Street Use Service Fee	Street Utility Charge	Transportation User Fee	Street Utility Fee
Status	Implemented 1990	Implemented 1992	Implemented 1992	Implemented 1992
Fee Basis	Flat fee	Flat fee for residential uses Employees for non-residential uses	Trip generation	Trip generation
Use of Funds	Road maintenance	Road construction Road maintenance	Road maintenance	Road maintenance
Land-Use Categories	Single-family Multifamily 1 Non-residential category	1 Residential category 1 Non-residential category	4 Single-family categories 4 Multifamily categories 31 Non-residential categories	Single-family Multifamily Senior housing 16 Non-residential categories
Exemptions	Undeveloped land Disabled persons Occupants of public housing or federally subsidized low-income rental housing	Undeveloped land Tax-exempt properties Low-income elderly Low-income disabled	Undeveloped land Public properties Vacant residences Autoless households Elderly	Undeveloped land Developed land without water service
Fee per Single Family Unit	$3.00 per month	$2.00 per month	$1.70 per month	$2.00 per month
Billing Mechanism	Monthly utility bill	Monthly utility bill	Monthly utility bill	Monthly utility bill
Individual Billed	Owner or occupant	Owner or occupant	Owner or occupant	Owner or occupant
Enforcement Mechanism		Creation of lien Legal action	Discontinuance of utilities Legal action	Discontinuance of utilities Legal action
Discounts (upon application)	Elderly living in single-family units		Properties which generate less than the assigned traffic level	

TABLE 2 Eligible Uses of Funds

	Capital Improvements	Operation & Maintenance
Needs of New Development	Impact Fees	N/A
Needs of Existing Development	Utility Fees	Utility Fees

Economic Efficiency

If road users could be charged for the marginal costs of their trips, including delay imposed on other road users and environmental damage, then trips valued at less than their social costs would not be made. Economic theory tells us such charges would maximize net benefits to society.

The problem is that variable pricing requires that road use be metered. Absent metering, a TUF must appear as a fixed charge to the individual user. A TUF neither moderates travel demand nor encourages the use of more economically efficient modes of travel. It provides no demand signals that can be used in service planning.

Even so, the political process of fee setting may lead to a more efficient allocation of resources than is likely when roads compete for general tax revenues (as they do with property taxes). And a TUF appears more politically feasible at this point than the theoretically elegant road pricing schemes advocated by economists.

WHAT IS IN A NAME?

The name given a revenue source does not determine whether it is a tax, assessment, or fee. What counts are the characteristics of the source. The object in structuring a TUF is to make it as much like a user fee and as little like a tax or special assessment as possible.

Legal guidance is provided by the Supreme Court of Colorado, which upheld Fort Collins's TUF as a special fee (*Bloom v. City of Fort Collins*, 784 P.2d 304); the Supreme Court of Idaho, which struck down Pocatello's TUF as a disguised tax (*Brewster v. City of Pocatello*, 768 P.2d 765); the Attorney General of Oregon, who opined that Ashland's TUF is a property tax (Oregon Attorney General Opinion OP-6091, June 26, 1987); and the courts and legislatures of other states, which have drawn the line between user fees, taxes, and special assessments through their case law and statutes.

Taxes, Assessments, and Fees

In Florida as in most states, localities may levy taxes only if specifically authorized by state law. A TUF would be illegal if judged to be a tax.

Odds are that a TUF would also be illegal if judged to be a special assessment. Special assessments may be levied only on properties that realize some "special benefit" from expenditures. Traditionally, this meant that properties had to physically abut improvements financed with assessments. Such improvements tended to be limited in scale and scope (e.g., subdivision street paving and street lighting).

Increasingly, special assessments are being used to finance services and facilities beneficial to larger areas and a broader public. The courts in Florida are among the most receptive to such applications. Yet even in Florida, a special assessment for citywide road repaving was recently declared invalid because the benefits conferred by the program were not "special" enough (*Hanna v. City of Palm Bay*, 579 So.2d 320).

Only if a TUF is judged to be a user fee will it likely be upheld. "Home rule" states grant localities all powers that are not precluded or preempted by the state constitution or state statutes. As a rule, the power to impose user fees for public purposes is not preempted; quite the opposite, state statutes specifically grant localities such power under the mantle of home rule.

Costs Occasioned

To qualify as user fees, government charges must be reasonably related to use of public facilities or services. From a legal standpoint, the safest basis for TUFs is costs occasioned by the use of roads. By this I mean costs incurred by government in meeting the needs of specific classes of road users.

Cost occasioned is the most common basis for allocating highway costs among vehicle classes and establishing corresponding fuel tax rates. Cost of service, a concept analogous to cost occasioned, is the most common basis for utility rate setting and one generally acknowledged to meet the "fair and reasonable" standard applied to utility rates. Cost occasioned is the basis for transportation impact fees in Florida and other states adopting the rational nexus standard. And cost of service was the basis for a favorable ruling in the only case to date upholding a TUF, the Fort Collins case. Quoting the Colorado Supreme Court:

> We are thus satisfied that where, as here, a municipality imposes a special fee upon owners or occupants of developed lots fronting city streets for the purpose of providing revenues for the maintenance of city streets, and where the fee is reasonably designed to *defray the cost of the service* provided by the municipality, such fee is a valid form of government charge within the legislative authority of the municipality. [*Bloom v. City of Fort Collins*, 784 P.2d 304, 311 (emphasis added)].

"Reasonableness" Standard

The fact that user fees must be reasonably related to the use of facilities or services does not mean that fees must be based on precise estimates of use or costs occasioned by use. One challenging the reasonableness of a fee must overcome a presumption of validity by demonstrating that the fee structure was established arbitrarily.

In the Fort Collins case, the court stated:

> The amount of a special fee must be reasonably related to the overall cost of the service. . . . Mathematical exactitude, however, is not required, and the particular mode adopted by a city in assessing the fee is generally a matter of legislative discretion. (*Bloom v. City of Fort Collins*, 784 P.2d 304, 308)

The same standard has been applied by courts in other states. In Florida, for example, a court held that "mathematical exactitude" is not required in setting trash collection fees to exactly reflect the cost of service (*Pinellas Apartment Association, Inc. v. City of Petersburg*, 294 So.2d 676, 678).

WHOM TO CHARGE, AND HOW MUCH

In establishing a fee structure, it is first necessary to decide which roadway costs are occasioned by feepayers, then to decide which classes of feepayers occasion specific costs, and finally to decide how to divide such costs among individual feepayers within the classes occasioning them.

Existing Versus Future Traffic

If a TUF is truly a fee, feepayers may be charged only for the costs they themselves occasion. With only two exceptions (La Grande, Oregon, and Soap Lake, Washington), localities implementing TUFs to date have utilized funds strictly for operation and maintenance of roads. Clearly, such costs are occasioned by existing users because they are recurring costs that cannot be avoided if facilities are open to traffic.

That is not the case with capital improvements. Some capital improvements meet the needs of new rather than existing development; the costs of these improvements could be avoided in the absence of growth.

In setting transportation impact fees under the court-imposed rational nexus standard, new construction costs must be divided into two pots: costs incurred to eliminate any backlog of road needs at current traffic volumes and costs incurred to accommodate new trips caused by anticipated development. Only the latter may be funded with impact fees.

> The cost of new facilities should be borne by new users to the extent new use requires new facilities, but only to that extent. When new facilities must be built in any event, looking only to new users for necessary capital gives old users a windfall at the expense of new users. (*Contractors and Builders Association of Pinellas County v. City of Dunedin*, 329 So.2d 314, 321)

Conversely, only the costs of eliminating existing deficiencies could be funded with a TUF because new users would otherwise benefit from a "windfall" at the expense of old users. If a road user tax were being imposed, the proceeds could be used for any road-related purpose. However, a TUF is a fee and hence subject to a different standard. The connection between who pays and who benefits is stronger for a fee than for a tax.

Local Versus Through Traffic

Being an open system, roads serve not only those who reside or do business in the community but anyone who might wish to travel through the community. Both groups occasion costs through road use. Yet, again, feepayers may be charged only for their fair share of total costs.

The issue of road use by nonfeepayers loomed large in the Pocatello case. In ruling against Pocatello's TUF, the Idaho District Court held:

> Accepting the legal definition that a fee is a charge for a direct public service, while a tax is a forced contribution by the public at large to meet public needs, the City cannot charge and collect fees for the restoration and maintenance of public streets . . . fees collected from the owners/occupiers of the individual premises go to support a proprietary service that is *shared by the public at large*. [*Brewster v. City of Pocatello*, District Court, County of Bannock, State of Idaho, Register #39971-A, p. 11 (emphasis added)]

Similar concerns surfaced when impact fees were first used to finance infrastructure available to the general public. In a landmark case, feepayers argued that since anyone could use roads financed with impact fees, there was too wide a gap between those who paid and those who benefited, making the fee, in reality, a tax. The court disagreed, upholding impact fees as long as they do not exceed the costs of improvements required by new developments (*Home Builders v. Board of County Commissioners of Palm Beach County*, 446 So.2d 140).

By analogy, any benefit accruing to the public generally should not invalidate a TUF as long as fees meet the "fair-share" test. The balance of road costs, occasioned by those driving through the community, must be covered by other revenues.

The practical significance of this issue may be nil. It is likely that the great majority of those using city streets have at least one trip end within the city; their trips are generated (produced or attracted) by properties subject to the TUF. Those traveling through the city pay no fee, but in most communities, through traffic is confined to higher-level roads under county or state jurisdiction.

Arterials Versus Local Roads

In jurisdictions with TUFs, high-volume retailers have argued with some success that it would be unfair to make fees a straight function of traffic generation. On the basis of rates in the ITE *Trip Generation* manual, a convenience store or fast food restaurant could pay thousands of dollars annually in fees; a shopping center, tens of thousands. Yet, as these retailers know, the lion's share of local road expenditures only indirectly benefit them since they, by necessity, locate on high-volume county or state roads.

Jurisdictions have responded to such arguments in different ways. Austin, Texas, has capped its "traffic generation factor" at five times the residential rate; food stores and health care facilities pay the same amount per acre of development, although food stores generate almost four times as many trips per acre. Pocatello assigned 79 percent of its costs to residential users, the balance to other users; residential properties ended up paying 13 times as much per "average weekday vehicle trip end" as did nonresidential properties.

An arbitrary cap or allocation may solve a political problem but also jeopardizes the status of a TUF as a fee reasonably related to use of city streets. There is a better way to deal with this issue.

Arterials, collectors, and local roads perform different functions in the road hierarchy, and thus, occasion costs differently. Because the primary function of local roads is to provide access to land, costs of such roads are occasioned entirely by fronting properties. On the other hand, since the primary function of arterials is to provide mobility, their costs are occasioned by all developed properties, whether they front on a jurisdiction's roads or not. Collectors fall somewhere between the two extremes.

In Port Orange's ordinance, we distinguish among roads in different functional classes and allocate their costs separately, still on the basis of trip generation rates. This approach poses less legal risk than arbitrary caps or allocations, yet it results in moderate fees for nonresidential uses located on arterials or collectors.

Trip Generation Estimates

Two of the nine jurisdictions surveyed, La Grande, Oregon, and Beaumont, Texas, have flat fee schedules independent of road use; their TUFs almost certainly would not meet the legal standard for user fees (if ever challenged). Two other jurisdictions peg fees to surrogates for road use: the number of required off-street parking spaces in Ashland, Oregon, and the number of full-time workers employed by businesses in Soap Lake, Washington. Such fees might or might not pass legal muster.

Most jurisdictions have chosen to base fees on the number of trips generated by developed properties—a measure of road use, albeit a crude one. Fee schedules have typically grouped ITE land use codes into broad categories and then applied average ITE rates to all land uses in a given category. The categories often are those already used to classify ratepayers in the city's utility billing system.

This can lead to inequities. Tualatin, Oregon, has defined six nonresidential groups (plus a catchall group) and applied the average trip generation rate for an entire group to all uses that fall within a broad rate band. Thus, a strip shopping center of 10,000 to 20,000 ft^2, falling in Group 4, pays \$11.08/1,000 ft^2/month, whereas a center of 5,000 to 10,000 ft^2, falling in Group 5, pays \$29.51/1,000 ft^2/month. The latter generates only 27 percent more trips per 1,000 ft^2, according to the ITE manual, yet pays almost three times as much per 1,000 ft^2 of building area.

Inequities of this sort can wreak political havoc. Wholesalers in Medford, Oregon, refused to pay the city's new TUF, contending that they were unfairly lumped into the "retail/commercial" land use category with retailers that generate many more trips. Just weeks after its fee went into effect, Medford was forced to add new land use categories to its fee schedule.

The more precisely trip generation rates can be estimated, preferably with supporting traffic counts, the better fees will fare from a legal standpoint. The Oregon Attorney General declared a TUF proposed by the city of Monmouth to be a tax on property because it was not based on actual use of roads:

> The City of Monmouth has labeled its proposed exaction a "street user fee." That fee, however, does not relate to actual automobile trips generated by individual pieces of property. Rather, the city bases its fee categories upon estimates of trips generated by typical properties in each category of use. . . . Because the proposed Monmouth fee is not based on actual use, it is not a user fee. (Oregon Attorney General Opinion, OP-6091, June 29, 1987, p. 3)

In Port Orange, all residential and commercial land uses are classified as precisely the ITE manual permits; each property is assigned to the closest ITE land use category (if ITE has anything close), and the trip generation rate specific to that category is then applied to the property. In addition, traffic counts are performed to establish trip generation rates for properties that do not fall neatly into any ITE land use category and to validate or refine ITE rates for properties that do fall into ITE categories. Eventually, all nonresidential properties will be subject to counts, and fees will be adjusted accordingly. The relatively small number of nonresidential properties makes blanket counts possible.

Refinements

Even in jurisdictions using trip generation as the basis for fee setting, fees to date have not reflected road use or costs occasioned by use to the fullest extent possible. A trip-based fee schedule can be refined as follows:

- Households with varying trip generation rates can be distinguished. The ITE manual provides trip adjustment factors for households by size, automobile ownership, and density. Austin waives its TUF for automobileless households; Port Orange provides a discount corresponding to the ITE adjustment factor. An application process can be used to simplify the administration of a fee waiver or special fee.
- Trip generation estimates can be adjusted for pass-by trips that are attracted to commercial properties as intermediate stops on the way to primary destinations and for internal trips within mixed-use developments that never venture onto the public road system. Adjustments for pass-by and internal trips have become standard procedure in traffic impact studies and transportation impact fee calculations. Of the five jurisdictions surveyed with trip-based fees, only Medford includes an adjustment for pass-by trips. Port Orange adjusts for both pass-by and internal trips, on the basis of studies reported in the ITE manual.
- The volume of truck traffic generated by different land uses can be adjusted for. In terms of wear and tear on pavement, heavily loaded trucks are equivalent to hundreds or even thousands of automobiles. Accordingly, highway cost allocation studies have used equivalent single-axle loads (ESALs) as a basis for allocating pavement rehabilitation and main-

tenance costs. Some adjustment is necessary whenever trucks represent a significant portion of traffic on a jurisdiction's roads. Happily, they do not in Port Orange.

• Trip generation rates can be multiplied by average trip lengths to arrive at vehicle miles of travel (VMT) generated by developed properties. On its face, VMT is a better measure of local road use than is trip ends. However, the practical problems of estimating average trip lengths for a multitude of land uses and breaking out travel on city streets may preclude this refinement. It was judged to do so in Port Orange.

POLICIES AND PROCEDURES

Along with a fee schedule, operating policies and procedures must be established. Many have legal implications.

Table 3 summarizes relevant differences between user fees, taxes, and special assessments. Port Orange is careful to treat its TUF as a user fee in all respects.

Tax-Exempt Properties

All developed properties, whether tax exempt or not, should pay TUFs if they are truly user fees. This includes properties owned by state and federal agencies and school districts, as well as a local government's own properties.

Local governments can expect challenges from other levels of government contending that a TUF is a disguised tax or special assessment from which they are exempt. The Comptroller General has authorized federal agencies to pay charges for use of public utilities but would almost certainly object to roads being treated as a utility.

> The furnishing by a State or local government of a *quantum* of direct utility type services, such a sewer, water, trash, etc., to the Federal Government, for which payment may be authorized, is to be distinguished from the performance by a State or local government of governmental functions, such as police and fire protection, regulation of traffic and *road construction and maintenance*, generally supported by taxes, for which payment by a Federal agency, absent statutory authority, is not permissible. [51 Comp. Gen., p. 137 (emphasis added)]

Anticipating a challenge from a public agency, a locality would want to base its TUF on an actual traffic count at a public agency's property line. This would ensure that the fee, at least, reflects the actual "quantum" of road use. The first traffic counts in Port Orange were at public schools.

Vacant Land and Unoccupied Structures

Property taxes are levied on vacant land and unoccupied structures. Special assessments are also levied on vacant land and unoccupied structures to the extent that their value is increased by public improvements.

However, as a user fee, a TUF is premised on the use of facilities or services. A TUF must be based on costs occasioned by use, benefits derived from use, or other use-related criteria. Since vacant land and unoccupied structures do not generate trips, it would be inconsistent to impose a TUF on them.

A fee waiver can be granted to owners of unoccupied structures either by means of an application process or on the basis of negligible water use. Tualatin used applications originally but, to simplify administration, switched to automatic waivers based on water use. The Port Orange ordinance provides for a discount, upon application.

TABLE 3 Characteristics of Revenue Sources Under Florida Law

	User Fee	Property Tax	Special Assessment
Basis for Charge	Use	Property value	Special benefits
Who Pays?	Property owners or occupants	Property owners	Property owners
Revenues Necessarily Spent Where Collected?	NO	NO	YES
Public Property Exempt?	NO	YES	YES, unless statutes authorize payments
Nonprofit Property Exempt?	NO	YES, if used for charitable, religious, literary or scientific purposes	NO
Vacant Land/Vacant Structures Exempt?	YES	NO	NO
Billing Mechanism	Utility bill or separate bill	Property tax bill	Property tax bill or separate bill
Enforcement Mechanism	Discontinuance of utility services	Tax deed procedure	Lien foreclosure or tax deed procedure

Characterizations of user fees and special assessments are based upon: Attorney General Opinion (AGO) 076-137; AGO 90-39; AGO 90-47; AGO 91-27; 49 Comp. Gen. 72 (1969); 48 Fla.Jur.2d *Special Assessments*.

Owners and Occupants

Whenever practical, TUF ordinances should impose fees on occupants of rental property. It is the occupants of property that use the road system. The billing of nonowner occupants was a factor in the court decision upholding Fort Collins's TUF as a special fee rather than a tax.

Still, it was only one of many factors persuading the Colorado Supreme Court to rule as it did. Moreover, the court seemed unperturbed by the fact that property owners were ultimately liable for unpaid fees in Fort Collins. Thus, in the interest of fee collection and enforcement, it may be advantageous to make property owners contingently liable for fees if renters fail to pay them, even at the risk of affecting the status of a TUF. Port Orange and most other localities have done so.

TUF Collection

A TUF ordinarily will be collected as part of a unified municipal utility bill. In Florida and many other states, the property tax bill is limited to assessments and ad valorem taxes. In these states, any locality collecting a TUF via the property tax bill is inadvertently compromising the TUF's status as a user fee.

Orlando's storm water utility fee was judged by the Florida Attorney General to be a special assessment rather than a user fee. His opinion was based, in part, on Orlando's use of the property tax bill to collect the "fee" (Attorney General Opinion 90-47).

Even in states without such restrictions, a locality may create a presumption that the TUF is a fee for service by adding it to the unified utility bill along with water and sewer fees, trash collection fees, and storm water utility fees. Making the TUF part of the monthly utility bill could also help defuse public opposition. In explaining public preferences for the sales tax over the property tax, commentators often note that sales taxes are collected in small amounts with daily purchases whereas the property tax is collected annually in one lump sum. Port Orange's TUF will be added to the city's monthly utility bill (once collection begins).

Enforcement

Payment of fees should be enforced primarily by cutting off other public utility services billed with the TUF. Given a unified utility bill, partial payments may be applied either to the TUF before water and sewer charges or to all utility charges proportionally. In either case, the water payment will come up short, and water service may be discontinued as a method of enforcing payment of other utility fees.

Discontinuance of services is the most effective means of enforcing payment of delinquent charges; however, it is not immune to challenge. It has long been accepted that a public utility may not refuse to render service for some collateral matter unrelated to service. Accordingly, courts in Florida and certain other states have adopted an "interlocking" test to decide whether one utility service can be cut off for nonpayment of service charges due another utility (*State v. City of Miami*, 27 So.2d 118).

Water service and road maintenance are not sufficiently interrelated to meet the interlocking test. Even so, it may be possible to tie the two together through a unified utility bill. As local governments in other states have added new service charges to their utility bills, a new legal test has emerged: Does the legislative scheme that permits unified billing and termination of any or all such services for failure to make payment in full bear a reasonable relation to the goal of public health protection? If the answer is yes, unified billing has been upheld (*Perez v. City of San Bruno*, 616 P.2d 1287). Port Orange will discontinue water service for failure to pay the TUF.

Appeals

To ensure that due process requirements are met, a TUF ordinance should provide an avenue for appeal of fee calculations. In Tualatin, road users may appeal the land-use classification or square footage assigned to a property in the fee calculation. In Fort Collins, users were able to appeal in special cases or cases of error in the fee calculation. Port Orange has provided for various appeals.

The calculation of a TUF is analogous to the calculation of a road impact fee. Under applicable case law, property owners are entitled to independently calculate road impact fees on the basis of their own trip generation studies (*Home Builders v. Board of County Commissioners of Palm Beach County*, 446 So.2d 140, 145).

Parenthetically, only users subject to variable (parcel-specific) fees need be given the option of conducting independent fee studies. For users subject to flat fees, appeals may be limited to the issue of appropriate land use classification because the fee is based on average trip-making characteristics for an entire class of users.

Dedication of Funds

TUF funds should be dedicated to road-related purposes, without exception. While upholding Fort Collins's TUF ordinance in the main, the Colorado Supreme Court struck down a severable provision authorizing the city council to transfer excess revenues not required for road maintenance to any other fund of the city. If transferred, the fee would no longer bear a "reasonable relationship" to services rendered by the city, for the nexus between who pays and who benefits would be broken. The fee would become the "functional equivalent of a tax" (*Bloom v. City of Fort Collins*, 784 P.2d 304, 311).

Even granting that TUF funds may be used only for road-related purposes, the courts will likely hold that a portion of a city's general administrative expenses are "road related" and thus eligible for TUF funding. Even as it struck down Fort Collins's "transfer provision," the Colorado Supreme Court made it clear that the ruling applied only to the use of excess revenues for general governmental expenses "unrelated to the maintenance of city streets." The use of funds to defray an equitable share of overhead was left intact. Port

Orange and most other localities provide for recovery of administrative expenses.

CONCLUSION

Under applicable law, a TUF must be reasonably related to the use of roadways or to costs occasioned by such use. A reasonable basis for fee setting is a property's trip generation rate, as reported in the ITE manual. Local traffic counts should be conducted in cases in which the ITE manual has no suitable land use code and as a means of validating or refining ITE rates generally.

Road costs should be allocated separately for arterials, collectors, and local roads, because they perform different functions and occasion costs differently. To more accurately reflect costs occasioned, adjustments may be made for household automobile ownership, pass-by and internal trips, truck traffic, and differential trip lengths.

If they are truly user fees rather than taxes or assessments, TUFs should be imposed on governmental and other tax-exempt traffic generators and on occupants of rental property. By the same token, they should not be imposed on vacant land and unoccupied structures because these land uses are not significant trip generators. Fees should be collected with other utility charges through a unified billing system and should be enforced by discontinuing other utility services for failure to pay the utility bill in full. Funds must be spent for the benefit of feepayers and may not be diverted to nonroad purposes.

These and other elements have been incorporated into a TUF ordinance for Port Orange, Florida. The city adopted the ordinance in May 1992, becoming the nation's 10th locality to do so, and is proceeding to test the legality of the TUF by validating bonds to be repaid with fee proceeds. If a TUF is upheld in Florida, it will bode well for localities in other strong home rule states that might wish to tap this new funding source.

ACKNOWLEDGMENTS

Research for this paper was completed while the author was a visiting faculty member at the Center for Urban Transportation Research, University of South Florida. A national search for localities with TUFs was conducted by Lori Burns, a research assistant at the center. Leagues of Cities in all 48 continental states were contacted.

REFERENCES

1. C. K. Zorn. User Charges and Fees. In *Local Government Finance: Concepts and Practices* (J. E. Peterson and D. R. Strachota, eds.). Government Finance Officers Association, Chicago, 1991.
2. J. L. Mikesell. *Fiscal Administration: Analysis and Applications for the Public Sector*. Brookes/Cole Publishing Co., Pacific Grove, Calif., 1991.

Publication of this paper sponsored by Committee on Local Transportation Finance.

PART 2

Economic Analysis

Inferring Variations in Values of Time from Toll Route Diversion Behavior

TERJE TRETVIK

The use of tolls to finance new road infrastructure has become widespread in many countries, and at many locations around the world the introduction of road pricing is being considered. The question of what values of time (VOT) should be used in toll road and road pricing studies has thus become increasingly important. Evidence is presented from a recent toll road diversion study, in which behavioral data were collected before and after a substantial price increase. The main emphasis is on the variations in willingness to pay for small time savings. Interesting changes over time in the relationships between background factors, and the difference between stated and measured time savings, are also discussed. VOT increased nearly linearly, but less than proportionately, with mean income. In general, short or frequent trips showed lower values than long or infrequent ones. Most drivers revealed a 2 to 4 percent increase in VOT corresponding to a 10 percent increase in their gross personal income. VOT was found to be a smaller percentage of the drivers' gross hourly wage rate as income increased. The ranges were 100 to 50 percent for commuting trips, 165 to 90 percent for business trips, and 175 to 75 percent for other trips.

The adoption of road user tolls to help finance interconnected or single sections of high-quality road infrastructure is common in many countries. The basic motivation is that government investment budgets are too tight and that users are willing to pay for time savings and better driving conditions. When there are alternative routes, however, the toll rates have to be set right for projects of this kind to be successful.

In the first section of this paper the general situation concerning the use of road tolls in Norway is outlined. Subsequently, detailed results from a toll road diversion study are given. Descriptive results about variations in toll road usage with background factors and new evidence concerning the relationships between stated and measured time savings are presented.

The last section addresses the question of modeling route choice under the influence of tolls. Logit models based on pooled data from 2 consecutive years are developed, and these are applied to study variations in values of time.

CHARGING FOR USE OF ROAD SPACE IN NORWAY

In Norway there is a long tradition of financing sections of road infrastructure, especially bridges and tunnels, by combining road user tolls and public funds. Most of the projects have no free-of-charge competitive routes in terms of distance or travel time. This is because the tolled sections either replace existing ferry crossings or establish new links in the road network. In times of steady traffic growth, the creation of enough revenues to defend the private sector involvement usually went according to plan. Often these types of projects generated trips exceeding the overall growth in traffic, and the charging period could in these instances be shortened.

Figure 1 shows the recent trend of increasing the private sector share of investments in national (state) highways. For 1991 the contribution from toll companies was expected to be 1.8 billion kroner, which is about one-third of the total investments.

One explanation for this large increase in private sector involvement is the introduction of the urban toll rings in Bergen (January 1986), Oslo (February 1990), and Trondheim (October 1991). The original political agreement was to raise extra private sector money, to be matched by extra government money, to fulfill urban road building programs in a much shorter time than was previously possible. The contents of the investment packages and the design of the schemes have, however, changed in line with increasing environmental awareness and developments in technology.

Although the focus of the original argument for the Bergen toll ring was entirely on road building, the emphasis widened to include infrastructure investments for public transport, cyclists, and pedestrians in the Oslo and Trondheim schemes. The Trondheim toll ring is the first scheme to have no monthly or yearly passes that allow an unlimited number of crossings. Tolls are charged per vehicle Mondays through Fridays from 6:00 a.m. to 5:00 p.m. for all inbound traffic. Charge levels during the morning peak are higher than they are later in the day, which indicates that the payment scheme is not entirely fiscal. It is also designed to influence car drivers' choice of mode and departure time.

As a result of a liberal credit policy and no government control on the issue of bonds, economically more marginal toll projects have been started. Some have even been financed entirely by borrowing. Others have been built in areas where competitive (old) free-of-charge routes existed.

The question of traffic diversion from the new route has thus become important. Environmental objectives of the new projects may not be met, and toll companies risk running into financial difficulties. This is exactly the case for the project that we now turn our attention to.

TOLL ROAD STUDY

The first tolled section was opened in 1988 on the E6 national highway route east of Trondheim in the direction of the air-

SINTEF Transport Engineering, N-7034 Trondheim, Norway.

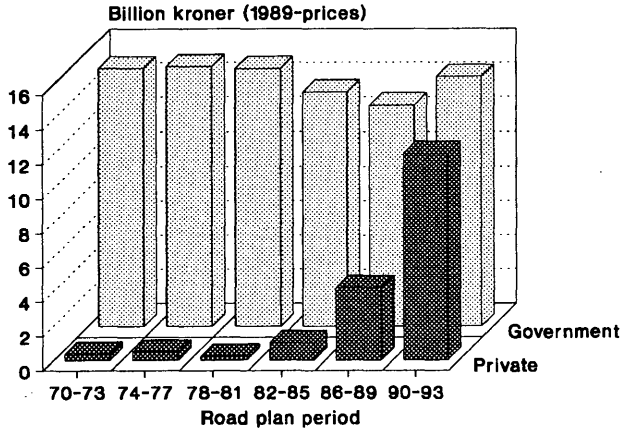


FIGURE 1 Private and government investments in national highways (*1*).

port. The motivation for building a new road was to divert through traffic from the heavily built-up area of the old route for environmental and traffic safety reasons and to provide a faster connection between the city and the airport.

The toll project has since been in operation 24 hr/day, and drivers passing through the toll plaza located at the periphery of the city have to pay in both directions. The charge was 10 kroner (early January 1993: 1 U.S. dollar was equivalent to 7 kroner) for light vehicles and 25 kroner for heavy vehicles in 1988 and 1989. This was increased to 20 kroner and 40 kroner in 1990, in conjunction with the latest extension of the tolled route.

Choice Situations

A special feature of the payment scheme is that drivers can deposit money in their personal toll accounts and pass through the toll plaza without any delay, being identified as bona fide account holders by the identity of their personal electronic tag mounted inside the windscreen. Tag holders can also be billed via their bank at the end of each month according to actual use by direct debiting.

In 1989 a second section was opened, and the toll company offered motorists 12.5 km of motorway driving conditions. The old route had a much lower standard and passed through built-up areas with several local speed limits of 50, 60, or 70 km/hr. Its length was roughly equal to that of the new route, and it was available free of charge. Choosing the old route in the direction of the city during the busiest time of the morning peak also implied the risk of some queuing.

During 1990 the new motorway was lengthened by 7.5 km, thereby presenting long-distance traffic with the choice of "buying" larger time savings than in 1989, but at a higher price. The old route was still similar in length to the motorway route for long-distance traffic and was available free of charge.

Interview surveys were conducted on users of both routes in November 1989 and November 1990, and average driving times between key origins and destinations were measured. To cover most trip purposes during a week, questionnaires relating to the drivers' current trip were handed out at certain times during 3 consecutive days in both years (Sunday, Monday, and Tuesday).

Total average daily traffic on the two routes passing the cross section where the toll plaza is situated was around 18,000 vehicles in both interview periods. Two-thirds of the returned forms came from choosers, that is, time-versus-money traders, in the sense that the tolled route represented the shortest (measured) time route, given the drivers' own statements about origin and destination.

Time Savings and Costs for Light Vehicles

Time savings depended on the drivers' origins and destinations and on whether it was a trip during the morning peak toward Trondheim. A small time delay was imposed on drivers who did not possess a tag because of time that was, or would have been, spent in money transactions at the toll plaza.

Figure 2 shows that the number of minutes to be gained by choosing the tolled section was quite modest. For the choosers represented in the samples, the average time savings increased from 4.4 min in 1989 to 6.8 min in 1990. The mode of the distributions increased from 5 to 8 min.

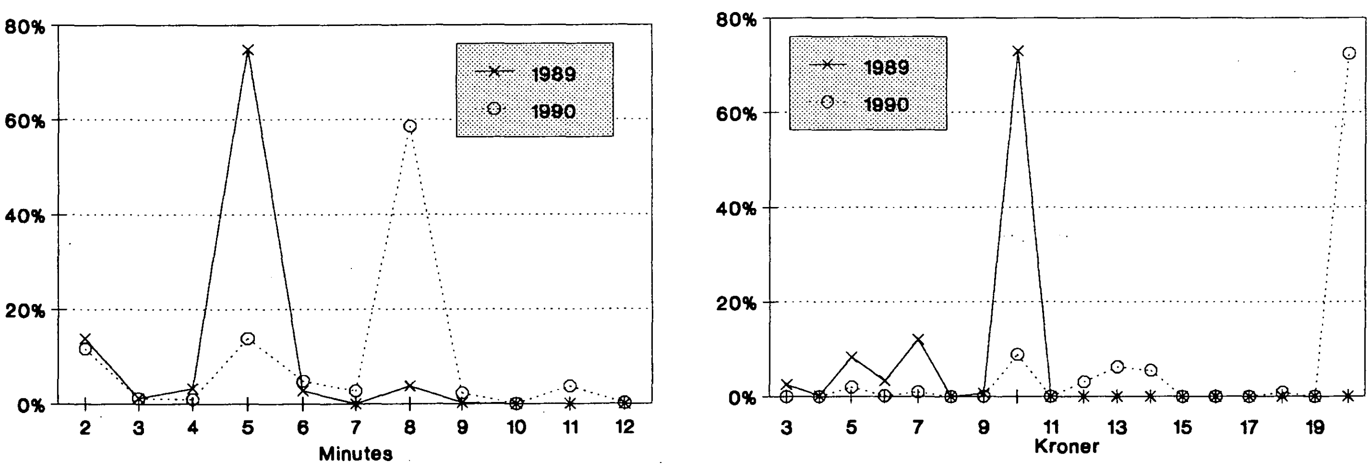


FIGURE 2 Distribution of actual time savings (*left*) and actual costs (*right*), rounded to whole numbers.

The mean costs were 8.85 kroner in 1989 and 17.63 kroner in 1990. Drivers with no tags had to pay the full price of 10 kroner in 1989 and 20 kroner in 1990. Slightly less than 30 percent of drivers in both years possessed a tag. These had variations in their cost per trip, depending on how many trips they had prebought. For instance, in 1990 the price per trip was reduced to 18, 16, 14, 12, or 10 kroner if the number of trips bought in advance was 25, 50, 100, 250, or 500, respectively.

If the driver stated on the questionnaire that others had, or would have, contributed to the payment (e.g., cost sharing with passengers or some kind of company car arrangement), the cost variable was reduced by 50 percent for nonbusiness travel purposes. The rationale for doing this was that company car usage for private purposes is taxed in Norway, so the marginal cost of a private trip is never zero. If the toll was paid by the employer and it was a business trip, the actual cost was not reduced, since for this trip purpose it is as much the employer's willingness to pay for time savings that is revealed.

It is suspected that the possession of a tag and the size of the rebate per trip for tag owners were related to income. The correlations between the final cost variable and gross personal income had the expected signs but were modest in size: −0.189 in 1989 and −0.214 in 1990. For the subsamples having a tag, the correlations were −0.162 in 1989 and −0.116 in 1990.

Choices

Figure 3 shows that usage of the tolled section dropped from 54 percent in 1989 to 40 percent in 1990. Groups labeled as "Commuting" and "Other" reacted more sharply to the price increases than did "Business." In general, drivers' reactions reveal that they did not find the extra time savings worth the double price.

In Table 1 usage of the tolled section by market segments is shown, together with the percentage distribution for each variable in brackets. The choices show a very clear pattern. First, in both years there is increased usage with (a) increased income, (b) lower frequency of traveling the section, (c) others paying, (d) owning a tag, and (e) increased length of the journey. Second, in each cell, usage is down in 1990 compared with that in 1989. It is evident from these results that travel-related factors, as well as income and details in connection with money transactions, play key roles in determining travelers' choice of route. We return to this point in another section.

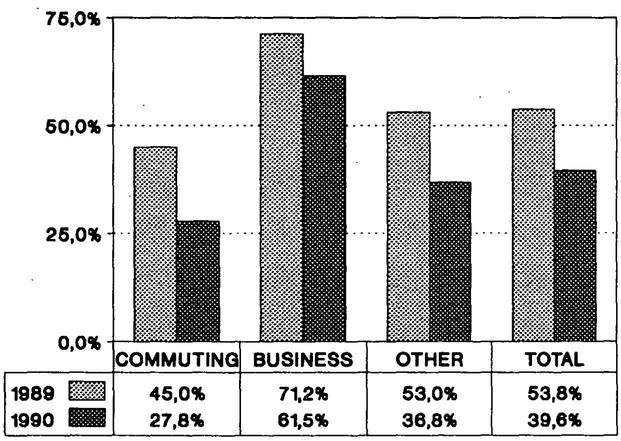


FIGURE 3 Usage of the tolled section in 1989 and 1990.

TABLE 1 Usage by Market Segments (Percentage of Observations in Brackets)

VARIABLE	1989	1990
INCOME GROUP (KRONER/YEAR)		
0 - 100 000	43.4 (16.4)	23.9 (15.9)
101 - 150 000	46.3 (19.7)	34.3 (18.0)
151 - 200 000	53.7 (33.9)	37.0 (32.3)
201 - 250 000	60.8 (15.9)	45.2 (16.9)
251 - 300 000	67.4 (7.6)	53.9 (9.1)
> 300 000	70.3 (6.5)	64.4 (7.8)
FREQUENCY OF CHOICE SITUATION		
Daily	46.1 (56.8)	30.7 (52.3)
Weekly	61.6 (18.7)	43.2 (19.0)
Monthly	69.1 (14.2)	51.3 (15.1)
More seldom	60.1 (10.4)	54.0 (13.6)
WHO PAYS THE TOLL?		
Car driver alone	47.2 (70.7)	29.8 (71.5)
Others, partly or completely	72.5 (29.3)	68.3 (28.5)
WAY OF PAYING		
Cash	44.0 (72.0)	29.4 (72.1)
Tag	80.6 (28.0)	67.4 (27.9)
TRIP LENGTH		
Short/local	44.8 (67.2)	30.0 (64.9)
Long	72.8 (32.8)	57.5 (35.1)

The distributions of the background variables were reasonably stable. A small shift toward higher-income classes can be noticed. Mean annual incomes in the samples increased by 3 percent from 183,000 kroner in 1989 to 189,000 kroner in 1990, which was close to the inflation rate that year (4 percent).

Subjective Versus Objective Time Savings

Choosers of both routes were asked to estimate the amount of time savings associated with use of the tolled section. Drivers on the old route were requested to estimate how many minutes of travel time they thought they would have saved if they had chosen the tolled route for their current trip. Drivers on the tolled route were asked how many minutes they thought they had saved by choosing the tolled route for their current trip.

Table 2 shows that all subgroups believed that the savings in travel time earned by choosing the tolled section were greater than they really were, as measured by observers using the car-following method.

In 1989 the average subjective time savings was 6.7 min, compared with the objective value of 4.4 min. This changed to 7.6 min subjectively in 1990, compared with 6.8 min objectively. The overestimation thus improved from +2.3 min to only +0.8 min—in percentage terms from +57 percent to +23 percent. We think that the effect of learning and possibly a more realistic view of the time benefits motivated

TABLE 2 Overestimation of Time Savings on Tolled Section by Subgroups

VARIABLE	1989 (minutes)	1990 (minutes)
FREQUENCY OF CHOICE SITUATION		
Daily/weekly	2.1	0.8
Less frequent	2.5	0.8
WHO PAYS THE TOLL?		
Car driver alone	2.8	1.2
Others, partly or completely	2.1	0.7
WAY OF PAYING		
Cash	2.1	0.7
Tag	2.6	1.1
TRIP LENGTH		
Short/local	2.1	0.9
Long	2.4	0.6
PURPOSE GROUP		
Commuting	2.1	0.7
Business	2.6	1.1
Other	2.2	0.8
CHOSEN ALTERNATIVE		
Old section	1.2	0.3
Tolled section	3.1	1.6

by the steep price increases are the main explanations for this improvement.

Drivers on the tolled section overestimated most seriously in both years, which indicates a sort of selection bias. It is almost surprising that 1990 toll road choosers did not overestimate even more, because of the effect of rationalizing their payment of twice the charge from the previous year.

People who travel the routes often or whose origins or destinations are local are bound to know better the attributes of the choice alternatives, and the results show that their estimates are more accurate. In addition, drivers who pay the toll completely out of their own pockets, or who have made the effort of acquiring a tag, are more likely to exaggerate their time savings. This phenomenon could be taken as evidence of attempts to correct the psychological strain referred to by Festinger (*2*) as cognitive dissonance.

Attempts at finding simple well-fitting linear relationships between objective and subjective time savings were not successful. Correlations (Pearson's R) between these two variables even for subgroups defined by travel purpose and year or choice were in general low. Groups labeled Commuting and Other segmented by either year or actual choice returned the highest coefficients, and these were around $+0.4$.

In Tretvik (*3*) simple binary logit models were estimated from 1989 data and applied in prediction exercises to forecast 1990 usage. Aggregate information about prices ($+100$ percent) and time savings ($+50$ percent) that would have been known in advance of the 1990 situation were used. The utility functions were

$$V_{\text{tolled route}} = \text{constant} + b_1 \times \text{actual cost} + b_2 \times \text{time saved (measured or stated)} \quad (1)$$

$$V_{\text{free route}} = 0 \quad (2)$$

The constants (b_1 and b_2) were estimated separately for each purpose group, and the probability for choosing the tolled route is given by the logit formula

$$P(\text{tolled route}) = 1/[1 + \exp(-V_{\text{tolled route}})] \quad (3)$$

The results showed that 1989 models using a stated time savings had a better fit than did the corresponding models using a measured time savings. This was to be expected because stated values measure people's perceptions much more closely than the more objective engineering values. However, whereas the 1989 objective models performed well in predicting 1990 usage, the 1989 subjective models seriously underestimated the actual 1990 demand. In short, the models did not respond well to the simple assumption that all drivers' perceived values for the time savings changed by $+50$ percent.

This result, and the previous discussion about the biases in people's perceptions about the actual time savings, underlines observations made by Small (*4*) about the problems of using stated values in prediction models. Even if reported values accurately measure the perceptions that determine choice, the resulting models cannot be used for prediction, unless one can predict how a given change will alter those perceptions.

GENERALIZATIONS ON THE VALUES OF TIME

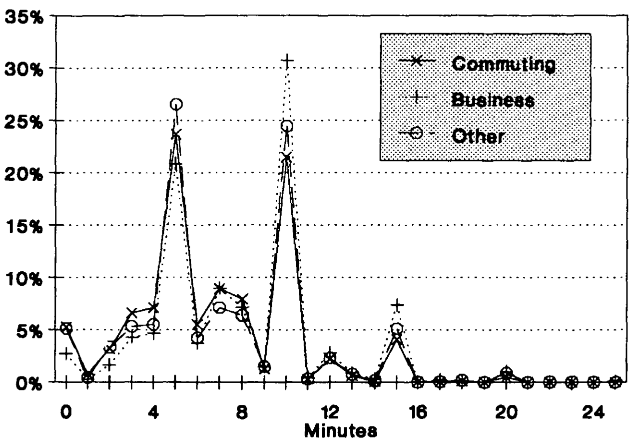
In this section the focus is on variations in the car drivers' willingness to pay for perceived marginal time savings, rather than forecasting future demand. The 1989 and 1990 samples are added together, and we use stated instead of measured time savings as the explanatory variable. Figure 4 shows that the range is wider than that for measured time (Figure 2).

A rounding effect is noticeable in peoples' estimates of the time savings, causing distinct peaks at the values of 5, 10, and 15 min. Notice that small minorities (5 percent) of the Commuting and Other groups have the impression that there are no time savings associated with the tolled route. Figure 4 also shows that drivers with the purpose labeled as Other have the highest propensity for paying the full charges and that drivers with the purpose labeled as Commuting have the lowest propensity for paying the full charges.

The pooled sample allowed the estimation of more complete model specifications. The questionnaire did not contain variables such as sex, household composition, personal occupation, or age group. However, it was possible to take into account the effects of the length of the trip, whether the driver was a frequent traveler in the area, and whether he or she covered the cost privately. The effects of income were modeled by segmentation into six gross personal income groups.

Table 3 gives the results for the whole sample and for each purpose group. All parameters are attached to the utility function of the tolled alternative. The utility is specified in such a way that all parameters having names like "+ Cost if . . ." are additive corrections to the base actual cost parameter. This way of specifying the effects of income group on the cost variable is adopted from the Dutch value of time study (*5*).

Purpose group differences between the base actual cost parameters and between the stated time-savings parameters are modest. Differences in preferences between purpose groups are mainly accounted for by the additive cost parameters. A test statistic for the null hypothesis (H_0) of no taste variations across the purpose group segments can be computed as twice



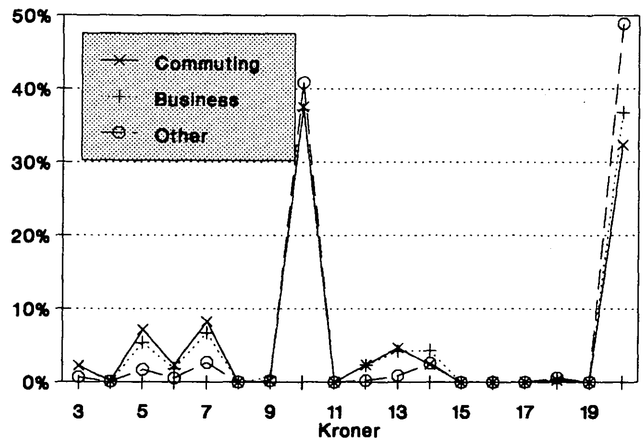


FIGURE 4 Distribution of stated time savings (*left*) and actual costs (*right*), rounded to whole numbers.

the difference in final likelihoods between the whole sample and the sum of likelihoods for the segments (*6*). It is χ^2-distributed with, in this case, $3 \times 11 - 11 = 22$ degrees of freedom. χ_{test} works out at 95.0, compared with $\chi_{22,.01} = 40.3$, so H_0 can be firmly rejected at the 1 percent level.

In Table 4 implied values of time resulting from the estimations are shown. Values for each income segment are calculated, with additive and independent percentage adjustments for the effect of other factors.

The values in the top section of Table 4 thus apply for travelers that are on a short trip, travel in the area often, and pay the toll themselves. These conditions are satisfied by 62 percent of the commuters but only by 11 percent of the drivers on business trips and by 13 percent of the drivers with other purposes. At the bottom of Table 4 average computed values across each sample are given, both per vehicle and per person in the vehicle.

In Norway, officially recommended values of time per person in the vehicle for use in cost-benefit analyses are given as standard percentages of the average wage rate in industry; 35 percent for Commuting trips, 134 percent for Business trips, and 20 percent for Other trips (*7*). This average wage rate was 90 kroner in the first quarter of 1990. The behavioral values revealed in this study, expressed in similar fashion, are

TABLE 3 Pooled Estimation Results (*t*-Values in Brackets)

VARIABLES AND KEY STATISTICS	WHOLE SAMPLE	COMMUTING	BUSINESS	OTHER
Constant	0.4071 (4.6)	0.4870 (3.4)	0.8105 (4.4)	0.1863 (1.3)
Actual cost on the tolled section (base)	-0.2488 (-31.7)	-0.2744 (-21.6)	-0.2452 (-15.5)	-0.2259 (-14.7)
Stated time saving on the tolled section	0.2355 (27.4)	0.2408 (17.4)	0.2424 (13.5)	0.2226 (15.7)
+ Cost if income ≤ 100 000	-0.03432 (-5.7)	-0.02971 (-2.6)	-0.00886 (-0.6)	-0.03146 (-3.7)
+ Cost if income 101 - 150 000	-0.00424 (-0.7)	-0.00177 (-0.2)	0.01167 (1.0)	-0.00866 (-1.0)
+ Cost if income 201 - 250 000	0.01039 (1.7)	0.02355 (2.0)	0.00222 (0.2)	0.00507 (0.5)
+ Cost if income 251 - 300 000	0.03758 (5.0)	0.05795 (4.2)	0.02291 (1.8)	0.03472 (2.8)
+ Cost if income > 300 000	0.03954 (4.8)	0.05921 (3.4)	0.04310 (3.4)	0.01284 (0.9)
+ Cost if on a long distance trip	0.05882 (14.4)	0.06782 (8.3)	0.03688 (4.8)	0.06475 (10.7)
+ Cost if infrequent traveller in the area	0.07129 (16.2)	0.07347 (7.4)	0.05708 (6.7)	0.06825 (6.5)
+ Cost if others contribute to toll payment	0.06420 (13.7)	0.01117 (0.8)	0.06390 (8.0)	0.03959 (3.8)
Sample size	8197	3464	2051	2682
Final likelihood	-4073.6	-1575.2	-997.3	-1453.6
Rho-squared (0)	0.2830	0.3440	0.2985	0.2181

TABLE 4 Values of Time (kroner/hr) per Vehicle (Percentage of Observations in Brackets)

	COMMUTING		BUSINESS		OTHER	
Base values by gross personal income group (kroner/year):						
0 - 100 000	47.51	(18.5%)	57.25[a]	(6.7%)	51.90	(22.4%)
101 - 150 000	52.32[a]	(18.5%)	62.28[a]	(13.2%)	56.94[a]	(21.8%)
151 - 200 000 (base)	52.65	(33.9%)	59.31	(32.6%)	59.12	(29.8%)
201 - 250 000	57.60	(15.2%)	59.85[a]	(21.6%)	60.48[a]	(14.3%)
251 - 300 000	66.75	(8.2%)	65.43	(11.5%)	69.85	(7.2%)
> 300 000	67.14	(5.6%)	71.96	(14.4%)	62.69[a]	(4.6%)
Adjustments for other factors:						
Trip length						
Short/local (base)	...	(76.9%)	...	(57.4%)	...	(59.8%)
Medium/long	+32.8%	(23.1%)	+17.7%	(42.6%)	+40.2%	(40.2%)
Frequency of choice situation						
Daily/weekly (base)	...	(88.1%)	...	(40.1%)	...	(18.1%)
Less frequent	+36.6%	(11.9%)	+30.4%	(59.9%)	+43.3%	(81.9%)
Who pays the toll?						
Car driver alone (base)	...	(87.4%)	...	(31.9%)	...	(89.2%)
Others, partly or completely	+4.2%[a]	(12.6%)	+35.3%	(68.1%)	+21.3%	(10.8%)
Average value across the sample per vehicle	73		138		120	
Average value across the sample per person in the vehicle	52		89		53	

[a] *Estimate not significantly different from base group (|t| < 1.8)*

smaller for Business trips (99 percent), larger for Commuting trips (57 percent), and considerably larger for Other trips (59 percent).

CONCLUSIONS

The data that were available for this study made it possible to establish that systematic variations existed in car drivers' willingness to pay for small time savings, with purpose group, income, and some key characteristics of the journey.

The results concerning private trips confirm a trend found in recent Norwegian studies from which values of time can be deduced, namely significantly higher values for Commuting and Other trips than the ones that are recommended officially. The results are also consistent with a summary of international empirical studies in Small (*4*). He concludes that 50 percent of the gross wage rate is a reasonable average for a journey to work, that business travel seems to have a higher value than commuting travel, but not necessarily equal to the wage rate, and that Other travel may have a value higher or lower than Commuting. He also notes that values on weekends are higher than values on weekdays, and that observation suggests one reason for the relatively high values for Other trips found in this study: because Sunday was one of the three interview days, we have picked up a large share of weekend social and recreation trips.

Figure 5 shows the average values of time per vehicle hour across the samples for mean incomes in the income segments. In the first three boxes, values are segmented by trip length and frequency of the choice situation. Values of time per vehicle hour are seen to vary with income in the range 50 to 200 kroner. In general, short or frequent trips have lower values than long or infrequent ones.

In the last box of Figure 5 the average values for each purpose group are plotted, and fitted linear regression lines are shown. The data give good support to a hypothesis about a straight-line relationship between value of time and income. The regression lines for value of time (VOT) in kroner/vehicle hour as a function of gross personal annual income in thousands of kroner were given by Equations 4, 5, and 6 for Commuting, Business, and Other, respectively:

$$\text{VOT} = 59 + 0.084 \times \text{income} \qquad R^2 = 87.1\ \text{percent} \tag{4}$$

$$\text{VOT} = 89 + 0.226 \times \text{income} \qquad R^2 = 89.8\ \text{percent} \tag{5}$$

$$\text{VOT} = 100 + 0.122 \times \text{income} \qquad R^2 = 91.9\ \text{percent} \tag{6}$$

Values of time per vehicle for purpose groups, expressed as percentages of the gross hourly wage rate, are shown in Figure 6. In cases in which there is a lack of information about the respondent's working hours, the wage rate for each individual is computed as the reported annual income divided by 2,000. The values are decreasing percentages of income as income increases, and the ranges are 100 to 50 percent for Commuting, 165 to 90 percent for Business, and 175 to 75 percent for Other.

Equations 4, 5, and 6 imply that VOT increases less than proportionately with income. In Figure 7 the variations in percentage change in VOT, for a 10 percent change in income, are shown. Most travelers reveal an increase of 2 to 4 percent in VOT for a 10 percent increase in their gross income.

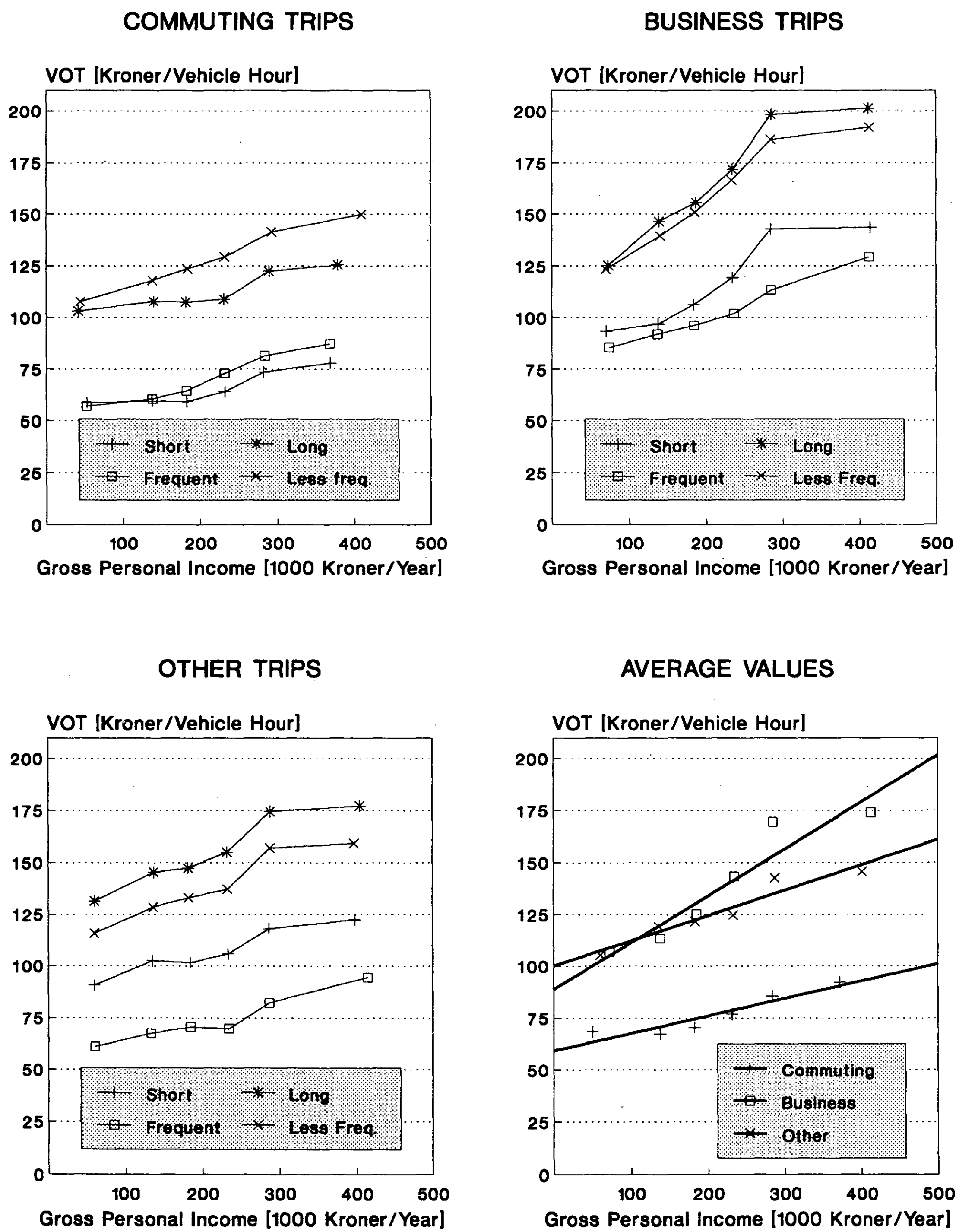


FIGURE 5 Variations in values of time with income.

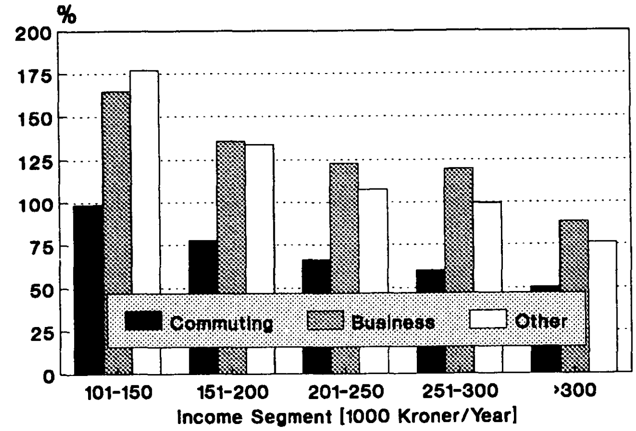


FIGURE 6 Values of time per vehicle as percentage of gross wage rate.

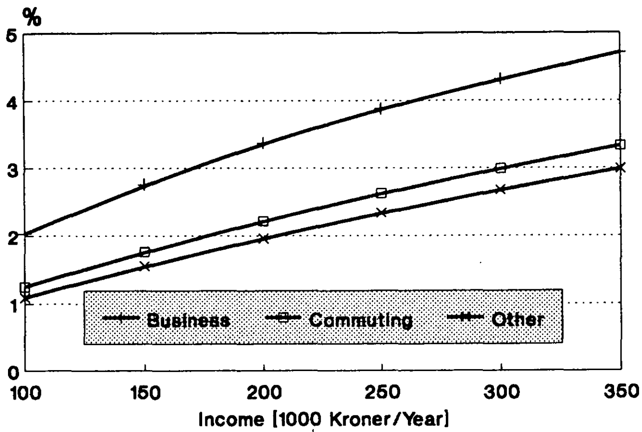


FIGURE 7 Percentage change in VOT per vehicle with 10 percent change in income.

Discussions are underway in Norway and among the Nordic countries, in preparation for coordinated efforts to supplement this and other recent studies of behavioral values of time. The aim is to provide a comprehensive body of evidence as a basis for recommending revised values for the assessment of transportation improvements.

ACKNOWLEDGMENTS

The writing of this paper was financed by PROSAF, a research program of the Royal Norwegian Council for Scientific and Industrial Research. The 1989 and 1990 data collections were planned and executed as a joint effort between ASPLAN Trondheim and SINTEF and financed by the Public Roads Administration. The author has benefited from the suggestions of Tore Sager of the Norwegian Institute of Technology.

REFERENCES

1. *St.meld.nr.46 (1990–91).* Samferdseldepartementet, Oslo, Norway, 1991.
2. L. Festinger. *Conflict, Decision, and Dissonance.* Stanford University Press, Stanford, Calif., 1964.
3. T. Tretvik. The Toll Road Alternative: Variations in Choice Behaviour and Values of Time. *Proc., 6th World Conference on Transport Research*, Lyon, France, 1992, Report STF63 S92001, SINTEF Samferdselsteknikk, Trondheim, Norway.
4. K. A. Small. *Urban Transportation Economics.* Harwood Academic Publishers, Chur, Switzerland, 1992.
5. *The Netherlands' Value of Time Study: Final Report.* Hague Consulting Group, Den Haag, The Netherlands, 1990.
6. M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis. Theory and Applications to Travel Demand.* MIT Press, Cambridge, Mass., 1985.
7. *Kjørekostnadshandboka.* Transportøkonomisk Institutt, Oslo, Norway, 1989.

Publication of this paper sponsored by Committee on Transportation Economics.

Derivation of Value of Time and Traffic Demand Curves in Bangkok

YUICHIRO MOTOMURA AND SURAPONG LAOHA-UNYA

A value of time has been calculated by calibrating a traffic assignment model to match toll expressway usage in Bangkok. This model was developed to reflect the situation in Bangkok, in which extreme traffic congestion and the resulting suppressed traffic demand are common. The values of time derived for models are considerably higher than those derived from income data using either government statistics or home interview survey data, suggesting the unreliability of statistics obtained in the normal manner. A demand curve for expressway travel was then established by extending this method and applied to forecast use of a proposed expressway. The curve derived from the model suggests a demand shift (an upward and clockwise shift of the demand curve) in the future with declining demand elasticity with respect to toll levels as a consequence of increased traffic congestion and real value of time.

Determining the value of time is critical in making urban transport investment decisions. The value of time heavily affects both demand and quantification of economic benefits. The level of demand determines the amount of needed investment; the level of benefit determines the feasibility of the investment. The value of time as perceived by transport facility users is an important determinant of users' behavior (behavioral value of time), whereas the value of time as a limited economic resource is considered to be the basis of economic feasibility determination (resource value of time).

Much theoretical and practical work has been undertaken, mostly in developed countries, to determine the value of time. The emphasis of such work is now moving to limited but carefully controlled and detailed field surveys.

Rapid urbanization throughout the developing world is saturating the limited transport infrastructure of its major cities, requiring solutions. Funds are in short supply, as are planning data, including those required to calculate the value of time. This paper presents the approach to estimating the value of time (behavioral) that was used in a major urban toll expressway planning exercise in Bangkok, Thailand. As an extension of the approach, this paper also presents a method of determining a demand curve for the urban toll expressway.

ESTIMATING THE VALUE OF TIME

Alternative Approaches

Many methods to define and quantify the value of time have been proposed. A classification and discussion of the alternatives in Figure 1 are provided by Suzuki et al. (*1*).

Y. Motomura, PADECO Co., Ltd., Ichibancho, Chiyoda-ku, Tokyo 102, Japan. S. Laoha-unya, PADECO (Thailand) Ltd., 22nd Floor Silom Complex Building, 191 Silom Road, Bangkok 10500, Thailand.

Practical Problems

The income method is normally applied as a shortcut in developing countries for determining not only resource value but also behavioral value. This method is used to explain travel behavior with some modifications by analogy to data developed in other countries. Income data in developing countries, however, often are unreliable. Top incomes are understated, and the poor are to a large extent outside of the monetary economy. Data on individuals' travel patterns are lacking, and the limited study resources available are inadequate to research them.

The study of a proposed toll expressway in Bangkok presented the authors with such problems. A new urban toll expressway was being planned for the Bangkok metropolitan area, and a major issue was the sensitivity of demand to the toll rate, which would determine the financial viability of the project. The toll on the existing expressway has remained unchanged since the system opened in 1981, so the price elasticity of demand has never been tested. No survey data relating travel patterns to journey times and costs were available. However, areawide origin and destination data did exist. Therefore, the authors attempted to overcome the obstacle by taking advantage of traffic data available for the existing toll expressway.

Methodology

A variant of the route choice model method shown in Figure 1 was applied. Ordinarily, the route choice model method attempts to establish an explanatory function that includes travel time and costs as variable data of trip makers taking a set of competing routes. The ratio of coefficients to travel time and travel cost variables is taken as the value of time. However, direct application to an urban area where there are many alternative routes is difficult.

A simple and direct approach was devised, taking advantage of the situation in Bangkok. The existing toll expressway network basically covers only the central part of the large metropolitan area, as shown in Figure 2. Traffic volume data are available by toll plaza location. Also available to the authors (courtesy of the government of Thailand) were the results of large-scale home and roadside interview surveys undertaken by the Japan International Cooperation Agency in 1989 (*2*). Origin and destination (O-D) tables were developed from these data. A cross-sectional approach was adopted because these detailed O-D data were available only for 1989.

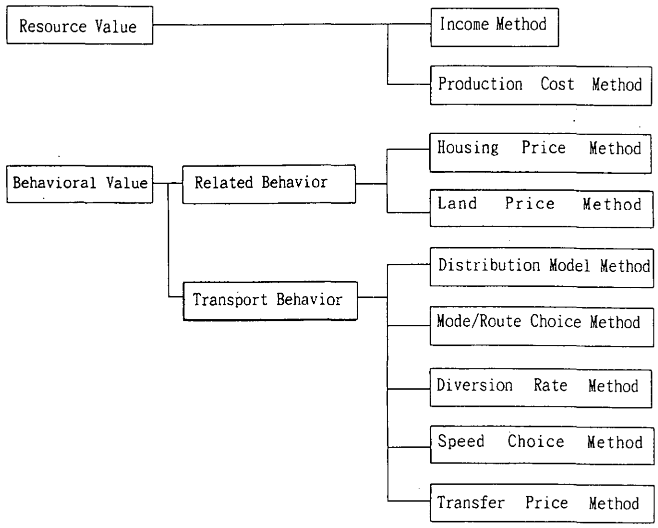


FIGURE 1 Classification of measurement methods.

The six-step procedure adopted was as follows:

1. Vehicular O-D tables were developed and calibrated to match actual traffic volumes throughout the metropolitan area, excluding those on the toll expressway network. The routes of the existing toll expressway and their connections to streets are such that the final calibrated O-D tables could be considered accurate for the purpose of estimating expressway usage.
2. The resultant O-D tables were assigned to the network, with the expressway links included. The expressway toll, 10 Baht (U.S. $0.40) for light vehicles and 20 Baht (U.S. $0.80) for heavy vehicles, was converted to a time penalty for the assignment step. After assignment, model-estimated traffic volumes on each of the toll plazas were checked with actual records to ascertain the validity. Link speeds and travel times were also checked with actual records, and necessary modifications were made to attain a reasonable level of matching.
3. The process was repeated using different time values to convert the monetary toll cost into a different time penalty.
4. A curve relating time value and expressway usage was established.
5. A behavioral value of time was determined by matching the actual traffic volume totals by toll plaza location with the value of time versus traffic volume curve developed above.
6. Steps 1 through 5 above were repeated using the above-obtained value of time to ascertain the results.

Traffic modeling was carried out using the MOTORS suite of programs (microcomputer-based software), originally developed by Steer Davies & Gleave of the United Kingdom. This MS-DOS–based suite has been upgraded and enhanced by PADECO to simulate the extraordinary congestion in Bangkok.

The traffic assignment method used for this study was an incremental assignment on an hourly basis. Link travel times were updated after each incremental assignment and the next incremental assignment loaded traffic onto the shortest path using the updated link travel times. The degree of congestion in Bangkok is such that over 30 iterative assignments were necessary for each hour. Later assignment iterations amounted to less than 1 percent of the total demand. Because the model was primarily calibrated against actual traffic volumes on network links, assigned link volumes matched actual traffic volumes well. Travel times or speeds on each of the network links produced by the model were found to correspond closely with the actual values. Figure 3 shows survey and model results for major arterials. It is evident from this figure how horrendous the level of congestion is in Bangkok.

The selection of the toll expressway by the model as part of the shortest path depends on how the toll expressway network is connected to the zone centroids in the model network representation. Traffic volumes assigned to the toll expressway, therefore, could vary depending on the level of detail of the modeled network. Two levels of detail were tested to examine the effect of this factor. First, the model was run for a network with 90 zones representing the entire Bangkok metropolitan area. Then another network with 170 zones was developed, after dividing zones (except external ones) into two or three smaller zones and recoding the appropriate zone connectors to the road network.

Results

Figures 4 and 5 show the resulting relationships between the toll equivalent time and the total combined traffic assigned to the expressway for light vehicles and heavy vehicles, respectively. Actual recorded traffic volumes indicate that the existing toll of 10 Baht for light vehicles is equivalent to about 2.5 min (i.e., the value of time is 4 Baht/min for light vehicles) and that the toll of 20 Baht for heavy vehicles is equivalent to 4.3 min (i.e., the value of time is 4.7 Baht/min for heavy vehicles).

Figure 4 indicates that the division of the area into smaller zones resulted in less traffic on the expressway. As a larger number of alternative routes became available with a more detailed zone representation, some of the short trips found routes that exclude the expressway, resulting in less traffic on the expressway. However, this is not the case for heavy vehicles, as shown in Figure 5, because heavy vehicle trips include few short trips. Further subdivision of the area into even smaller zones did not yield results that were significantly different from the results of the case of 170 zones because such subdivisions produced no significant alternative routes, considering the limited road link length.

The value of time of 4 to 5 Baht/min or 16 to 20 cents/min that was obtained is much higher than even the monetary value of 1 min derived from personal income figures as reported in government statistics or field surveys.

The value of 1 min as calculated from government statistics (and required assumptions) amounts to less than 2 Baht:

1. Average Bangkok household monthly income in 1986: 7,410 Baht (from government household survey in 1986).
2. Estimated 1989 Bangkok household monthly income: 11,322 Baht (nominal increase: 1986–1987, 7.1 percent; 1987–1988, 14.9 percent; and 1988–1989, 24.2 percent).
3. Average number of income earners/household: 1.5 (from government household survey).
4. Average number of income earners/vehicle: 1.3 (assumed on the basis of home interview survey).

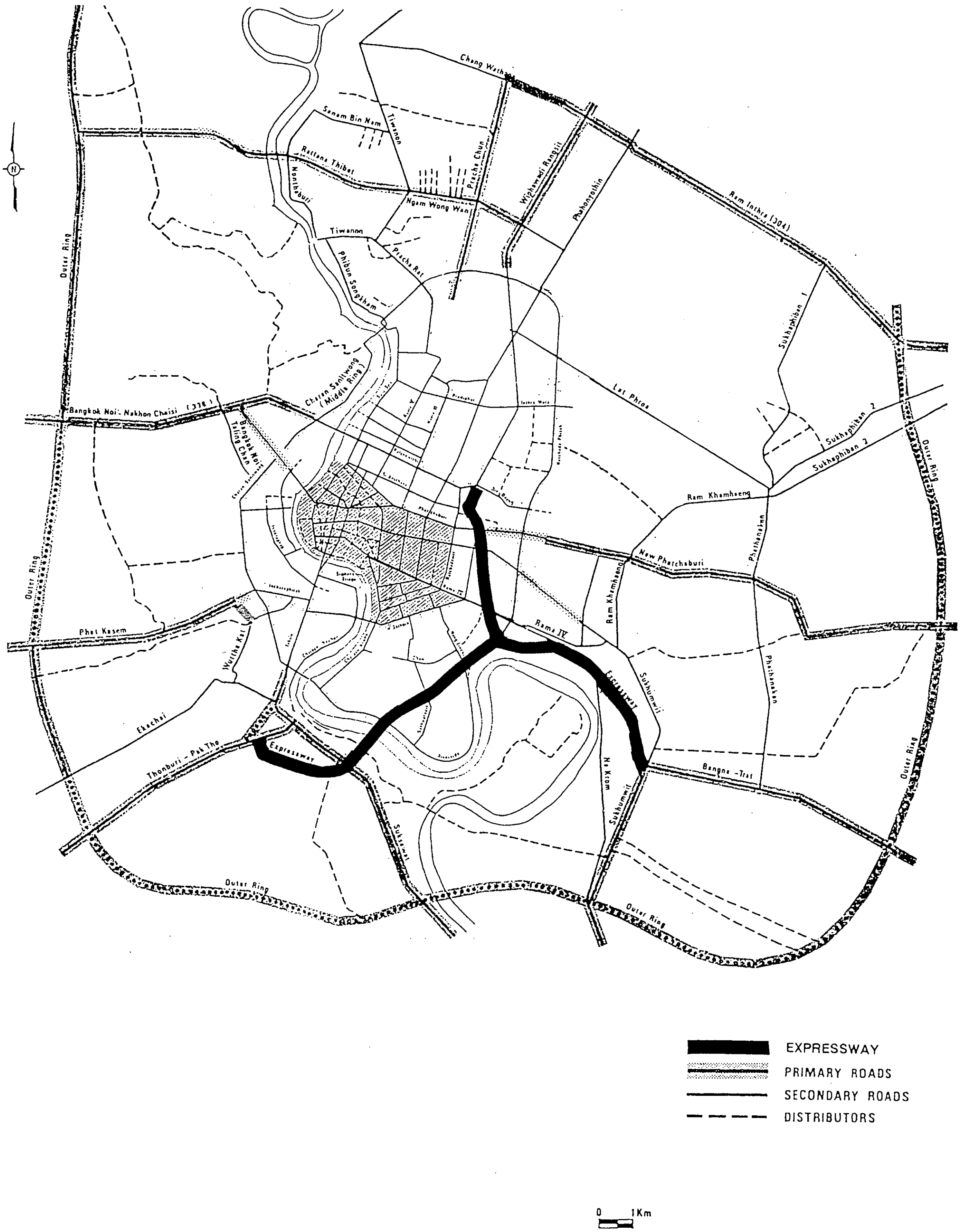


FIGURE 2 Road network in Bangkok.

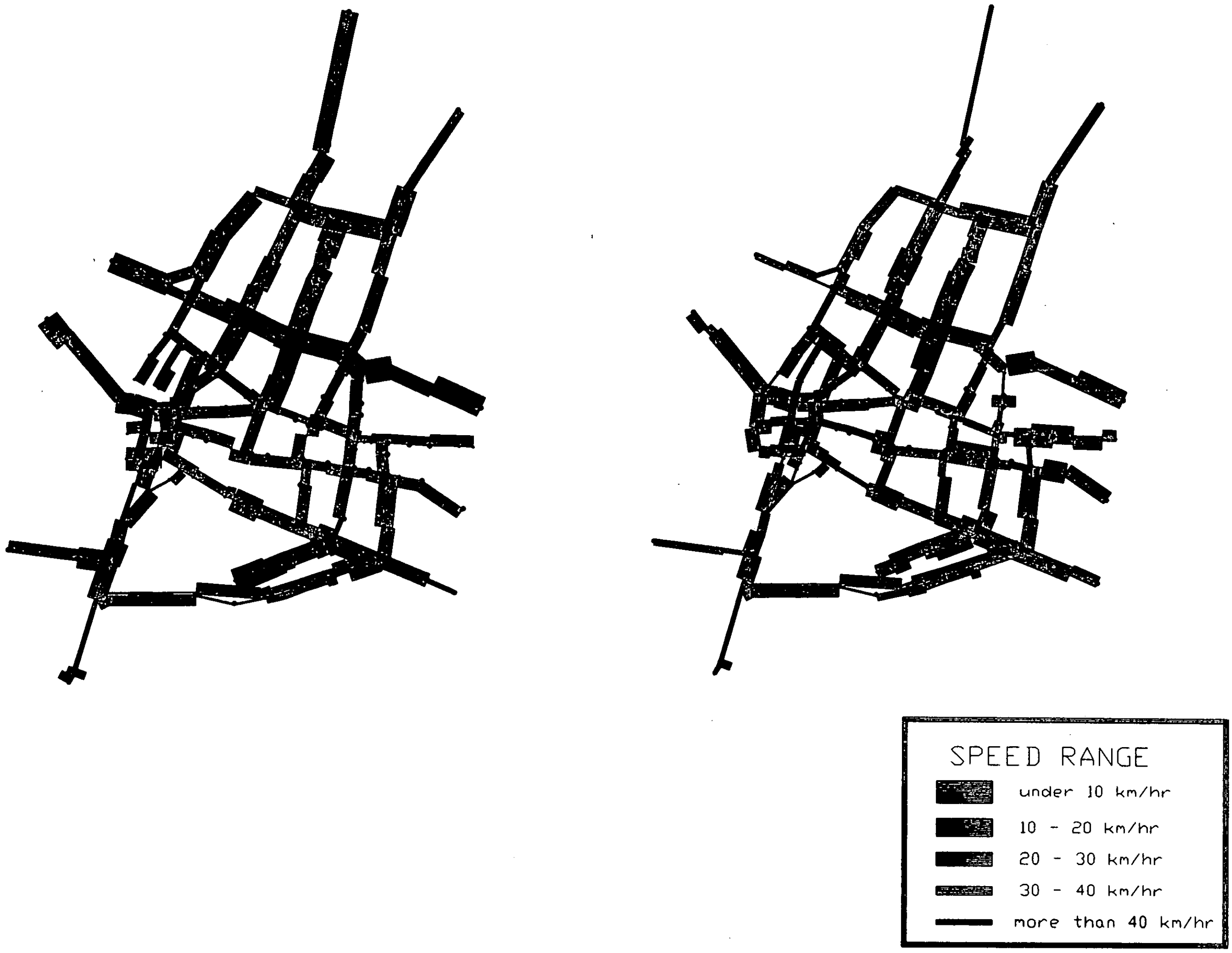


FIGURE 3 Comparison of actual (*left*) and simulated (*right*) network link speed, morning peak, 1989.

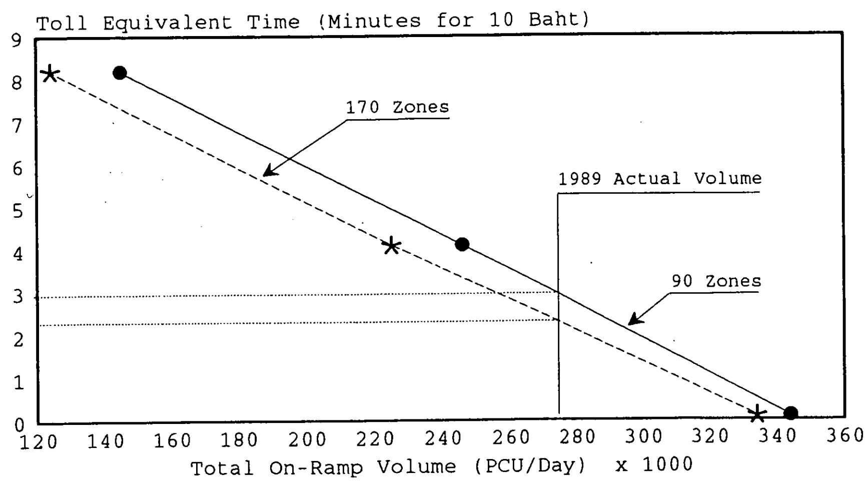


FIGURE 4 Toll equivalent time versus expressway traffic (light vehicles).

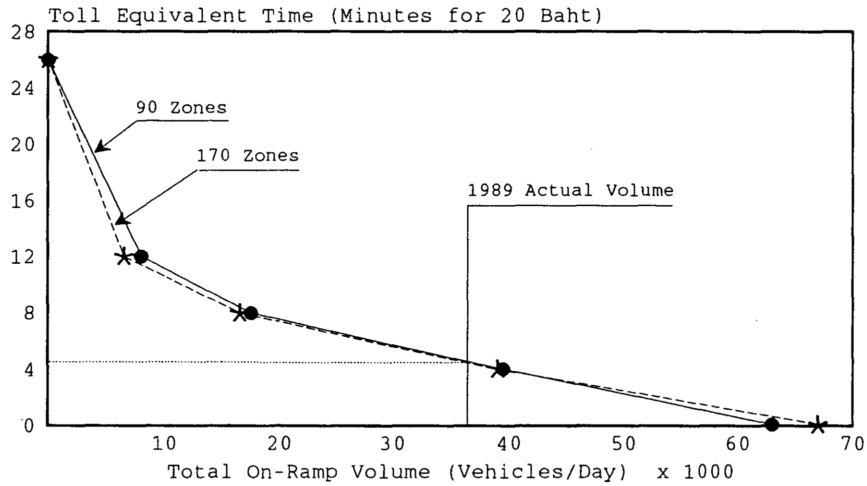


FIGURE 5 Toll equivalent time versus expressway traffic (heavy vehicles).

5. Average number of working hours/month: 170 hr (assumed on the basis of Labor Department data).
6. Ratio of average incomes of top one-third to overall average: 1.63.
7. Hourly passenger income/vehicle: 1.56 Baht/min. (Note: one in three households in Bangkok owns at least one passenger car.)

Many studies in developed countries found that the perceived value of time was less than the hourly wage although the conventional practice of valuing work trip time higher than nonwork trip time is suspect (*3*). The value derived from the actual usage of the expressway turned out to be higher than the hourly wage that was calculated from government statistics and field interview survey results.

Personal income data obtained by a private-sector household survey carried out in 1989 (*2*) also showed lower figures than those expected from the value of time. The average personal income of the three highest-earning occupation groups (i.e., professionals, administrative staff, and shop owners) was 7,665 Baht per person. Assuming 1.3 income earners per vehicle, passenger income per vehicle amounts to less than 1 Baht/min.

The high value of time estimated by the model can primarily be explained by two factors: underestimation of income and truly horrendous traffic congestion. In addition to the underreporting of actual income, the social structure in Bangkok oriented toward extended families may contribute to a higher level of disposable income of individuals than in other societies with less extended-family reliance. Extreme congestion in Bangkok makes journey times utterly unpredictable because of the highly unstable nature of oversaturated traffic flow. A study in the United Kingdom suggested that time savings in highly congested conditions might be valued 40 percent more highly than in less congested conditions (*3*). Values derived in Bangkok suggest such a phenomenon, but even more pronounced. Journey times in Bangkok vary greatly by day. An interview survey conducted by the authors revealed that people tend to report journey times on bad rather than average days. Perceived gains by more predictable alternatives, such as expressway, seem to be much higher than the actual average time savings.

Supporting Evidence

Estimating personal income in Bangkok is an extremely difficult task, and consequently so is estimating the value of time. A recent study attempted to estimate the income levels of private vehicle owners from vehicle purchase prices and purchasing conditions (*4*).

Automobiles and pickups typically are purchased with loans. In July 1990, car loan interest rates ranged from 10 to 15 percent, depending on the model and whether the vehicle was new or used. A down payment of at least 20 percent was required, and repayment of the remaining portion plus interest was to be made over a period of 24 to 48 months. Typically, lenders required that monthly installments not exceed 25 percent of the respective borrower's monthly income. The minimum monthly incomes to satisfy the above conditions were found to be 13,000 Baht for a small 3-year-old (i.e., used) sedan and 41,000 Baht for the new medium-sized sedan that was popular at the time. The model-estimated value of time of 4 to 5 Baht/min corresponds closely with the above-estimated family income of 41,000 Baht/month (3.5 Baht/min).

DEMAND CURVES FOR EXPRESSWAY TRAVEL

Demand Curve, 1989

The value of time obtained above is the average value of all vehicle users that gives the best fit for assigned expressway usage against actual usage. The toll equivalent times (time savings needed to compensate for the toll) in Figures 4 and 5 can be expressed as monetary tolls by multiplying these equivalent times by the average value of time. The point on the vertical axis corresponding to the 1989 actual volume would become 10 Baht and the scale of the vertical axis could be redefined proportionately. The resultant figure would be a demand curve showing the relationship between toll levels and traffic volumes. For light vehicles the curve indicates that a 10 percent increase in the toll level would reduce traffic by 2.3 percent corresponding to a demand elasticity of -0.23.

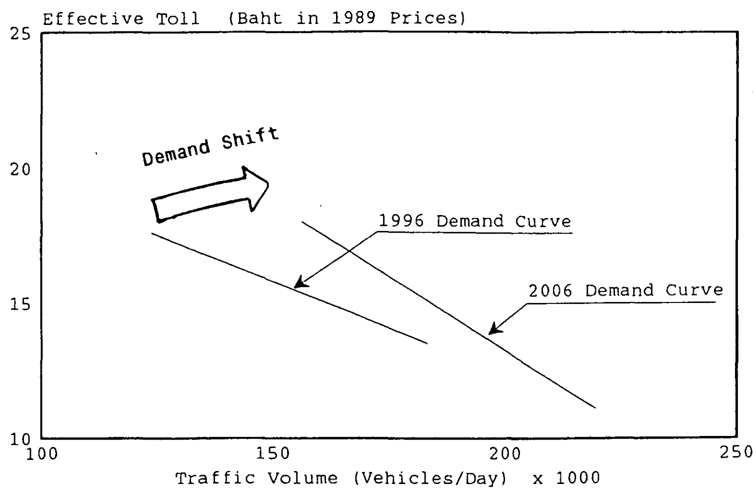


FIGURE 6 Expressway demand curve (northern route, light vehicles).

Future Demand Curves and Demand Shift

This methodology has been applied to estimate loadings on a planned expressway in Bangkok. The forecast O-D trips were assigned to the future network, which included the proposed expressway, for alternative toll levels using the 1989 value of time adjusted for the assumed increase in per capita real income (standard practice in developing countries). The demand curves obtained for 1996 and 2006 are shown in Figure 6. As demand grows between 1996 and 2006 the curve itself rotates clockwise (demand shift), as a result of higher real-time values and greater congestion on alternative nontollway routes. From the future demand curves obtained, future traffic volumes were forecast by applying the proposed toll level expressed in real constant prices.

CONCLUSIONS

A behavioral value of time suitable for forecasting toll expressway demand can be obtained when accurate O-D tables and traffic data on an existing toll system are available. The value determined for Bangkok is higher than that obtained by conventional methods. The methodology provides demand curves (current and future, with respect to toll level) that indicate that the demand elasticity in Bangkok is declining markedly as a result of rapidly increasing real incomes and growing congestion on the nontollway network.

ACKNOWLEDGMENTS

The authors are grateful to the Obayashi Corporation of Japan for the use of study data and to the Expressway and Rapid Transit Authority of Thailand, the Bangkok Metropolitan Administration, and the Japan International Cooperation Agency for making survey results available. The authors' colleagues, Chiaki Kuranami, Peter Mansell, Bruce Winston, and Kozo Yagi, provided helpful comments and assistance. Bruce Winston's work on determining income levels and vehicle acquisition costs and Kozo Yagi's assistance with model operation are particularly acknowledged.

REFERENCES

1. S. Suzuki, N. Harata, and K. Ota. A Study on the Value of Time in Highway Planning (in Japanese). *Highways and Automobiles*, Oct. 1987.
2. *Short and Medium Term Improvement of Roads in Greater Bangkok*. Japan International Cooperation Agency, March 1990.
3. The Value of Time Savings. *Policy Journals*, The MVA Consultancy, 1987.
4. *The Survey of Urban Transport Costs and Fares in the SEATAC Region, Phase I*. PADECO Co., Ltd., Oct. 1990.

Publication of this paper sponsored by Committee on Transportation Economics.

Interchange Versus At-Grade Intersection on Rural Expressways

JAMES A. BONNESON, PATRICK T. MCCOY, AND DUANE S. EITEL

The economic benefits and costs of replacing a two-way stop-controlled intersection on a rural expressway with either a signalized intersection or a conventional diamond interchange were compared. Economic benefits were based on the difference in road user costs among alternatives. Road user costs were composed of five components: delay, idle fuel, acceleration-deceleration delay, speed-change running costs, and accident costs. The benefit-cost analysis of the signalized intersection and interchange under rural expressway conditions indicated that the interchange was a more economically viable alternative than the signalized intersection. The signalized intersection's main benefit is a reduction in accidents; however, this benefit is generally negated by the signal's higher operational costs whenever the minor road demand is less than one-half that of the major road. Three geometric scenarios were formulated for the intersection and interchange. The first considered a four-leg junction with a two-lane minor road. The second considered a four-leg junction with a four-lane minor road. The third considered a three-leg junction with a two-lane minor road. Three figures were developed relating the major and minor road daily traffic demands that would economically justify an interchange in terms of a benefit-cost ratio. Whenever the major road demand is about 4,000 vehicles per day (vpd) or more, the minor road demands that provide a 2.0 benefit-cost ratio are about 4,000, 6,500, and 8,000 vpd for the three scenarios, respectively.

The ability to assess the cost-effectiveness of a construction project is particularly important when an agency is operating under a limited construction budget. Some state agencies have adopted general procedures and conditions under which an interchange is warranted; however, these criteria do not typically consider accident history or road user costs. An economic analysis of road user benefits (e.g., reduced delay, fuel, stops, and accidents) and project costs would justify a project from a pragmatic standpoint and could facilitate alternative selection and project prioritization.

The objective of this research was to develop guidelines for use in determining when a signalized intersection or an interchange would be a cost-effective corrective measure for a problematic unsignalized intersection on a rural expressway. Such an intersection might be experiencing operational or safety problems. The guidelines described in this paper are based on an economic assessment of the operational and safety benefits realized by upgrading a two-way stop-controlled intersection to a signalized intersection and to a conventional diamond interchange for a range of traffic demand levels.

J. A. Bonneson and P. T. McCoy, Department of Civil Engineering, University of Nebraska-Lincoln, Lincoln, Nebr. 68588-0531. D. S. Eitel, Nebraska Department of Roads, Lincoln, Nebr.

LITERATURE REVIEW

Interchanges are typically justified for one of two reasons. An interchange may be necessary to maintain a consistency with the major road's functional classification. Alternatively, an interchange may be needed to efficiently serve existing or future traffic demands. The general trend among highway departments is to deal with candidate interchange locations on a case-by-case basis.

A recent survey of state departments of transportation by the Nebraska Department of Roads (NDOR) (*1*) indicated that most states do not have policies or guidelines that identify the conditions needed to justify an interchange. Many states indicated that the decision to develop a grade separation or interchange is a direct consequence of the decision to provide an access-controlled roadway.

Two of the states that responded to the survey indicated that they design all new rural expressways that bypass cities as full-access control facilities. In this regard, they construct interchanges at all major intersections along the bypass. The justification offered for this policy was the poor safety record of existing bypasses with at-grade intersections.

One research effort at establishing interchange warrants was conducted by Ockert and Walker (*2*). On the basis of a series of simulation studies for a range of economic and traffic characteristics, they concluded that the operational benefits of an interchange begin to outweigh its cost at traffic volume levels that meet or exceed the *Manual on Uniform Traffic Control Devices*' (MUTCD) (*3*) Signal Warrants 1 and 2 (Warrants 9 and 11 did not exist at the time of this study). In other words, the benefit-cost ratio of the interchange is higher than that of a conversion to signal control at demand levels that meet MUTCD signal warrants, although both ratios are greater than 1.0. This implies that signal control is not an economically justifiable alternative on rural expressways.

Another effort at establishing interchange warrants was conducted by Van Every (*4*). Van Every calculated the road user benefits for a conventional diamond interchange compared with a two-way stop-controlled intersection. He then compared these benefits with the incremental cost of constructing an interchange and made recommendations on major and minor road volume thresholds that would yield benefit-cost ratios of 1.0 and 2.0. Van Every (*4*) reported volume warrants for several combinations of traffic growth rates and turn percentages. Examination of these warrants indicates a strong similarity to MUTCD Signal Warrant 11 [64 km/hr (40-mph) major road speed], which is consistent with the findings of Ockert and Walker (*2*).

ANALYSIS METHODOLOGY

The analysis methodology was based on the benefit-cost comparison of three design alternatives. The base alternative was defined as the two-way stop-controlled intersection. The two alternatives included a signalized intersection and a conventional diamond interchange with stop-controlled off ramps. The benefit-cost approach used in this research was based on the methods described in *A Manual on User Benefit Analysis of Highway and Bus-Transit Improvements* (i.e., Red Book) (*5*).

To assess the relative benefits of the three design alternatives, procedures for quantifying the operational and safety performance of the typical intersection and diamond interchange were established. The operational assessment procedure is sensitive to a variety of design factors, such as traffic demand, traffic composition, design speed, and traffic control. This procedure is based on the quantification of various measures of motorist cost, such as motorist delay and vehicle fuel consumption. The assessment of operational performance is consistent with the methodology published in the 1985 *Highway Capacity Manual* (*6*) (i.e., 1985 HCM).

The safety assessment procedure developed for this research is sensitive to traffic demands, traffic control type, and junction type. This procedure was developed using data from the Highway Safety Information System (HSIS) (*7*). The procedure is composed of several regression models for predicting the expected frequency and severity of accidents at a junction on the basis of its traffic demand and traffic control conditions.

The attractiveness of each design alternative was assessed by comparing the present worth of its benefits and costs with the base alternative. Costs considered for this analysis included construction cost, maintenance cost, and residual worth (a negative cost) of the property at the end of the alternative's design life.

Economic Factors

Discount Rate

The discount rate is needed to calculate the worth of all benefits and costs in terms of present dollars. The discount rate used in this analysis is based on the market rate of return adjusted for inflationary effects (i.e., a "constant dollar" approach) and is assumed to be 4.0 percent, based on a market rate of return of about 9.2 percent and a rate of inflation of 5.0 percent.

Cost Updating Procedure

Dollar values for all unit prices and project cost components have been updated to January 1991 levels using appropriate consumer and wholesale price indices. All of the unit prices used in this study were extracted from the Red Book (*5*) with the exception of accident costs, which were obtained from FHWA (*8*).

Value of Travel Time

For this study, travel time rates for passenger cars, single-unit trucks, and tractor-trailer trucks were obtained from the Red Book (*5*,pp.15–19). All values were updated to January 1991 levels. The updated value of time for passenger cars was estimated as \$0.84/hr. This estimate is based on an "average" value for all trip types (e.g., social and work), a relatively low time savings due to any operational improvements (less than 5 min), and an average vehicle occupancy of 1.56 persons. The updated value of time for single-unit and tractor-trailer trucks is estimated as \$15.50/hr and \$17.72/hr, respectively.

Analysis Period

For this study, the lifetimes of the geometric improvements associated with the alternative designs are expected to range from 20 to 40 years. However, because of the uncertainties in the predictions of future traffic demands, travel patterns, and land use, the duration of the analysis period was established as 20 years. This duration was believed to be more defensible in terms of our greater confidence in analysis assumptions made for shorter periods. A 20-year period was also believed to yield more conservative results by making the alternative justify its entire construction cost with benefits accrued over the first 20 years.

Traffic Characteristics

Directional Distribution

The directional distribution of traffic demand on each junction approach was established as 50 percent inbound and 50 percent outbound.

Annual Growth Rate

NDOR's *1990 State Highway Plan and Highway Needs Report* (*9*) was consulted to determine a range of typical growth rates on rural Nebraska highways. Present and future average daily traffic values (ADTs) were selected from this document for 19 segments of six state highways. These ADTs were used to calculate the annual growth rate for each segment. The calculated growth rates were found to range from 2.4 to 4.6 percent per year. The average rate of 3.5 percent per year was established as a typical value.

Turn Percentages

Turn movement percentages were obtained for 24 rural intersections where at least one of the intersecting routes was a state highway. An examination of these turn percentages indicated a relationship between turn percentage, traffic pattern, and demand levels. To accurately model this relationship for the full range of demand conditions, an algorithm for predicting turn percentages was incorporated into the benefit-cost analysis program. This algorithm was developed by Hauer et al. (*10*) but enhanced to calculate the "balanced" approach and departure volumes for specified ADTs. By incorporating this algorithm in the analysis program, all that was needed as input was an estimate of the major and minor road ADT; hourly turn movement volumes could then be estimated by the algorithm.

Application of the algorithm to the 24 intersections, using initial "seed" turn percentages of 10 percent left and right on the major road and 30 percent left and right on the minor road, indicated that it could predict the actual turn percentages reasonably well. In fact, the algorithm was able to explain 87 to 92 percent of the variation in turn percentages on the major road and 61 to 68 percent of the variation on the minor road.

Traffic Composition

NDOR's 1990 report (*9*) was consulted to determine the percentage of truck traffic on rural Nebraska highways. The percentage of trucks in the daily traffic stream was selected from this document for 19 segments on six state highways. These percentages were found to range from 6 to 25 percent, with a representative value established as 14 percent (7 percent single-unit trucks and 7 percent tractor-trailers).

Hourly Volume Frequencies

The analysis program was written to estimate user costs on an hourly basis. However, instead of analyzing all 8,760 hr in each year of the analysis period, only five representative hourly volumes were considered for each year. Each hourly volume level is estimated by multiplying an hourly volume factor by the ADT for the corresponding year. The total user costs for each year are then calculated as the sum of the user costs for each of the five hourly volumes multiplied by the corresponding number of hours represented. This process is then repeated for each analysis year.

The five hourly factors used by the program were developed by first ranking and then dividing the frequency distribution of hourly flows for 1 year into five intervals and calculating an average hourly volume for each interval. The frequency distribution was obtained from the hourly volumes recorded at seven automatic traffic recorder (ATR) stations on rural highways in Nebraska (*11*). The interval widths, in hours, were selected to include hourly flows of similar magnitude. As a result, the interval widths were small for the few peak hours and larger for the many nonpeak hours.

Five hourly factors were calculated for each ATR station for the same hour intervals. Each factor was calculated by dividing the average hourly volume for the corresponding interval by the ADT. These factors were then averaged over all seven stations. The resulting hourly volume factors (expressed as a percentage of ADT) and associated hour intervals are 9.5 percent for 250 hr, 8.6 percent for 250 hr, 7.6 percent for 1,000 hr, 5.9 percent for 2,500 hr, and 1.6 percent for 4,760 hr.

Average Travel Speed

The average travel speed was estimated as 88 percent of the 85th percentile speed on the basis of information provided elsewhere (*12*). The 85th percentile speed was, in turn, assumed to equal the posted speed limit. The posted speed limits on the major and minor roads were assumed to be 89 and 64 km/hr (55 and 40 mph), respectively. The ramp speed for the conventional diamond interchange was assumed to be 72 km/hr (45 mph). The right-turn speed was assumed to be 21 km/hr (13 mph). All left-turning vehicles were assumed to stop.

Project Characteristics

Geometric Configuration

The geometry of the intersection and interchange included the following elements:

- A four-lane expressway/major road with a 12-m (40-ft) median.
- The signalized and unsignalized intersections have left-turn bays on the major road. Only the signalized intersection has left-turn bays on the minor road.
- The on- and off-ramps for the interchange intersect with the major road about 549 and 488 m (1,800 and 1,600 ft) from the minor road, respectively. The ramp/minor road junctions are offset 183 m (600 ft) from the major road centerline.
- The major road is constructed at grade through the interchange and the minor road is elevated above the major road. The bridge is 61 m (200 ft) in length, supported by a bent in the major road median.

Construction Costs

For this study, cost estimates were formulated for each of the three junction types. The construction cost for the unsignalized intersection includes only the costs of upgrading the minor road approaches and adding left-turn bays on the major road. The construction cost for the signalized intersection includes upgrading and widening the minor road approaches, providing left-turn bays on all approaches, and installing a traffic signal. Two cost estimates were made for the interchange: one for a two-lane overpass and one for a four-lane overpass. The cost estimates included costs for mobilization, earthwork, subgrade preparation, pavement, and drainage.

The following assumptions are included in the cost estimates: (a) the cost of the four-land major road is common to all junction types and can be excluded; (b) contingency costs would amount to 20 percent of the construction cost; and (c) the cost of engineering is 14 percent of the contingency and construction costs. The estimated 1991 construction costs for the alternatives considered in this study are as follows: unsignalized intersection, $156,000; signalized intersection, $363,000; conventional diamond interchange (two-lane), $2,020,000; and conventional diamond interchange (four-lane), $3,060,000.

Maintenance Costs

Estimates of annual maintenance costs for 1991 were obtained from NDOR. These data indicate annual maintenance costs of $2,000 for unsignalized intersections and $7,000 for interchanges with two-lane minor roads. Maintenance costs for interchanges with four-lane minor roads were not obtained; however, these costs were assumed to be $10,000 per year.

The annual maintenance cost for a signalized intersection was estimated to be about $1,000 higher than that of the unsignalized intersection; the annual operating cost of the signal is $1,000. Thus, the operating and maintenance costs of the signalized intersection were estimated at $4,000 per year.

Residual Value

The residual value of the alternative facilities was not included in the analysis. This omission resulted in a more conservative analysis in that the alternative would have to justify its total construction cost in terms of road user benefits.

Operational Costs

Time Cost Components

Time cost components include all traffic control or geometric factors that delay motorists. These costs represent the excess travel time (i.e., delay) incurred by motorists because of the junction. Thus, they represent the difference between the actual travel time and the travel time that would have been incurred had all movements been served by exclusive through lanes or directional ramps. The most significant time cost components are

1. The delays from having to transition from the main lane speed to a turn speed and back to the main lane speed,
2. The delays to minor movements that are stop controlled, and
3. The delays to all movements that are signal controlled.

In those instances in which the predicted delays exceed reasonable values (e.g., 30 min), it is assumed that the drivers will actually divert to alternative routes or postpone their trips to other times. Thus, delays in excess of 30 min could conceptually represent a notional cost that reflects the added inconvenience of diversion or postponement. For this analysis, the predicted delays were limited to the duration of the period of analysis (i.e., 1 hr). The maximum delay of 1 hr reflects a reasonable upper limit to the concept of notional cost.

Excess Delay for Speed Change The excess delay for a speed change cycle is calculated using constant rates of acceleration and deceleration. The rates used are for normal operating conditions and for speeds up to 64 km/hr (40 mph). The rates for passenger cars and single-unit trucks are estimated as 5.3 km/hr/sec (3.3 mph/sec) for acceleration and 8.0 km/hr/sec (5.0 mph/sec) for deceleration (*13*). The rates for tractor-trailer trucks are estimated as 1.6 km/hr/sec (1.0 mph/sec) for acceleration and 6.4 km/hr/sec (4.0 mph/sec) for deceleration.

Excess delay for a speed change is calculated using the following equations:

$$t_{ex} = t_{ad} - t_{ff} \tag{1}$$

$$t_{ad} = \frac{v_1 - v_0}{d} + \frac{v_2 - v_1}{a} \tag{2}$$

$$x_t = \frac{v_1^2 - v_0^2}{2d} + \frac{v_2^2 - v_1^2}{2a} \tag{3}$$

$$z = \frac{v_2^2 - v_0^2}{2x_t} \tag{4}$$

$$t_{ff} = \begin{cases} x_t/v_0 & \text{if } z = 0 \\ (-v_0 + \sqrt{v_0^2 + 2zx_t})/z & \text{otherwise} \end{cases}$$

where

- t_{ex} = excess delay for a speed change cycle (sec/vehicle),
- t_{ad} = travel time for the acceleration-deceleration maneuver (sec/vehicle),
- t_{ff} = travel time at free flow speed uninterrupted by junction (sec/vehicle),
- v_0 = initial speed (mps),
- v_1 = speed decelerated to (mps),
- v_2 = speed accelerated back to (mps),
- a = acceleration rate (mpss),
- d = deceleration rate (a negative value) (mpss),
- x_t = travel distance during deceleration and acceleration (m), and
- z = average acceleration needed to transition from v_0 to v_2 (mpss).

Unsignalized Intersection Delay Average delay to a stop-controlled movement is estimated using the following equation derived by Tanner (*14*):

$$w = \frac{q_1 e^{\beta q_1}(e^{\alpha q_1} - \alpha q_1 - 1) + q_2 e^{\alpha q_1}(e^{\beta q_1} - \beta q_1 - 1)}{q_1[q_1 e^{\beta q_1} - q_2 e^{\alpha q_1}(e^{\beta q_1} - 1)]} \tag{5}$$

where

- w = average delay to minor road vehicle (sec/vehicle),
- q_1 = major road flow rate (vehicles/sec),
- q_2 = minor road flow rate (vehicles/sec),
- e = base of natural logarithms (2.7183 . . .),
- α = critical acceptance gap (sec), and
- β = follow-up gap (sec).

The length of the critical gap for passenger cars was taken to be 6.0 sec for the higher speed conditions found at rural expressway intersections (*6*). On the basis of the work of Fitzpatrick et al. (*15*), tractor-trailer trucks were assumed to have an 8.0-sec critical gap. The follow-up gaps for passenger cars and trucks are assumed to be 2.0 and 3.0 sec, respectively.

Signalized Intersection Delay Delays for signal-controlled movements associated with the signalized intersection alternative were calculated using the methodology described in the 1985 HCM (*6*). For this analysis, the procedures described in the 1985 HCM (Chapter 2 of Appendix II) were used to calculate the signal cycle length and phase splits. It was assumed that the intersection operated in a semiactuated mode with all excess time allocated to the major movements.

The delay to stopped vehicles is calculated using the following formulas:

$$d = \frac{d_1 + d_2}{1.3} \tag{6}$$

$$d_1 = \frac{C\left(1 - \frac{g}{C}\right)^2}{2\left(1 - \frac{g}{C}X\right)} \tag{7}$$

$$d_2 = 900\ TX^2\left((X - 1) + \sqrt{(X - 1)^2 + \frac{4X}{Tc}}\right) \tag{8}$$

where

- d = average stopped delay (sec/vehicle);
- d_1 = uniform arrival delay component (sec/vehicle);
- d_2 = random arrival delay component (sec/vehicle);
- C = cycle length (sec);
- T = duration of the period of analysis (1 hr) (hr);
- g = effective green time (sec);
- X = v/c ratio for movement;
- c = sg/C, capacity of the movement (vph); and
- s = saturation flow rate of the movement (assumed as 1,800 vphgpl) (vphg).

Running Cost Components

Excess Running Cost for Speed Change The running costs for a speed change cycle are presented in Tables B-10, B-11, and B-12 in the Red Book (*5*) for a range of speeds and vehicle types. Least-squares regression techniques were used to develop the following equations for predicting the trends shown in these tables:

$$R_p = -2.07 + 0.477S_i - 0.411S_r + 0.00222S_i^2 + 0.00397S_r^2 - 0.00674S_iS_r \tag{9}$$

$$R_{su} = -7.55 + 1.467S_i - 1.077S_r - 0.00196S_i^2 + 0.00133S_r^2 - 0.00401S_iS_r \tag{10}$$

$$R_{wb} = -27.33 + 4.578S_i - 3.355S_r + 0.00606S_i^2 + 0.0221S_r^2 - 0.0396S_iS_r \tag{11}$$

where

- R_p = passenger car speed change cost per 1,000 cycles (dollars),
- R_{su} = single-unit truck speed change cost per 1,000 cycles (dollars),
- R_{wb} = tractor-trailer truck speed change cost per 1,000 cycles (dollars),
- S_i = initial and final speed (mph) (1 mph = 1.61 km/hr), and
- S_r = speed reduced to and returned from (mph) (1 mph = 1.61 km/hr).

In those instances in which the initial and final speeds are not the same, their average value is used as an approximation for S_i.

These costs represent the running costs per 1,000 speed changes at January 1975 price levels. They were inflated to the 1991 (base) analysis year by multiplying by the ratio of the 1991 price indexes to those for 1975. These ratios are 2.52 for passenger cars and 2.21 for single-unit and tractor-trailer trucks.

Idling Costs at an Intersection Idling costs are primarily dependent on the composition of the stopped traffic queue. Trucks typically consume less fuel and oil while idling and thus have lower idling costs than passenger cars. The costs for 1,000 hr of idling were obtained from the Red Book (*5*). These costs at 1991 levels are \$790, \$613, and \$427 per 1,000 idling hr for passenger cars, single-unit trucks, and tractor-trailer trucks, respectively.

Accident Costs

Accident prediction models were developed for this research that are specific to junction type. The accident data base used to calibrate these models was obtained from FHWA (via the Highway Safety Research Center at the University of North Carolina). This data base was subset from HSIS (*7*) to contain only nonurban junctions. The key feature of the HSIS is its inclusion of roadway design and traffic volume data. This added information permits the investigation of cause-and-effect relationships between geometry and traffic demand and accident frequency and severity.

The approach taken in developing an accident prediction model for this study was based on procedures described by Hauer et al. (*16*). Hauer has argued against using traditional least-squares regression of accident data because of violations of several assumptions on which this type of analysis is based. Instead, Hauer advocates the use of a general linear model [e.g., GLIM (*17*)] wherein these assumptions are removed, thereby yielding a better predictor of accident frequency as influenced by other factors.

Several factors were examined for their effect on accidents, including junction type, average daily traffic demand, speed limit, and median width. Unfortunately, correlation analysis among these variables indicated strong interrelationships between traffic demand, speed, and median width, which precluded their combined use in one accident model. Because traffic demand had the strongest correlation with accident frequency, it was selected for inclusion in the final accident model. The models developed using this approach for two-way stop-controlled intersections, signalized intersections, and interchanges are given by Equations 12, 13, and 14, respectively.

$$E(m)_U = 0.6503 * \left(\frac{T_m}{1{,}000}\right)^{0.2925} * \left(\frac{T_c}{1{,}000}\right)^{0.7911} \quad (12)$$

$$E(m)_S = 0.3603 * \left(\frac{T_m}{1{,}000}\right)^{0.7213} * \left(\frac{T_c}{1{,}000}\right)^{0.3663} \quad (13)$$

$$E(m)_I = 0.04864 * \left(\frac{T_m}{1{,}000}\right)^{1.337} \quad (14)$$

where

$E(m)$ = expected number of accidents per year,
T_m = major road traffic demand (vehicles/day), and
T_c = minor (cross) road traffic demand (vehicles/day).

Accident costs were obtained from an FHWA technical advisory (*8*). The technical advisory provides estimates of accident costs as of 1986 and is based on a willingness-to-pay approach. This approach includes the direct and indirect costs associated with the accident as well as the amount the typical individual is willing to pay to avoid harm. The costs per incident recommended in this advisory are as follows: fatality, $1,500,000/person; injury (overall average), $11,000/person; and property damage only (PDO), $2,000/vehicle.

These costs were combined with accident severity data to predict average accident costs for each junction type. The property cost component of the average accident cost was calculated by first inflating the PDO cost by 50 percent. This inflation stems from the fact that there is an average of 1.5 vehicles per accident on rural roadways in Nebraska (*18*). As a result of these computations, the average accident costs (updated to 1991 price levels) were determined as $45,500 at two-way stop intersections, $23,000 at signalized intersections, and $19,800 at interchanges.

The accident cost at two-way stop-controlled intersections is about twice that of the other two junction types. This disparity stems from the greater severity of accidents found at this type of intersection (*19*).

Road User Benefit Analysis

Sensitivity Analysis

Several variables were initially considered in this examination of sensitivity to determine which had the greatest potential impact on user benefits. The variables found to have the most significant impact were minor road traffic demand and discount rate. To a lesser extent, major road demand, traffic composition, and the annual traffic growth rate also had some influence on the amount of user benefits. Because it was the most influential, minor road demand was included in the sensitivity analyses of the other variables.

The sensitivity of road user benefit to minor road demand and discount rate is shown in Figure 1. The benefits shown are those derived from the operation of an interchange instead of a two-way stop-controlled intersection. Major road demand was set at 10,000 vpd. In general, an increase in the discount rate significantly decreases the road user benefits. Alternatively, the minimum minor road traffic demand needed to achieve the level of user benefit that justifies an interchange would be increased (on the basis of equating road user benefits to the incremental construction cost of the interchange alternative). As a result of this sensitivity, a common discount rate must be used when making comparative assessments of the economic worth of competing projects.

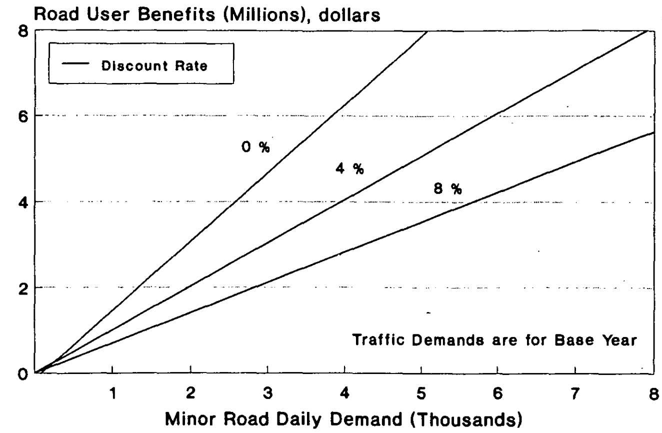


FIGURE 1 Road user benefits as affected by minor road demand and discount rate.

Road User Cost Components

Road user costs are composed of the operational and accident costs associated with the particular junction type. Operational costs can be further categorized as stopped delay, acceleration-deceleration delay, idling fuel consumption, and acceleration-deceleration running costs (e.g., fuel, brakes, and tires). The percentage contribution of each component to the total user cost is shown in Figure 2.

As this figure indicates, acceleration-deceleration running costs constitute the major contributor to user costs. These costs generally range from 40 to 75 percent of the total road user costs, depending on junction type. Accident costs are the next biggest contributor, ranging from 15 to 50 percent. In contrast, stopped delay, acceleration-deceleration delay, and idling fuel collectively do not contribute more than 20 percent to the road user costs.

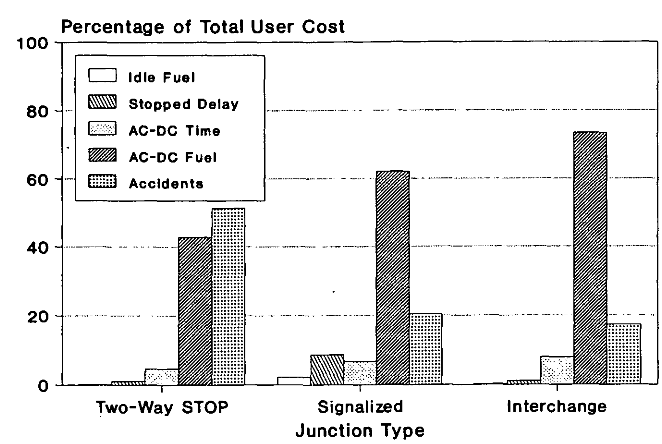


FIGURE 2 Typical road user cost components for each junction type.

INTERCHANGE GUIDELINES

Approach

The approach taken in developing the interchange guidelines is based on a benefit-cost (B/C) ratio analysis. This approach was preferred to a net present value calculation because the dimensionless character of the B/C ratio made it more applicable to the development of the interchange guidelines. On the basis of the results of the sensitivity analysis, it was determineu that the guidelines should include both major and minor traffic demands in the base year as independent variables. The base year is defined as the year the new interchange or intersection would be opened to traffic.

The analysis indicated that interchanges were more economically viable than signalized intersections as an alternative to two-way stop-controlled intersections on rural expressways. The signalized intersection's main benefit is a reduction in accidents; however, this benefit appears to be negated by the signal's higher operational costs whenever the minor road demand is less than one-half that of the major road. Moreover, even when this demand ratio is exceeded, the interchange consistently yields about $4 million more in benefits than the signalized intersection. This incremental benefit is generally sufficient to yield a B/C ratio of 2 to 3 over the signalized intersection under the full range of traffic demands. As a result, the interchange appears to be a more viable alternative than the signalized intersection on the basis of the economic and traffic conditions assumed for this study.

In recognition of the large degree of uncertainty in the estimate of the road user cost components, a range of B/C ratios was considered in the development of the guidelines. In particular, a range of major/minor road traffic demand combinations were evaluated using the computerized analysis methodology to find combinations that would yield B/C ratios of 1.0 and 2.0. Demand combinations that result in a B/C ratio of 2.0 imply that the road user benefits associated with a specific alternative design outweigh the incremental costs of this alternative by a factor of 2.0. Experience with the methodology suggests that this "factor of safety" of 2.0 should provide a traffic demand threshold that minimizes the uncertainty associated with the assumptions made for the analysis. Traffic demand combinations that result in B/C ratios between 1.0 and 2.0 suggest that more detailed, site-specific examination is needed to determine whether special circumstances exist that are contrary to the assumptions made in the analysis.

Guidelines

The benefit-cost approach used to develop the guidelines focuses only on economic factors. The decision to implement a particular alternative should also consider its social and environmental implications. In this regard, the results of the economic analysis should be considered as only one component of a more comprehensive alternative evaluation process. The guidelines herein are directed toward rural expressways and are based on economic and traffic conditions in Nebraska. All of the aforementioned assumptions and unit prices (particularly those for accidents) should be reviewed before application in other areas.

The analysis considered three geometric scenarios for the junctions. The first assumes that the intersection and the interchange have four approach legs and a two-lane cross section on the minor road. This is the most common configuration for both junction types in rural areas. The traffic demand-based guideline for this interchange scenario is shown in Figure 3.

Two lines are shown in this figure, which divide the graph into three regions. The lower line represents the base year traffic demand combinations that result in a B/C ratio of 1.0. The upper line represents the demand conditions coinciding with a B/C ratio of 2.0. The region in which the combination of major and minor road traffic demands fall indicates the recommended action. If the combination falls above the upper line, the interchange would be economically justified in most situations. If the demand combination falls below the lower line, there is not sufficient traffic demand to economically justify the cost of the interchange.

If the demand combination falls between the upper and lower lines, a more detailed examination is needed to ascertain the economic need for an interchange. This examination should determine if the assumptions made in this analysis are representative of the particular location. If they are, the interchange may be economically justified. If they are not, a benefit-cost analysis using the site-specific conditions may be needed to determine whether an interchange is justified.

Examination of Figure 3 indicates that road user benefits are relatively insensitive to major road traffic demand. Minor road demands of 2,000 and 4,000 vpd appear to define minimum threshold values for junctions with major road demands ranging from 4,000 to 15,000 vpd. These findings are similar in trend and magnitude to those of Van Every (*4*), who recommended using minor road demands of about 3,000 and 4,000 vpd for B/C ratios of 1.0 and 2.0, respectively.

The second scenario assumes that the conventional diamond interchange will have four approach legs and a four-lane cross section on the minor road. The traffic demand-based guideline for this scenario is shown in Figure 4. The general trend of the B/C lines in Figure 4 is similar to those in Figure 3; however, the higher cost of the four-lane cross section has shifted the lines upward toward higher minor road

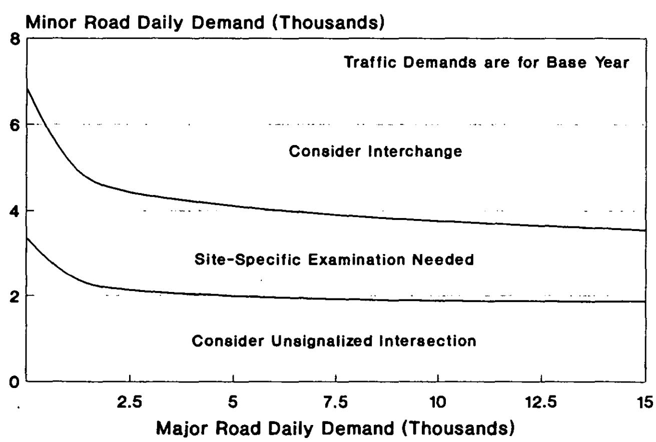


FIGURE 3 Interchange guidelines based on four approach legs and a two-lane minor road.

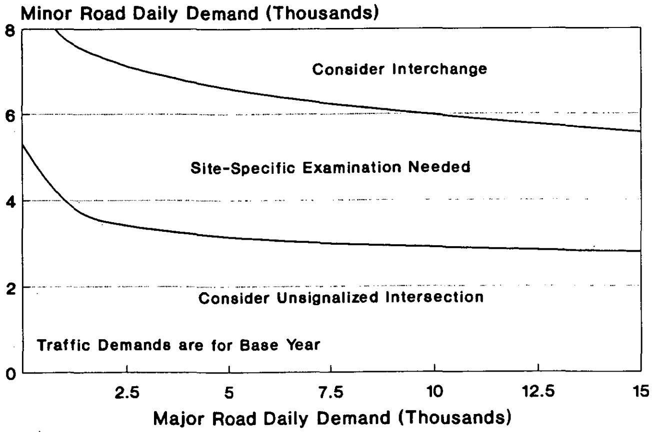


FIGURE 4 Interchange guidelines based on four approach legs and a four-lane minor road.

volumes. This shift stems from the need for more minor road demand (and related operating costs) to justify the higher construction cost of a four-lane bridge.

The third scenario assumes that the intersection and the interchange have three approach legs (i.e., a T-junction) and a two-lane cross section on the minor road. The traffic demand-based guideline for this scenario is shown in Figure 5. The general trend in the B/C lines in Figure 5 is also similar to those in Figure 3; however, the absence of a through movement on the minor road eliminates some of the user costs that are incurred at a four-leg intersection. As a result, more total minor road traffic (representing only turn movements) is needed to justify the cost of the interchange.

The analysis of a three-leg junction was conducted using the same methodology as that applied to the four-leg junction. The only difference was that one approach leg was specified as having zero demand. The equations for calculating the operating costs are valid for this application; however, the equations for predicting accidents become more of an approximation because they were derived from data for four-leg intersections. In spite of this limitation, the results appear to be generally consistent with experience and should be considered representative.

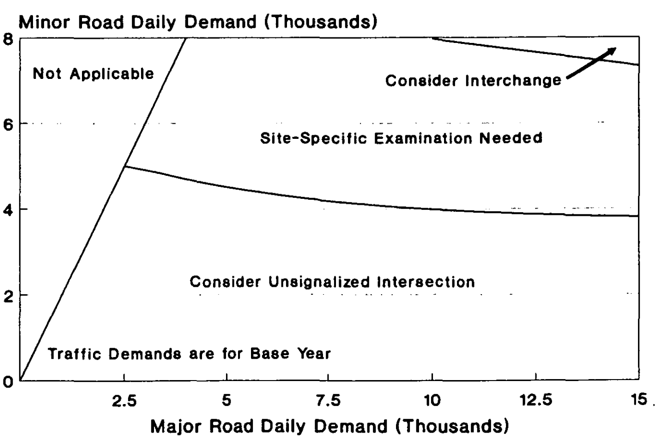


FIGURE 5 Interchange guidelines based on three approach legs and a two-lane minor road.

A fourth region, labeled "Not Applicable," is also shown in Figure 5. This region represents the traffic demand combinations that are not possible at T-junctions on the basis of the assumed 50/50 directional split in traffic demand on both intersecting roadways.

CONCLUSIONS

Examination of the benefit-cost analysis procedure indicated a strong sensitivity to minor road demand and discount rate. In general, road user benefits increased with increasing minor road demand and decreasing discount rate. Major road demand, traffic composition, and annual traffic growth rate did not have as strong an influence on user benefits.

The benefit-cost analysis of the signalized intersection and interchange indicated that the interchange was a more economically viable alternative than the signalized intersection. The signalized intersection's main benefit is a reduction in accidents; however, this benefit appears to be negated by the signal's higher operational costs whenever the minor road demand is less than one-half that of the major road.

Three geometric scenarios were formulated for the intersection and interchange. The first considered a four-leg junction with a two-lane minor road. The second considered a four-leg junction with a four-lane minor road. The third considered a three-leg junction with a two-lane minor road. Three figures were developed relating the major and minor road daily traffic demands that would economically justify an interchange in terms of benefit-cost ratios of 1.0 and 2.0. When the major road demand exceeds about 4,000 vpd, the minor road demands that provide a 2.0 benefit-cost ratio are about 4,000, 6,500, and 8,000 vpd for the three scenarios, respectively.

ACKNOWLEDGMENTS

This paper presents a portion of the findings from a research project conducted by the authors for the Nebraska Department of Roads. The authors wish to express their appreciation to the individuals and agencies who assisted in making the project successful. These individuals include Rick Ernstmeyer and Richard Ruby of the Nebraska Department of Roads, who provided the traffic and maintenance cost data. The traffic accident data were provided by Jeffrey F. Paniati of FHWA and Forrest M. Council of the University of North Carolina. A special thank-you is extended to them for their efforts. The authors would also like to acknowledge the assistance of Brian Moen and Jim Kollbaum, students at the University of Nebraska-Lincoln.

REFERENCES

1. J. A. Bonneson, P. T. McCoy, and J. E. Truby. Intersections on Rural Expressways. Presented at 70th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1991.
2. C. W. Ockert and C. R. Walker. Criteria To Be Used in Developing Warrants for Interchanges on Rural Expressways. In *Highway Research Record 244*, HRB, National Research Council, Washington, D.C., 1968, pp. 69–89.

3. *Manual on Uniform Traffic Control Devices*. FHWA, U.S. Department of Transportation, Washington, D.C., 1988.
4. B. E. Van Every. A Guide to the Economic Justification of Rural Grade Separations. *Australian Road Research*, Vol. 12, No. 3, Australian Road Research Board, Australia, pp. 147–154.
5. *A Manual on User Benefit Analysis of Highway and Bus-Transit Improvements*. AASHTO, Washington, D.C., 1977.
6. *Special Report 209: Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 1985.
7. F. M. Council and J. F. Paniati. The Highway Safety Information System. *Public Roads, A Journal of Highway Research and Development*, Vol. 54, No. 3. FHWA, Washington, D.C., Dec. 1990, pp. 234–239.
8. *Technical Advisory T7570.1*. FHWA, U.S. Department of Transportation, Washington, D.C., June 1988.
9. *1990 State Highway Plan and Highway Needs Report—Volume 3: Inventory*. Nebraska Department of Roads, Lincoln, 1990.
10. E. Hauer, E. Pagitsas, and B. T. Shin. Estimation of Turning Flows from Automatic Counts. In *Transportation Research Record 795*, TRB, National Research Council, Washington, D.C., 1981, pp. 1–7.
11. *1990 Continuous Traffic Count Data and Traffic Characteristics on Nebraska Streets and Highways*. Nebraska Department of Roads, Lincoln, April 1991.
12. J. L. Pline (ed.). *Traffic Engineering Handbook*, 4th ed. Prentice-Hall, Englewood Cliffs, New Jersey, 1992, p. 54.
13. W. S. Homburger, L. E. Keefer, and W. R. McGrath (eds.). *Transportation and Traffic Engineering Handbook*, 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey, 1982, pp. 166–168.
14. J. C. Tanner. A Theoretical Analysis of Delays at an Uncontrolled Intersection. *Biometrica*, Vol. 49, 1962, pp. 163–170.
15. K. Fitzpatrick, J. M. Mason, and D. W. Harwood. Comparison of Sight Distance Procedures for Turning Vehicles from a Stop-Controlled Approach. Presented at 70th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1991.
16. E. Hauer, J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
17. P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, New York, 1983.
18. *Standard Summary of Nebraska Motor Vehicle Traffic Accidents*. Highway Safety Division, Nebraska Department of Roads, Lincoln, 1983–1988.
19. J. A. Bonneson and P. T. McCoy. *Interchange vs. At-Grade Intersection on Rural Expressways*. Research Report TRP-02-25-91. Nebraska Department of Roads, Lincoln, May 1992.

Publication of this paper sponsored by Committee on Transportation Economics.

Welfare Maximization with Financial Constraints for Bus Transit Systems

SHYUE KOONG CHANG AND PAUL M. SCHONFELD

A bus system with time-dependent demand and supply characteristics is analytically optimized to maximize a welfare objective, subject to financial constraints. With some approximations, equations for optimal route spacing, headways in various periods, and fares are obtained for unconstrained, break-even, and subsidy cases. The relationships between the optimized decision variables and system parameters are thus identified analytically. A numerical example is given for a bus transit system with three service periods. In the vicinity of the maximum welfare solution, the welfare is found to be relatively flat with respect to subsidies. Since subsidy increments yield disproportionately smaller welfare increments, break-even or constrained subsidy solutions may be preferable to pure welfare maximization. A minimum allowable ratio of welfare change to subsidy change is suggested as a criterion for optimizing individual bus systems and efficiently allocating resources among alternatives.

Multiple-period analytic optimization models have been developed for analyzing bus systems (*1*). To extend that work, this paper presents analytic models for optimizing bus systems using a maximum welfare objective subject to constraints on allowable subsidies, including break-even constraints.

Studies on analytic optimization models have been extensively reviewed by Chang and Schonfeld (*1,2*). Such models for public transportation system optimization have most often assumed a perfectly inelastic demand (*2–12*). This assumption may be reasonable for some systems and may simplify models to the point at which analytic solutions may be obtained. However, this assumption may preclude the model from analyzing pricing policy and subsidy issues or from optimizing the system for objectives, such as maximum net social benefit or profit.

Kocur and Hendrickson (*13*) have analyzed bus services with demand elasticity and developed closed-form solutions for optimal route spacing, headway, and fare with various objective functions. This was accomplished with some approximations, most notably with a linear approximation of a logit mode split model. Nash (*14*) assessed alternative objectives for bus transit service in terms of fares, service levels, and financial results. Frankena (*15*) investigated the conditions under which a maximum ridership objective would be economically efficient and analyzed the effects of subsidy on system efficiency. Else (*16*) analyzed optimal fares and subsidies while considering various externalities. Bly and Oldfield (*17*) investigated analytically the effects of subsidy on bus operation. An analytic model considering demand elasticity, financial constraints, and congestion effects also has been developed by Oldfield and Bly (*18*) to determine the optimal vehicle size for urban bus systems. None of these studies considered time-dependent supply and demand characteristics or tradeoffs between subsidies and a welfare objective.

Many studies have used numerical instead of analytic methods to optimize public transportation systems and to investigate issues covered in this paper (*19–22*). However, an analytic approach is used here to find closed-form solutions for the decision variables and objective function and to identify some related optimality conditions.

In this paper, the analytic models for bus systems assumed by Chang and Schonfeld (*1*) are applied to analyze unconstrained, break-even, and subsidy cases. Three decision variables—namely, route spacing, headway, and fare—are optimized jointly in those models. The system assumptions are briefly reviewed in the next section, and then the welfare objective function is formulated. The following section presents the analytic results for the various cases and discusses relations and implications of the analytic results. The section after that presents a numerical case. The final section presents conclusions.

SYSTEM ASSUMPTIONS

The assumed bus system, which can represent a variety of transit operations, is taken from Chang and Schonfeld (*1*). A brief description of the assumed system follows.

Bus System Characteristics

In this analysis, a branched-zone bus system is assumed to provide service for a rectangular area with dimensions $L \times W$, from which trip ends are assumed to access a single point, such as a mass transit station or activity center. Figure 1 illustrates this bus system. The variables and their typical baseline values are defined in Table 1.

The service area has N zones, each of length L and width $r = W/N$. A vehicle round trip to zone j during period t consists of (a) a line haul distance J traveled at express speed yV_t from the starting point to a corner of service area; (b) a distance of W_j km traveled at local nonstop speed bV_t from the corner to the assigned zone; (c) a collection route L km traveled at local speed V_t along the middle of the zone stopping for passengers every d km; and (d) a retracing in reverse order of the first three stages.

S. K. Chang, Department of Civil Engineering, National Taiwan University, 1 Roosevelt Road, Section 4, Taipei, Taiwan, R.O.C. P. M. Schonfeld, Department of Civil Engineering, University of Maryland, College Park, Md. 20742.

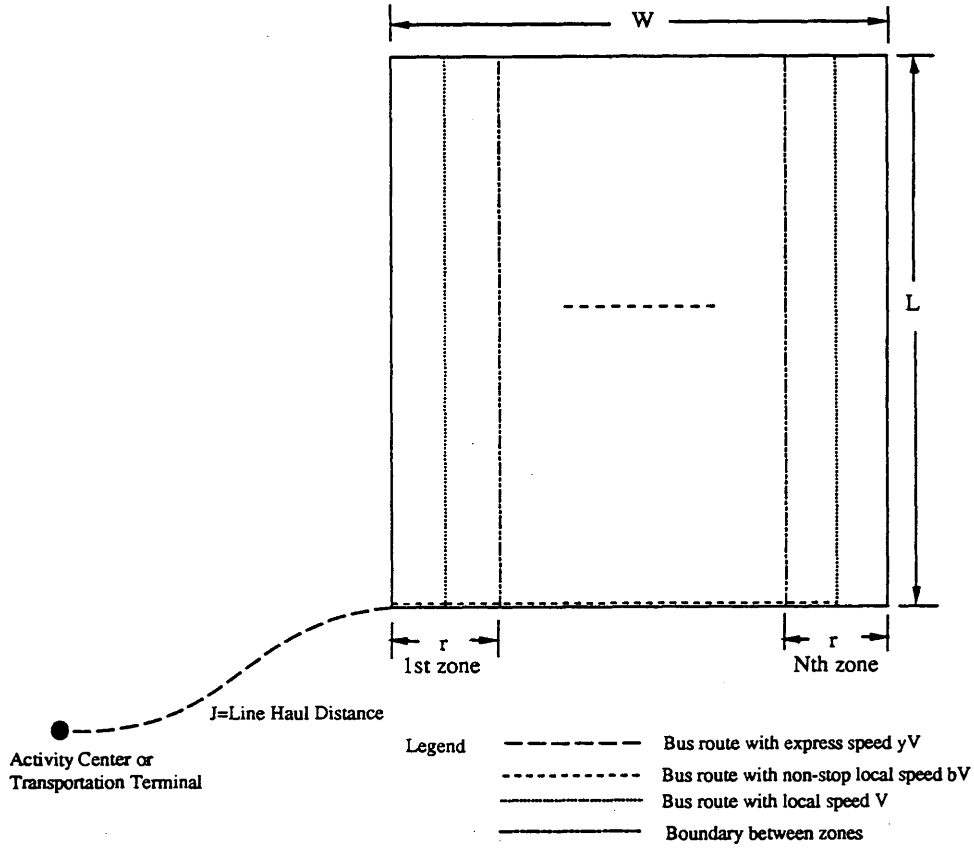


FIGURE 1 Bus system configuration.

Passengers are assumed to walk at speed g between their trip ends and the nearest bus stop along a rectangular street network (parallel and perpendicular to the feeder route) with negligible street spacing. This assumption implies an access distance of $(r + d)/4$ and an access time of $(r + d)/4g$.

The bus network assumed in this paper has been previously analyzed for maximum profit and welfare objectives without financial constraints (*1*). The bus route structure and model formulations may be used to analyze a wide variety of bus service types, including feeder services to and from transfer stations, zone structure services, and radial services to activity centers. They may also be used in some cases to analyze bus systems with many-to-many demand patterns if the system can be separated into subsystems in which many-through-one analysis is applicable [as discussed by Newell (*9*)].

Demand Functions

A linear demand function in which the demand density is sensitive to various travel time components and the fare is formulated as

$$Q_t = q_t\left[1 - e_w z_1 h_t - e_x z_2\left(\frac{r + s}{g}\right) - e_v M_t - e_p f\right] \qquad (1)$$

where

q_t = potential demand density of the bus service during each period;
$z_1 h_t$ = wait time, which is assumed to be constant factor z_1 (usually $z_1 = 0.5$ for uniform passenger arrivals at bus stops) multiplied by headway h_t;
$z_2[(r + s)/g]$ = average access time (usually $z_2 = 0.25$ for grid street networks with negligible street spacing);
M_t = average in-vehicle travel time;
f = fare; and
e_w, e_x, e_v, e_p = elasticity factors.

The optimizable decision variables are the headway (h_t) for each period, the route spacing (r), and the fare (f). This implies that the optimized bus route structure and fare are assumed to be the same in all periods, whereas the headways are optimized separately for each period.

The values of elasticity factors e_w, e_x, e_v, and e_p are not the actual elasticities in such a linear function. In addition, the ratios between the elasticity factors for wait time and fare (e_w/e_p), for access time and fare (e_x/e_p), and for in-vehicle time and fare (e_v/e_p) determine the implied values of wait time, access time, and in-vehicle time, respectively.

It is assumed that the potential demand q_t in each period could be determined from the time distribution of demand shown in Figure 2. A step demand distribution (Figure 2*b*) relating monotonically volume levels and their durations can be directly obtained from the empirical demand distribution (Figure 2*a*). Although only three periods are used as an example of the demand distribution in Figure 2*e*, the number

TABLE 1 Variable Definitions

Variable	Definition	Baseline Value
B_t	bus operating cost in period t; B_1, B_2, and B_3 = 50, 25, and 25 dollars/veh. hr.	
b	non-stop ratio = non-stop speed/local speed	2
C	system total cost (\$/period)	-
C_o	total operator cost (\$/period)	-
D_t	bus avg. round trip time during period t (hrs.) $=2L/V_t+W/bV_t+2J/yV_t$	
e_p	demand elasticity parameter for fare	0.07
e_v	demand elasticity parameter for in-vehicle time	0.35
e_w	demand elasticity parameter for wait time	0.7
e_x	demand elasticity parameter for access time	0.7
F_t	fleet size in period t (vehicles)	-
f	fare (\$)	-
G	consumer surplus (\$/period)	-
g	average walk speed (km/hour)	4.0
h_t	headway in period t (hrs/vehicle)	-
I_1	the lagrange multiplier associated with the break-even constraint	-
I_2	the lagrange multiplier associated with the subsidy constraint	-
I_s	$(I_2+1)/(2I_2+1)$	-
J	line haul distance (km)	6.4
K	subsidy (\$/period)	-
k_t	invariant components of the demand function $=1-e_x d/4g-e_v M_t$	-
L	length of service area (km)	4.8
M_t	passenger avg. trip time during period t (hrs.) $= L/2V_t+W/2bV_t+J/yV_t$	
m	number of time periods	3
P	profit (\$/period)	-
Q_t	demand function for period t (Eq. 1)	-
q_t	potential demand density in period t q_1, q_2, and q_3 = 50, 20, and 5 trips/sq. km/hour, respectively.	-
R	total revenue (\$/period)	-
r	route spacing (km)	-
s	stop spacing (km)	0.4
T_t	service time in period i (hours); T_1, T_2, and T_3 = 3, 3, 4 hours, respectively.	
V_t	local speed during period t (km/hour)	24.0
W	width of service area (km)	3.2
X	composite variable $= \Sigma_t T_t(D_t B_t q_t)^{1/2}/\Sigma_t T_t q_t$	-
Y	welfare (\$/period)	-
y	express ratio = express speed/local speed	2.0
z_1	ratio of wait time/headway	0.5
z_2	geometric factor for determining access time	0.25

and duration of periods are unlimited in the models and may be selected to represent variations over time with whatever precision is desired. Other models (*8,18*) have used smoothed functions (e.g., Figure 2*c*) to represent demand variation over time. However, such smoothing is not necessary to formulate objective functions that are twice differentiable and hence appears to be an unnecessary complication (*23*). The approach used in this work relies on step functions for demand, costs, speeds, and other variables that are obtained directly from the empirical data.

Operator Costs

The operator costs per analysis period (e.g., per day) is the fleet size multiplied by the hourly operating cost and total daily service time. The fleet size is the bus round-trip time divided by the headway. The bus round-trip times D_t are assumed for various periods because different traffic conditions are represented by different speeds:

$$D_t = \frac{2L}{V_t} + \frac{W}{bV_t} + \frac{2L}{yV_t} \tag{2}$$

The hourly operating costs B_t are assumed for different periods. The operator costs per day are

$$C_o = \sum_{t=1}^{m} F_t B_t T_t = \sum_{t=1}^{m} \frac{W D_t B_t T_t}{r\, h_t} \tag{3}$$

MAXIMUM WELFARE OBJECTIVE

Various objective functions have been considered appropriate for optimizing bus transit systems (*24*). In this paper maximum social welfare, also known as the net social benefit or simply welfare, is used as the objective function.

The welfare Y is the sum of the consumer surplus G and producer surplus P:

$$Y = G + P \tag{4}$$

The producer surplus, also known as profit, is the total revenue R minus the operator cost C_o:

$$P = R - C_o \tag{5}$$

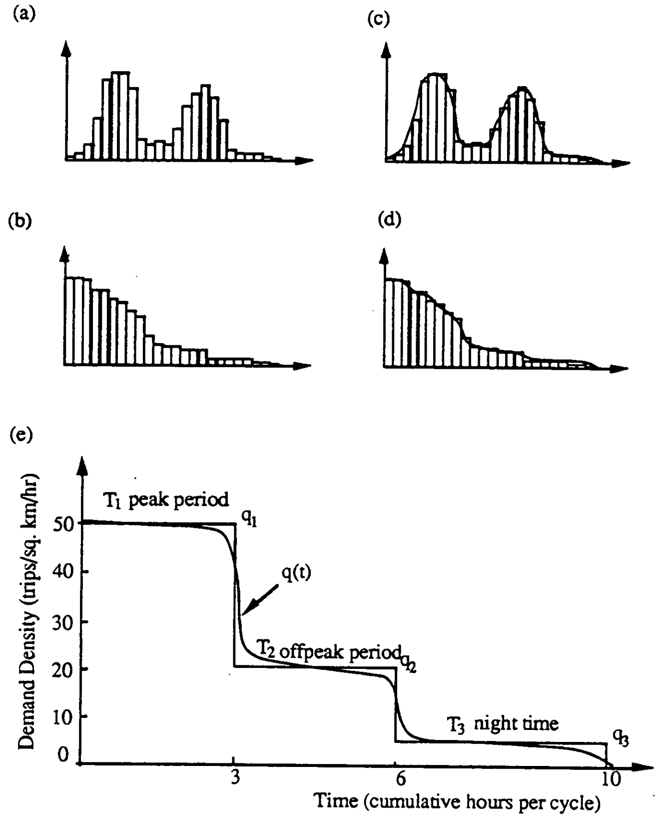


FIGURE 2 Distributions of demand over time: *a*, empirical demand distribution; *b*, transformed demand distribution; *c*, smoothed demand function; *d*, smoothed and transformed demand distribution; *e*, demand distribution assumed in numerical example.

The operator cost has been defined in Equation 3. The total revenue is the fare multiplied by the total demand:

$$R = \sum_{t=1}^{m} fLWT_t Q_t \tag{6}$$

Therefore the total profit can be formulated as

$$P = \sum_{t=1}^{m} fLWT_t q_t (k_t - e_w z_1 h_t - e_x z_2 r/g - e_p f) - \sum_{t=1}^{m} \frac{WE_t T_t}{r\, h_t} \tag{7}$$

where k_t is a constant representing a component of the potential demand that is insensitive to the decision variables optimized here:

$$k_t = 1 - e_x z_2 s/g - e_v M_t \tag{8}$$

By inverting the demand function shown in Equation 1 to find fare as a function of demand and by integrating the inverted function over the demand, the total social benefit can be obtained. Then, the consumer surplus G can be derived as the total social benefit minus the total cost that the users actually pay (*13,25*):

$$G = \left(\frac{LW}{2e_p}\right) \sum_{t=1}^{m} T_t q_t (k_t - e_w z_1 h_t - e_x z_2 r/g - e_p f)^2 \tag{9}$$

Therefore the welfare can be formulated by adding Equations 7 and 9. In the following sections the major analytic results for the various objectives are presented.

ANALYTIC RESULTS

The objective here is to maximize the welfare:
Maximize

$$Y = G + R - C_o \tag{10}$$

In solving this problem, a deficit constraint is considered that can be generally expressed as follows:

$$C_o \leq R + K \tag{11}$$

It means that the operator cost C_o should not be larger than the sum of the revenue R and subsidy K.

The problem is solved for three cases with different assumptions about the constraint and subsidy: (a) without the constraint; (b) with the constraint $K = 0$, that is, a break-even requirement; and (c) with the constraint $K \neq 0$. These three cases are presented separately.

Unconstrained Results

In the first case, the first-order conditions for an optimum are

$$\partial Y/\partial r = 0 \tag{12}$$

$$\partial Y/\partial h_t = 0 \qquad t = 1, 2, \ldots, m \tag{13}$$

$$\partial Y/\partial f = 0 \tag{14}$$

The optimized fare can be immediately obtained from Equation 14:

$$f^* = 0 \tag{15}$$

The zero fare result is not surprising because the marginal operator cost is 0 in the bus systems considered here. Similar zero-fare results have been discussed previously (*13,24*). However, the marginal costs would become positive if a capacity constraint were binding (*23*) or if certain congestion effects were modeled.

By substituting this result into Equations 12 and 13, the relation between the optimal headway h_t^* and optimal route spacing r^* is found to be

$$h_t^* \cong \left(\frac{E_t}{q_t}\right)^{1/2} \left(\frac{z_2 e_x r^*}{Xgz_1 e_w}\right) \tag{16}$$

The approximation errors in Equation 18 and some further results below are slight (*2*). Using this optimality result in

Equation 16 and the optimized fare obtained, the optimized route spacing is found to be

$$r^* \cong \left(\frac{X^2 g^2 z_1 e_w e_p}{z_2^2 e_x^2 L k}\right)^{1/3} \quad (17)$$

The optimized headway for each period can then be obtained by substituting Equation 17 into Equation 16:

$$h_t^* \cong \left(\frac{E_t}{q_t}\right)^{1/2} \left(\frac{z_2 e_x e_p}{X g L z_1^2 e_w^2 k}\right)^{1/3} \quad (18)$$

Since the optimal fare in this case is 0, the optimized welfare is simply the consumer surplus minus the operator cost.

Results with a Break-Even Constraint

In the previous case, the optimized fare is 0, which implies a deficit for the operator. Therefore, in the second case, a break-even constraint $C_o \leq R$ is considered.

The problem can be stated as follows:

Maximize

$$Y = G + R - C_o$$

Subject to

$$C_o - R \leq 0$$

A break-even solution would not exist if the demand function were always below the average operator cost function. That situation would imply a negative profit. Therefore, it is assumed in the following analysis that the travel demand is sufficient to yield a positive profit in some circumstances for the bus operation considered.

The Lagrange multipliers method is used here for constrained optimization, and the Lagrangian α is formulated as follows:

$$\alpha = G + R - C_o - I_1(C_o - R) \quad (19)$$

where I_1 is the Lagrange multiplier associated with the break-even constraint.

Solving the first-order conditions, as shown by Chang and Schonfeld (*2*), the following approximate relationship between headway and route spacing can be obtained:

$$h_t^* \cong \left(\frac{E_t}{q_t}\right)^{1/2} \left(\frac{z_2 e_x r^*}{X g z_1 e_w}\right) \quad (20)$$

This relationship is identical to that obtained for the previous case.

The following results for optimized route spacing and headway can also be derived (*2*):

$$r^* \cong 1.12\left(\frac{X^2 g^2 z_1 e_w e_p}{z_2^2 e_x^2 L k}\right)^{1/3} \quad (21)$$

$$h_t^* \cong 1.12\left(\frac{E_t}{q_t}\right)^{1/2} \left(\frac{z_2 e_x e_p}{X g L z_1^2 e_w^2 k}\right)^{1/3} \quad (22)$$

However, the optimized fare is no longer 0. It is found to be

$$f^* = \left(\frac{e_x}{e_p}\right)\left(\frac{z_2 r^*}{g}\right) \quad (23)$$

Since the ratio (e_x/e_p) represents the value of access time and $z_2 r^*/g$ is the average lateral access distance, the result indicates that the fare is identical to the lateral access cost in the optimized system.

The optimal revenue R^* and operator cost C_o^* are equal at break even, and their solutions are shown by Chang and Schonfeld (*2*).

Results with Subsidy

To maximize welfare subject to a subsidy constraint, the problem can be stated as follows:

Maximize

$$Y = G + R - C_o$$

Subject to

$$C_o - R \leq K$$

This problem is also solved using the Lagrange multipliers method. The Lagrangian α is formulated as

$$\alpha = G + R - C_o - I_2(C_o - R - K) \quad (24)$$

in which I_2 is the Lagrange multiplier associated with the subsidy constraint, and K is subsidy or maximum allowable deficit. Solving the first-order conditions, as shown by Chang and Schonfeld (*2*), the relationship between the optimal headway and route spacing can be obtained:

$$h_t^* \cong \left(\frac{E_t}{q_t}\right)^{1/2} \left(\frac{z_2 e_x r^*}{X g z_1 e_w}\right) \quad (25)$$

The relations among the fare, the route spacing, and the Lagrange multiplier are also derived:

$$f = \left(\frac{I_2}{2I_2 + 1}\right)\left(\frac{k}{e_p} - \frac{2 z_2 e_x r}{g e_p}\right) \quad (26)$$

$$r = \left(\frac{X^2 g^2 z_1 e_w e_p (2I_2 + 1)}{z_2^2 e_x^2 L k (1 - 3 z_2 e_x r/gk)(I_2 + 1)}\right)^{1/3} \quad (27)$$

Equation 25 shows that the proportionality relation between the optimized headway and route spacing can still be obtained. Solving with an approximation (*2*), the Lagrange multiplier I_2 is found to be

$$I_2^* = \frac{-\mu_2 + [\mu_2^2 - 4\mu_3(\mu_2 - \mu_1)]^{1/2}}{2(\mu_2 - \mu_1)} \quad (28)$$

in which μ_1, μ_2, and μ_3 are defined as follows:

$$\mu_1 = \left(\frac{X^2 z_1 z_2 e_w e_x}{g k L e_p^2}\right)^{1/3} \quad (29)$$

$$\mu_2 = \frac{4K}{kQ_T} + \left(\frac{k}{e_p}\right)[1 - 2e_p\mu_1]^2 - 3\mu_1 \qquad (30)$$

$$\mu_3 = \frac{K}{kQ_T} - \mu_1 \qquad (31)$$

The positive root in Equation 28 is the solution of the quadratic equation whenever the constraint is binding, because the negative root yields negative values of shadow price I_2.

Therefore, the optimized route spacing can be obtained by substituting Equation 28 into Equation 27. The optimized headway and fare can then be obtained by substituting the optimized route spacing into Equations 25 and 26, respectively. The optimal operator cost C_o^*, revenue R^*, and consumer surplus G^* can also be obtained (*2*). The welfare can be derived as the consumer surplus minus the subsidy: $Y^* = G^* - K$.

Discussion of Analytic Results

Analytic results concerning the design variables, including route spacing, headway, or fare, or all of these, are summarized in Table 2. These results show more clearly than any numerical results what the relationships among variables should be for optimized bus services under various objectives.

To a large extent the sensitivity of optimized design variables to the various system parameters can be determined by visually inspecting the functions rather than by numerical analysis. Taking, for example, systems optimized for maximum welfare subject to a break-even constraint, we can observe that the headway is proportional to the ½ power of the bus operating cost and round trip time ($E_t = B_tD_t$) and to the ⅓ power of the access time and fare elasticity factors. Similar sensitivity relations for the optimal route spacing and fare also can be observed (*2*).

For unconstrained welfare maximization, the optimized fare is 0, whereas for break-even welfare maximization the optimized fare is equal to the lateral access cost. When marginal costs are 0, such zero-fare results are expected and are similar to those obtained in previous works (*1,13,24*).

The results also show the optimality of a constant ratio between route spacing and headway. Such results are very similar to findings in several previous transit system optimization studies (*8,10,13*), except that a time-dependent factor $(E_t/q_t)^{1/2}$ can now be incorporated. This means that the following relation for feeder systems always holds for all periods:

$$h_t^*\left(\frac{q_t}{D_tB_t}\right)^{1/2} = \left(\frac{z_2e_x}{Xgz_1e_w}\right)r^* \qquad t = 1, 2, \ldots, m \qquad (32)$$

Equation 32 was rewritten from Equations 16, 20, and 25; time-dependent headways and parameters, such as q_t, B_t, and D_t, are combined on the left side of Equation 32.

The analytic models presented here have not considered a vehicle capacity constraint. Vehicles may be overloaded in some cases unless a vehicle capacity constraint is applied. Chang and Schonfeld (*2*) show how these results can be modified to satisfy vehicle capacity or load factor constraints.

NUMERICAL EVALUATION

Numerical examples for various cases under different objectives are presented and compared to illustrate the applicability of the models developed. The baseline parameter values shown in Table 1 are used in numerical examples. The numerical results are presented in Table 3. These results are computed for a 6.4- × 4.8-km rectangular service area with a three-period demand pattern in which potential demand densities are 50, 20, and 5 trips/km²/hr during service periods of 3, 3, and 4 hr, respectively. These conditions have been shown in

TABLE 2 Analytically Optimized Decision Variables

unconstrained case	break-even case	subsidy case
$r^* = \left(\frac{X^2g^2z_1e_we_p}{z_2^2e_x^2Lk}\right)^{\frac{1}{3}}$	$r^* \cong 1.12\left(\frac{X^2g^2z_1e_we_p}{z_2^2e_x^2Lk}\right)^{\frac{1}{3}}$	$r^* \cong \left(\frac{X^2g^2z_1e_we_p}{z_2^2e_x^2LkI_s^*}\right)^{\frac{1}{3}}$
$h_t^* = \left(\frac{E_t}{q_t}\right)^{\frac{1}{2}}\left(\frac{z_2e_xe_p}{XgLz_1^2e_w^2k}\right)^{\frac{1}{3}}$	$h_t^* \cong 1.12\left(\frac{E_t}{q_t}\right)^{\frac{1}{2}}\left(\frac{z_2e_xe_p}{XgLz_1^2e_w^2k}\right)^{\frac{1}{3}}$	$h_t^* \cong \left(\frac{E_t}{q_t}\right)^{\frac{1}{2}}\left(\frac{z_2e_xe_p}{XgLz_1^2e_w^2kI_s^*}\right)^{\frac{1}{3}}$
$h_t^* = \left(\frac{E_t}{q_t}\right)^{\frac{1}{2}}\left(\frac{z_2e_xr^*}{Xgz_1e_w}\right)O_t$	$h_t^* = \left(\frac{E_t}{q_t}\right)^{\frac{1}{2}}\left(\frac{z_2e_xr^*}{Xgz_1e_w}\right)\Psi_t$	$h_t^* = \left(\frac{E_t}{q_t}\right)^{\frac{1}{2}}\left(\frac{z_2e_xr^*}{Xgz_1e_w}\right)\Psi_t$
$f^* = 0$	$f^* = \frac{z_2e_xr^*}{ge_p}$	$f^* = \left(\frac{I_2^*}{2I_2^*+1}\right)\left(\frac{k}{e_p} - \frac{2z_2e_xr^*}{ge_p}\right)$

TABLE 3 Numerical Results for Bus Systems

	unconstrained		break-even	subsidy [a] 1,500	subsidy [a] 3,000
Route Spacing (km)	1.715	(1.715)[b]	1.933	1.904	1.808
Fare ($)	0		1.21	0.82	0.51
Headway (hours)	0.199	(0.152)	0.225	0.220	0.116
	0.203	(0.203)	0.260	0.226	0.209
	0.364	(0.364)	0.411	0.403	0.324
Fleet Size (no. of vehicles)	23	(30)	18	19	21
	19	(19)	13	15	17
	9	(9)	7	7	8
Demand Density (trips/sq. km/hour)	29.26	(30.04)	24.00	25.83	27.19
	12.31	(12.31)	10.19	10.96	11.50
	2.91	(2.91)	2.46	2.61	2.76
Avg. Cost ($/trip)	6.538		7.123	6.973	6.758
Avg. User Cost ($/trip)	5.213		5.915	5.754	5.414
Avg. Operator Cost ($/trip)	1.325		1.208	1.219	1.284
Avg. Wait Cost ($/trip)	1.072		1.208	1.190	1.130
Operator Cost ($/day)	5,565	(6,627)	4,250	4,523	5,016
Revenue ($/day)	0		4,243	3,043	1,992
Profit ($/day)	-5,565		-7	-1,480	-3,024
Consumer Surplus ($/day)	19,025	(19,684)	13,297	14,861	16,461
Welfare ($/day)	13,464	(13,057)	13,290	13,381	13,437
Bus Load (passengers/veh.)	64		67	67	66
	25		33	30	29
	12		13	13	13
Shadow Price	-		0.19	0.07	0.11

a. Results for subsidies of 1,500/3,000 dollars per day, respectively.
b. Results with vehicle capacity constraint.

Figure 2*e*. The bus operating costs during these three periods—peak, offpeak, and night—are assumed to be $50, $25, and $25/vehicle-hr, respectively.

1. The optimal fare is 0 in the unconstrained case, $1.21 in the break-even case, and $0.82 if a subsidy of $1,500/day is provided. That subsidy represents about 30 percent of the daily operator cost. When the subsidy increases to $3,000/day, the optimal fare becomes $0.51.
2. For break-even welfare maximization, the consumer surplus and welfare are $13,297 and $13,290/day, respectively. Deficits for the constrained cases have small deviations from the theoretical results. For example, deficit (or profit) for the break-even case should be 0, whereas we obtain $7/day; when subsidy is constrained to $3,000 the profit should be −$3,000, whereas we obtain −$3,024. These deviations are the result of minor approximations in the analytic solutions.
3. In comparing unconstrained and break-even cases, we find that when the break-even constraint is removed, the deficit increases from 0 to $5,565, whereas the welfare rises by $174 (1.3 percent of $13,290). Thus, at least for the typical parameter values used in this analysis, the financial and political advantages of a break-even policy are quite strong.
4. When the subsidies increase, the optimal route spacing and headway decrease, and the fleet size increases. The equilibrium demand, consumer surplus, and welfare also increase.
5. The shadow price associated with the break-even constraint is $0.19, indicating that welfare would be increased by $0.19 if the deficit were increased from 0 to $1. The shadow prices associated with the subsidy constraints are $0.11 and $0.07 for subsidies of $1,500 and $3,000, respectively. These indicate that the welfare would be increased by $0.11 and $0.07 if the subsidies of $1,500 and $3,000 were increased to $1,501 and $3,001, respectively.

Basically, in the vicinity of the unconstrained welfare maximization solution, that is, as the subsidy approaches $5,565, the shadow price approaches 0. These relationships can be shown conceptually in Figure 3. It is shown that the optimal welfare Y^* is obtained with a subsidy of K^*. In the numerical results (Table 3) the optimal welfare of $13,464 is obtained with a deficit (and subsidy) of $5,565 in the unconstrained case. In the break-even case, the subsidy is 0, and the welfare is Y_B, which in the numerical results is $13,290. In Figure 3 the welfare becomes Y_1 when the subsidy is K_1. The subsidy constraint is not binding whenever it is to the right of the maximum social welfare point (e.g., $K_2 > K^*$). As long as a subsidy K_1 is binding, the slope of the curve $\Delta Y/\Delta K$ at $K = K_1$ may be interpreted as the shadow price of the subsidy

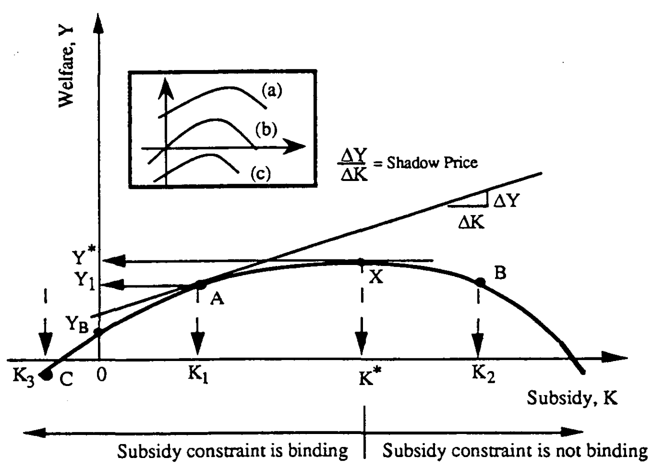


FIGURE 3 Conceptual relation between subsidy and welfare.

constraint. This shadow price $\Delta Y/\Delta K$ indicates how much opportunity there is to increase the welfare Y per increment of subsidy ΔK. Given some information about society's willingness to subsidize, which may also be expressed as a $\Delta Y/\Delta K$ ratio (i.e., a minimum acceptable welfare increment per unit of subsidy, or a minimum acceptable rate of return to subsidy), this indicator may be used as an allocation criterion to determine how much to subsidize a particular activity. Thus for our bus system the subsidy K would be increased as long as the slope $\Delta Y/\Delta K$ in Figure 3 exceeded a minimum acceptable $\Delta Y/\Delta K$ rate. This approach allows an efficient allocation of resources among various transportation and nontransportation activities. Figure 3 (Curve *a* in the inset) shows that this approach may sometimes imply a negative subsidy (i.e., a profit) in cases in which the welfare Y is positive at negative values of the subsidy K.

Figure 3 (Curve *a* in the inset) also indicates that if the social welfare function would shift downward to reflect higher operator cost functions or lower demand functions (e.g., from *a* to *c* in Figure 3), the maximum possible solutions for profit (Point *C*), break-even welfare (Y_B), subsidy-constrained welfare (*A*), and unconstrained welfare (*X*) would gradually become negative, in that order.

The results in Table 3 suggest that the break-even case may well be preferable to the unconstrained welfare maximization case, because by removing the break-even constraint the welfare rises by only $174 (1.3 percent), whereas the deficit increases from 0 to $5,565. It is desirable to examine the sensitivity of this result to the elasticity factors used in the demand function.

The peak-period busloads in Table 3 exceed the capacity of standard buses. The corresponding capacity constrained results can be obtained with the analytic models developed by Chang and Schonfeld (*2*) and are also presented for pure welfare maximization cases in Table 3. These results eliminate the overload problem in the peak period.

Table 4 shows the welfare results for the two cases with several fare elasticity factors. The comparison shows that the welfare rises by only 1.3 and 1.6 percent, whereas the deficits rise from 0 to $5,565 and $6,971 for elasticity factors of 0.07 and 0.05, respectively. These results suggest that in such cases, in which large increases in subsidies are required for such smaller increases in welfare, operators (and taxpayers) may find the break-even objective preferable.

Table 4 also indicates the effects of the fare elasticity factor on the optimal results. As expected from analytic results for route spacings, when the fare elasticity factors decrease from 0.07 to 0.05, the optimal route spacings decrease from 1.715 to 1.534 km for the unconstrained case, and from 1.933 to 1.718 km for the break-even case. The optimal fares are all 0 for the unconstrained case and increase from $1.21 to $1.50 when the fare elasticity factors decrease from 0.07 to 0.05.

It is also worth presenting the effects of subsidy on the optimized results, because these effects may not be visually perceived from the analytic results in the subsidy case. Figure 4 shows that in maximizing welfare the optimal fare varies inversely with the subsidy. Figure 5 shows the effects of subsidy on consumer surplus and welfare. The consumer surplus increases with the subsidy, whereas the producer surplus (profit) is the subsidy with a negative sign. These relationships have the net effect that the optimal welfare is very flat over a wide range of subsidy values. These results indicate that the break-even objective in welfare maximization may be quite acceptable, because it yields a zero deficit and only slightly less welfare than the unconstrained case. The discussion of Figure 3 also suggests how other solutions, with less than maximum welfare, corresponding to smaller levels of subsidy, or even profits, may be found preferable on the basis of the minimum

TABLE 4 Effects of Fare Elasticity Factors

(a) Case 1: Unconstrained Welfare Maximization

fare elast. factor	route spacing	consumer surplus	profit	operator cost	welfare	fare
0.05	1.534	27,989	-6,971	6,971	21,018	0
0.06	1.630	22,724	-6,137	6,137	16,587	0
0.07	1.715	19,029	-5,565	5,565	13,464	0

(b) Case 2: Break-Even Welfare Maximization

fare elast. factor	route spacing	consumer surplus	profit	operator cost	welfare	fare
0.05	1.718	20,682	0	5,557	20,682	1.50
0.06	1.826	16,375	0	4,921	16,375	1.33
0.07	1.933	13,297	-7	4,243	13,290	1.21

(c) Comparison

fare elasticity factor	unconstrained case welfare (1)	unconstrained case operator profit (2)	break-even case welfare (3)	break-even case operator profit (4)	change in welfare (1) - (3)
0.05	21,018	-6,971	20,682	0	336(1.6%)
0.06	16,587	-6,137	16,375	0	212(1.3%)
0.07	13,464	-5,565	13,290	0	174(1.3%)

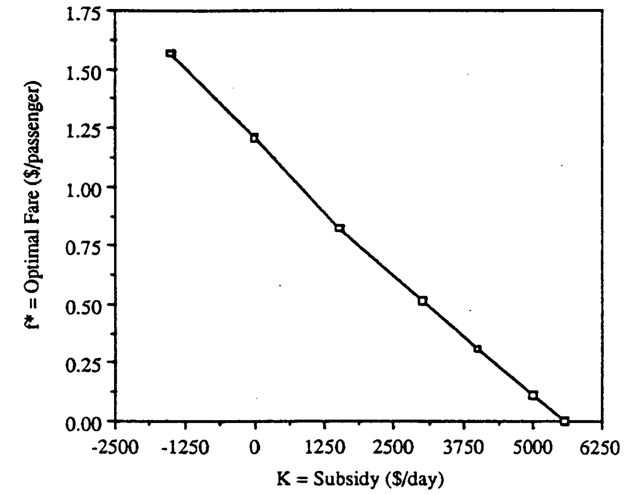


FIGURE 4 Effects of subsidy on optimal fare.

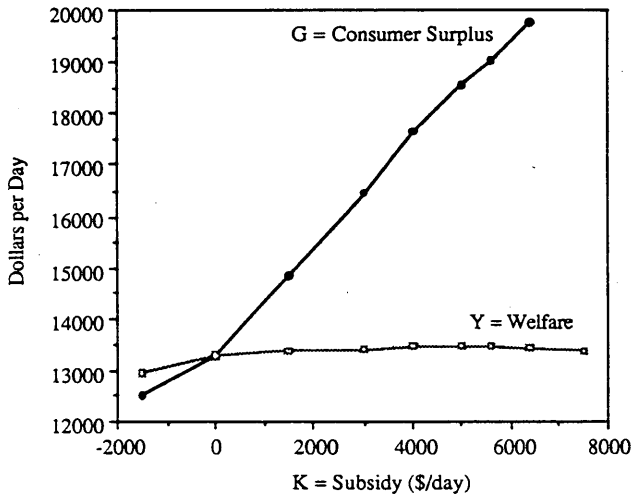


FIGURE 5 Effects of subsidy on consumer surplus and welfare.

acceptable $\Delta Y/\Delta K$ criterion. It would then be desirable to optimize the welfare subject to a constraint on the ratio of Δ welfare/Δ subsidy, as discussed earlier.

CONCLUSIONS

A fairly general bus system is analytically optimized to maximize welfare with various financial constraints. Closed-form solutions are derived for the optimal design variables (e.g., route spacing, headway, and fare), as summarized in Table 1. Interrelationships among the optimized design variables, the objective functions, and the system parameters are identified for various cases.

The optimality of a constant ratio between route spacing and headway, which has been found in previous studies for various bus network and demand conditions, is also found to be maintained with a multiperiod adjustment factor for all cases considered. It is not surprising that the optimal fare for welfare maximization is 0 in systems in which, in the absence of vehicle size constraints and congestion effects, the marginal operator cost is 0.

The effects of subsidy on the optimized results are presented. In maximizing welfare the optimal fare varies inversely with the subsidy. The consumer surplus increases with the subsidy, whereas the profit is the subsidy with a negative sign. These relationships have the net effect that the optimal welfare is very flat over a wide range of subsidy values. The effects of the bus operating costs on the optimized welfare, route spacing, fare, and fleet size are also evaluated (*2*).

The most interesting finding of this study is that the welfare-versus-subsidy function is very flat over a wide range in the vicinity of the optimum, as suggested in Figure 3 and shown numerically in Figure 5. Furthermore, our sensitivity analysis and preliminary results for very different kinds of transit system, such as flexible route paratransit (*26*), suggest that this is not an isolated case based on an accidental combination of parameters but a typical situation. This implies that subsidies can be greatly reduced below those required for maximum welfare, with only slight sacrifices in welfare. However, the relative funding burden would shift from subsidies (i.e., taxpayers) to fares (users). The changes in optimized systems associated with subsidy reductions, such as the changes in ridership, fares, network headways, and service levels can be analyzed with the analytic models presented here.

Comparisons of the various cases considered indicate that the effects of subsidy on welfare are very small over a wide range around the maximum welfare case. Thus, welfare maximization with a break-even constraint yields a zero deficit with very small reductions in welfare, compared with the subsidized cases (including unconstrained welfare maximization). Such results imply that the break-even objective might be preferable to subsidized cases whenever break-even solutions exist on a relatively flat part of a welfare function. In other words, where large increases in subsidies are required for much smaller increases in social welfare, operators (and taxpayers) may find the break-even objective preferable.

More generally, the conceptual discussion of Figure 3 suggests how a minimum acceptable $\Delta Y/\Delta K$ (i.e., ratio of welfare change to subsidy change) criterion may be used to determine the proper amount of subsidy or profit, and to efficiently allocate resources among various activities, including the bus systems modeled here.

REFERENCES

1. S. K. Chang and P. M. Schonfeld. Multiple Period Optimization of Bus Transit Systems. *Transportation Research*, Vol. 25B, No. 6, 1991, pp. 453–478.
2. S. K. Chang and P. M. Schonfeld. *Welfare Maximization with Financial Constraints for Bus Transit Systems.* TSC Report 92-24. Department of Civil Engineering, University of Maryland, College Park, 1992.
3. A. Black. *The Design of Urban Mass Transit Systems: A General Model.* Ph.D. dissertation. Cornell University, Ithaca, N.Y., 1975.
4. B. Byrne. Public Transportation Line Position and Headways for Minimum System and User Cost in A Radial Case. *Transportation Research*, Vol. 9, 1975, pp. 97–102.
5. G. C. Clarens and V. F. Hurdle. An Operating Strategy for a Commuter Bus System. *Transportation Science*, Vol. 9, No. 1, 1975, pp. 1–20.

6. N. S. A. Ghoneim and S. C. Wirasinghe. Optimum Zone Configuration for Planned Urban Commuter Rail Lines. *Transportation Science*, Vol. 21, No. 2, 1987, pp. 106–114.
7. E. M. Holroyd. The Optimal Bus Service: A Theoretical Model for A Large Uniform Urban Area. *Proc., 3rd International Symposium on the Theory of Traffic Flow*, New York, N.Y., 1965, pp. 308–328.
8. V. F. Hurdle. Minimum Cost Locations for Parallel Public Transit Routes. *Transportation Science*, Vol. 7, No. 4, 1973, pp. 340–350.
9. G. F. Newell. Dispatching Policies for a Transportation Route. *Transportation Science*, Vol. 5, No. 1, 1971, pp. 91–105.
10. S. Tsao and P. M. Schonfeld. Branched Transit Services: An Analysis. *Journal of Transportation Engineering*, ASCE, Vol. 110, No. 1, 1984, pp. 112–128.
11. R. Vaughan. Optimum Polar Networks for an Urban System with a Many-to-Many Travel Demand. *Transportation Research*, Vol. 20B, No. 3, 1986, pp. 215–224.
12. V. Vuchic and G. Newell. Rapid Transit Interstation Spacing for Minimum Travel Time. *Transportation Science*, Vol. 2, No. 4, 1968, pp. 359–374.
13. G. Kocur and C. Hendrickson. Design of Local Bus Service with Demand Equilibrium. *Transportation Science*, Vol. 16, No. 2, 1982, pp. 149–170.
14. C. A. Nash. Management Objectives, Fares and Service Levels in Bus Transport. *Journal of Transport Economics and Policy*, Vol. 12, No. 1, 1978, pp. 70–85.
15. M. W. Frankena. The Efficiency of Public Transport Objectives and Subsidy Formulas. *Journal of Transport Economics and Policy*, Vol. 17, No. 1, 1983, pp. 67–76.
16. P. K. Else. Optimal Pricing and Subsidies for Scheduled Transport Services. *Journal of Transport Economics and Policy*, Vol. 19, No. 3, 1985, pp. 263–279.
17. P. H. Bly and R. H. Oldfield. An Analytic Assessment of Subsidies to Bus Services. *Transportation Science*, Vol. 20, No. 3, 1986, pp. 200–212.
18. R. H. Oldfield and P. H. Bly. An Analytic Investigation of Optimal Bus Size. *Transportation Research*, Vol. 22B, No. 5, 1988, pp. 319–337.
19. P. G. Furth and N. H. M. Wilson. Setting Frequencies on Bus Routes: Theory and Practice. In *Transportation Research Record 818*, TRB, National Research Council, Washington, D.C., 1981, pp. 1–7.
20. E. K. Morlok and P. A. Viton. Feasibility of Profitable Transit Service in Radial Urban Corridors. In *Transportation Research Record 980*, TRB, National Research Council, Washington, D.C., 1984, pp. 46–54.
21. P. A. Viton. Privately-Provided Urban Transport Services. *Journal of Transport Economics and Policy*, Vol. 16, No. 1, 1982, pp. 85–94.
22. P. A. Viton. Pareto-Optimal Urban Transportation Equilibria. In *Research in Transportation Economics*, Vol. 1, Jai Press, Inc., Greenwich, Conn., 1983, pp. 75–101.
23. S. K. Chang. *Analytic Optimization of Bus Systems in Heterogeneous Environments*. Ph.D. dissertation. UMCP-TSC-DS-90-2. Department of Civil Engineering, University of Maryland, College Park, 1990.
24. C. A. Nash. *Economics of Public Transport*. Longman Inc., New York, N.Y., 1982.
25. M. Wohl and C. Hendrickson. *Transportation Investment and Pricing Principles*. John Wiley and Sons, Inc., New York, N.Y., 1984.
26. S. K. Chang and C. J. Lee. Welfare Comparison of Fixed- and Flexible-Route Bus Systems. In *Transportation Research Record 1390*, TRB, National Research Council, Washington, D.C., 1993.

Publication of this paper sponsored by Committee on Transportation Economics.

Improving Michigan's Border-Crossing Railroad Infrastructure: Implications for Metropolitan Detroit

JOSEPH P. SCHWIETERMAN

Along the Michigan-Canada border, government officials and business leaders are engaged in a highly politicized and divisive debate over Canadian National–North America's railroad tunnel project under the St. Clair River. The 6,000-ft tunnel, to link Port Huron, Michigan, and Sarnia, Ontario, is to be Michigan's first transborder facility capable of handling double-stack containers and other oversized rail cars. However, Detroit officials, concerned about their city's status as a rail hub, favor an alternative tunnel location in the Detroit-Windsor area. The economic and social implications of the two tunnel alternatives for the Detroit-Windsor metropolitan area are assessed. Using a methodological approach developed by FRA, the results show that the metropolitan Detroit area stands to gain \$5.5 million annually if the tunnel is completed as scheduled and \$4.5 million annually if the tunnel is built in the immediate Detroit area. Broad lessons are discussed about the municipal implications of rail infrastructure projects—lessons relevant in the analysis of rail projects across the country.

Railroad infrastructure projects are gaining unprecedented attention as municipalities, counties, and states vie to attract new industry. By developing comprehensive railroad plans, providing technical assistance, and offering capital grants to rail carriers, state and local governments are working to build private-public partnerships to ensure the timely completion of vital transportation projects.

Along the Michigan-Canada border, however, government officials and business leaders are engaged in a highly politicized and divisive debate over one railroad project—Canadian National (CN)–North America's tunnel project under the St. Clair River, linking Port Huron, Michigan, and Sarnia, Ontario. The 6,000-ft tunnel, located 60 mi north of Detroit, would be the most expensive privately financed railroad improvement in Michigan's recent history. As Michigan's first transborder facility capable of handling double-stack containers and other oversized rail cars, it could have a significant economic impact on the Detroit area. Detroit officials vigorously oppose the project and have released a study concluding that it will undermine Detroit's position as an expanding international rail gateway. In fact, they are pursuing regulatory actions that could delay or halt the Port Huron–Sarnia project and favor an alternative tunnel project in the Detroit-Windsor area.

This study assesses how the two alternative tunnel locations will affect the economy of the Detroit-Windsor metropolitan area. Using a methodological approach developed by FRA, the results offer broad lessons about the municipal implications of rail infrastructure projects—lessons relevant in the analysis of rail projects across the country.

Public Services Management Program, DePaul University, 243 South Wabash Avenue, Chicago, Ill. 60604.

BACKGROUND

Most rail traffic between Michigan and eastern Canada travels through one of two aging tunnel facilities—a century-old tunnel between Port Huron and Sarnia, and an aging twin-tube concrete tunnel between Detroit and Windsor built in 1910. Michigan's railroad companies have long recognized that neither facility adequately meets their changing operational needs. Because of low clearances, the facilities do not accommodate many modern types of rail cars, including double-stack container trains, high-cube box cars, or trilevel automobile rack cars. In fact, existing clearances are so low that the tunnels cannot even accommodate conventional "piggyback" intermodal cars—equipment that has been used for more than half a century. (Intermodal trains handle freight in containers or truck trailers that are suitable for conveyance by several modes of transportation.)

Railroad companies are responding to the tunnel clearance problems in several ways. First, to the extent possible, they are rerouting much of their transborder freight around Michigan. For example, the Canadian Pacific Railroad (CP) has entered a marketing agreement with the Norfolk Southern Railroad (NS) to send many double-stack intermodal cars over the Niagara Bridge, near Buffalo, New York. Second, they are using ferry services to transport oversized rail cars across the border. Often, however, this alternative is excessively costly: not only are federal harbor fees nearly \$300/car, but the process delays delivery by 12 to 24 hr (*1*). Finally, a new type of rail car, dubbed a "well car," is being used to allow certain truck-on-flat-car (TOFC) services to use the existing tunnels.

The two railroad companies that own the aging Detroit-Windsor tunnel, CN and CP, also have embarked on a partial enlargement project. They recently agreed to spend \$25 million to deepen one of the twin-tube tunnels to allow the passage of high-cube box cars and standard TOFC equipment. However, even when this project is completed, clearances will remain insufficient to accommodate the double stacking of containers 9 ft 6 in. and trilevel automobile rack cars 20 ft 2 in. in size.

Demand for the double stacking of containers is expected to grow rapidly throughout the decade because it virtually doubles a train's capacity. Most important, it lowers the costs of shipping containerized freight and allows railroads to compete with motor carriers for freight movements as short as 250 mi. By contrast, conventional single-stack services generally are competitive with trucks only over distances 800 mi or more (*1*).

CN and its U.S. subsidiary, the Grand Trunk Western Railroad (GTW), which use the CN–North America name for marketing purposes, believe the new rail tunnel will enhance its competitive position in the movement of transcontinental intermodal freight. Even though CN–North America's intermodal traffic is rising (up 15 percent between 1990 and 1991), intermodal revenues are falling (down nearly 4 percent over the period) because of escalating price competition—and additional price cuts are anticipated as technological innovations are made. The carriers maintain that low-cost transborder double-stack service is a necessary step in restoring profitability (*2*).

CN–North America will construct the $128 million ($155 million in Canadian dollars) Port Huron-to-Sarnia tunnel without government funds, with completion scheduled for late 1994. The Port Huron–Sarnia site is located on the GTW-CN main line linking Chicago with Toronto and Canadian seaports near Montreal. Because of the configuration of the 510 mi of track GTW operates in Michigan, railroad personnel believe that routing trains via Detroit would be difficult, adding 45 mi and 3 to 4 hr of travel time to the journey—thus eliminating much of the tunnel's marketing and cost-saving benefits (Figure 1). The northern route is more direct for Chicago-bound freight and bypasses substantial congestion at Detroit.

On completion of the Port Huron-to-Sarnia tunnel, CN–North America also plans to improve service to Detroit by expanding its $20 million investment in local terminal facilities, including its Moterm, Brownstown, and Mazda terminals. The largest of these facilities, Moterm, in north suburban Ferndale, is scheduled for additional enlargement into adjacent Detroit in upcoming months. From these terminals and those in Chicago, Lansing, Battle Creek, and Durand, CN–North America currently transports about 75,000 oversized rail cars over the Port Huron–Sarnia ferry each year and 210,000 containers through the existing tunnel.

As some Detroit officials observe, CN–America may have another motive for selecting a more remote tunnel location. The northern tunnel could preclude other railroads—including rivals CP, NS, and CSX—from sharing in the tunnel's benefits (these carriers' intermodal routes are also shown in Figure 1). Because these competitors have not yet demonstrated a financial commitment to a double-stack tunnel project, however, such allegations have been largely discounted by state rail officials. For example, the Michigan Department of Transportation (MDOT) remains supportive of the planned CN–North America tunnel.

The Detroit-Windsor Port Corporation (DeWin) has been a leading critic of CN–North America's tunnel. DeWin proposes an alternate tunnel site between Detroit and Windsor at a cost of $172 million (U.S. dollars), not including the cost of the approach infrastructure, ultimately to be financed through railway use charges. It also seeks to convert the existing twin-tube rail tunnel between Detroit and Windsor to a motor

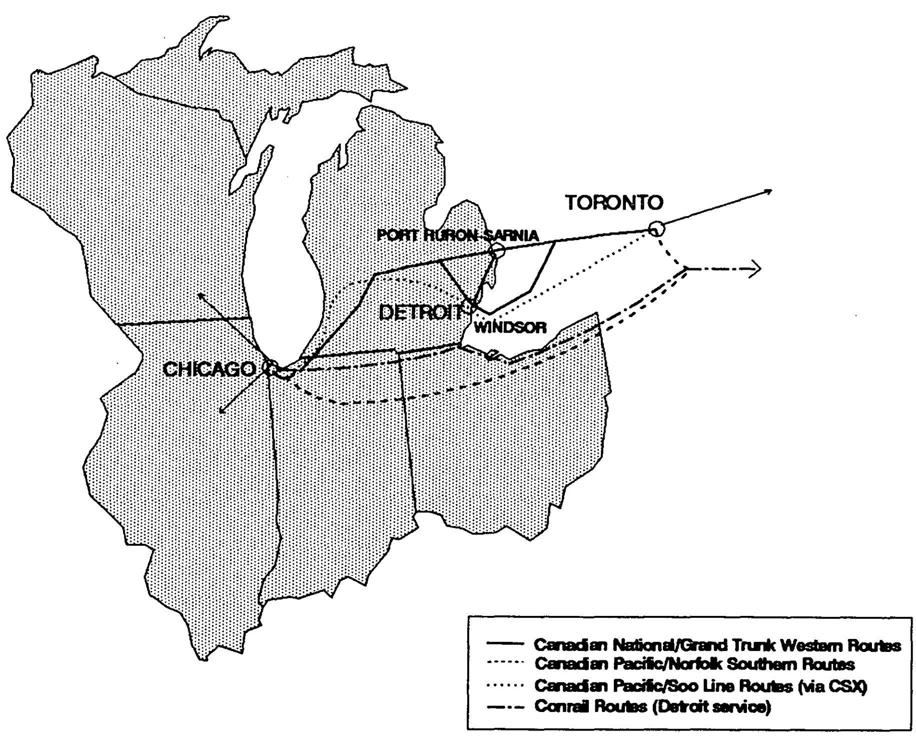


FIGURE 1 Principal intermodal railroad services: Chicago to Toronto.

carrier facility that would help relieve transborder highway congestion on the Ambassador Bridge. Although it adds $65 million to the project's overall cost, DeWin considers the motor tunnel an important ancillary component of the proposal (*3*).

IMPLICATIONS FOR THE DETROIT-WINDSOR ECONOMY

Unfortunately, the debate over the tunnel's implications for the Detroit-Windsor economy is proceeding without the benefit of detailed economic analysis. This section explores the limitations of previous studies and outlines an approach to assess the benefits and costs of the alternative tunnel locations for the Detroit-Windsor area.

Limitations of Previous Research

An MDOT-sponsored study released in December 1991 (*1*) explores in detail the project's implications for Michigan. The study's author concludes after extensive research that "without double-stack service, Michigan will sit at the northern terminus of a rail system which increasingly relies on auto industry volume to absorb costs." The study's financial analysis indicates that the Port Huron–Sarnia site offers a much faster payback for both Michigan and the United States. Still, it outlines unique advantages associated with each location. For example, although the study concludes that CN–North America has a strong interest in the more direct northern facility, it notes that the Chrysler Corporation, which ships trilevel automobile rack cars from its plants in Ontario, would benefit most from a tunnel in the immediate Detroit area. The following analysis presented relies extensively on this earlier investigation.

Because of the MDOT study's exclusive focus on the project's statewide implications, it has done little to lessen the specific concerns of Detroit officials about the planned Port Huron–Sarnia tunnel. Detroit officials maintain that the study contradicts MDOT's earlier position—a position that MDOT publicized in the form of an informational brochure—that Detroit urgently needs its own double-stack tunnel.

Recognizing the need for additional research, DeWin commissioned a study to measure these local effects of a Port Huron–Sarnia Tunnel. That study concludes that Detroit would suffer significant economic losses if the northern tunnel were built (*3*). Of primary concern was "the potential migration of existing industries from Detroit-Windsor areas" to Port Huron–Sarnia. The study forecasts that more than 85,000 area jobs in the manufacturing, service, and retail sectors would be lost. As supporting evidence, it cites the availability of 8,100 acres of industrial parks in the Port Huron–Sarnia area. The northern tunnel location, it maintains, would "seriously reduce the economic and social viability of the greater Detroit-Windsor area."

DeWin's study also concludes that the Port Huron–Sarnia tunnel would affect the ability for the Detroit-Windsor region to attract new business by denying local shippers access to double-stack services. The study states: "If Detroit-Windsor is relegated to the status of 'bridesmaid' while Port Huron–Sarnia becomes the bride, the Detroit-Windsor area will be placed in the arduous position of striving to finance, plan, and construct a double-stack tunnel." However, this conclusion does not agree with MDOT's conclusion that the Port Huron–Sarnia tunnel would be able to handle almost all transborder double-stack freight traveling between metropolitan Detroit and Canada. MDOT finds that most shippers in the Detroit area will be well served by either tunnel. In fact, many shippers using the Detroit area's largest intermodal terminal—Moterm—might prefer a Port Huron–Sarnia route to a Detroit-Windsor route because it offers shorter travel distances to Canadian seaports.

Another point of contention is whether CN–North America will establish competing intermodal terminal facilities in the Port Huron–Sarnia area, encouraging companies to relocate their terminals out of the Detroit vicinity. The carrier denies this claim. Regardless of the ultimate location of the tunnel, CN–North America maintains that the local Moterm facility will remain the region's primary loading area for Canada-bound freight. The carrier points to its recent expansions and financial commitment in support of this claim.

Analytical Techniques

The validity of these conflicting claims about the expected impact of the tunnel can be addressed using government-approved analytical techniques. FRA has developed a methodology specifically for the evaluation of railroad projects; it weighs a project's effects on local employment, shippers costs, railroad operating costs, and other factors (*4*). Many state departments of transportation, including MDOT, use this methodology extensively in project analysis.

MDOT uses the FRA methodology as the basis of a public benefits model that measures the effects of rail projects on the state economy across a 20-year planning horizon (*5,6*). Although this methodological approach has not been applied previously to projects as large as the double-stack tunnel—most of the projects MDOT considers cost only a few million dollars and involve low-density branch lines—it provides a useful framework for understanding the economic effects of the alternative tunnel locations.

The FRA-MDOT approach provides important lessons about the economic effects of rail projects:

1. *The principal beneficiaries of rail projects are shippers and their employees.* Rail projects affect business conditions in a community by changing transportation costs and the quality of rail services available to shippers. By forestalling abandonments, improving services, and lowering prices, rail projects can lower freight costs and mitigate the need for layoffs by on-line shippers. The methodology also suggests that the gains and losses that municipalities experience because of employment changes often are relatively modest. According to the FRA methodology, job losses impose only temporary costs—and job additions provide only temporary gains—for a metropolitan area's economy because affected workers typically can find other employment after sustained job searches. In MDOT's analysis, for example, the length of time required

for workers to find suitable alternate employment is based on Michigan Employment Security Corporation estimates.

2. *Some states, including Michigan, consider regional multiplier effects to be relevant.* MDOT's public benefits model assumes that job losses in primary industries "multiply," resulting in additional losses in secondary industries, thus affecting employment in the service and retail sectors. The department's publications suggest that approximately 1.3 secondary jobs are lost for every primary worker displaced (*6*). (The opposite relationship holds true for employment gains.) However, Michigan is the only state in the Midwest to consider multiplier effects, rendering its assessments of the employment-related benefits of rail projects higher than average (*7,8*).

3. *Growth in transportation-related employment is not a likely benefit associated with rail projects.* Railroad projects rarely lead to significant gains through expanded job opportunities for transportation workers. In the current environment, in which railroad jobs are rapidly being eliminated through work rule changes and technological innovation, large-scale job growth in railroading is an improbable project benefit, particularly for intermodal services in which two-person crews are becoming the industry standard. Moreover, new railroad jobs often are offset by lost employment opportunities in the trucking and waterway sectors. For example, officials in Kansas found that each new railroad job eliminated almost three trucking jobs, leading to temporary unemployment in that state (*8*).

4. *"Bridge traffic" is of little economic benefit to a municipality.* Additional bridge traffic, consisting of freight that originates or terminates on other rail lines, has obvious benefits to a railroad, allowing fixed costs to be apportioned over a larger traffic base (*1*). However, most state officials concede that municipalities receive no direct benefit from bridge traffic that passes through their borders. Despite this, officials in Detroit contend that their city has much to gain if transborder double-stack trains were routed through a Detroit-Windsor tunnel. This optimism stems from their assumption that Detroit would become a major national classification facility for double-stack trains after the tunnel's completion. Although this possibility cannot be ruled out definitively, it is more likely that intermodal trains will simply pass through Detroit with only changes of crew—just as they do in Cleveland, Port Huron, Fort Wayne, and other intermediate points on the Chicago-to-Canada route. Only 235 mi from the Canadian border, Chicago remains America's premier hub for east-west intermodal freight and is expected to retain this status well into the next century. Large and capital-intensive terminal additions in either Detroit or Port Huron are unlikely outcomes of the tunnel project.

The benign effects of bridge traffic on local economies are documented in earlier governmental analyses of rail projects. For example, when MDOT analyzed a $3 million trackage and bridge rehabilitation project between Marion and Ashley, Michigan, the department identified no local benefits associated with freight passing through the affected communities. Municipalities reaped no significant benefits through property tax growth, employment growth, or industrial expansion from this traffic. In fact, the analysis of 55 railroad infrastructure projects recently undertaken in Illinois, Indiana, Iowa, Kansas, Michigan, North Dakota, and Ohio—leaders in benefit-cost analysis—shows that no municipalities were expected to benefit from increased bridge traffic. Local communities were affected primarily by changes in the quality of transportation service provided to their shippers (*8*).

Bearing in mind these lessons from earlier projects, the following section measures the implications of the tunnel's location for the Detroit-Windsor economy.

ASSESSMENT OF BENEFITS AND COSTS

The benefits and costs to the Detroit-Windsor area as a result of either tunnel's construction can be divided into three categories. Transportation efficiency benefits are the reduced transportation costs realized by shippers and carriers that will use the tunnel. Secondary benefits are the benefits associated with improved employment opportunities, property tax revenues, and other ancillary effects. Finally, direct project costs for the Detroit-Windsor area are the expenditures necessary to build the tunnel and rebuild the right-of-way in the vicinity of the tunnel. (Because private companies would finance the tunnel project, area shippers would pay for these costs only indirectly through rail service fees.)

Transportation Efficiency Benefits

Transportation efficiency benefits associated with the tunnel are calculated separately for existing and newly generated rail freight.

Benefits for Existing Freight

Most existing shippers in the Detroit-Windsor area would realize immediate benefits from a tunnel in either location, because it would allow the initiation of double-stack service between Detroit and eastern Canada and accommodate oversized rail cars that now must be transported by ferries. However, as stated earlier, their use of the tunnel will not be cost-free: shippers will help pay for the tunnel's construction indirectly through the transportation fees they pay.

The Port Huron–Sarnia tunnel is expected to cost $128 million (U.S. dollars) from summit to summit, inclusive of modifications to the approach infrastructure and track realignment costs. The Detroit-Sarnia tunnel is expected to cost $172 million (U.S. dollars) from portal to portal. Conservatively estimated, an additional $20 million will be needed for the approach infrastructure and trackage work on supporting rail lines. For example, CN–North America will be required to undertake extensive work on its 110-mi single-track line between Windsor and London, Ontario, because it lacks the signal equipment and right-of-way enhancements necessary for high-density double-stack railroading. Additional work also may be necessary on CSX and CP lines radiating from Detroit.

At either location, the tunnel would handle approximately 180,000 fully loaded, oversized rail cars a year. The Detroit-Windsor tunnel would serve more railroad companies, but it would handle much less CN–North America freight, offsetting this numerical advantage. On the basis of a 20-year planning horizon and an 8 percent discount rate, the fully allocated cost of the tunnel for each oversized rail car, including articulated double-stack cars, would be about $72.43 at Port Huron–Sarnia and $108.64 at Detroit-Windsor. (These estimates are based on total construction costs for the northern and southern tunnel alternatives of $128 million and $192 million, respectively.) As a result of the high costs, shippers in Detroit probably would pay slightly higher rates if the southern alternative were selected.

Of these 180,000 loaded, oversized cars, about 40,000 would originate or terminate in Detroit and another 4,000 would originate or terminate in Windsor. For freight shipped to and from Detroit, the new double-stack tunnel would reduce operating costs by approximately 10 percent—about $200/carload (or $100/container). [This estimate is based on the MDOT study (*1*).] Thus, after construction costs, Detroit shippers would save approximately $127.57/carload (i.e., $200.00 minus $72.43 for capital cost of the tunnel) from a Port Huron–Sarnia tunnel and approximately $91.36/carload (i.e., $200.00 minus $108.64 for capital cost of the tunnel) from a Detroit-Windsor tunnel. [This approach is similar to that used by MDOT (*1*).] Cumulatively, shippers would save an estimated $5.1 million a year by using a Port Huron–Sarnia tunnel and $3.7 million a year by using a Detroit-Windsor tunnel (Table 1).

For shippers of freight originating or terminating in Windsor, however, the location of the tunnel is a more pivotal issue. Although freight that originated in Detroit would travel economically through either tunnel, freight that originated in Windsor would not be able to use the Port Huron–Sarnia tunnel in a cost-effective manner. If the northern tunnel were built, Windsor shippers would continue to transport their freight via motor carrier to terminals in Detroit at an estimated additional cost of $200/carload (*1*). Altogether, Windsor shippers would benefit by $0.4 million annually from a Detroit-Windsor tunnel but probably would not benefit directly from a Port Huron–Sarnia tunnel.

Of course, regardless of the location of the tunnel, neither Windsor- nor Detroit-area shippers would benefit measurably from the estimated 140,000 annual carloads of bridge traffic. The benefits associated with this traffic, ranging from $11 million to $14 million a year, would be captured by out-of-town consumers and businesses.

Benefits from Newly Generated Traffic

The tunnel also will generate new intermodal rail business in the Detroit-Windsor area, providing transportation savings to shippers who otherwise would ship their freight via motor carrier—or not ship it at all. These benefits also are considered.

Although double-stack services typically reduce operating costs by as much as 10 percent, these cost estimates suggest that intermodal freight rates in the Detroit-Windsor area are likely to fall by only about 4.6 percent if a Detroit-Windsor tunnel is built and 6.4 percent if a Port Huron–Sarnia tunnel. Moreover, as discussed earlier, Windsor shippers will realize no savings if the northern tunnel is built.

Research by Winston (*9*) allows estimation of the volume of additional rail freight likely to be generated as a result of

TABLE 1 Annual Transportation Efficiency Benefits (thousands of dollars)

	PORT HURON - SARNIA TUNNEL			DETROIT - WINDSOR TUNNEL		
	Approx. Carloads*	Benefit per Carload	Total Benefit	Approx. Carloads*	Benefit per Carload	Total Benefit
Existing Rail Freight						
Detroit-area freight	40,000	$127.6	$5,103	40,000	$91.4	$3,654
Windsor-area freight	-	-	-	4,000	$91.4	$365
Newly Generated Rail Freight						
Detroit-area freight	6,838	$63.8	$436	4,897	$45.7	$224
Windsor-area freight	-	-	-	490	$45.7	$22
Total	**46,838**		**$5,539**	**49,386**		**$4,265**
Net Present Value - 20 year period			*$54,383*			*$41,882*

* Based on carload and market share data provided by CN-North America

Rows or columns may not add due to rounding.

these rate reductions. Winston finds that for every 1 percent drop in railroad rates for truck-competitive freight, traffic increases by 2.8 percent. (He estimates that the cross-price elasticity of demand for truck-competitive rail freight with respect to rail prices is -2.8.) On the basis of Winston's estimate, the Port Huron–Sarnia tunnel could be expected to generate about 6,838 additional carloads of local freight; the Detroit-Windsor tunnel could be expected to generate an additional 5,386 carloads. If the demand curve for this intermodal freight is assumed to be linear, the average benefit will be exactly half of the rail price reduction, averaging $45.70/carload for the Detroit-Windsor tunnel to $63.80/carload for the Port Huron–Sarnia tunnel. Therefore, the annual gain for local shippers using the tunnels would be $436,000 and $246,000, respectively.

Altogether, considering both newly generated and existing freight, metropolitan Detroit would enjoy total transportation efficiency benefits of $5.5 million as a result of the Port–Sarnia tunnel and $4.3 million as a result of the Detroit-Windsor tunnel. Clearly, these are upper-bound estimates because some of these benefits will be passed on to out-of-town manufacturers or consumers through lower prices for intermediate and finished goods.

Secondary Benefits

The double-stack tunnel also might produce secondary benefits through employment growth in the Detroit-Windsor area. The potential job growth among shippers and railroads is measured below.

Shipper Employment

The tunnel project is likely to stimulate a limited number of new jobs among Detroit-area shippers. However, because tunnel-related savings to the average shipper represents only about 0.3 percent of total production costs, the associated employment growth is not likely to be large. (Virtually all of the growth in rail traffic volume will represent merely a shifting of modes.) Moreover, considering that either tunnel will offer similar rate reductions to shippers, retailers, small businesses, and consumers, these employment-related benefits should flow to the metropolitan Detroit area regardless of where the tunnel is built. Because these benefits will not affect the relative attractiveness of the alternatives, no attempt is made to quantify them.

Transportation Employment

Local employment by railroad companies, however, is likely to increase marginally if a tunnel is built between Detroit and Windsor. This tunnel would handle approximately three additional CN–North America intermodal trains in each direction daily. Considering railroad work rules, and assuming that three employees are needed to operate each train (most intermodal trains have only two-person crews), about 18 CN–North America jobs would be relocated to Detroit. Administrative personnel and facilities already are available to handle these employee relocations.

It is also likely that CP and NS would reroute double-stack trains (eight per week) that currently use the Buffalo gateway through Detroit if the Detroit-Windsor tunnel were built. Similarly, a local tunnel also might help forestall the possible relocation to other gateways of CP/Soo Line intermodal trains (15 per week) that currently use CSX trackage through Michigan. This suggests that a local tunnel would preserve or produce an additional 36 jobs for metropolitan Detroit. Other local railroads, such as Conrail, are unlikely to relocate any personnel to the area because their crews already originate in Detroit.

How much would the Detroit-Windsor area benefit from these 54 jobs? Using FRA's methodology, MDOT publications suggest that a reasonable estimate of the social value of each new job created (or each layoff averted) is $40,000 (*8*). (This estimate is inclusive of a regional multiplier.) This suggests a total, one-time secondary benefit of $2.2 million to Detroit-Windsor if the tunnel were to be built locally.

Four factors suggest that this estimate overstates the actual secondary gains associated with the Detroit-Windsor tunnel:

1. New railroad jobs are likely to be filled by employees transferring from other locations instead of by Detroit-area residents. Therefore, they would alleviate local unemployment problems less than MDOT's estimates suggest.
2. MDOT's estimate of the social benefits of job creation is relatively high by Midwestern standards. Illinois uses a higher benefit estimate in its analysis, but Indiana and Iowa assume that cities derive no local employment benefits as a result of a rail project.
3. The estimate is based on the supposition that CN–North America would reroute all its intermodal trains through the Detroit-Windsor tunnel. Considering the difficult and circuitous nature of the Detroit routing, the company has a strong incentive to continue using more northern lines.
4. The estimates ignore probable losses in employment by local barge operations and trucking companies from improved rail services.

The tunnel's construction may produce other ancillary benefits, such as improved safety and reduced pollution. Most important, the tunnel will greatly lessen highway maintenance costs and improve traffic flows on local expressways (*10*). However, these benefits are not considered because they would be realized under either tunnel scenario; they do not affect the relative attractiveness of either alternative. Also ignored are the possible benefits from DeWin's proposal to convert the existing Detroit-Windsor rail tunnel for trucking use; MDOT-sponsored research indicates that this project may not be worth its $65 million cost (*1*).

The cumulative value of transportation efficiency and secondary benefits for the Detroit-Windsor metropolitan area are $5.5 million annually from the Port Huron–Sarnia tunnel and $4.5 million annually from the Detroit-Windsor tunnel (Table 2). The difference is small, attributable primarily to the higher transportation benefits to local shippers from the lower-cost Port Huron–Sarnia tunnel. These benefits outweigh the benefits associated with additional employment from the Detroit-Windsor tunnel.

TABLE 2 Net Present Value of Project Benefits (thousands of dollars)

	Port Huron - Sarnia	Detroit - Windsor
Transportation Efficiency	$54,383	$41,882
Secondary Benefits	$0	$2,160
Total Gain -- 20 years	$54,383	$44,042
Annualized Gain	*$5,539*	*$4,486*

* based on 20-year planning horizon @ 8 percent interest

CONCLUSIONS AND POLICY IMPLICATIONS

As the analysis shows, the location of the tunnel is not nearly as important for the Detroit-Windsor area as is its prompt completion. Considering that the Detroit-Windsor metropolitan area stands to lose more than $5 million a year from delays in the completion of the Port Huron–Sarnia tunnel, the state should pursue a fast-track regulatory approval process, which will involve 55 U.S. and Canadian regulatory agencies. Regulatory issues are likely to delay needlessly the project's completion, adversely affecting trade volume of $21.6 billion between Michigan and Ontario.

Although the lower-cost double-stack tunnel between Port Huron and Sarnia is the best alternative for metropolitan Detroit, the region's deteriorating rail system remains a growing problem for state transportation officials. Congestion and antiquated rail facilities are serious obstacles to expanded rail service in Detroit, and recent funding decisions have largely focused on rural facilities. Thus, MDOT should consider allocating a larger portion of its discretionary rail funds as part of the Michigan Rail Plan (*5,6*). This will improve southeastern Michigan's access to transborder double-stack services, which will be available as early as 1994, and enhance shippers' use of the existing Detroit-Windsor tunnel, which will be enlarged by 1993. These efforts, together with CN–North America's tunnel investment, will provide a much-needed catalyst for private-sector investments in the metropolitan Detroit area.

REFERENCES

1. J. C. Taylor. *Michigan-Ontario Railroad Border Crossing Infrastructure: An Evaluation of Modernization Proposal Benefits and Recommendations for Future Action.* Michigan Department of Transportation, Lansing, Dec. 1991.
2. *Canadian National Annual Report.* Canadian National, Toronto, Ontario, 1991.
3. F. Spaniola. *Preliminary Economic Impact to the Greater Detroit-Windsor Area of the Canadian National Railway's Construction of a New Tunnel at Port Huron–Sarnia.* Detroit Windsor Port Corporation, Mich., Feb. 1992.
4. *Standard Benefit-Cost Methodology.* FRA, U.S. Department of Transportation, July 1990.
5. *Michigan Railroad Plan: February 1990.* Michigan Department of Transportation, Lansing, 1990.
6. *Michigan Railroad Plan: January 1992.* Michigan Department of Transportation, Lansing, 1992.
7. *Ohio State Rail Plan 1990–1991.* Ohio Department of Transportation, Columbus, June 1990.
8. J. Schwieterman. States Railroad Branchline Investments and Community Employment: Problems in Benefit-Cost Analysis. *Journal of the Transportation Research Forum* (in press).
9. C. Winston, T. Corsi, C. Grimm, and C. Evans. *The Economic Effects of Surface Freight Deregulation.* Brookings Institution, Washington, D.C., 1990.
10. J. Schwieterman. *Highway User Fees and the Efficient Movement of Heavy Freight: Implications for Michigan's Railroads.* Working paper, 1992.

Publication of this paper sponsored by Committee on Transportation Economics.

User Cost Methodology for Investment Planning and Maintenance: Management of Roads and Highways

WAHEED UDDIN AND K. P. GEORGE

A user cost and life-cycle analysis methodology is developed to quantify cost-effectiveness of alternative strategies for investment planning and maintenance management of roads and highways. The proposed methodology evaluates user costs based on the vehicle operating cost data from FHWA in conjunction with the deterioration models developed in the World Bank and NCHRP studies. Various scenarios of new construction alternatives and maintenance and rehabilitation strategies for asphalt-surfaced or portland cement concrete pavements can be analyzed to establish cost-effectiveness of competing alternatives. The methodology considers a user-specified maintenance intervention policy. The results of the present model compare well with those of the World Bank's Highway Design and Maintenance-PC program, especially for pavements in poor condition. For well-maintained pavements in good condition, the proposed methodology predicts relatively lower user costs. Life-cycle analysis with user cost demonstrates that timely maintenance treatments can be very cost-effective. The USER microcomputer program is well suited for investment analysis, maintenance programming, and network-level pavement management applications.

The development of a rational annual work program and budget for roads and highways in the jurisdiction of a public work agency requires life-cycle analysis of all agency cost streams and user costs associated with pavement conditions during the analysis period. Various construction and maintenance, rehabilitation, and reconstruction (MR&R) strategies will result in different life-cycle costs. This approach can lead to cost-effective investment planning of new roads and to MR&R work programming for the existing network. The life-cycle analysis approach based on agency and user costs is equally applicable in project-level pavement designs in establishing priority ranking of candidate roads and in selecting competing MR&R treatment alternatives for network-level pavement maintenance management applications. Another application is in providing a rational basis for comparison of effectiveness of innovative maintenance strategies developed in the Strategic Highway Research Program (SHRP) with other standard MR&R treatments. The quantification of user cost savings as a result of improved pavement condition, following the implementation of an appropriate MR&R treatment, can be used as a surrogate for "benefit." Vehicle operating cost (VOC) is the most significant component of road user cost. The other quantifiable component of user cost stream is traffic delay cost (TDC), resulting from temporary road closure or traffic diversion during MR&R activities.

Department of Civil Engineering, Universty of Mississippi, University, Miss. 38677.

The life-cycle analysis for investment planning, new pavement design, overlay design, and selection of pavement type has been a recognized approach since the 1970s (*1–7*). Toward this end the World Bank and other organizations abroad have developed pavement deterioration and VOC models (*8*). The World Bank Highway Design and Maintenance Standards Model (HDM) is probably the most comprehensive user cost and life-cycle analysis methodology that considers the existing pavement distress and roughness condition and includes a rational procedure to calculate VOC components of the user cost (*8*). It is noteworthy that HDM-VOC parameters are derived from road user cost studies conducted during the 1970s and 1980s in Brazil, Kenya, and India. However, the HDM methodology is currently applicable to asphalt surfaced and unpaved roads only.

This paper describes a comprehensive user cost methodology for life-cycle analysis of asphalt surface and portland cement concrete (PCC) pavements. The methodology incorporates the results of the state-of-the-art pavement deterioration and performance models developed by the World Bank (*8,9*) for asphalt pavements, PCC pavement deterioration models developed in an NCHRP study (*10*), and the VOC studies (*11,12*) conducted by FHWA. The methodology is directly applicable in planning and budget analysis using the pavement management system (PMS) data bases as well as maintenance management system (MMS) data bases. The proposed methodology can be readily incorporated in (a) project-level pavement design applications and (b) economic evaluation of competing scenarios for capacity and level-of-service improvement studies.

USER COST METHODOLOGY DEVELOPMENT

The user cost is computed in four steps:

1. Cumulative traffic prediction by vehicle type over the analysis period and hourly traffic prediction for the period of MR&R treatment intervention.
2. Pavement deterioration prediction over the analysis period for each pavement type on the basis of the current condition, with cumulative traffic being the primary causal factor.
3. Calculation of VOC stream for each year of the analysis period as a function of pavement condition, vehicle type, and associated VOC parameters and geometrical characteristics.

4. Calculation of traffic delay cost streams for each year of MR&R treatment intervention.

Figure 1 illustrates the key components and logic used in compiling the USER cost and life-cycle analysis program. The methodology is primarily developed as an application program that can be used alone or with a PMS/MMS data base and analysis software. The USER software can analyze a single homogeneous road section with up to nine alternatives or all sections in a road/highway network.

Input Data Requirements

The input data for user cost calculations generally are available from the inventory and condition data base files of a PMS/MMS data base. The input data can be categorized into four groups.

1. *Section-Specific Data*. These data elements include road section identification data, section length and width, pavement type, grade, curvature, last construction date, traffic volume [annual daily traffic (ADT)] and year of traffic count, growth rate, 18 kips (80 kN) equivalent single-axle load (ESAL) factors, subgrade California bearing ratio, structural number, joint spacing (for PCC pavements only), selected environment data, rural or urban area designation, typical initial running speed and speed adjustment factor (for congestion and posted speed), average section international roughness index (IRI) (meters per kilometer) and equivalent present serviceability rating (PSR) or present serviceability index (PSI), traffic de-

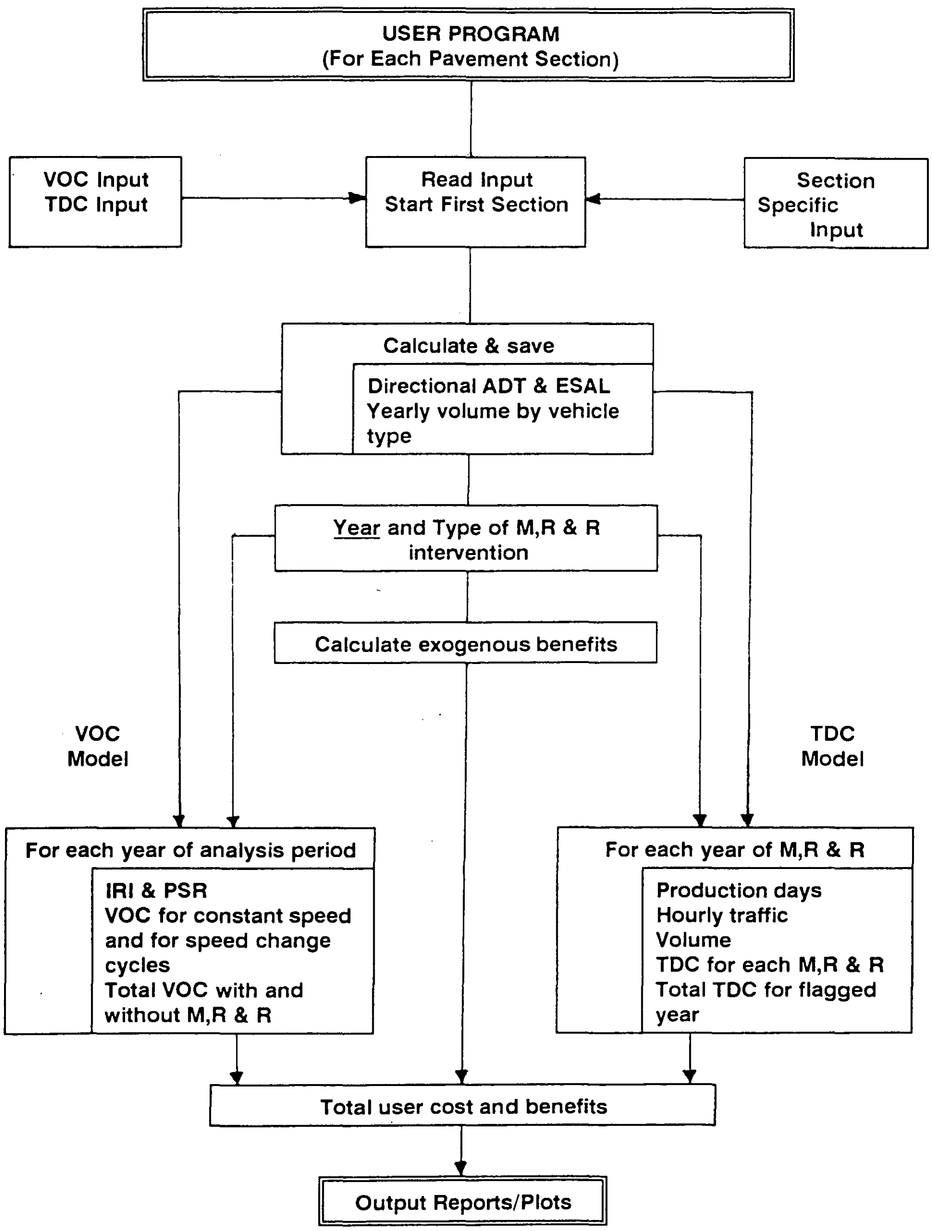


FIGURE 1 USER methodology.

tour model, construction/widening and MR&R unit costs, and data on travel time and accident cost savings.

2. *Decision Criteria.* The decision criteria applicable to each pavement type include analysis period, maintenance intervention criteria (minimum acceptable PSR) at which major MR&R treatment is triggered, and the adjusted PSR (or IRI) after MR&R treatment.

3. *Input for VOC Model.* The inputs for the VOC model include typical distribution of vehicle types in the traffic stream, 18-kip (80-kN) equivalency factor for each vehicle type, and the following VOC unit costs, for each vehicle type:

- –Repair unit cost (for routine vehicle maintenance) in dollars per 1,000 mi,
- –Oil unit cost (for oil consumption) in dollars per quart,
- –Fuel unit cost (for fuel consumption) in dollars per gallon, and
- –Cost of new vehicle (for depreciation) in dollars per new vehicle.

4. *Input for Traffic Delay Cost Model.* Inputs for the traffic delay cost model apply to all sections and include data on detour distance, time of traffic control, number of open lanes, percentage of vehicles affected, average vehicle delays, and hourly traffic distribution.

Procedure for User Cost and Benefit Analysis

A simplified flow chart of user cost and benefit analysis is shown in Figure 1.

- Step 1. Assign section input, as described in Item 3 of the preceding list.
- Step 2. Assign common user cost input, as described in Items 2 through 4.
- Step 3. Calculate directional ADT and corresponding ESAL and number of vehicles of each vehicle type for each year of analysis.
- Step 4. Check and identify year and applicable MR&R type from the condition prediction model and maintenance intervention criteria.
- Step 5. Predict IRI/PSR in each year; adjust for MR&R intervention of sections flaged in Step 4; calculate VOC stream for each year considering the pavement condition and other consumption parameters as identified in Item 3.
- Step 6. Calculate hourly traffic volume during the predicted period of MR&R activity in the intervention year identified in Step 4; calculate total traffic delay cost (TDC) in the intervention year.
- Step 7. Calculate user benefits resulting from a reduction in total VOC caused by MR&R intervention compared with a "do-nothing" policy. Also, calculate exogenous user benefits resulting from travel time savings and reduction in traffic accident cost.
- Step 8. Perform present worth analysis of agency cost and user cost and benefits, priority rank the analyzed sections and alternatives, and generate reports.

PAVEMENT DETERIORATION MODELS

Pavement deterioration prediction is basic to the user cost methodology and long-term budget analysis. Ideally, a pavement deterioration model (or pavement performance model) should include the effects of traffic, pavement strength, age, environment, and initial pavement condition. The time-series empirical models (*13*) are applicable only for local conditions and cannot account for all of the stated effects. The state-of-the-art deterioration models incorporated into the USER program are selected from the literature, ensuring that all of the important causal factors are accounted for.

Deterioration Models for Asphalt Surfaced Pavements

The World Bank road deterioration and maintenance model used in the HDM program is a comprehensive roughness progression model for asphalt road performance with independent variables of road surface distresses, environment, subgrade strength, traffic loads, and time. The model has been used in many field applications around the world (*8,9,14*) for life-cycle economic analysis. This incremental roughness prediction model has the following functional form:

$(IRI)_t = f$ (pavement structure and strength, traffic loading, extent of cracking, thickness of cracked layer, and rut depth variation)
$+ f$ (increment in cracking, patching, and pothole area)
$+ f$ (nontraffic parameters such as environment parameter, initial roughness, and time) (1)

Using a data base generated from numerous HDM computations, Paterson and Attoh-Okine (*9*) proposed a new simplified roughness progression model for applications in flexible pavements, particularly in roads with an area of cracking not exceeding 30 percent. The model is applicable to the full roughness range of up to 12 m/km of IRI as expressed below;

$$(IRI)_t = 1.04e^{mt}[(IRI)_0 + 263 \times (1 + SNC)^{-5} (CESAL)_t] \quad (2)$$

where

$(IRI)_t$ = roughness at time t (m/km IRI);
$(IRI)_0$ = initial roughness at time t equal to 0 (m/km IRI);
t = time since last construction/rehabilitation (years);
m = environmental coefficient, which varies between 0.01 and 0.70 (from dry, nonfreeze to wet, freeze environment);
SNC = structural number modified by subgrade strength, as used in HDM program; and
$(CESAL)_t$ = cumulative ESAL applications at time t (millions).

This simplified World Bank deterioration model has been incorporated in the USER methodology. Several road sections at different distress levels were investigated using both HDM and USER programs, and the results show that the USER program predicts a relatively slower rate of deterioration, as illustrated in Figure 2, for an asphalt concrete road

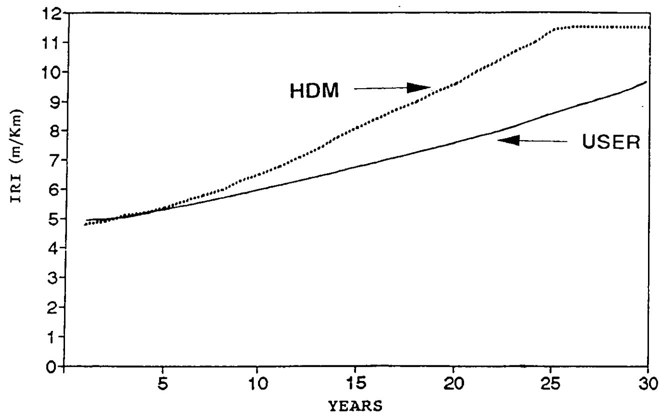


FIGURE 2 Comparison of pavement deterioration predictions using HDM and USER programs: County 4, SR12E, mile 0.000.

in poor condition, without MR&R. Table 1 presents pertinent inventory and condition data for the two asphalt road sections, SR 12 E/0.000 and SR 14 E/14.614, that are used for illustration throughout this paper.

Deterioration Models for Concrete Pavements

The COPES jointed concrete pavement performance models, developed in a nationwide NCHRP study (*10*), have been incorporated in the user cost methodology. These models predict PSR at time *t*, $(PSR)_t$ as a function of initial PSR or $(PSR)_0$, cumulative ESAL applications at time *t* in millions or $(CESAL)_t$, environment parameters and pavement strength. Separate COPES models are included for jointed reinforced concrete pavements (JRCP) and jointed plain concrete pavements (JPCP).

JRCP Model

$(PSR)_t = f$ [$(PSR)_0$, $(CESAL)_t$, edge stress/PCC modulus of rupture, dummy variables (for transverse joint spacing, reactive aggregates and base type), freezing index, average annual precipitation, average monthly temperature] (3)

JPCP Model

$(PSR)_t = f$ [$(PSR)_0$, $(CESAL)_t$, edge stress/PCC modulus of rupture, average annual precipitation, average monthly temperature] (4)

Continuously Reinforced Concrete Pavement Model

The USER program incorporates a simplified stepwise linear model (rate of deterioration with years) for continuously reinforced concrete (CRC) pavements. The default user-defined model is based on Texas data (*6*).

Effect of Maintenance Intervention on Pavement Deterioration

The USER program accounts for the effects of current and future maintenance treatment intervention on the pavement condition by initializing the $(IRI)_0$ or $(PSR)_0$ to the appropriate expected condition after the overlay or other major maintenance treatment. The following three decision criteria are used in the algorithm:

1. Inbuilt asphalt concrete overlay option when $(PSR)_t$ reaches 2.0, or $(IRI)_t$ approaches 5.10. (This is the default

TABLE 1 Summary Inventory and Condition Data for Two Asphalt Surface Road Sections Used in Illustrative Examples

Summary Data	Road Sections	
	Poor Condition	Good Condition
Road/Beginning Mile	SR 12 E/0.000	SR 14E/14.614
Section Length	1.96 km (1.22 mile)	8.32 km (5.17 mile)
Number of Lanes	2	2
Lane Width	3.36 m (11 ft.)	3.05 m (10 ft.)
Pavement Surface Type	Asphalt	Asphalt
Date of Last Construction	1972	1982
Structural Number (SN)	3.2	3.2
CBR	10%	10%
Assumed Daily Traffic Volume	5000 ADT	5000 ADT
Assumed Traffic Growth	5%	5%
Truck Percentage	15%	15%
1991 Condition Survey		
IRI (m/km)	4.60 m/km	2.50 m/km
(PSR)	(2.17)	(3.31)

criterion; it can be modified by the user.) The default PSR after major MR&R treatments is 4.0.

2. Other major MR&R treatment options with the year of intervention, as generated by the maintenance policy specified by the user.

3. Annual routine and emergency maintenance and localized minor maintenance treatment not considered because these do not have significant effect on pavement performance.

Because the program uses IRI and PSR deterioration models, respectively, for flexible and rigid pavements, and because the intervention criterion is in terms of PSR, the World Bank IRI-PSR relationship, as listed in Equation 5, is utilized in the USER program.

$$IRI = 5.5 \log_n (5/PSR) \quad (5)$$

The effects of maintenance intervention on two asphalt pavements, each at various condition levels, are shown in Figure 3. In the good-condition pavement [Figure 3 (*top*)], with the initial IRI of 2.5 m/km, the HDM model prescribes a rehabilitation treatment of some 11 years compared with 27 years according to the USER model. On the contrary, in a poor-condition pavement [Figure 3 (*bottom*)] both models predict MR&R intervention in Year 3.

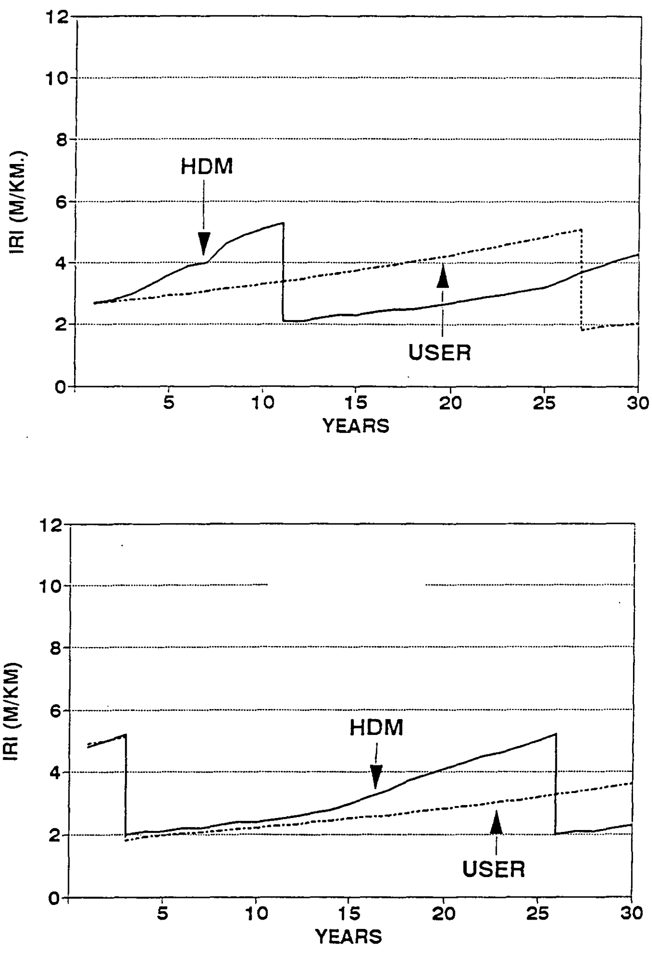


FIGURE 3 Comparison of maintenance intervention effect on pavement deterioration predictions by HDM and USER programs: *top*, County 4, SR14E, mile 14.614; *bottom*, County 4, SR12E, mile 0.000.

VOC COMPUTATIONS

The VOC associated with fuel consumption, tire wear, repair and maintenance of vehicles, depreciation, and pavement condition history over the analysis period are calculated by the user cost program relying on the built-in VOC consumption rate tables. These tables were originally developed in a comprehensive FHWA study of VOCs (*11*). The consumption rate for each VOC attribute is a function of the vehicle type and yearly volume, constant running speed, speed change cycles, and grade and curvature of the road section. The average running speed and consumption rates are adjusted for the prevailing pavement condition. These VOCs are calculated for each analysis year using the following generalized relationships.

$$(VOC)_{cy} = \sum_{i=1-5} \sum_{j=1-8} [(CONSTANT)_i (CONSUMPTION)_j (DADTY)_j (UNITCOST)_j]_i \quad (6)$$

where

$(VOC)_{cy}$ = total VOC for year y, at constant speed;
i = ith VOC attribute, where
i = 1, VOC associated with fuel consumption,
i = 2, VOC associated with oil consumption,
i = 3, VOC associated with tire wear,
i = 4, VOC associated with vehicle repair, and
i = 5, VOC associated with vehicle depreciation;
j = jth vehicle type, where
j = 1, small car,
j = 2, medium car,
j = 3, large car,
j = 4, pick-up and buses,
j = 5,6, single-unit trucks, and
j = 7,8, semitrailer and combination;
CONSTANT = a constant for constant-speed VOC calculation, a function of section length and consumption rate unit for each VOC attribute;
CONSUMPTION = consumption rate from VOC consumption tables, a function of vehicle type, speed, grade, curvature, and pavement condition;
DADTY = directional traffic volume for year y and vehicle type j; and
UNITCOST = unit cost associated with each VOC attribute.

The VOC calculation for speed-change cycles $(VOC)_{sy}$ is calculated using Equation 7.

$$(VOC)_{sy} = \sum_{i=1-5} \sum_{j=1-8} [(CHANGE)_j(CONSUMPTION)_j(DADTY)_j(UNITCOST)_j]_i \quad (7)$$

where CHANGE is a constant for speed-change cycle VOC calculation for each vehicle type. It is a function of vehicle type and associated number of speed-change cycles. The method of estimating speed-change cycles is based on FHWA's Highway Performance Monitoring System program (*12*).

$$\text{total VOC for each year } (VOC)_y = (VOC)_{cy} + (VOC)_{sy} \quad (8)$$

The VOC analysis programmed in the study enables us to calculate user costs arising from maintenance deferment and differential user costs (or indirect benefit) for improved pavement condition (major maintenance intervention to improve PSR/IRI) and improved road capacity and traffic flow (relieving traffic congestion and increasing the vehicle running speed).

TRAFFIC DELAY COSTS CAUSED BY MAINTENANCE INTERVENTION

The proposed methodology calculates TDC arising from expected traffic delays or interruptions, or both, as a result of maintenance intervention in the maintenance intervention year. The user cost referred to here is associated with overlay placement, reconstruction, or any other major MR&R treatment that requires traffic control and diversion or closure of one or more lanes, or both of these. The model developed in Texas (*4,6*) and used in the LCC1 program (*7*) is adopted in the USER software.

The TDC model considers a number of user-specified traffic diversion scenarios depending on the road classification and geometry. It first predicts the delay times incurred by each vehicle as it passes through the restricted work zones of MR&R treatment. These times are calculated using the production rates and quantity of work. TDC is calculated as a function of the calculated delay time, traffic volume, and inbuilt user delay unit cost (per unit time).

TOTAL USER COST AND VOC BENEFIT

After the VOC is calculated at both constant speed and according to speed change cycles and the traffic delay cost, the total user cost is estimated as follows:

$$\text{total user cost (for each year } y\text{)}, (USCOST)_y = (VOC)_y + (TDC)_y \quad (9)$$

The user cost methodology is also used as a surrogate for user benefit. Accordingly, the VOC user benefit is calculated by comparing the total life-cycle user cost (USCOST) for the base "do-nothing" alternative (with no improvement in pavement condition, capacity, or traffic flow) to the reduced life-cycle user cost for the recommended alternative strategies.

EXOGENOUS USER BENEFITS

Inputs can be provided for calculation of the following exogenous user benefits associated with each strategy being analyzed. These user benefit calculations are particularly useful in transportation investment planning of new facilities, network expansion, alternate routes for the existing facilities, or in selecting the most cost-effective maintenance strategy to improve pavement condition. Two categories of exogenous benefits are recognized:

- *Travel time cost savings*—Estimated from total travel time savings for each vehicle, travel time unit costs, and traffic volume.
- *Accident cost savings*—Estimated from expected reduction in the number of accidents (fatal, nonfatal, and property damage–only) unit cost per each accident type, and traffic volume.

APPLICATION OF USER SOFTWARE

The user cost and benefit methodology, assembled in the USER software described herein, can be used as a standalone program for project-level applications of transportation planning and pavement maintenance management. The USER software can also be integrated with the existing network-level PMS/MMS software packages with appropriate interface programs. The USER software provides customized reports for project-level applications and output files for integration with PMS/MMS software.

Figures 4 and 5 illustrate the VOC predictions of the USER software as compared with the HDM predictions for an existing pavement in good condition and another in relatively poor condition, respectively. The HDM program overpredicts the VOC in both pavements; however, the difference is substantial in the pavement in good condition.

Compared in Figures 6 and 7 are the VOC predictions (considering maintenance intervention) for the good- and poor-

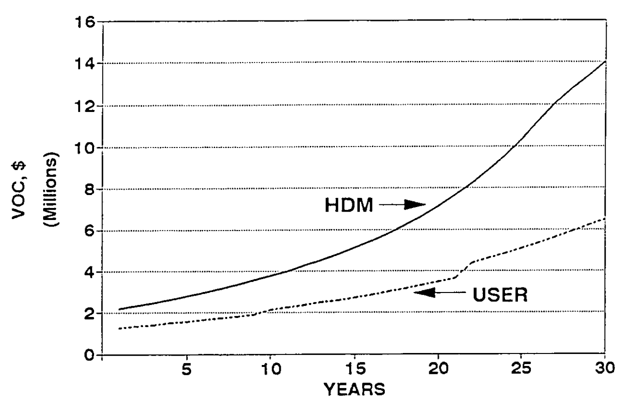


FIGURE 4 Comparison of VOC predictions by HDM and USER programs (without maintenance intervention) for a good pavement: County 4, SR14E, mile 14.614.

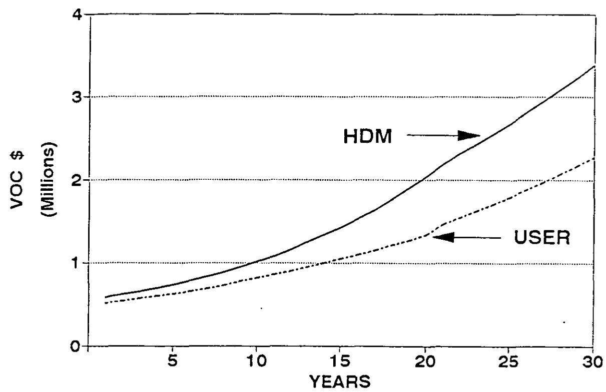


FIGURE 5 Comparison of VOC predictions by HDM and USER programs (without maintenance intervention) for a poor pavement: County 4, SR12E, mile 0.000.

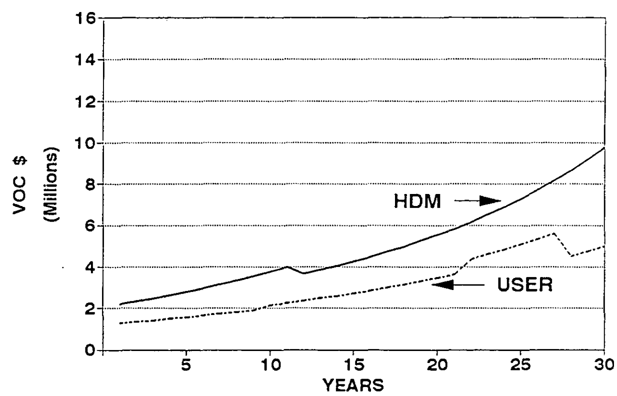


FIGURE 6 Comparison of maintenance intervention effect on VOC predictions by HDM and USER programs for a good pavement: County 4, SR14E, mile 14.614.

condition pavements, respectively. Again, the HDM overpredicts the VOC in both cases with the difference less pronounced in the poor-condition pavement.

The costs in dollars per vehicle mile, according to HDM and USER, are tabulated in Table 2. Comparing the VOC figures with and without maintenance, the life-cycle VOC decreases when the pavement is repaved at the opportune time as per the specified maintenance intervention policy. The USER analysis further indicates a substantial reduction in VOC if timely maintenance is performed on a pavement sec-

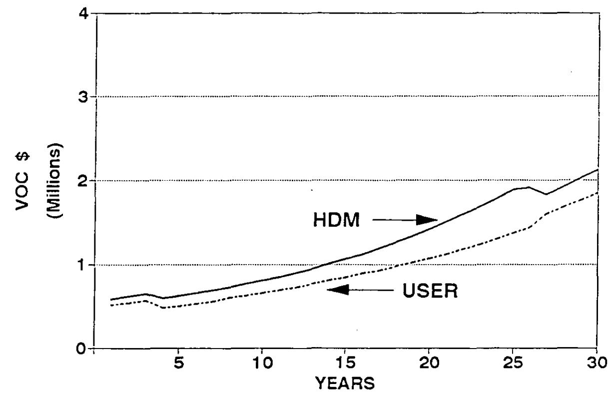


FIGURE 7 Comparison of maintenance intervention effect on VOC predictions by HDM and USER programs for a poor pavement: County 4, SR12E, mile 0.000.

tion in poor condition compared with that for a pavement in good condition. As predicted by the USER program, the user benefits from a VOC reduction in the case of adequately maintained pavements are $5.4 million/mi and $0.49 million/mi for the two scenarios (poor and good condition), respectively. These VOC user benefits are reduced by a small amount because of the traffic TDC associated with the maintenance treatments.

SUMMARY AND CONCLUSIONS

A comprehensive user cost and life-cycle analysis methodology is developed and coded in a microcomputer software USER for stand-alone project-level applications or for integration with existing network-level PSM/MMS software packages. Applicable in both flexible and rigid pavements, it incorporates the state-of-the-art pavement deterioration models, VOC parameters, and traffic delay cost methodology. The proposed life-cycle analysis methodology is applicable for quantifying the cost-effectiveness of timely and improved maintenance and rehabilitation alternatives and for calculating benefits for improved transportation investment planning scenarios associated with capacity and traffic flow and congestion and safety management.

Comparisons are made with the VOC analysis of flexible pavements by using the World Bank's HDM program. The results compare reasonably well in pavements in poor condition. However, the HDM program overestimates VOC for pavements in good condition.

TABLE 2 Summary of VOC Analysis

VOC Analysis Condition	Vehicle Operating Cost, $/Vehicle Mile			
	Good Pavement		Poor Pavement	
	HDM	USER	HDM	USER
(a) Without Maintenance	0.58	0.29	0.65	0.46
(b) With Maintenance Intervention	0.46	0.28	0.47	0.37

REFERENCES

1. W. R. Hudson, F. N. Finn, B. F. McCullough, K. Nair, and B. A. Vallerga. *Systems Approach to Pavement Design, System Formulation, Performance Definition and Materials Characterization.* Final Report, NCHRP Project 1-10. Materials Research and Development Inc., March 1968.
2. F. Moavenzadeh, E. Fernandez, and B. Brademeyer. *The Road Investment Analysis Model: General Framework for Capital Budgeting and Programming.* Series on Road Transport Planning Program for Developing Countries, Vol. 4. TAP Report 77-7. Technology Adaptation Program, Cambridge; Massachusetts Institute of Technology, Cambridge; Agency for International Development, U.S. Department of State, 1977.
3. W. R. Hudson, R. K. Kher, and B. F. McCullough. A Working Systems Model for Rigid Pavement Design. In *Highway Research Record 407*, HRB, National Research Council, Washington, D.C., 1972, pp. 130–145.
4. R. F. Carmichael and B. F. McCullough. *Modification and Implementation of the Rigid Pavement Design System.* Research Report 123-26. Center for Highway Research, University of Texas, Austin, Jan. 1975.
5. E. S. Lindow and J. J. Brown. A System for Life-Cycle Design of Pavement. Presented at International Conference on Concrete Pavement Design, U.S. Army Construction Engineering Research Laboratory, Champaign, Ill., 1977.
6. B. F. McCullough, W. Uddin, and J. P. Zaniewski. Design and Evaluation of Resurfacing Alternatives for Pavement Rehabilitation, *Proc., 6th International Conference on the Structural Design of Asphalt Pavements*, Ann Arbor, Mich., July 1987, Vol. 1, pp. 859–873.
7. W. Uddin, R. F. Carmichael III, and W. R. Hudson. A Methodology for Life-Cycle Cost Analysis of Pavements Using Microcomputer. *Proc., 6th International Conference on the Structural Design of Asphalt Pavements*, Ann Arbor, Mich., July 1987, Vol. 1, pp. 773–794.
8. H. Watanatada et al. *The Highway Design and Maintenance Standards Model.* Johns Hopkins University Press, Baltimore, Md, 1987.
9. W. D. Paterson and B. Attoh-Okine. Summary Models of Paved Road Deterioration Based on HDM-III. In *Transportation Research Record 1344*, TRB, National Research Council, Washington, D.C., 1992.
10. *NCHRP Report 277: Portland Cement Concrete Pavement Evaluation System.* TRB, National Research Council, Washington, D.C., 1985.
11. J. P. Zaniewski, B. C. Butler, G. Cunningham, G. E. Elkins, M. S. Paggi, and R. Machemehl. *Vehicle Operating Costs, Field Consumption and Pavement Type and Condition Factors.* Final Report. FHWA, U.S. Department of Transportation, June 1982.
12. *Highway Performance Monitoring System Analytical Process*, Vol. 1–4. Office of Highway Planning, FHWA, U.S. Department of Transportation, March 1983.
13. *Pavement Maintenance Management for Roads and Streets Using the PAVER System.* Technical Report M-90/05. U.S. Army Corps of Engineers, July 1990.
14. P. Bein, J. B. Cox, R. W. Chursinoff, G. H. Heiman, and G. A. Huber. Application of HDM3 Pavement Deterioration Model in Saskatchewan Pavement Management Information System. In *Transportation Research Record 1215*, TRB, National Research Council, Washington, D.C., 1989, pp. 60–69.

The contents of this paper reflect the views of the authors, who are responsible for the facts, findings, and the accuracy of the data presented herein.

Publication of this paper sponsored by Committee on Transportation Economics.

Cost Analysis of Paved Shoulders

BENJAMIN H. COTTRELL, JR.

A cost analysis on paved shoulders was performed. The scope was limited to shoulders made of an asphalt plant mixture used to extend the mainline pavement. The literature review generally supported the notion that paved shoulders are economically justifiable under certain conditions. However, there was no consensus on the specific conditions. The survey results showed that 91.4 percent of the state departments of transportation (DOTs) surveyed use paved shoulders on two-lane roads to some degree. Most or all shoulders were paved by 42.9 percent of these DOTs, and 40.0 percent have threshold values to warrant paved shoulders. For the average new two- and four-lane road projects, the initial cost increases 16.7 and 8.3 percent, respectively, and there is a service life increase of 14.3 percent with 0.61-m (2-ft) paved shoulders. For a resurfacing project, initial cost increases of 72.0 and 36.0 percent are realized with a 0.61-m paved shoulder on two- and four-lane roads, respectively. Through an economic analysis using the equivalent uniform annual cost method, it was revealed that 0.61-m paved shoulders for new two-lane roads are economically justifiable under certain average daily traffic volumes that depend on the road's functional classification and terrain type. Paved shoulders of 0.61 m are not economically justifiable for most existing two-lane roads. For four-lane and six-lane roads, 0.61-m paved shoulders are economically justifiable for all new roads and for existing roads above certain average daily traffic volumes.

The literature on paved shoulders presents evidence that paved shoulders reduce maintenance costs and accidents and are economically justifiable under certain conditions. However, there is no consensus on the specific conditions under which paved shoulders should be used.

OBJECTIVE AND SCOPE

The objective of the study was to perform a cost analysis of paved shoulders. The scope of the analysis was limited to shoulders made of asphalt plant mixture used to extend the mainline pavement. Because there are only a small number of primary roads with portland cement concrete surfaces, concrete pavement and shoulders were omitted. By limiting the analysis to arterial and collector roads, the impact of paved shoulders on arterial and collector roads may be evaluated and reviewed before considering implementation on local roads. Moreover, the potential savings in the cost of maintenance and the possible reduction in accidents are higher for arterial and collector roads. Although the use of paved shoulders for bicyclists is receiving much attention, the focus of this effort was on cost savings related to maintenance and accidents. Other research efforts are under way to address methods to accommodate bicyclists.

Virginia Transportation Research Council, Box 3817 University Station, Charlottesville, Va. 22903.

METHODOLOGY

Three tasks were undertaken to achieve the study's objective:

1. Literature on paved shoulders was reviewed.
2. A survey of other state departments of transportation (DOTs) on the use of paved shoulders was conducted.
3. A cost analysis on paved shoulders was performed, the methodology of which is described in following sections.

Literature Review

Although many studies of paved shoulders have been performed, many are either outdated (i.e., more than 20 years old) or flawed because of questions about the reliability of the data (study design and quality of the data) or the analysis of the data and results (statistical tests and interpretation of findings). Seven pertinent reports were reviewed (*1*).

There are numerous similarities and differences in the methods of analysis in and in the findings of these studies. The notion that paved shoulders are economically justifiable under certain conditions was generally supported. However, there is no consensus on what these conditions are.

Survey of State DOTs on Paved Shoulders on Two-Lane Roads

State DOTs were surveyed to determine their policies on paved shoulders on two-lane roads. Thirty-five state DOTs responded to the survey for a 70 percent response rate.

Paved Shoulder Use

Thirty-two of the thirty-five state DOTs (91.4 percent) used paved shoulders on two-lane roads in their design standards to some degree. Minimum paved shoulder widths of at least 0.61 m (2 ft) were used by 21 of the 32 state DOTs (65.6 percent). A 0.61-m minimum paved shoulder is used by 10 of the 32 state DOTs (31.3 percent). Most or all shoulders were paved by 15 state DOTs (42.9 percent). Paved shoulder criteria and the corresponding number of state DOTs are given in Table 1.

From a review of demographics and paved shoulder use, it was revealed that all of the northwestern state DOTs pave all shoulders. The winters are long and snowplows are used frequently, and paved shoulders provide a smoother, safer place for plows to operate. The ability of paved shoulders to keep water out of the base and subgrade is even more im-

TABLE 1 Paved Shoulder Criteria of State DOTs

Paved Shoulder Criteria		Number of State DOTs (percent)
Most or all shoulders paved		15 (42.8)
ADT and functional classification		4 (11.4)
ADT only		3 (8.6)
All principal arterials only		3 (8.6)
ADT and truck volume		3 (8.6)
Generally no paved shoulders		3 (8.5)
Criteria for RRR and construction/reconstruction only		2 (5.7)
ADT for new and reconstruction		1 (2.9)
Truck volume		1 (2.9)
	Totals	35 (100.0)

portant when snow is plowed onto the shoulders. No other demographic trend was noted.

Paved Shoulder Benefits

State DOTs were asked to identify the benefits of paved shoulders. Of the 35 state DOTs, 21 (60 percent) responded to this question. The responses are shown in Table 2. Ten state DOTs (28.6 percent) identified lateral support to the highway (longer service life) and reduced maintenance costs as benefits. In addition, other benefits mentioned that are of interest to this study are improved drainage, provision of a recovery area, edge raveling and pavement drop-off control, and decreased accident rates.

Cost Analysis

Cost analysis examines costs under the current Virginia DOT (VDOT) design policy of using unpaved shoulders and the proposed use of 0.61-m asphalt paved shoulders. The analysis focuses on two-lane minor arterials and collector roads and four-lane principal and minor arterials on the basis of functional classification. VDOT data are collected by administrative classification. Consequently, data for the primary system were used because the major target groups are in the primary system.

TABLE 2 Benefits of Paved Shoulders

Benefits	No. of State DOTS
Lateral Support to the Highway (Longer SVC Life)	10
Reduced Maintenance Costs	10
Accommodating Stopped Vehicles/Emergency Parking	6
Improved Drainage of Roadway	6
Providing a Recovery/Maneuvering	6
Edge Raveling/Pavement Drop-off Control	5
Decreased Accident Rate, Protecting Errant Vehicles	5
Bicycle Safety	4
Reduced Damage by Encroachment of Vehicles	3
Providing a Traffic Lane During Highway Rehab Work	2
Increased Safety for Pedestrians	2
Smoother, Safer Snow Plow Operation	2
A Cleaner Highway/Aesthetic Value	2
Providing for Agricultural Equipment	1
Providing a Sense of a Safe, Open Highway	1
Increased Sight Distance at Horizontal Curves	1
Maintain Capacity	1
Compensation of Off-Tracking	1
Providing a Bus Stop Area	1

Initial Cost

Initial costs were considered for constructing a new road and for resurfacing an existing road. The initial pavement cost for an average project [7.32-m (24-ft) width] and an average project with full-depth 0.61-m paved shoulders is given in Table 3 for a new road with a typical VDOT asphalt concrete mixture at 152.4-mm (6-in.) depth and a typical surface treatment at 38.1 mm (1.5 in.) over a 152.4-mm cement-treated aggregate base. The cost for resurfacing an existing 7.32-m-wide road and resurfacing with 0.61-m paved shoulders through trench widening provide for 38.1 mm of a typical surface treatment and traffic control. The trench widening consists of cutting out sod to make a 0.61-m trench on each side, filling the trench with 152.4-mm aggregate stone (152.4 mm of a typical VDOT-type asphalt concrete mixture), and overlaying with 38.1 mm of a typical surface treatment to make the shoulder even with the existing pavement. Service life information is also provided in Table 3. The initial cost increase on the average new road project is $21,163, or 16.7 percent. The initial cost increase for the average resurfacing project is $22,356, or 72.0 percent. Trench widening accounts for the increase. The corresponding increase in service life is 14.3 percent for both types of roads. The 7-year service life is based on historical data that reveal a 7.3-year average service life for primary roads. The expected 8-year service life with paved shoulders was based on the experiences of the North Carolina DOT and engineering judgment.

Maintenance Costs

Two types of maintenance activities are related to the use of paved shoulders: general shoulder maintenance (primarily for aggregate shoulders) and edge-of-pavement patching. Computerized printouts of maintenance expenditures for FY 1987–1988 for shoulder maintenance and patching on the primary system were provided by the maintenance division.

The total number of centerline kilometers of primary roads was used to determine the cost per kilometer. On the primary system, 69.5 percent of the kilometers are two-lane roads.

TABLE 3 Cost and Service Life Information

	New Road	Resurfacing Existing Road
Average project cost/kilometer	$214,292	$50,000
Cost/kilometer with 0.61-m shoulders	$249,974	$86,000
Cost/kilometer increase	$ 35,682	$36,000
Percent increase	16.7	72.0

Average project service life = 7 yrs
Service life with 0.61-m shoulders = 8 yrs
Service life increase = 1 yr
Percent increase = 14.3

The actual maintenance expenditures on two-lane primary roads are not available from the computerized records.

The North Carolina DOT staff conservatively estimates that a 75 percent reduction in shoulder maintenance is realized when the pavement is extended into the shoulder 0.61 m. From the state DOT survey, Iowa DOT listed the following annual costs per kilometer: for unpaved shoulders [earth/granular 1.8 to 3.1 m (6 to 10 ft) wide], shoulder maintenance cost was $350; for paved shoulders, shoulder maintenance cost was $130. Maintenance cost saving per kilometer resulting from paved shoulders was $220, or 62.9 percent. Consequently, shoulder maintenance cost saving of 62.9 percent for paved shoulders was used in the analysis. An estimated 25 percent reduction in the cost of pavement patching could be realized as a result of less raveling, less cracking, and improved drainage. The potential cost savings for asphalt shoulders on two-lane primary roads are calculated below using these estimated reductions.

VDOT average annual shoulder maintenance cost per kilometer = $369
Expected savings with paved shoulder per kilometer = $232
Annual shoulder maintenance cost with paved shoulder per kilometer = $137

VDOT average annual pavement patching cost per kilometer = $343
Expected savings with paved shoulders per kilometer = $86
Annual pavement patching costs with paved shoulders per kilometer = $257

VDOT total average annual shoulder-related maintenance cost per kilometer = $712
Expected savings with paved shoulders per kilometer = $318
Annual shoulder-related maintenance cost per kilometer with paved shoulders = $394

Accident Analysis

There were two objectives in the accident analysis: (a) to determine the expected reduction in accidents attributable to the use of 0.61-m paved shoulders, and (b) to determine the expected cost savings from the reduction in accidents.

Accident Reduction The accident prediction model developed by Zegeer et al. was selected because it was developed in a recent study based on an extensive sample size of 3,075 km (4,951 mi) of two-lane roads in seven states (*2,3*). This model was selected because (a) it includes head-on and sideswipe accidents, as well as single-vehicle accidents (all of which logically should be affected by roadway geometric features); (b) the coefficients and the R^2-value, 0.456, appear to be reasonable and consistent with the literature; and (c) terrain effects (flat, rolling, or mountainous) are incorporated into the model (*2*). The accident prediction model/equation is (*2*)

$$\begin{aligned} AO/KM/Y = {} & 0.0019(\mathrm{ADT})^{0.8824}(0.8786)^{3.28W} \\ & \times (0.9192)^{3.28\mathrm{PA}}(0.9316)^{3.28\mathrm{UP}} \\ & \times (1.2365)^{H}(0.8822)^{\mathrm{TER1}}(1.3221)^{\mathrm{TER2}} \\ & \div 1.61 \end{aligned} \quad (1)$$

where

$AO/KM/Y$ = related accidents (i.e., single-vehicle plus head-on plus opposite-direction sideswipe plus same-direction sideswipe accidents) (per km/year);
ADT = average daily traffic;
W = lane width (m);
PA = average paved shoulder width (m);
UP = average unpaved (i.e., gravel, stabilized, earth, or grass) shoulder width (m);
H = median roadside hazard rating (scale of 1 to 7 with 7 as the highest hazard rating);
TER1 = 1 if flat, 0 otherwise;
TER2 = 1 if mountainous, 0 otherwise.

The conditions for use are as follows:

1. Two-lane rural roads with an ADT of 100 to 10,000.
2. Lane widths of 2.4 to 3.6 m (8 to 12 ft).
3. Shoulders 0 to 3.6 m (12 ft) wide, that are paved or unpaved (or partly paved and partly unpaved).

Because the current concern is not site-specific, a median roadside hazard rating in the middle (4 to 6) was assumed, and 5 was selected as recommended in the informational guide (*3*). Although a confidence interval for $AO/KM/Y$ was desired, it was not determined because the standard error of the estimate was unknown.

A four-step process was used to determine the accident reduction and cost savings on the basis of the reduction in accident frequency for 0.61-m paved shoulders compared with the existing unpaved shoulder standards.

1. VDOT's road and bridge standards (*4*) were used to select ADT ranges (with some expansion), two functional road classes (arterials and collectors), and three terrain types, which in turn determine the lane and shoulder widths.
2. For improved accuracy, the equation was entered on a microcomputer spreadsheet program in lieu of using nomographs provided in the information guide. The equation was entered twice: (a) for the current unpaved shoulder design width, and (b) for the proposed 0.61-m paved shoulder plus the remaining shoulder design width unpaved. All other variables are the same for a given road design.
3. The reduction in the number of related accidents per kilometer per year was determined for the 0.61-m paved shoulder versus the standard shoulder design. This difference and the corresponding cost savings were calculated automatically with entry of the variables in the model.
4. Two matrix tables were developed: one for each road class for the accident frequency reduction and one for the related cost savings for various ADT and terrain types.

When the current road design unpaved shoulder width is changed to a 0.61-m paved shoulder and the remaining width is left unpaved, a 2.6 percent reduction in accident frequency is realized.

Accident Cost Savings FHWA's recommended approach for determining motor vehicle accident costs was used (*5*). FHWA's recommended accident costs are $1.5 million/fatality, $11,000/injury, and $2,000/vehicle for property damage–only (PDO) accidents. These costs per incident were used instead of cost per accident to include accident experience in Virginia for the specific accident types and accident severity. The following equation was used:

average cost per accident
= (percentage of fatal accidents
× number of fatalities/fatal accident
× cost/fatality + percent of injury accident
× number of injuries/injury accident
× cost/injury + percent of PDO accidents
× number of vehicles/PDO accident
× cost/vehicle) ÷ 100 (2)

Average cost per accident was determined for head-on, sideswipe same-direction and opposite-direction, and fixed-object, off-the-road accidents on the primary system. Using Virginia accident data from 1985 to 1987 (*6–8*), the equation yields

$$\text{average cost per accident} = (1.9401 \times 1.1812 \times 1{,}500{,}000 + 42.1759 \times 1.4613 \times \$11{,}000 \times 55.8840 \times 1.4703 \times \$2{,}000) \div 100 = \$42{,}797$$

FHWA's approach also states that the accident costs should be updated at least every 2 years. Consequently, a 7.8 percent increase was used on the basis of the increase in the consumer price index (CPI) from 1986 (the base year) to 1988 (*9*). Consequently, the average cost per accident becomes $46,135.

The cost saving per kilometer per year was determined by multiplying the average cost per accident and the accident frequency per kilometer per year. The accident cost savings range from $54 to $2,954, depending on type of highway, ADT, and terrain.

For principal arterials and four- and six-lane divided minor arterials, the average cost per accident was determined for sideswipe same direction, and fixed-object, off-the-road accidents on the primary system. Including increases from the CPI, the average cost per accident for divided roads was $33,186.

Analysis

The two alternatives were analyzed using the equivalent uniform annual cost (EUAC) method (*10*) as follows:

$$EUAC_A = -I(CR - i \text{ percent} - SL) - SM - PC \qquad (3)$$

where

$EUAC_A$ = equivalent uniform annual cost for Alternative A,
I = initial cost,
CR = capital recovery factor,
i percent = interest rate,
SL = service life (years),
SM = annual shoulder maintenance cost, and
PC = annual pavement patching cost.

An interest rate of 5.0 percent is used because the real-time value of money is 4.5 to 5.0 percent.

When comparing the EUAC of the current design (no paved shoulder) with the 0.61-m paved shoulder design, the latter has an EUAC $701 higher than the current design before accident cost savings are considered. In other words, an annual accident savings of $701 or more is necessary to economically justify the use of paved shoulders. The next step is to determine the ADT threshold that will result in the accident cost savings being equal to $701 for the three terrain types for each functional classification. At this ADT value, the costs of the two alternatives are equal; any ADT greater than the threshold will yield a savings for the 0.61-m asphalt paved shoulder alternative.

RESULTS

Two-Lane Roads

The results of the analysis (including ADT threshold values) are shown in Table 4 for new roads and for resurfacing existing roads. The ADT threshold values for new roads can be expected to be exceeded by some minor arterials and collectors.

The ADT threshold values for resurfacing existing roads are so high that almost all two-lane roads will not exceed the threshold values; therefore, the use of 0.61-m paved shoulders

TABLE 4 Analysis Results for Two-Lane Roads with a 0.61-m (2-ft) Paved Shoulder

	New Road	Resurfacing Existing Road
Difference in EUAC (current-proposed)	701	2,579
Minor Arterial		
ADT Threshold by Terrain Type		
Mountainous	3,705	16,210
Rolling	5,085	22,240
Level	5,860	25,635
Collector Roads		
ADT Threshold by Terrain Type		
Mountainous	2,690	11,755
Rolling	3,690	16,130
Level	4,250	18,595

by trench widening with the resurfacing of existing two-lane roads is not economically justified.

Four-Lane Roads

The analysis results for one direction of a four-lane road are presented in Table 5. For a new road, a savings of $1,031 is realized with a 0.61-m paved shoulder. The increase in service life more than offsets the initial cost increase and accounts for $695 or 67 percent of the savings. The remaining savings ($318) is from maintenance cost reductions. These savings are realized without considering accident reductions. Paved shoulders that are 0.61 m wide are economically justified for all new four-lane roads.

The ADT thresholds for four-lane existing roads should be used with caution. The accident model used was developed for two-lane roads. The model was used to determine the reduction in accidents expected when a paved shoulder exists. It is assumed that this accident reduction for four-lane roads would be similar to the accident reduction for two-lane roads. The same primary system accident data were used for both two- and four-lane roads. Moreover, there was no model available to predict such accident reductions specifically for four-lane roads.

Many four-lane undivided and divided roads exceed the threshold values. Therefore, the use of 0.61-m paved shoulders is economically justified with the resurfacing of existing roads with certain ADT volumes.

Limitations for Use of ADT Thresholds for Paved Shoulders

A 0.61-m paved shoulder provides a benefit by removing the pavement edge away from the travel lane. Consequently, reductions in shoulder maintenance and pavement edge raveling repairs are realized. To ensure that the 0.61-m paved shoulder is not used as part of a wider travel lane, it is required that all roads eligible for paved shoulders have a road width of 6.1 m or greater and have edgeline and centerline pavement markings. To be effective, edgeline markings must be installed to maintain a 0.61-m paved shoulder. In other words, the lane width must remain the same after installation of the paved shoulders.

Summary

This analysis of the alternatives was conducted with the available data. The analysis has a reasonable level of confidence. Although they are not exact, maintenance-related costs and accident costs are supported by the information available.

DISCUSSION OF RESULTS

Paved Shoulders for All Roads Versus Selected Roads

A policy to pave 0.61 m of shoulders on all roads with pavement widths of 6.1 m or wider would provide the most wide-

TABLE 5 Analysis Results for One Direction of a Four-Lane Road with 0.61-m (2-ft) Right Shoulder

	New Road	Resurfacing Existing Road
Difference in EUAC	−1,013	850
Undivided Road		
One Direction ADT Threshold by Terrain Type		
Mountainous	0	4,605
Rolling	0	6,320
Level	0	7,285
Divided Road		
ADT Threshold by Terrain Type		
Mountainous	0	5,700
Rolling	0	9,180
Level	0	10,580

spread impact. The design and programming process would be facilitated compared with a process with a decision-making step to determine whether paved shoulders are required. Blanket use of paved shoulders would yield statewide uniformity and consistency. Fifteen state DOTs (42.9 percent) pave most or all shoulders on arterials/primary/state roads.

By limiting paved shoulders to selected roads, usage may be restricted to roads that yield lower EUAC. Fourteen state DOTs (40 percent) use a threshold to determine when to use paved shoulders. The analysis indicated that paved shoulders provide a savings compared with the current design for new and existing roads with an ADT equal to or above those identified in the previous section. However, 0.61-m paved shoulders are not economically justified for existing two-lane roads. Paved shoulders are economically justified for all new four-lane roads, selected new two-lane roads, and existing four-lane roads.

Opposition from Subdivision Developers

If subdivision streets are required to have 0.61-m paved shoulders, then VDOT can expect to receive a considerable number of protests from developers. The additional costs will likely be passed on to home buyers. The costs can be justified based on lower maintenance costs for VDOT and safer roads for the subdivision residents. Paved shoulders should be used on new subdivision collector streets that exceed the ADT threshold values.

Paved 2-ft Shoulders Versus 1-ft Wider Lane and 1-ft Paved Shoulder

When a current unpaved shoulder is changed to a 0.61-m paved shoulder with the remainder unpaved, a 2.6 percent reduction in accident frequency is realized. When the lane width is increased by 0.31 m (1 ft) and 0.31 m of the shoulder width is paved, a 6.9 percent reduction in accident frequency is realized. An additional 4.3 percent reduction in accident frequency is realized for a 0.31-m wider lane and 0.31-m paved shoulder compared with the 0.61-m paved shoulder. When the design lane width is less than 3.6 m, substantial additional accident cost savings may be experienced without an increase in the initial cost. For example, a $701 accident cost savings increases to $1,844. The lane widening and paved shoulder combination is promising. Another alternative is to widen the lanes to 3.6 m and provide 0.61-m paved shoulders. On the other hand, based on the road design standards (*4*) and ADT threshold values, a 3.6-m lane width [3.36 m (11 ft) for selected mountainous areas] is expected at most locations that justify a 0.61-m paved shoulder.

One Direction of a Six-Lane Road

Because the conditions under which 0.61-m paved shoulders are economically justifiable on two- and four-lane roads have been identified, it is suspected that there may be some interest in identifying such conditions for six-lane roads. This analysis is presented in Table 6. As with four-lane roads, the accident analysis must be used with caution.

A 0.61-m paved shoulder is economically justified for all new six-lane roads. VDOT does not typically design new six-lane undivided roads. However, six-lane undivided roads sometimes result from the widening of a four-lane undivided road. The resurfacing of existing roads with 0.61-m paved shoulders is economically justifiable for many six-lane roads.

Paved 4-ft Shoulders for Bicyclists

It was suggested that a minimum paved shoulder of 1.22 m be used to accommodate bicyclists. This recommendation was based in part on VDOT's interest and support for accommodating bicyclists. The VDOT Bicycle Advisory Committee was established to examine the extent to which VDOT policies and standards accommodate bicyclists. The preferred method for accommodating bicyclists, be it a paved shoulder, a wider right lane, or other alternative, has not been identified. Nevertheless, the results of the analysis for 1.22-m (4-ft) paved shoulders are shown in Table 7 for two-, four-, and six-lane roads, respectively.

TABLE 6 Analysis Results for One Direction of a Six-Lane Road with 0.61-m (2-ft) Right Shoulder

	New Road	Resurfacing Existing Road
Difference in EUAC	− 2,217	568
Undivided*		
ADT Threshold by Terrain Type		
Mountainous	0	2,920
Rolling	0	4,005
Level	0	4,620
Divided		
ADT Threshold by Terrain Type		
Mountainous	0	4,240
Rolling	0	5,820
Level	0	6,710

*VDOT typically does not design new six-lane undivided roads.

TABLE 7 ADT Threshold Values for 1.22-m (4-ft) Paved Shoulders

	Mountainous	Rolling	Level
New Two-Lane Minor Arterials	12,800	17,565	20,245
Existing Two-Lane Minor Arterials	19,670	29,995	31,115
New Collector	9,285	12,740	14,685
Existing Collector	14,270	19,580	22,565
New Four-Lane Undivided Road	1,460	2,355	2,715
Existing Four-Lane Undivided Road	7.505	10,295	11,865
New Four-Lane Divided Road	2,125	3,420	3,940
Existing Four-Lane Divided Road	9,280	14,955	17,235
All New Six-Lane Undivided, and Divided Roads	0	0	0
Existing Six-Lane Undivided	6,580	9,030	10,410
Existing Six-Lane Divided	9,560	13,120	15,120

Notes: 1. Principal arterials are four- and six-lane divided roads. Multilane minor arterials are either divided or undivided.

2. For two-lane roads, the ADT threshold values are for total ADT and for a 0.61-m paved shoulder on both sides of the roadway. For multilane highways, the ADT threshold values are for one direction only and for a 0.61-m paved right shoulder.

3. VDOT typically does not design six-lane undivided roads.

Only on a limited number of new two-lane roads and practically on no existing two-lane roads can 1.22-m paved shoulders be economically justified (see Table 7). On the majority of new four-lane roads and a limited number of existing four-lane roads, 1.22-m paved shoulders can be economically justified. On all new six-lane roads and a limited number of existing six-lane roads, a 1.22-m paved shoulder can be justified. Paved shoulders are economically justified for a greater number of road kilometers than 1.22-m paved shoulders. Consequently, the potential cost savings are greater with implementation of the 0.61-m paved shoulder.

CONCLUSIONS

1. From a survey of state DOTs, it was found that

–Paved shoulders were used to some degree by 32 of 35 state DOTs (91.4 percent);

–Most or all shoulders are paved by 15 state DOTs (42.9 percent);

–Fourteen state DOTs (40.0 percent) have threshold values to warrant paved shoulders;

–Minimum paved shoulder widths greater than or equal to 0.61 m were used by 21 of 32 state DOTs (65.6 percent);

–Paved shoulders of 0.61 m are used by 10 of the 32 state DOTs (31.3 percent); and

–Ten state DOTs (28.6 percent) each noted lateral support of the highway and reduced maintenance costs as benefits of paved shoulders;

2. From the cost analysis, it was found that

–Initial cost increases on the average new road project are 16.7 and 8.3 percent with a corresponding service life increase of 14.3 percent when 0.61-m asphalt paved shoulders are used on two- and four-lane roads, respectively;

–When the current road design unpaved shoulder width is changed to a 0.61-m paved shoulder and the remaining width is unpaved, a 2.6 percent reduction in accident frequency is realized;

–An annual total maintenance cost savings of $512 is expected when using 0.61-m paved shoulders; therefore, an annual accident cost savings per kilometer of $1,129 is needed to economically justify 0.61-m paved shoulders on new roads;

–Paved shoulders 2 ft wide are economically justifiable on (a) all new four-lane roads and (b) new two-lane roads and existing four-lane roads that exceed ADT threshold values. They are not economically justifiable on most existing two-lane roads; and

–Roads that are eligible for paved shoulders must be greater than 6.1 m wide and have edgeline and centerline markings. For the paved shoulders to be effective, edgeline markings must be installed to maintain a 0.61-m paved shoulder.

3. Discussions on other issues concluded that

–Paved shoulders of 0.61 m are economically justified on all new six-lane roads and many existing six-lane roads;

–Paved shoulders of 1.22 m are economically justified on (a) all new six-lane roads, (b) a majority of new four-lane roads, and (c) a limited number of new two-lane roads and existing four- and six-lane roads; and

–The potential for installation and subsequent cost savings for 0.61-m paved shoulders is much greater than that for 1.22-m paved shoulders.

RECOMMENDATIONS

On the basis of the results of this study, the following recommendations are made: VDOT should consider using 0.61-m asphalt paved shoulders (mainline pavement extended) for all new four- and six-lane roads and for all roads that have ADT values that exceed those shown in Table 8. For existing roads that exceed the threshold, 0.61-m paved shoulders should be installed when resurfacing is scheduled. For paved shoulders to be considered, the roadway width must be 6.1 m or wider, and the road must have edgeline and centerline pave-

TABLE 8 Recommended ADT Threshold Values for 0.61-m (2-ft) Paved Shoulders

	Mountainous	Rolling	Level
New Two-Lane Minor Arterials	3,705	5,085	5,860
New Collector	2,690	3,690	4,250
All New Four- and Six-Lane undivided, and divided roads	0	0	0
Existing Four-Lane Undivided Road	4,605	6,320	7,285
Existing Four-Lane Divided Road	5,700	9,180	10,580
Existing Six-Lane Undivided	2,920	4,005	4,620
Existing Six-Lane Divided	4,240	5,820	6,710

Notes: 1. Principal arterials are four- and six-lane divided roads. Multilane minor arterials are either divided or undivided.

2. For two-lane roads, the ADT threshold values are for total ADT and for a 0.61-m paved shoulder on both sides of the roadway. For multilane highways, the ADT threshold values are for one direction only and for a 0.61-m paved right shoulder.

ment markings. After installation of the 0.61-m paved shoulders, centerline markings must be installed to keep the lane width unchanged and to maintain a 0.61-m paved shoulder. It may be desirable to round up the threshold values or otherwise simplify these values. For existing roads that have lane widths less than 3.6 m, it is suggested that the need to widen the lanes be determined through the appropriate VDOT process.

ACKNOWLEDGMENT

The study was supported by HPR funds administered through FHWA.

REFERENCES

1. B. H. Cottrell, Jr. *Review of Paved Shoulders*. Virginia Transportation Research Council, Charlottesville, Va., 1990.
2. *Safety Effects of Cross-Section Design on Two-Lane Roads*, Vol. 1. Final Report FHWA-RD-87/008. FHWA, U.S. Department of Transportation, 1987.
3. *Safety Cost-Effectiveness of Incremental Changes in Cross-Section—Informational Guide*. Report FHWA/RD-87/094. FHWA, U.S. Department of Transportation, 1987.
4. *Road and Bridge Design Standards*. Virginia Department of Transportation, Richmond, Va., 1989.
5. *Motor Vehicle Accident Costs*. FHWA Technical Advisory T7570.1. FHWA, U.S. Department of Transportation, 1988.
6. *1985 Summary of Accident Data*. Virginia Department of Transportation, Richmond, Va., 1985.
7. *1986 Summary of Accident Data*. Virginia Department of Transportation, Richmond, Va., 1986.
8. *1987 Summary of Accident Data*. Virginia Department of Transportation, Richmond, Va., 1987.
9. *1989 Annual Highway Safety Improvement Report*. Virginia Department of Transportation, Richmond, Va., 1989.
10. R. Winfrey. *Economic Analysis for Highways*. International Textbook Company, Scranton, Pa., 1969.

The opinions, findings, and conclusions expressed in this report are those of the author and not necessarily those of the sponsoring agencies.

Publication of this paper sponsored by Committee on Transportation Economics.

New Perspectives on Highway Investment and Economic Growth

GREGORY L. THOMPSON, BOB WELLER, AND E. WALTER TERRIE

In challenging the idea that highway investment leads to economic growth, it was hypothesized that both highway investment and economic growth are related to a third variable: decentralization. To test this idea with 1980 and 1990 data from the county level in Florida, three working hypotheses were postulated: (a) economic growth is a function not of highway investment but of the population growth rate, (b) highway investment is a function of the population growth rate, and (c) traffic congestion growth is a function of initial traffic congestion, growth in road capacity, and population growth. Equations to test the three hypotheses were estimated with county-level data from Florida. Data included population, jobs, income, traffic, and road growth between 1980 and 1990. Overall, the results generally failed to support the idea that both income growth and highway growth are related to the variable decentralization. Income growth appears to be weakly related to suburbanization; growth of the state highway system does not. On the other hand, traffic congestion, road construction, or the endowment of the road system did not influence population or job growth or growth in total income. The endowment of the state highway system may have influenced the growth of earned income, however, perhaps by promoting greater interaction. The results strongly show that building more roads induces greater vehicular use, although doing so somewhat reduces the amount of traffic on each mile of state highway in the short term.

Economists and planners increasingly debate the effect of highway investment on economic growth. Some argue that congestion and increased vehicle wear contribute to the widely reported falling rate of productivity of the United States workforce. They argue that congestion and vehicle wear result from reduced levels of highway spending in relation to demand, an imbalance that has occurred since the late 1960s. Thus, massive road-building programs would end congestion, reduce vehicle wear, and restore U.S. competitiveness (*1–5*).

Others challenge such claims. Schultze (*6*) and Winston (*7*) argue that previous theses (*1,2,4,5*) arise from spurious correlations. The slowdown in the growth of U.S. productivity that began in the early 1970s most likely spurred a slowdown in road investment rather than the opposite. Areas that grow rapidly can afford to build more roads. They explain that although economic growth requires roads and other infrastructure, a society can build too many roads in the wrong places. After a certain level of investment, greater productivity growth would result from investments in other sectors of the economy. Appealing to arguments of previous investigators (*8,9*), these economists and planners argue that society should use road pricing to determine the optimal level and location of road investment. Moreover, road pricing could eliminate productivity losses resulting from congestion.

Although Winston (*7*) and Small et al. (*8*) believe that road pricing would point toward some increased road construction in certain areas, others argue that too much road construction has led to U.S. economic decline. The increasingly decentralized, automobile-dependent organization of American life is responsible for reduced productivity (*10*). Pucher (*11*) concludes that large subsidies encourage automobile use and the organization of land uses that go with it, making the alteration of travel behavior difficult to impossible. It can be inferred from these previous arguments that, although decentralization has continued at a rapid rate since the 1960s, its pace would have been even faster had highways been built at a faster rate. Thus, increased highway spending would worsen rather than improve U.S. productivity because it would increase the pace of decentralization and thus contribute to increased travel to accomplish the same objective.

This paper addresses the debate by examining primarily cross-sectional relationships between highway capacity, economic growth, and decentralization in Florida. Data come from the early and late 1980s, providing two time points. Our hypotheses derive from major arguments in the literature, which we summarize first.

LITERATURE

Using national time-series data, Aschauer (*1*) concluded that public investments in core infrastructure explain a significant amount of labor productivity. Highways, transit, and water and sewers constitute core infrastructure. Declines in core infrastructure since the late 1960s can account for much of the reduction in the growth of labor productivity in recent years. In later work, Aschauer (*2*) explicitly tested the impact of highway capacity, measured as centerline miles of road per square mile, and congestion, measured as vehicle registrations per centerline mile, on productivity, measured as the growth rate of income per capita. His data consisted of pooled time series and cross-sectional data for the various states. He found that road capacity explains a significant amount of the growth of real per-capita income.

Munnell (*4*) found results similar to those of Aschauer. Core infrastructure investment explains a significant amount of economic growth. In follow-up work (*5*), she explicitly examined the contributions of highway compared with water and sewer investments on economic growth. She concluded

G. L. Thompson, Department of Urban and Regional Planning, R-117, Florida State University, Tallahassee, Fla. 32306. E. W. Terrie, Center for the Study of Population, Florida State University, Tallahassee, Fla. 32306. B. Weller, National Institutes of Health, Bethesda, Md. 20892.

that both stimulated economic growth but that water and sewer investments tended to complement private investment in the economy, whereas highway investments tended to substitute for private investment.

Basing their arguments on compelling empirical evidence showing that the marginal cost of highway use far exceeds the average costs in congested or worn-out environments, several economists criticize Aschauer and Munnell, who assume that the marginal cost of road use is zero. Schultze (*6*) examined individual factors contributing to the growth of the U.S. economy, concluding that infrastructure investments play a relatively insignificant role. Efficient highway pricing would do much more to stimulate growth than blind investment in highways in areas of congestion or heavy wear. In making these arguments he relies on the work of transportation economists, particularly Small et al. (*8*), which in turn relies on theoretical work of Mohring (*9*) and others. These economists generally conclude that some additional highway investment in high-demand corridors would improve public welfare, but determining where investment would yield net benefits poses difficulties without efficient highway pricing. In most areas efficient highway pricing also would alleviate congestion more efficiently than capacity expansion would because expansion leads to increased driving and ultimately to the levels of congestion approaching those before the expansion.

Some policy analysts argue that inefficient road pricing leads to inefficient land uses and government expenditures. Pucher (*11*) cites work by Lee (*12*) in discussing large subsidies to the U.S. road system. Newman and Kenworthy (*10*) use empirical evidence to show that greater road investments promote additional driving, which more than compensates for less pollution and energy consumed per vehicle mile from free-flowing vehicles. Altshuler (*13*) argues that increased road capacity does not relieve congestion but promotes greater decentralization. Putman (*14*) uses mathematical models to show similar results. Schultze also dismissed the strong correlation between economic growth and core infrastructure investment found by Aschauer. Schultze suggested that both were related to exogenous variables changing over time rather than to each other directly.

HYPOTHESES

The results that Aschauer and Munnell present do not entirely support their arguments. Aschauer's results showed, surprisingly, that congestion also seems to explain part of the growth of productivity—that is, productivity grows faster in areas that are more congested than those that are not. Aschauer's theoretical position predicts such an occurrence only if highways are loaded below the bottleneck level. His results thus imply that highways are not loaded to the bottleneck level, meaning that they are not congested severely enough to depress economic growth. If this is so, then it also would appear that additional highway investment would not stimulate economic growth. Munnell suggests that highway investment drives out other forms of investment. These results and the arguments of critics lead us to suspect a spurious correlation between economic growth and highway investment that we test in this paper.

We suspect that population decentralization accounts for a spurious relationship between growth rate and highway density. As people and jobs migrate from older central cities, income rises in the low-density areas receiving the population. Part of the income growth derives from greater productivity of new capital investment. The same activity undertaken in a new plant is likely to be more productive than when undertaken in the old plant left behind. However, a good part of the growth may be illusory and is merely the relocation of economic activity from one geographic area to another.

At the same time, funding formulae for highways dictate that most funding goes to areas with the most vehicle miles traveled. Such funds can build more roads where urban activity is dispersed, such as in the suburban environments of the Sun Belt states. Thus, the greatest road construction goes on in the areas receiving the most migration.

Thus, we suspected a spurious correlation between economic growth and centerline-mile density. Growth in both variables likely resulted from a third variable, which is population growth. We explore this idea by testing two hypotheses:

1. Economic growth is a function of population growth and is not related to highway investment; and
2. Highway investment is a function of population growth.

We also developed a third, related, hypothesis: road construction does not reduce traffic congestion. The third hypothesis is based on the idea that in congested environments the construction of new roads merely allows more travel to take place.

METHODS

We tested the three hypotheses in one state—Florida—with the use of county data on population, jobs, income, traffic, and road growth between 1980 and 1990. Data sources include the Florida Department of Transportation (FDOT), which compiles at the county level statistics on various categories of state highways, including centerline miles, lane miles, and vehicle miles traveled. The FDOT also made records available to us from which we tabulated lane miles of all roads in each county. The *Florida Statistical Abstract* yielded information on population and jobs.

We first tested the hypothesis that growth in real per-capita income between 1980 and 1990 was related to population and job growth more than to the presence of road capacity. We estimate two models:

$$\begin{aligned}\text{GIPC} = b_1 &+ b_2 * \text{IPC80} + b_3 * \text{AREA} \\ &+ b_4 * \text{POPD80} + b_5 * \text{BEACH} \\ &+ b_6 * \text{SUBDUM} + b_7 * \text{TRAF80} \\ &+ b_8 * \text{GPOP} + b_9 * \text{GJOBS} \\ &+ b_{10} * \text{LMD80} + b_{11} * \text{GLMD} \\ &+ b_{12} * \text{TLMD80} + b_{13} * \text{GTLMD} \qquad (1)\end{aligned}$$

where

GIPC = growth in real county per-capita income between 1980 and 1989;
IPC80 = per-capita income in 1980;
AREA = area of county (mi^2);
POPD80 = population density of county (people/mi^2);
BEACH = dummy variable denoting whether county has sandy ocean or gulf beach frontage;
SUBDUM = dummy variable denoting whether county is adjacent or bedroom to traditional urban counties of Dade, Hillsborough, Pinellas, Orange, Duval, and Escambia;
TRAF80 = state highway traffic congestion in 1980 (daily vehicle-mi traveled on state highways/lane-mi of state highway);
GPOP = population growth rate by county between 1980 and 1989;
GJOBS = county job growth rate;
LMD80 = density of state highways in county in 1980 (lane mi/mi^2 of county);
TLMD80 = similar measure for all state and local roads;
GLMD = growth rate in size of state highway system between 1980 and 1990; and
GTLMD = growth rate of entire road system over same period.

Equation 1 is similar to the Aschauer model, except that we use one time period and we include measures of growth.

We also reestimated Equation 1 with a different measure of income:

$$\begin{aligned} \text{GRIPC} = {} & b_1 + b_2 * \text{RIPC80} + b_3 * \text{AREA} \\ & + b_4 * \text{POPD80} + b_5 * \text{BEACH} \\ & + b_6 * \text{SUBDUM} + b_7 * \text{TRAF80} \\ & + b_8 * \text{GPOP} + b_9 * \text{GJOBS} \\ & + b_{10} * \text{LMD80} + b_{11} * \text{GLMD} \\ & + b_{12} * \text{TLMD80} + b_{13} * \text{GTLMD} \qquad (2) \end{aligned}$$

where GRIPC is the growth between 1980 and 1990 of earned income (total income less transfer payments, interest, and dividends). We deflated earned income with the consumer price index and the index of price variation between Florida counties for 1980 and 1990. RIPC80 is real earned income per capita in 1980.

Finally, we explained growth rates in county population and county jobs with the following equations:

$$\begin{aligned} \text{GPOP} = {} & b_1 + b_2 * \text{POPD80} + b_3 * \text{AREA} \\ & + b_4 * \text{RIPC80} + b_5 * \text{TRAF80} \\ & + b_6 * \text{LMD80} + b_7 * \text{TLMD80} \\ & + b_8 * \text{GLMD} + b_9 * \text{GTLMD} \\ & + b_{10} * \text{SUBDUM} + b_{11} * \text{BEACH} \\ & + b_{12} * \text{GJOBS} \qquad (3) \end{aligned}$$

and

$$\begin{aligned} \text{GJOBS} = {} & b_1 + b_2 * \text{JOBD80} + b_3 * \text{AREA} \\ & + b_4 * \text{RIPC80} + b_5 * \text{TRAF80} \\ & + b_6 * \text{LMD80} + b_7 * \text{TLMD80} \\ & + b_8 * \text{GLMD} + b_9 * \text{GTLMD} \\ & + b_{10} * \text{SUBDUM} + b_{11} * \text{BEACH} \\ & + b_{12} * \text{GPOP} \qquad (4) \end{aligned}$$

In testing the second hypothesis we estimated an equation explaining the growth in density of state highway lane miles:

$$\begin{aligned} \text{GLMD} = {} & b_0 + b_1 * \text{LMD80} + b_2 * \text{TLMD80} \\ & + b_3 * \text{AREA} + b_4 * \text{POPD80} \\ & + b_5 * \text{RIPC80} + b_6 * \text{GJOBS} \\ & + b_7 * \text{GPOP} + b_8 * \text{BEACH} \\ & + b_9 * \text{SUBDUM} + b_{10} * \text{TRAF80} \\ & + b_{11} * \text{GTLMD} + b_{12} * \text{GVMD} \qquad (5) \end{aligned}$$

where

GLMD = growth rate in state highway lane-mile density in each county between 1980 and 1989;
POP80 = county population in 1980;
VMD89 = state vehicle miles traveled per square mile of county; and
LMD89 = state highway lane-mile density in 1989.

In testing the third hypothesis, we estimated two equations. Equation 6 explains the growth in traffic congestion, whereas Equation 7 explains growth in vehicle miles traveled. In Equation 6

$$\begin{aligned} \text{GTRAF} = {} & b_0 + b_1 * \text{VMD80} + b_2 * \text{POPD80} \\ & + b_3 * \text{GPOP} + b_4 * \text{GJOBS} \\ & + b_5 * \text{GRIPC} + b_6 * \text{LMD80} \\ & + b_7 * \text{GLMD} + b_8 * \text{TLMD80} \\ & + b_9 * \text{GTLMD} + b_{10} * \text{BEACH} \\ & + b_{11} * \text{SUBDUM} \qquad (6) \end{aligned}$$

where

GTRAF = growth rate in state highway traffic congestion (daily vehicle-mi/lane-mi of state highway) between 1980 and 1989;
VMD80 = vehicle-mile density in 1980 (state highway vehicle-mi ÷ county area);
GLMD = growth rate in state lane-mile density in each county; and

GTLMD = growth rate in total lane-mile density in each county.

In Equation 7

$$\begin{aligned} GVMD = {} & b_0 + b_1 * VMD80 + b_2 * POPD80 \\ & + b_3 * GPOP + b_4 * GJOBS \\ & + b_5 * TRAF80 + b_6 * AREA \\ & + b_7 * LMD80 + b_8 * GLMD \\ & + b_9 * TLMD80 + b_{10} * GTLMD \\ & + b_{11} * BEACH + b_{12} * SUBDUM \\ & + b_{13} * GRIPC \end{aligned} \quad (7)$$

where GVMD is the growth rate of daily vehicle-mile density in a county on the state highway system.

Table 1 presents a description of the variables used in the study, and Table 2 presents correlations.

TABLE 1 Data Summary: 67 Cases

Variable	Mean	S.D.	Maximum	Minimum
RIPC80	5575.143	1369.725	9157.227	3079.103
AREA	809.731	395.260	2054.000	240.000
POPD80	0.182	0.375	2.758	0.005
JOBD80	0.074	0.153	1.057	0.001
BEACH	0.388	0.491	1.000	0.000
SUBDUM	0.194	0.398	1.000	0.000
TRAF80	2.813	1.764	8.656	0.726
GPOP	0.353	0.211	1.182	0.029
GJOBS	0.573	0.333	1.412	0.000
LMD80	0.620	0.444	3.311	0.141
TLMD80	4.044	3.990	30.670	0.772
GLMD	0.075	0.182	0.740	-0.152
VMD80	2.218	3.310	23.004	0.015
GTLMD	0.183	0.343	2.118	-0.160
GVMD	0.567	0.246	1.238	0.104
DVM	0.063	0.166	0.933	0.000

RESULTS

Tables 3 through 6 present the results of estimating Equations 1 through 4. Table 3 (Equation 1) only weakly explains total income growth, with growth in jobs (GJOBS), the initial level of income, and the suburban dummy variable (SUBDUM) having the greatest explanatory power. Population growth (GPOP) has no explanatory power; neither does traffic (TRAF80) or the endowment of roads (TLMD80 and LMD80) or the growth rates in roads (GLMD and GTLMD). Table 4 (Equation 2) shows similar results in explaining the growth of earned income, except that the endowment of the state highway system is significant in explaining county earned-income growth at the 10 percent level. Although the standard error for the coefficient is relatively high, the estimated coefficient yields an elasticity of income growth with respect to the endowment of state highway lane miles of somewhat less than 1. Tables 5 and 6, explaining GJOBS and GPOP, show that traffic congestion, the suburban and beach dummy variables, the road endowment variables, and the road growth variables have little explanatory power. Together the four equations weakly support Hypothesis 1—that income growth is related to suburbanization and not road construction. Except in one instance, road variables have little explanatory power in explaining growth in total income, jobs, or population. Traffic congestion appears to have no impact on any of these growth rates. The endowment of the state highway system, however, may have some power explaining growth in earned income.

Table 7 shows the estimate of Equation 5. The table shows that growth in state highway lane miles took place primarily where traffic was heavier in 1980, where traffic grew the most rapidly, where the endowment of state highway lane miles was weaker in 1980, and in beach areas. Neither the growth rate in population or jobs nor the suburban dummy variable (SUBDUM) had much of an impact on growth in state highway lane miles. These results show that the supply of state highway facilities was not related to growth rates in population or jobs, disproving the second hypothesis.

Tables 8 and 9 present the estimation of Equations 6 and 7, respectively, and speak to the third hypothesis that road construction does not reduce congestion but instead induces greater amounts of driving. The tables suggest that the first hypothesis is wrong, at least in the short term, but that the second hypothesis is true. Expansion of the state highway system did slow the rate of congestion in the short term, but it also induced greater volumes of travel. Looking at Table 8 we see that the greater the vehicle-mile density in 1980, the less congestion grew over the next 10-year period. State highway construction over the same 10 years mitigated the growth of congestion, whereas growth in total lane-mile density had

TABLE 2 Correlation Matrix

	RIPC80	AREA	POPD80	JOBD80	BEACH	SUBDU	TRAF80	GPOP	GJOBS	LMD80	TLMD80	GLMD	VMD80	GTLMD	GVMD
RIPC80	1.00	0.37	0.34	0.42	0.23	0.13	0.56	-0.04	-0.15	0.37	0.24	0.35	0.41	0.01	0.040
AREA		1.00	-0.00	0.06	0.14	0.24	0.31	0.08	0.00	-0.11	-0.14	0.21	0.01	-0.01	0.065
POPD80			1.00	0.99	0.33	0.06	0.70	-0.13	0.04	0.89	0.94	0.36	0.97	-0.01	-0.069
JOBD80				1.00	0.33	0.01	0.74	-0.15	0.00	0.90	0.91	0.39	0.98	-0.02	-0.060
BEACH					1.00	0.23	0.38	0.21	0.25	0.29	0.31	0.53	0.31	0.11	0.445
SUBDUM						1.00	0.30	0.12	0.23	0.03	0.00	0.12	0.07	0.04	0.028
TRAF80							1.00	0.05	0.14	0.61	0.62	0.65	0.76	-0.03	0.098
GPOP								1.00	0.61	-0.10	-0.08	0.03	-0.10	0.41	0.268
GJOBS									1.00	0.02	0.10	0.10	0.03	0.30	0.271
LMD80										1.00	0.88	0.20	0.94	0.01	-0.032
TLMD80											1.00	0.30	0.91	-0.08	-0.005
GLMD												1.00	0.34	0.30	0.488
VMD80													1.00	-0.01	-0.095
GTLMD														1.00	0.096
GVMD															1.000

TABLE 3 Growth in Real Per-Capita Total Income

VARIABLE	COEFFICIENT	T-STAT.
C	0.54194	4.80646
IPC80	-0.04003	-2.49325
AREA	-0.00001	-0.11169
POPD80	0.18074	0.95192
BEACH	0.03829	0.76600
SUBDUM	0.10825	2.10500
TRAF80	-0.02068	-0.92454
GPOP	-0.07492	-0.55289
GJOBS	0.22892	3.03368
LMD80	-0.01744	-0.15525
GLMD	0.22259	1.15965
TLMD80	-0.00811	-0.49216
GTLMD	-0.02584	-0.39107
R-squared		0.35808
Adjusted R-squared		0.21543
F-statistic		2.51021

TABLE 4 Growth in Real Per-Capita Earned Income

VARIABLE	COEFFICIENT	T-STAT.
C	0.40410	2.61491
RIPC80	-0.00006	-2.32519
AREA	-0.00000	-0.04832
POPD80	0.00205	0.00759
BEACH	0.05780	0.84191
SUBDUM	0.20064	2.75878
TRAF80	-0.01997	-0.60676
GPOP	-0.11583	-0.66267
GJOBS	0.22356	2.05319
LMD80	0.28015	1.75033
GLMD	-0.00621	-0.02624
TLMD80	-0.01782	-0.76307
GTLMD	-0.14659	-1.56082
R-squared		0.360166
Adjusted R-squared		0.217980
F-statistic		2.533060

TABLE 5 Job Growth Rate

VARIABLE	COEFFICIENT	T-STAT.
C	0.41723	2.04472
JOBD80	-0.45911	-0.55374
AREA	0.00001	0.08054
RIPC80	-0.00006	-1.86766
TRAF80	0.04538	1.10925
LMD80	-0.10586	-0.49795
TLMD80	0.02936	1.15772
GLMD	-0.10530	-0.35592
GTLMD	0.11812	1.03536
SUBDUM	0.08893	0.96659
BEACH	0.07837	0.92550
GPOP	0.76217	3.94354
R-squared		0.479059
Adjusted R-squared		0.374870
F-statistic		4.598009

TABLE 6 Population Growth Rate

VARIABLE	COEFFICIENT	T-STAT.
C	-0.00796	-0.06681
POPD80	-0.37706	-1.86956
AREA	0.00006	0.96198
RIPC80	0.00000	0.12765
TRAF80	0.01981	0.78476
LMD80	0.00295	0.02390
GLMD	-0.11598	-0.63784
TLMD80	0.02274	1.28091
GTLMD	0.18971	2.79873
SUBDUM	0.01564	0.27897
BEACH	0.04352	0.82688
GJOBS	0.29236	3.94189
R-squared		0.498973
Adjusted R-squared		0.398767
F-statistic		4.979499

TABLE 7 Growth in Lane-Mile Density

VARIABLE	COEFFICIENT	T-STAT.
C	-0.10686	-1.39549
LMD80	-0.27469	-3.85851
TLMD80	-0.00406	-0.35742
AREA	-0.00007	-1.64245
POPD80	0.22970	1.71649
RIPC80	0.00001	0.46938
GPOP	-0.07587	-0.90489
GJOBS	-0.04427	-0.84346
BEACH	0.07259	2.25842
SUBDUM	-0.02006	-0.57474
TRAF80	0.07474	5.77836
GTLMD	-0.00330	-0.07290
GVMD	0.29502	4.81128
R-squared		0.743714
Adjusted R-squared		0.686762
F-statistic		13.05854

TABLE 8 Growth in Traffic

VARIABLE	COEFFICIENT	T-STAT.
C	0.23615	3.28854
VMD80	-0.13909	-3.29714
POPD80	0.30081	0.92077
GPOP	0.15863	1.15443
GJOBS	0.02999	0.34189
GRIPC	0.16789	1.67139
LMD80	0.60526	3.35681
GLMD	-0.35690	-2.31062
TLMD80	0.00544	0.30831
GTLMD	0.01993	0.27566
BEACH	0.04858	0.89102
SUBDUM	-0.05076	-0.87405
R-squared		0.554720
Adjusted R-squared		0.465664
F-statistic		6.228894

TABLE 9 Growth in Vehicle-Mile Density

VARIABLE	COEFFICIENT	T-STAT.
C	0.06027	0.52841
VMD80	-0.19361	-3.42050
POPD80	0.39477	1.07586
GPOP	0.14734	1.04761
GJOBS	0.03057	0.34122
TRAF80	0.01675	0.50285
AREA	0.00008	1.15677
LMD80	0.81384	4.06194
GLMD	0.89035	4.61408
TLMD80	0.01397	0.73454
GTLMD	0.04799	0.62467
BEACH	0.04825	0.86495
SUBDUM	-0.07784	-1.22199
GRIPC	0.19454	1.86752
R-squared		0.617919
Adjusted R-squared		0.524201
F-statistic		6.593382

no impact on traffic congestion of the state highway system. Table 9 shows that the endowment of the state highway system in 1980 had a great impact on stimulating growth in total driving, whereas the growth rate in highway construction between 1980 and 1990 had an even greater impact. The magnitude of vehicle-mile density in 1980 significantly depressed further traffic growth.

CONCLUSIONS

In challenging the idea that highway investment leads to economic growth, we hypothesized that both highway investment and economic growth are related to a third variable: decentralization. To test this idea with 1980 and 1990 data from the county level in Florida, we postulated three working hypotheses:

1. Economic growth is not a function of highway investment but is a function of the population growth rate;
2. Highway investment is a function of the population growth rate; and,
3. Traffic congestion growth is a function of initial traffic congestion, growth in road capacity, and population growth.

Our results generally confirm the first hypothesis, although this statement is qualified by the definition of real income growth per capita. If income growth is defined as the growth of total income per capita, including transfer and investment income, its growth was not influenced by highway investment at the 1, 5, or 10 percent levels of significance. Although population growth rate also had no effect on the growth of total real per capita income, job growth rate did, as did a dummy variable denoting suburbanization.

On the other hand, if income growth is defined as real earned income per capita (no transfer or investment income included), highway investment had no impact on its growth at the 1 and 5 percent significance levels, but it did at the 10 percent level. Highway investment may have some impact on worker productivity, but the growth rate in jobs and the suburbanization dummy variable had greater explanatory power. We also found that highway investment had no explanatory power in the job or population growth rates of counties.

Our results disprove the second hypothesis, if highway investment is defined as the growth in state highway lane miles per square mile. The magnitude of traffic congestion and the growth rate in vehicle miles per square mile had the greatest explanatory power in the growth rate of state highway lane mile density; population and job growth rates had no explanatory power. However, the population growth rate does partly explain the growth rate in total highway lane mile density, which consists mostly of access roads.

In regard to the third hypothesis, we found that the greater the magnitude of driving per square mile in 1980, the less traffic grew on each mile of road. This confirmed part of our hypothesis. We also found that the greater the extent of the state highway road system in 1980, the greater the growth rate in traffic congestion. On the other hand, construction of additional highway capacity between 1980 and 1990 reduced the growth rate of the volume of traffic on each mile of state highway, which is a finding contrary to our hypothesis. Overall, however, adding miles of state highways stimulated additional driving.

Overall, the results generally fail to support the idea that both income growth and highway growth are related to decentralization. Income growth appears weakly related to suburbanization; growth of the state highway system does not. It is more influenced by the magnitude of road use and the growth of road use. Although both road use and the growth of road use may be greatest in areas that are undergoing suburbanization, most of the variables that we used to identify decentralization (GPOP, GJOBS, SUBDUM, and BEACH) do not support this notion. Only BEACH had a significant effect on explaining the growth of state highway lane miles.

On the other hand, neither traffic congestion, road construction, nor the endowment of the road system influenced population or job growth or growth in total income. The endowment of the state highway system may have influenced the growth of earned income, however—perhaps by promoting greater interaction. The results strongly show that building more roads induces greater vehicular use, although doing so reduces somewhat the amount of traffic on each mile of state highway in the short term.

Our results show little support for the idea that road construction leads to economic growth. This is not counter-intuitive: although transportation investment clearly is important to economic growth, it has diminishing returns. Early canals and railroads stimulated economic growth because their introduction into regions without improved transportation had a huge impact on regional accessibility. On the other hand, most regions today enjoy an abundance of improved transportation facilities. The addition of a new road further improves regional accessibility only marginally and may not be worth the well-documented direct and indirect costs associated with road construction, operation, and maintenance. It is possible that a society could have too many highways rather than too few. Another view is that more roads promote the proliferation of low-value, unproductive travel. The demand for low-value travel may be elastic, which means that little of it occurs when prices are high but lots of it occurs when prices are low, as in the U.S. context. The congestion that this type of travel causes unfortunately impedes high-value, productive travel, whose demand is inelastic. Because prices have little impact on high-value travel, increased road construction will not stimulate it very much. Instead, increased road construction would stimulate large increases in unproductive travel and the dispersed land uses that go with such travel. Although severe congestion may depress economic growth (although our aggregate results do not suggest this, as shown in Tables 3 and 4), severe congestion costs nothing with respect to capital outlay or destroyed neighborhoods, and it restricts growth in vehicle miles traveled, thus restricting growth in energy consumption and pollution. Expanding road capacity has the opposite effect. Only where it can be demonstrated that severe congestion depresses economic growth could road expansion be justified and then only if benefits outweighed costs. Cost-benefit analyses or efficient road pricing could determine when such conditions were met.

REFERENCES

1. D. A. Aschauer. Is Public Expenditure Productive? *Journal of Monetary Economics*, Vol. 23, 1989, pp. 177–200.
2. D. A. Aschauer. Highway Capacity and Economic Growth. In *Economic Perspectives*, Federal Reserve Bank of Chicago, Sept.–Oct. 1990, pp. 14–24.
3. D. A. Denslow. Highways and Economic Development in Florida. *Economic Leaflet*, Vol. 50, No. 3, March 1991, University of Florida, Gainesville.
4. A. H. Munnell. Why Has Productivity Growth Declined? Productivity and Public Investment. *New England Economic Review*, Jan.–Feb. 1990, pp. 3–22.
5. A. H. Munnell. How Does Public Infrastructure Affect Regional Economic Performance? *New England Economic Review*, Sept.–Oct. 1990, pp. 11–32.
6. C. L. Schultze. The Federal Budget and the Nation's Economic Health. In *Setting National Priorities* (H. Aaron, ed.), Brookings Institution, Washington, D.C., 1990, pp. 19–64.
7. C. Winston. Efficient Transportation Infrastructure Policy. *Journal of Economic Perspectives*, Vol. 5, 1991, pp. 113–127.
8. K. A. Small, C. Winston, and C. A. Evans. *Road Work: A New Highway Pricing and Investment Policy*. Brookings Institution, Washington, D.C., 1989.
9. H. Mohring. *Transportation Economics*. Ballinger Publishing Co., Cambridge, Mass., 1976, Chapter 3, p. 71.
10. P. W. G. Newman and J. R. Kenworthy. Gasoline Consumption and Cities. *Journal of the American Planning Association*, Vol. 54, Winter 1988, pp. 24–37.
11. J. Pucher. Urban Travel Behavior as the Outcome of Public Policy: The Example of Modal-Split in Western Europe and North America. *Journal of the American Planning Association*, Vol. 54, Autumn 1988, pp. 509–520.
12. D. Lee. *Highway Infrastructure Needs*. U.S. Department of Transportation, Transportation Systems Center, Cambridge, Mass., 1987, pp. 17–19, A-9.
13. A. Altshuler. *The Urban Transportation System*. MIT Press, Cambridge, Mass., 1979.
14. S. H. Putman. *Integrated Policy Analysis of Metropolitan Transportation and Location*. Report DOT-P-30-80-32. U.S. Department of Transportation, 1980.

Publication of this paper sponsored by Committee on Transportation Economics.

Evaluation of Alternative Network Preservation Strategies

EDWIN C. NOVAK, JR., WEN-HOU KUO, AND GILBERT Y. BALADI

A study was conducted to evaluate the effects of underfunding on the total long-term cost of preserving networks. Many state highway agencies are experiencing both declining revenues and declining trunkline conditions. This decline raises the following question: Which alternative has the lowest total cost of preservation—allowing the network to decline from its current condition and later restoring it or maintaining its current condition? The pavement management system developed for the Michigan Department of Transportation is a network management system that includes the ability to evaluate the effects any given funding scheme has on the long-term (40-year) total cost of network preservation. Five funding schemes were analyzed for the preservation of two networks consisting of more than 11,000 lane-mi of pavement. A manual version of Michigan's network management system was used because it provides a simple means of illustrating how a network management system is used to control the long-term relationship between funding streams and network condition. Program costs were estimated on the basis of 3 years of historical project cost data. Five alternative funding schemes were evaluated, ranging from maintaining current condition to doing nothing for the first 10 years and then restoring the current condition. The study showed that the total agency cost over a 40-year analysis period can be highest when networks are maintained in current condition and lowest when they are allowed to deteriorate for 10 years before restoring and then maintaining current condition.

The long-term decline of the condition of many highway networks raises questions about future funding and revenue needs. It has long been held that timely maintenance and rehabilitation of networks will reduce their preservation cost and that inadequate funding would result in large increases in the future total cost of preserving networks compared with the cost of preservation. In an effort to study the long-term effects of various funding schemes, a study was conducted of the three Michigan Department of Transportation (MDOT) Highway Districts 5, 6, and 7, shown in Figure 1. The analysis methods are based on MDOT's network management system (*1*). Network management principles are based on the remaining service life (RSL) concept (*2*), network strategy analysis (*1,3,4*) and network life-cycle cost (LCC) (*5*). Actual district network performance and historical cost data are used for this study.

Analysis of alternative funding schemes is based on a manual computation version (*6*) of MDOT's network management system. Network performance bar charts are used to illustrate network condition and its rate of deterioration, and a simple cost matrix based on historical cost data is used to estimate the cost of alternative funding schemes.

E. C. Novak, Jr. and W.-H. Kuo, Materials and Technology Division, Michigan Department of Transportation, P.O. Box 30049, Lansing, Mich. 48909. G. Y. Baladi, Department of Civil and Environmental Engineering, Michigan State University, East Lansing, Mich. 48824.

MDOT's application software system was used to obtain current network performance data. All other analysis products were manually computed in accordance with equations later explained. Typically, agency executives would need to evaluate alternatives in more detail than is presented in this paper. This can be accomplished within a reasonable time only by using application software designed for network management.

This paper is intended to illustrate how network management systems are used to evaluate alternative funding schemes before the preservation project and program development process, how they can simplify the process of allocating funds, and how there is economic advantage in placing network needs above the needs of individual projects.

BASIC NETWORK MANAGEMENT CONCEPTS

The performance of projects, networks, strategies, and programs is characterized by their lane-mile length and their average remaining service life (RSL). At the time of construction, the design service life (DSL) of projects and programs is the same as their RSL. Condition is considered to be poor or no longer acceptable when it deteriorates to an unacceptable level, referred to as the threshold value. The performance of networks is based on the RSL of the uniform sections that they consist of. The performance of programs is initially based on the DSL of the projects they consist of and later on the RSL of its projects. For networks, the sections of pavement of most concern are those in poor condition. They make up the majority of projects considered for annual programs. On the basis of remaining life methodology, Figure 2 illustrates the network deterioration process. The rehabilitation process simply moves projects from lower to higher RSL categories in accordance with their DSL. Network performance expressed in terms of RSL enables the use of an accounting process to keep track of the rate at which projects or uniform sections are deteriorating from each higher to each lower RSL category, the rate at which they are rehabilitated out of lower RSL categories, and to which higher RSL category the designers estimate of DSL would place them.

Relationship Between Network Performance, MR&R Programs, and MR&R Strategies

The condition of a network is simply the percentage of it that has an RSL of zero, which is the same as the percentage of network in poor or unacceptable condition. Network condition is a function of its rate of deterioration and the network

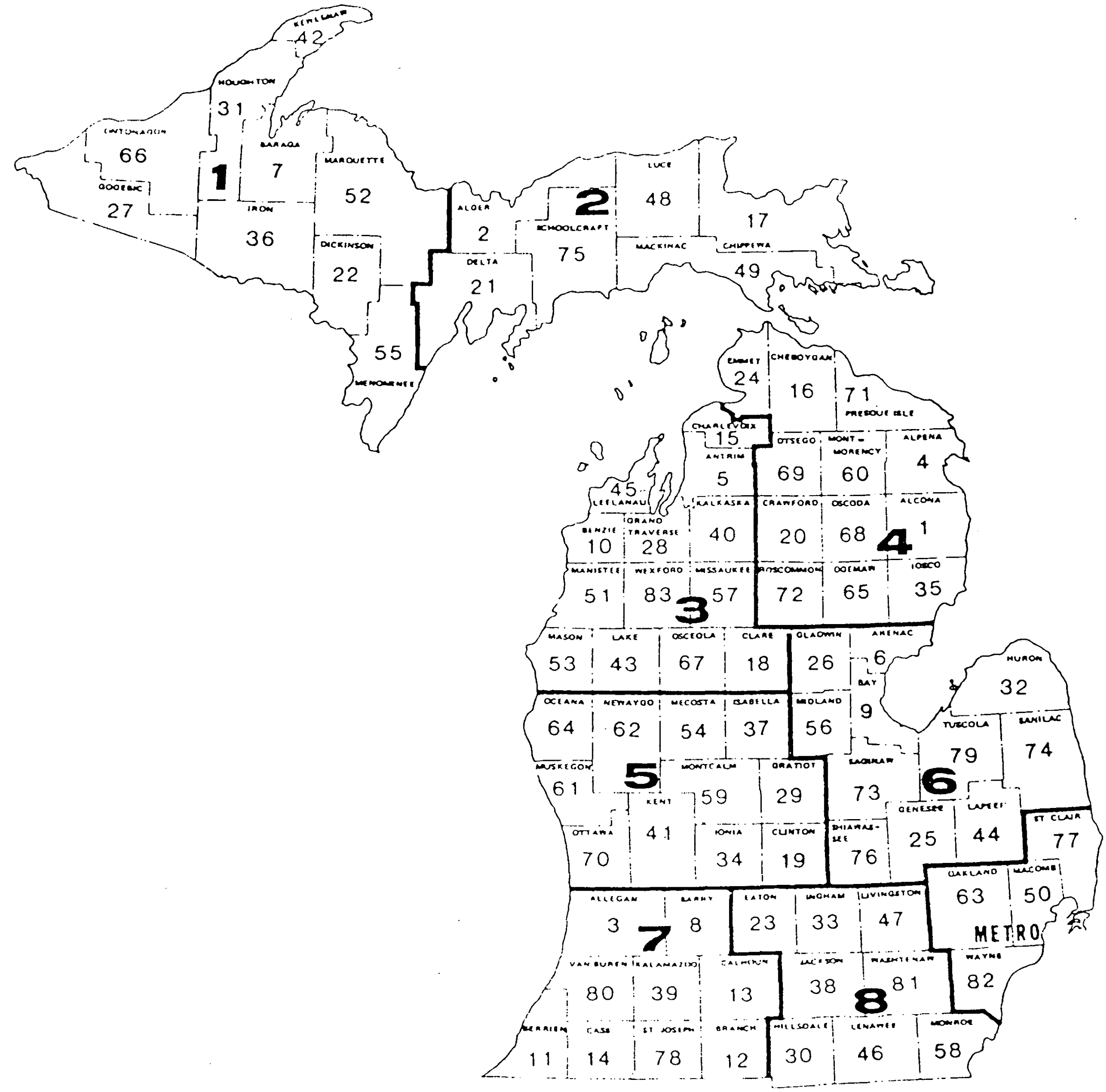


FIGURE 1 Map of Michigan showing locations of Districts 5, 6, and 7.

strategy used to preserve it. A maintenance, rehabilitation, and reconstruction (MR&R) strategy is defined as the percentage of network to be annually rehabilitated from each lower to each higher RSL category, and it is used as one of the MR&R program development constraints. For network management, it is beneficial to deal with network strategies rather than MR&R programs. This is so for two reasons: strategies eliminate the need to identify candidate projects, and the number of alternative funding schemes is not limited. For convenience, MR&R strategies can be generalized to the percentage of network annually preserved and its average DSL, in which case it is called a network strategy. The relationship between network condition (at equilibrium) and network strategy is as follows:

$$P_0 = 100 - (P \times \text{DSL}) \qquad (1)$$

where P_0 is the network condition (percentage of network in the zero RSL category) and the annual MR&R strategy consists of P as the percentage of network annually preserved and DSL as the strategy's average design service life.

Equation 1 can be used to estimate the resulting condition of any network, given the MR&R strategy that is to be followed by annual MR&R programs.

The network's average RSL is calculated as follows:

$$\text{network RSL} = \sum X_i Y_i / 100 \qquad (2)$$

where X_i is the RSL of the ith uniform section and Y_i is the percentage of network in the ith uniform section.

The products of Equations 1 and 2 were combined to form the chart shown in Figure 3. This chart relates resulting net-

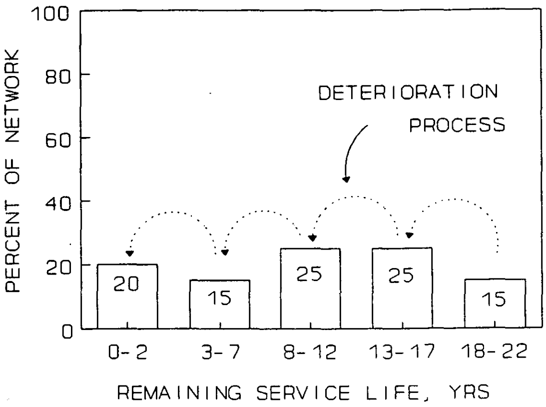


FIGURE 2 Network deterioration process.

work condition and RSL to the network strategies with which the MR&R program must comply. Its uses include the following:

- Given a network condition and RSL objective, what network strategy must each annual MR&R program comply with?
- Given a network strategy, what will the resulting network condition and RSL be?
- Given the desired network condition level, what RSL objective will maintain the desired condition level at lowest network LCC?

Cost of Alternative MR&R Programs

MR&R strategies provide the lane-mile lengths of projects to be designed into each RSL category. A simple cost matrix based on historical MR&R program cost data provides the relationship between average lane-mile cost of projects whose DSL is within each RSL category. The cost of alternative programs is the product of the lane-mile length of projects that the MR&R strategy requires to be designed into each RSL category and the corresponding cost per lane mile. Figure 4 is a simple cost matrix based on the average historical project cost data for District 5, 6, and 7 freeways and nonfreeways constructed from 1987 to 1989. Annual or 5-year MR&R program cost estimates are based on the strategy that would be used as a constraint for program development and the lane-mile cost data shown in Figure 4. Annual MR&R program cost estimates are based on the following equation:

$$\text{MR\&R program cost} = P/100 \times L \times C_x \quad (3)$$

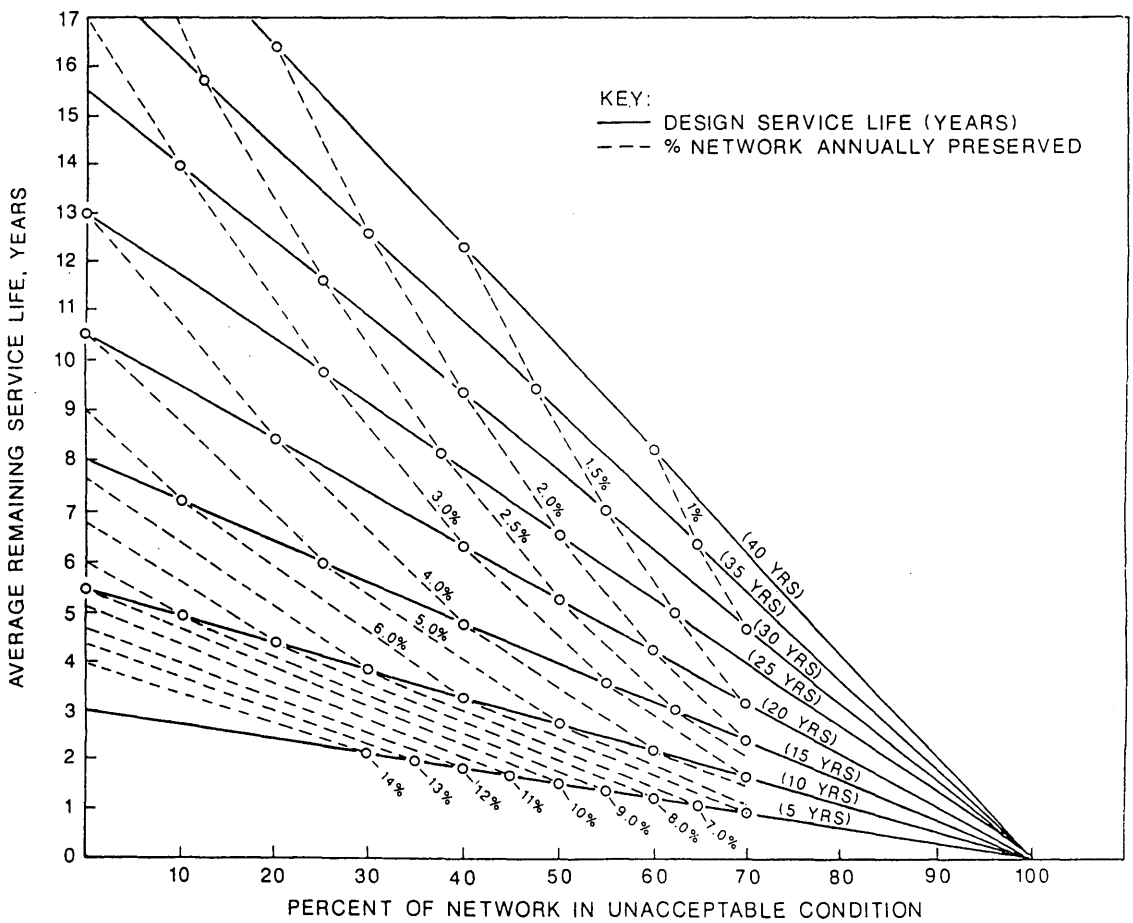


FIGURE 3 Network analysis chart relating alternative strategies (chart's interior) with which annual MR&R program must comply and resulting network condition and RSL.

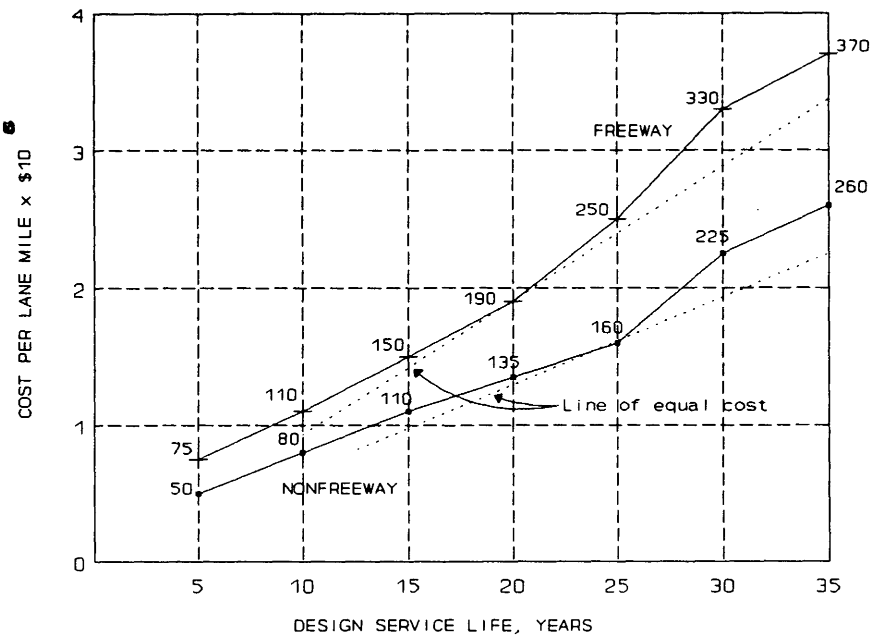


FIGURE 4 Simple cost matrix based on historical MR&R project cost data.

where L is the lane-mile length of the network, and C_x is the lane-mile cost of the DSL category corresponding to the program's DSL.

Reactive Maintenance Cost

The cost of reactive maintenance is based on procedures reported by Richardson (*7*). Simply, it is the product of the lane miles of pavement in unacceptable condition and the historical cost of reactive maintenance per lane mile of pavement in unacceptable condition. On the basis of 1989–1990 Michigan DOT maintenance cost data, the average cost of reactive maintenance for pavement in unacceptable condition is $1,200/lane-mi for nonfreeways and $4,600/lane-mi for freeways. Annual reactive maintenance cost (RMC) is computed on the basis of the following equation:

$$\$RMC = [P_0 + (P_5 - P)/2]/100 \times L \times C_x \qquad (4)$$

where P_5 is the percentage of network that annually deteriorates into the zero RSL category.

ECONOMIC ANALYSIS CONSIDERATIONS

Primary concerns when managing networks are knowing what minimum funding stream would be needed to maintain the desired condition and how this funding stream compares with anticipated revenues. The annual cost of a constant annual network strategy will be the same in 40 years as it is today, except as it is affected by the rate of inflation of construction costs. Agencies should consider this source of cost increase over time. For state highway agencies (SHAs), transportation revenues are largely a function of the relationship between the supply and demand for fuels, both of which are difficult to forecast and cannot be controlled.

Money is a productive resource, so there is a time value associated with its use. For SHAs, the difference between the earning power of money and the rate of construction cost inflation is considered the discount rate. Theoretically, it is reasoned, the cost of future investments should be discounted by an amount equal to the discount rate when compared with the cost of making the investment today. Discounting the value of money favors low-initial-cost alternatives and defers high-initial-cost investments. If decisions are made on the basis of the discounted value of money, the real cost of future pavement preservation programs will increase by an amount approximately equal to the rate of construction cost inflation plus the discount rate used to develop annual preservation programs (*8*). Therefore, the use of real today dollars is recommended because it presents a clearer picture of the relationship between revenues and funding streams and does not artificially increase the actual cost of future network preservation.

Typically, the selection of preservation treatments is based on which is best for the project or which has the lowest project LCC. The importance of the DSL of alternative treatments is unimportant except as it affects project LCC analysis. These methods of selecting treatments place project condition needs ahead of network condition needs and can result in the increased cost of network preservation (*5*). The basic idea is that at the project level, a 10-year DSL treatment does not have the same impact on the network as does a 20-year DSL treatment. The long-term impact of two projects with 10-year DSL treatments is equivalent to one project with a 20-year DSL, assuming all projects have the same lane-mile length. However, project LCC assumes that a string of short life treatments has the same impact on network condition as those having a longer life. This problem is created by thinking in terms of events (projects) rather than systems (networks), as discussed previously (*9*).

Economic analysis for preserving either projects or networks should include hundreds of alternative funding schemes, each of which is likely to have unequal costs and unequal benefits. This is the most complex configuration possible for economic analysis. Converting benefits to their dollar values and including the values in the project LCC analysis is a problem when considering benefits whose value is subjective or when their economic value is similar but their subjective value is not. The dollar value of benefits is difficult to estimate and of questionable accuracy and reliability. For this reason, the network management system developed for MDOT seeks to determine only the network strategy that will minimize the total long-term cost of network preservation given the target condition and RSL objectives. Benefits of alternative programs are addressed at the program development level by a program management system (*1,3,10*) whose objective is to maximize program benefits that have not been converted to dollar values.

This paper deals only with the network management perspective of economic analysis, that is, the relationship between the total cost of alternative 40-year strategies and the resulting network condition. For the sake of simplicity, no discount or inflation rate is considered. The cost of alternative funding schemes is expressed simply in terms of today dollars. The variable cost of reactive maintenance is included in the study.

TRUNKLINE SYSTEM PERFORMANCE DATA

The performance and lane-mile length of each district's networks are illustrated in bar chart form in Figure 5 for the nonfreeway and Figure 6 for the freeway. By use of manual pavement management system (PMS) analysis methods (*6*), all the products and information listed in the AASHTO guidelines for PMS (*11*) can be determined on the basis of the cost data shown in Figure 4, the average reactive maintenance cost per lane mile, the combined district pavement performance data, and the Figure 3 network analysis chart. For this paper, network condition and costs are determined on the basis of Equations 1 through 4. The Figure 3 network analysis chart can also be used to relate the target network condition (*X*-axis) and RSL (*Y*-axis) to the required network strategy by extending the *X*- and *Y*-values until they intersect. The point of intersection indicates the DSL and percentage of network values of the required network strategy.

FUNDING SCHEMES

Each of the five funding schemes listed in Table 1 is evaluated to determine the total cost of preservation—total cost being the cost of reactive maintenance plus the cost of the MR&R program for each of eight 5-year analysis periods. For simplicity, only the combined freeway and nonfreeway networks are analyzed using the funding schemes in Table 1.

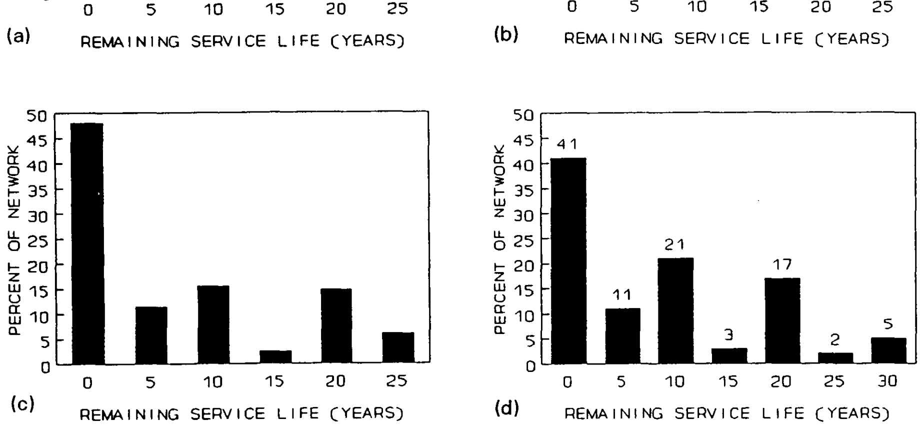


FIGURE 5 Current condition of nonfreeway networks: *a*, District 5, network length = 2,588 lane-mi; *b*, District 6, network length = 2,765 lane-mi; *c*, District 7, network length = 2,209 lane-mi; *d*, Districts 5 through 7, network length = 7,562 lane-mi, ARSL = 8.5.

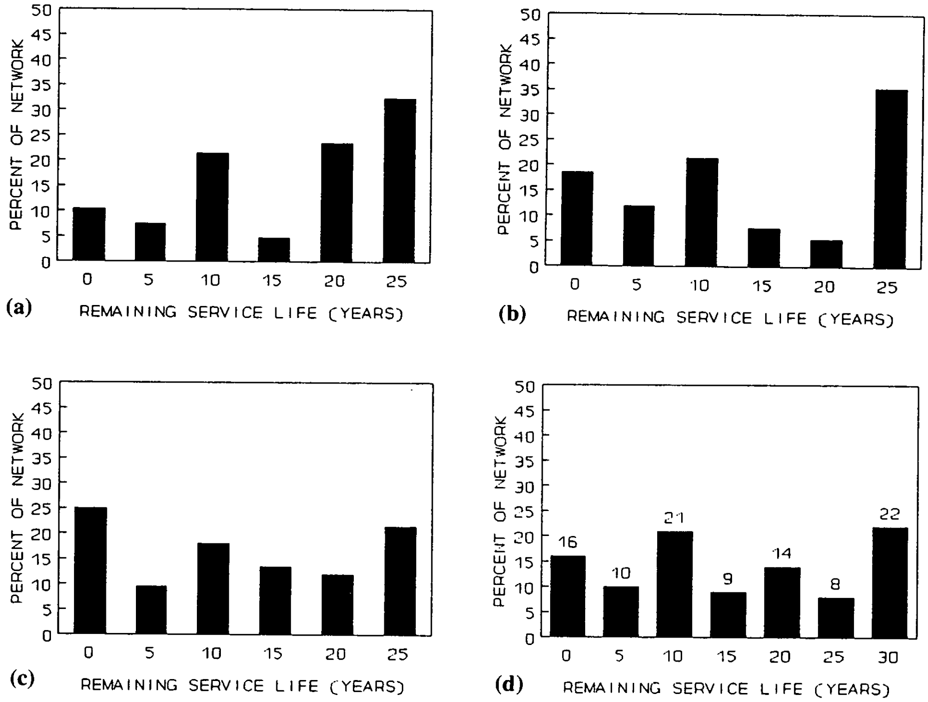


FIGURE 6 Current condition of freeway networks: *a*, District 5, network length = 1,343 lane-mi; *b*, District 6, network length = 1,152 lane-mi; *c*, District 7, network length = 1,253 lane-mi; *d*, Districts 5 through 7, network length = 3,748 lane-mi, ARSL = 15.3.

NETWORK LCC ANALYSIS

Analyses are based on the current condition status and RSL of each combined network (Figures 5*d* and 6*d*). Reactive maintenance cost is based on the percentage of network in the zero RSL category at the beginning of each 5-year period plus the percentage of network that deteriorates into it annually minus the percentage of network preserved into higher RSL categories (Equation 4). The MR&R strategy indicates the percentage of network that is to be preserved (moved or subtracted) from the zero RSL category and to what higher RSL category it is to be moved (added). The results explain what is done and present each calculation so as to illustrate methodology and results. Network LCC is simply the sum of the annual RMC and the annual MR&R program cost, over a 40-year analysis period, that is required to achieve and maintain a given network condition objective.

TABLE 1 Alternative Funding Schemes

Scheme Number	Action	Duration (years)
1	Do nothing but reactive maintenance	0–10
	Restore original condition	11–15
	Maintain original condition	16–40
2	Maintain original condition	0–40
3	Do nothing but reactive maintenance	0–10
	At least cost, restore and maintain original condition	11–40
4	Do nothing but reactive maintenance	0–10
	At least cost, eliminate all pavement in poor condition	11–40
5	At least cost, eliminate all pavement in poor condition	0–40

Scheme 1 Results

The condition of the networks after 10 years of doing nothing but reactive maintenance is shown in Figure 7*a* and *b*. Using the nonfreeway for an example, at the end of 10 years, the percentage of network in the zero RSL category is determined from Figure 5*d* data as the sum of the percentage of network in the following time periods: Period 0 (41 percent), Period 5 (11 percent), and Period 10 (21 percent) for a total of 73 percent. Computations necessary for determining total 40-year cost are listed to illustrate changes in cost and condition (P_0) over time.

The reactive maintenance cost for the first 10 years is as follows:

Nonfreeway

Period, $5[P_0 + (P_5 - P)/2]/100 \times L \times C_x$ = \$RMC

5 [41 + (11 − 0)/2]/100 × 7,562 × \$1,200 = \$21,100,000

10 5[52 + (21 − 0)/2]/100 × 7,562 × \$1,200 = <u>\$28,360,000</u>

Total = \$49,460,000

Freeway

5 [16 + (10 − 0)/2]/100 × 3,748 × \$4,600 = \$18,100,000

10 5[26 + (21 − 0)/2]/100 × 3,748 × \$4,600 = <u>\$31,460,000</u>

Total = \$49,560,000

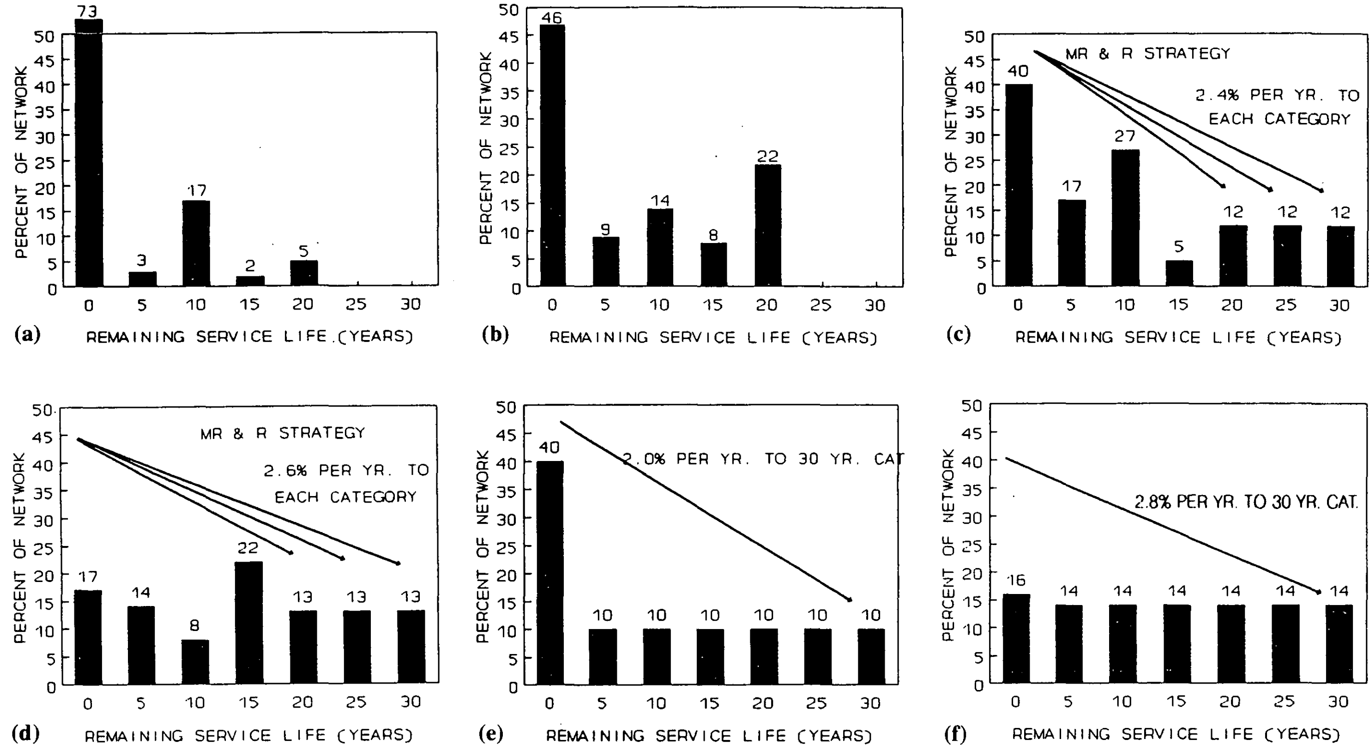


FIGURE 7 Condition of networks with *a*, after 10 years of do-nothing, ARLS = 3.2; *b*, after 10 years of do-nothing, ARLS = 7.6; *c*, 11 to 15 years of rapid preservation strategies, ARLS = 10.8; *d*, 11 to 15 years of rapid preservation strategies, ARLS = 14.6; *e*, 16 to 40 years of maintaining network condition with the preservation strategy shown, ARLS = 10.5; *f*, 16 to 40 years of maintaining network condition with the preservation strategy shown, ARLS = 14.7.

For Time Period 15, the MR&R strategy is to improve the network to its original condition. This means the percentage of network in the zero RSL category (unacceptable condition) must be reduced to the original level of 41 and 16 percent, respectively, for the nonfreeway and freeway networks. To do this the MR&R strategy (P) for the 5-year period must be equal to the percentage of network currently in the zero RSL category (P_0), as shown in Figures 7*a* and *b*, plus the percentage of network that will deteriorate into the zero RSL category in the 5-year period (P_5) minus the target percentage of the network that is to be in the zero RSL category (P_i) when the network's condition becomes stable:

$$P = P_0 + P_5 - P_i \quad (5)$$

The MR&R strategy (P) needed to restore the original condition of each network is as follows:

Network	$P_0 + P_5 - P_i = P$
Nonfreeway	73 + 3 − 41 = 35 percent
Freeway	47 + 9 − 16 = 40 percent

Based on Equation 4, the cost of reactive maintenance for Period 15 is as follows:

Nonfreeway

5 [73 + (3 − 35)/2]/100 × 7,562 × $1,200 = $25,860,000

Freeway

5 [47 + (9 − 40)/2]/100 × 3,748 × $4,600 = $27,150,000

The estimated cost of the MR&R program for Period 15 is based on the assumption that at the end of this period the 20-, 25-, and 30-year RSL categories would be void of pavements. Therefore, the strategy used is to fill each category with an equal percentage (length) of network. The selected MR&R strategy requires that at least 35 percent of the nonfreeway network be preserved, which, rounded up, would require rehabilitating 12 percent of the network into each category as shown in Figure 7*c*. This same procedure is used for the freeway network (Figure 7*d*). On this basis, the estimated MR&R program cost is as follows:

Nonfreeway

DSL	$P \times L \times C_i$ = MR&R cost
30	12 × 7,562 × $225,000 = $204,170,000
25	12 × 7,562 × $160,000 = $145,190,000
20	12 × 7,562 × $135,000 = $122,500,000
Total = $471,860,000	

Freeway

30	13 × 3,748 × $330,000 = $160,790,000
25	13 × 3,748 × $250,000 = $121,810,000
20	13 × 3,748 × $190,000 = $92,580,000
Total = $375,180,000	

For Periods 20 to 40, the MR&R strategy (P) that would maintain the restored network condition can be computed as follows:

$$P = (100 - P_0)/N \quad (6)$$

where N is the number of 5-year periods between the RSL category the pavements are improved to and the zero RSL category.

By assuming that the MR&R strategy will be to move pavements into the 30-year DSL category (the number of 5-year periods is 6), the 5-year MR&R strategy for each network is as follows:

Nonfreeway
$P = (100 - 41)/6 = 9.8$ percent or 10 percent

Freeway
$P = (100 - 16)/6 = 14$ percent

The resulting network performance would be as shown in Figures 7*e* and *f*. A strategy can be selected so as to change the network's RSL but not its condition. The strategy needed to maintain current network performance, including its current RSL, can be determined from Figure 3. From these strategies, the cost of reactive maintenance for each period from Periods 20 through 40 (five periods) is as follows:

Nonfreeway
Period $5[P_0 + (P_5 - P)/2]/100 \times L \times C_x$ = RMC
20 5[41 + (17 − 10)/2]/100 × 7,562 × $1,200 = $20,190,000
25 5[48 + (2 − 10)/2]/100 × 7,562 × $1,200 = $19,960,000
30 5[40 + (5 − 10)/2]/100 × 7,562 × $1,200 = $17,010,000
35 5[35 + (12 − 10)/2]/100 × 7,562 × $1,200 = $16,330,000
40 5[37 + (12 − 10)/2]/100 × 7,562 × $1,200 = $17,240,000
Total = $91,680,000

Freeway
20 5[16 + (14 − 14)/2]/100 × 3,748 × $4,600 = $13,790,000
25 5[16 + (8 − 14)/2]/100 × 3,748 × $4,600 = $11,210,000
30 5[10 + (22 − 14)/2]/100 × 3,748 × $4,600 = $12,070,000
35 5[18 + (13 − 14)/2]/100 × 3,748 × $4,600 = $15,090,000
40 5[17 + (13 − 14)/2]/100 × 3,748 × $4,600 = $14,220,000
Total = $66,380,000

The MR&R program cost for each 5-year period from Periods 20 to 40 (five periods) is as follows:

Nonfreeway
$P \times L \times C_{30}$ = MR&R cost
10 × 7,562 × $225,000 = $170,150,000
Period 20 to 40 cost = 5 periods × $170,150,000 = $850,750,000

Freeway
14 × 3,748 × $330,000 = $173,160,000
Period 20 to 40 cost = 5 periods × $173,160,000 = $865,800,000

Total 40-year cost for Preservation Scheme 1:

Nonfreeway
Period RMC + MR&R cost = total cost
5 and 10 $49,460,000 + $----------------- = $49,460,000
15 $25,860,000 + $471,860,000 = $497,720,000
20 to 40 $91,680,000 + $850,750,000 = $924,430,000
Totals $167,000,000 + $1,322,610,000 = $1,489,610,000

Freeway
5 and 10 $49,560,000 + $----------------- = $49,560,000
15 $27,150,000 + $375,180,000 = $402,330,000
20 to 40 $66,380,000 + $865,800,000 = $932,180,000
Totals $143,090,000 + $1,240,980,000 = $1,384,070,000

Figure 8*a* shows the network's condition over the 40-year analysis period.

Scheme 2 Results

To maintain the networks at their current condition, as shown in Figures 5*d* and 6*d*, the 5-year strategy (P) and DSL are, from Scheme 1, as follows:

Nonfreeway: P = 10 percent, DSL = 30 years.
Freeway: P = 14 percent, DSL = 25 years.

Total 40-year cost for preservation Scheme 2:

Network RMC + MR&R cost = total cost
Nonfreeway $163,110,000 + $1,361,200,000 = $1,524,310,000
Freeway $99,120,000 + $1,049,440,000 = $1,148,560,000
Totals $262,230,000 + $2,746,480,000 = $3,008,710,000

Figure 8*b* shows the network's condition over the 40-year analysis period.

Scheme 3 Results

The cost for the first 10 years is the same as the cost for Scheme 1. The least annual costs to restore and maintain the original condition based on Figure 4 costs are the 25-year DSL for the nonfreeway and the 20-year DSL for the freeway. The 5-year MR&R strategies needed to restore network conditions at least annual cost are as follows:

Nonfreeways $P = (100 - 41)/5 = 11.8$ percent or 12 percent
Freeways $P = (100 - 16)/4 = 21$ percent

Total 40-year cost for Preservation Scheme 3:

Network RMC + MR&R cost = total cost
Nonfreeway $203,950,000 + $871,140,000 = $1,075,090,000
Freeway $1,715,210,000 + $897,270,000 = $1,068,790,000
Totals $375,470,000 + $1,768,410,000 = $2,143,880,000

Figure 8*c* shows the network's condition over the 40-year analysis period.

Scheme 4 Results

The cost of doing nothing but reactive maintenance for Periods 5 and 10 is the same as that for Scheme 1. The least

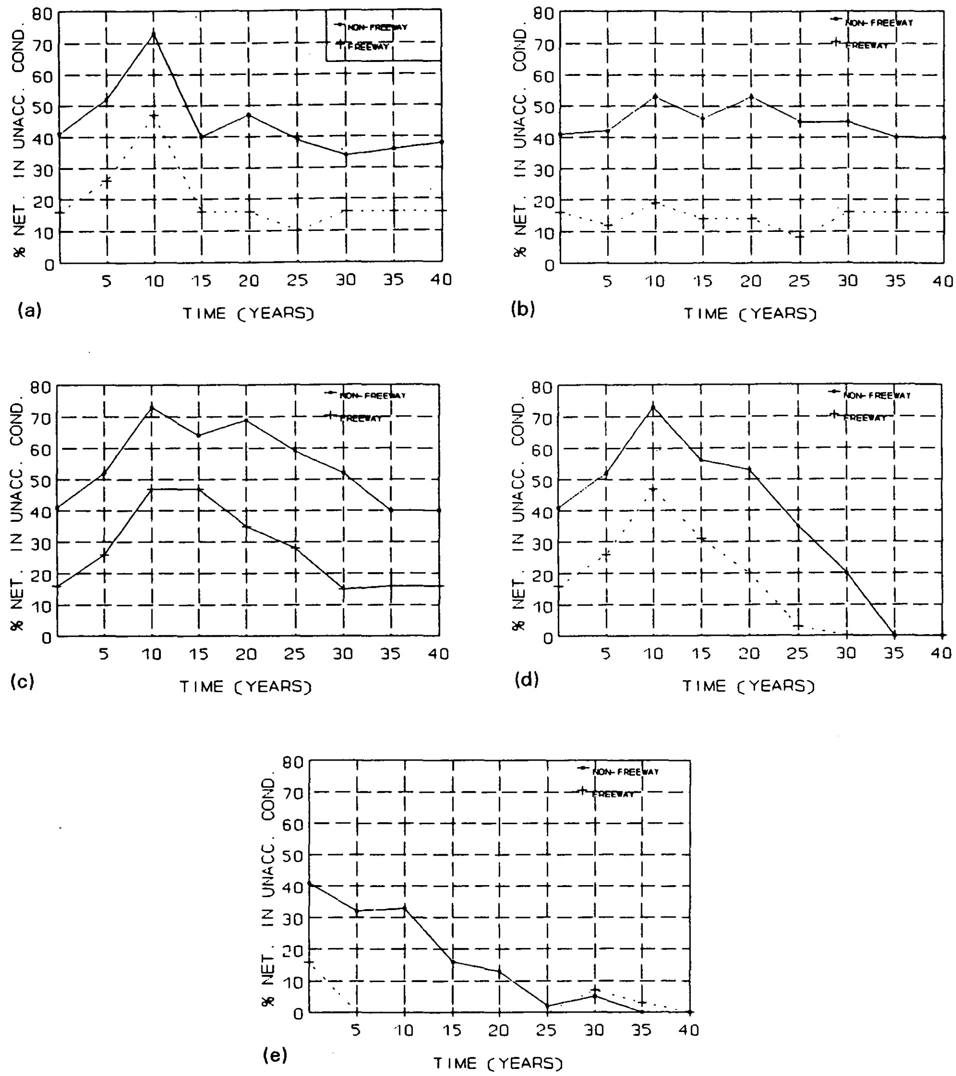


FIGURE 8 Change in network condition over 40-year analysis period as a result of funding schemes: *a*, Scheme 1; *b*, Scheme 2; *c*, Scheme 3; *d*, Scheme 4; *e*, Scheme 5.

annual cost to eliminate all pavements in unacceptable condition (0.0 percent of network in zero RSL category) would be to use the lowest lane-mile cost DSL for each network, which is the 25-year DSL for the nonfreeway and 20-year DSL for the freeway. The required 5-year MR&R straegies are as follows:

Nonfreeway P = (100 − 0)/5 = 20 percent (4 percent annually)
Freeway P = (100 − 0)/4 = 25 percent (5 percent annually)

Total 40-year network LCC for Preservation Scheme 4:

Network RMC + MR&R cost = total cost
Nonfreeway $140,430,000 + $1,451,900,000 = $1,592,330,000
Freeway $116,360,000 + $1,068,180,000 = $1,184,540,000
Totals $256,790,000 + $2,520,080,000 = $2,776,870,000

Figure 8*d* shows the network's condition over the 40-year analysis period.

Scheme 5 Results

The MR&R strategy for Scheme 5 would be the same as that for Scheme 4 except that it includes all eight 5-year periods. The 5-year MR&R program cost, based on Equation 3, for the 40-year analysis period is as follows:

Nonfreeway = $1,935,870,000
Freeway = $1,424,240,000

Total 40-year network LCC for Preservation Scheme 5:

Network RMC + MR&R cost = total cost
Nonfreeway $50,820,000 + $1,935,870,000 = $1,986,690,000
Freeway $7,330,000 + $1,424,240,000 = $1,431,570,000
Totals $58,150,000 + $3,360,110,000 = $3,418,260,000

Figure 8*e* shows the network's condition over the 40-year analysis period.

SUMMARY

The results of this study are presented in Table 2. Funding Scheme 1 reduces the cost of network preservation over that of Scheme 2. The cost of Scheme 2 is not the lowest-cost scheme for maintaining the current network condition; in addition, Scheme 1 does not generate as much agency savings as Scheme 3. Of Schemes 1 through 3, only Scheme 3 is a least-cost network strategy. These findings illustrate the following: (a) the cost of preserving networks does not have to be a function of the original network condition; (b) the cost of preserving networks at a given condition level is a function of the RSL at which it is maintained; and (c) Figure 4 indicates that the lowest-cost network strategy for preserving the network at any given condition level is when the DSL is 20 years for the freeway and 25 years for the nonfreeway.

Most of the greater cost of Scheme 4 compared with Scheme 3 and of Scheme 5 compared with Scheme 2 is attributed to the high cost of improving the heretofore neglected nonfreeway network. What is more significant is that the cost to improve the freeway network condition at least cost, Scheme 5, is essentially the same as that to maintain the current freeway condition with the Scheme 2 strategy. This illustrates how network management systems can be used to identify ways to improve network condition without increasing the funding level.

A do-nothing scheme should be a reasonable alternative in emergency situations and when declining revenues prevent maintaining target network condition levels. The duration of a do-nothing plan should be a function of rate of network deterioration, which is measured in terms of its RSL. Slowly deteriorating networks, characterized by large RSLs, can tolerate a longer period of do-nothing than can rapidly deteriorating networks. The impact on network condition of a 4-year do-nothing scheme if its RSL were 10 years is about the same as a 2-year do-nothing scheme if its RSL were 5 years. And somewhere in between do-nothing and full-funding (funding necessary to achieve target network condition objectives) alternatives are the conditions under which network management systems are most needed to learn how the economic efficiency of available funds can be improved.

More advantages and disadvantages of each funding scheme can be learned from Figure 8 and the data in Table 2 than are discussed. It is likely that decision makers would want to look even more deeply and at more alternatives. Because manual computations are too labor-intensive, this requirement is accomplished with a generic application software system that is surrounded with a utility software system designed to suit its users. The application software uses a comprehensive cost matrix (*1*) in place of the simple Figure 4 matrix. The advantages provided by the use of the software system are that it enables users to identify the network's patterns of behavior and underlying means by which preservation costs can be further reduced while maintaining condition objectives. The manual analysis methods used in this paper are not intended to be used to develop actual network strategies. And use of historical project cost data, like those shown in Figure 4, is not recommended to establish network strategies. MDOT's generalized or manual network management methods are intended to be used as training aides, for preliminary planning purposes, and as illustration of network management concepts.

For most agencies, timing of MR&R actions should be of economic importance only for pavements that are rapidly deteriorating and therefore will rapidly become a safety and reactive maintenance problem. However, timing can be of considerable economic importance to the users. And no economic justification can be found for maintaining pavements in good condition when only agency costs are considered. For agencies that do not yet have a good handle on user costs, the general practice is to subjectively set network condition objectives and then develop the lowest-cost programs necessary to achieve them. For condition objectives to be economically justified, they must include user as well as agency costs. That is, a pavement condition objective can be economically justified only when the annual agency-plus-user costs necessary to achieve network condition objectives are lower than the resulting agency-plus-user savings that are estimated to be derived from the annual preservation program. However, highway users seem to be willing to pay for good pavement serviceability whether or not they are economically justified. This subjective need for good pavement condition is similar to the choice between low-cost, low-comfort or high-cost, high-comfort automobiles. Hence, it is likely that condition objectives set for networks will be based on subjective as well as objective economic criteria.

TABLE 2 Performance and Total Cost of Funding Schemes 1 through 5

Scheme	Network	Network Performance(1) % Unacc.	ARSL(yrs)	Total Cost $ Billions
1	Non-Fr.	40	10.5	1.49
	Freeway	16	14.7	1.38 (2.87)
2	Non-Fr.	40	10.5	1.52
	Freeway	16	14.7	1.48 (3.00)
3	Non-Fr.	40	9.0	1.08
	Freeway	16	10.5	1.07 (2.15)
4	Non-Fr.	0	15.0	1.59
	Freeway	0	13.5	1.18 (2.77)
5	Non-Fr.	0	15.0	1.99
	Freeway	0	13.5	1.43 (3.42)

(1) Performance at the end of the 40 year analysis period.

CONCLUSIONS

This study demonstrates that the long-term performance of networks, when measured in terms of percentage of network in poor (unacceptable) conditions and RSL, is a function of only the network strategy (percentage of network and average DSL) with which the annual preservation programs comply. This approach is consistent with systems-thinking methodology: it is simple; it provides agencies with the ability to control budgets and network conditions over the long term, such as 40 years; it is compatible with any agency's project and program development process; and it facilitates communication among those representing political, managerial, and technical interests. Specific conclusions are as follows:

1. When managing networks, bar charts of the percentage of network in each 5-year remaining life category provide a

convenient means to communicate network patterns of behavior, condition status, and the effects of alternative funding schemes.

2. The cost of any alternative MR&R program can be estimated as the product of its lane-mile length and the historical cost per lane mile of projects having the program's average DSL.

3. The products and information in the AASHTO 1990 guidelines for PMS (*11*,pp.3–4) can be estimated on the basis of the data provided by the Figure 3 relationship between network performance and network strategies and the cost data provided by Figure 4.

4. The ultimate condition (percentage of network in poor condition) and RSL of any given network is dependent only on the lane-mile length and DSL (network strategy) with which the annual MR&R programs conform. This fundamental principle is based on the following assumptions: that the agency has the means to reliably estimate DSL of all feasible MR&R treatments, that actual pavement life correlates well with DSL estimates, and that the agency collects pavement condition data of sufficient detail to enable accurate identification of the boundaries and the RSL of all uniform sections that make up the designated network.

5. The cost to attain and then maintain any target network condition and RSL objective is dependent on the strategy that the annual MR&R programs must conform with and is independent of the network's original condition and RSL when the time required to reach the target condition level is not specified.

6. The diversion of funds from pavement preservation to other funding categories, such as Michigan has done with its District 5, 6, and 7 nonfreeway networks, is a reasonable method of meeting revenue shortfalls as long as the total agency-plus-user costs remain smaller than the total agency-plus-user savings derived from the annual preservation program.

7. The timing of network preservation should be more important for networks that rapidly deteriorate (low RSL) and least important when they slowly deteriorate (high RSL). For networks that have higher RSLs, such as those included in this study, timing of network preservation investments is not important.

8. Historical cost matrixes such as that shown in Figure 4 provide a simple means of identifying the approximate DSL of MR&R programs that should minimize preservation cost, and they provide a simple means to convey optimization concepts to nontechnical personnel.

9. It is often difficult to think in terms of preserving networks instead of projects. However, network management principles are simple, they provide the ability to improve economic efficiency, and they provide the relationship between any given funding level and the resulting network condition.

REFERENCES

1. W. H. Kuo, E. C. Novak, and G. Y. Baladi. Development of Long-Term Network Strategies Using Remaining Service Life. *ASTM STP-1121: Pavement Management Implementation*, ASTM, Philadelphia, Pa., April 1992.
2. G. Y. Baladi, E. C. Novak, and W. K. Kuo. Pavement Condition Index and Remaining Service Life. *ASTM STP-1121: Pavement Management Implementation*, ASTM, Philadelphia, Pa., April 1992.
3. E. C. Novak, W. H. Kuo, and G. Y. Baladi. Toward Standardization of a PMS Analysis Method. *ASTM STP-1121: Pavement Management Implementation*, ASTM, Philadelphia, Pa., April 1992.
4. *Advanced Course in Pavement Management.* FHWA, AASHTO, and TRB. Washington, D.C., 1991.
5. E. C. Novak and W. K. Kuo. Life-Cycle Cost Versus Network Analysis. In *Transportation Research Record 1344*, TRB, National Research Council, Washington, D.C., 1992.
6. K. Cooper. *Manual Methods for Using Program Strategy*. Michigan Department of Transportation, Lansing, Oct. 1990.
7. M. J. Richardson. *Manual Methods for Estimating Reactive Maintenance Cost.* Michigan Department of Transportation, Lansing, Jan. 1991.
8. E. C. Novak and W. K. Kuo. A Comparison of Life-Cycle Cost and Cost Effective Methods of Economic Analysis. *Proc., 2nd North American Conference on Pavement Management*, Toronto, Ontario, Canada, Vol. 3, Nov. 1987.
9. E. C. Novak and W. K. Kuo. There Are Two Categories of PMS Analysis Methods, Reactive and Generative. Presented at 72nd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 10–14, 1993.
10. E. C. Novak and W. K. Kuo. The Role of Pavement Management System Analysis in Preservation Program Development. In *Transportation Research Record 1344*, TRB, National Research Council, Washington, D.C., 1992.
11. *AASHTO Guidelines for Pavement Management Systems.* AASHTO, Washington, D.C., July 1990.

The entire contents of this paper are the product of the authors and in no way reflect the opinions, policies, or recommendations of the Michigan Department of Transportation.

Publication of this paper sponsored by Committee on Transportation Economics.

An Economic Theory of Travel Decisions

AAD RÜHL

An economic theory of travel decisions, including an introduction to the dimension of time, is presented. Four categories of travel were distinguished: travel that is (a) related to economic consumption, (b) related to noneconomic activities, (c) between places of production and consumption, and (d) for productive purposes. For travel related to economic consumption a utility function includes the utility derived from consumption, the effort of producing consumption goods or services in the household of the consumer and the utility (or disutility) of travel related to consumption. A budget constraint is presented on the basis of the cost of consumption goods and travel. Noneconomic activities do not lend themselves to economic valuation by definition; however, the budget constraint is extended to include travel for noneconomic activities. The choice of home and work locations is essentially a long-term one, depending on far more factors than the transportation situation; with regard to transportation, all types of journeys for all members of the household need to be considered. Travel for work appears to be far more similar to travel for private purposes than was generally believed. An introduction to the influence of the dimension of time is given, stating a time budget including time spent on work, economic consumption, noneconomic activities, and travel. Although the total time available per day is fixed, there are some options for using time for multiple purposes. Finally, it is stated that the study of transportation should be integrated into the study of human activities in general.

Consumption in an economic sense provides satisfaction, to be called utility, by satisfying needs of people with the help of scarce resources that may have other uses. These resources will be called goods and services, as opposed to resources that either are not scarce in relation to the need for them or have no alternative employment. The traditional example of the first category is air, essential for our life, but plentiful (but, unfortunately, not always of the best quality). An example of the second category is a letter that brings us news from a friend, which can be very important to us but is not of interest to other people, as opposed to a newspaper, which provides news of general interest.

TRANSPORT AS A COST OF CONSUMPTION

Theory of Consumption

The theory of economic consumption deals with the consumption of goods and services that can be obtained in the market. It does not generally consider private production of consumption goods. Preferences of individuals are taken as given, and no attention is paid to variations of preference scales of the same individuals over time.

In a market economy, goods and services are exchanged for a monetary price, and consumptive utility can be obtained by spending a certain sum of money. Making abstraction from indivisibilities, total utility can be maximized by equalizing the utility to be derived from the last unit of money spent on the consumption of each good or service. The proof of this can easily be given: keeping fixed the total budget to be spent, we can augment total utility by transferring money expenditure from one good to another as long as utility per unit of money spent on various goods is not equal. Applying this rule to individual commodities will not work: either we buy a certain book or we do not. If, however, we apply it to categories of goods or to goods that we use in larger quantities, it will hold at least approximately: in the equilibrium situation the last dollar we spend on books will give us approximately the same utility as the last dollar we spend on nonmental food.

In a very simple algebraic form, we can formulate

$$u = u(\mathbf{c})$$

denoting that the total utility derived from consumption depends on the quantities of all goods and services consumed ($\mathbf{c}$). If we use $\mathbf{p}_c$ as the vector of the prices of all consumption goods and Y as income, a budget constraint can be formulated as

$$\mathbf{p}_c'\mathbf{c} \leq Y$$

Home-Produced Consumption Goods

Up to now we have followed the traditional consumption theory. However, in the line of arguments it is necessary to stress that some consumption goods can be produced at home as well as bought in the market and that they can be substitutes for each other. If we have a garden, we can grow vegetables in it. Instead of having our laundry collected and delivered, we can either go to a launderette or do our own wash at home.

Producing consumption goods or services in our own household not only will change the demand for goods in the market; it also implies that we have to put in some effort that can be considered as a negative utility. Assuming that the quality of the laundry done at the launderette is equal to that done at home, a comparison should be made between buying the full service, buying the use of the launderette and putting in some effort, and buying a washing machine or other equipment and soap and putting in the effort. This effort is labor we do not sell to an employer but use for our own purposes.

The effort needed for labor for our own purposes is sometimes felt to be less than the effort of similar labor done for market purposes: we value work in our own home differently

Ministry of Transport, Public Works and Water Management, P.O. Box 20901, 2500 EX, The Hague, The Netherlands.

from work in other people's homes: we would rather knit a sweater for a relative than knit it to be sold in a shop. Some of these activities can even generate utility on their own right.

We can now write the general formulation of consumption as

$$u = u\,(\mathbf{c}, \mathbf{e}_c)$$

where $\mathbf{e}_c$ stands for the effort for the production of consumption goods and normally will have a negative sign—that is, it will be a disutility or cost—but can have a positive sign, which then would add to the utility derived from consumption itself.

The income constraint will not be influenced by the input of our own effort and therefore remains

$$\mathbf{p}_c'\mathbf{c} \leq Y$$

Transport and Consumption

Passenger transport for private purposes usually is considered a consumption good in accordance with these rules, but in doing so one is running into some difficulty. Normally, consumption goods are consumed only if they give satisfaction or positive utility. When utility or utility and effort combined are negative, we will not include the good concerned in our consumption basket. This relation, however, does not hold for transportation. In most cases transportation does not provide positive satisfaction; to the contrary, it is generally considered a negative element in total utility. It is in accordance with this situation that people generally will try to minimize transport, in that for reaching a given destination they will choose the shortest or least-difficult route.

When transportation is considered a consumption good, it is not possible to explain that people will travel at all. The obvious reason for the fact that people travel long distances while yet trying to minimize their travel is that it is not transportation itself, but reaching the destination, that provides utility. It is therefore often formulated that the demand for transportation is a derived demand and that the utility is a derived utility. It is not the journey to the theater that provides utility but seeing the play. When we explicitly or implicitly consider the pros and cons of going to the theater, the cost of the journey has to be added to the price of the ticket, and the disutility of travel has to be considered as a negative factor against the utility of seeing the play. It can now be seen why people who live far from theaters visit these buildings less often than those who live nearby (another reason will be introduced later, when we will discuss the influence of time).

When we carefully consider the various categories of consumption, we can see that frequently at least some travel is associated with it. We can therefore state generally that the satisfaction provided by consumption has to be compared not only with the money cost of the consumption goods but also with the money cost and dissatisfaction of travel. The tendency to equalize marginal utility of expenditures on all consumption goods therefore will not hold because individual consumption goods will involve different amounts of transportation. Only after the dissatisfaction resulting from travel is converted into monetary units can equalization of marginal utility of consumption per unit of money spent (including generalized cost of travel) be reintroduced.

For the moment, we will limit our exposé to restating the consumption utility function as

$$u = u\,(\mathbf{c}, \mathbf{e}_c, \mathbf{v}_c)$$

and the budget constraint as

$$\mathbf{p}_c'\mathbf{c} + \mathbf{p}_v'\mathbf{v}_c \leq Y$$

where $\mathbf{v}_c$ stands for all travel related to consumption and $\mathbf{p}_v$, for the price or monetary cost of travel.

Approach to Analysis of Consumption Decisions

The utility of consumption of certain goods and the utility (or disutility) of effort and travel are variable between people and, for individual people, over time. Utilities and disutilities therefore are not stable quantities that can be measured easily but are quantities variable with all the differences between individuals and for each individual with circumstances and moods.

This situation will not only make quantitative studies in the field of consumption behavior more difficult but also have an influence on the next step of our theoretical considerations: the treatment of travel as a cost of consumption. As we have seen, obtaining satisfaction from economic consumption may cost us the price of the goods—effort for own production and travel.

The first is measured in terms of money, the second in terms of satisfaction, and the last involves both money and satisfaction. According to the usual definition of economic cost, this notion includes the minimum sacrifice needed to obtain a certain result. So, if a certain article is sold for $2.50 in one shop and for $2.75 in another, we would assume that the cost of the article is $2.50, unless there are good reasons to pay the higher price: maybe it is not thought worthwhile to shop around until the cheapest price has been found, or that the shop selling the item at $2.75 needs to be visited anyway (then there is a trade-off between money cost and transport cost), or that the shopper does not like to visit (or to be seen in) the shop selling the article at the lower price (in this case the goods are in fact not completely equivalent).

For transportation, the determination of minimum sacrifice is still more complicated because for each trip there may be alternative ways to travel, each with a different money cost and level of satisfaction, and a comparison of combinations of these two categories is necessary to determine the best way to make a journey. For this comparison, we need to add in our minds money expenditures and degrees of satisfaction. In studies on passenger transport demand, the result of this addition is indicated as generalized cost. At this stage, only two observations will be made.

As with consumer preference in general, preferences in the field of transportation vary with individuals and between individuals over time. We cannot expect generalized cost of travel to depend on a simple and stable formula.

The second observation goes partly contrary to the first one, but it is also in agreement with general consumption behavior. There is an influence of habit formation and economy of search of optimal solutions in that for certain types of journeys a predetermined mode is chosen, such as a car for a visit to a suburban shopping center, and public transportation for shopping in the city center. Also, it has been shown that people are not always aware of all relevant aspects of journeys that they may have to make.

TRAVEL RELATED TO NONECONOMIC ACTIVITIES

Satisfaction derived from the consumption of economic goods and services is important to all of us, but this fact should not lead us to forget the role that noneconomic activities play in our lives.

Nobody can do without sleep and rest; we all need contacts with members of our families, friends, and other people with whom we share common interests or ideas; and many of us would rather do without theaters, concerts, cinemas, and other recreational services than miss the benefit given by spending at least some time in a natural environment, many of which are still plentiful.

Many social contacts need travel, and so does recreation in nature. In principle this travel is not different from travel related to economic consumption, but the satisfaction against which it must be traded off is of a noneconomic nature. That these trade-offs take place can be derived from the fact that we visit friends in faraway places less often than we do those nearby, and the same applies to other travel purposes in the noneconomic sphere.

Normally, noneconomic activities do not involve money expenditure and can therefore be left out of consideration in the budget constraint for economic consumption. When we consider travel not as a consumption activity in its own right but as ancillary to other activities that generally generate utility, whether economic or not, we have to enlarge the budget constraint:

$$\mathbf{p}_c'\mathbf{c} + \mathbf{p}_v'\mathbf{v}_c + \mathbf{p}_v'\mathbf{v}_n \leq Y$$

where $\mathbf{v}_n$ stands for travel related to noneconomic activities.

For the rest, the relation of travel to noneconomic activities is similar to the relation to economic consumption: to do both, we may have to incur travel cost and dissatisfaction of making journeys.

TRAVEL BETWEEN PRODUCTION PLACES AND CONSUMPTION PLACES

The journey between home and work traditionally has obtained much attention in theoretical studies and practical research in the field of transportation demand. This attention can be explained by the importance of the journey's purpose in peak demand and also by the relation it establishes between transportation and land use.

A particular property of the home-to-work relation is that it depends on decisions made over a long period. The locations being chosen provide two fixed places in daily travel patterns. Possibilities for variation do exist, however, because people can make stops between home and work, vary their route and mode of travel, and make trips for private purposes from their workplaces.

Choices of home and workplace sometimes are made simultaneously or nearly so; for example, people take jobs far away from their original homes and look around to find a suitable place to live as soon as possible. Sometimes people change either homes or jobs without intending to change the other also. Decisions become more complicated, however, if more than one member of a household has a job or when one person has two jobs.

As for consumption, the financial and psychological burdens of the daily trip to and from work have to be traded off against the advantages of certain combinations of home and job or combinations of a home and several jobs when there is more than one job per household. This trade-off is far more complicated than in the consumption case: in that case we compared a costly or unpleasant journey, giving access to cheaper or better-quality consumption, with a cheap and short trip to less attractive consumption activities. This is not so for the journey to work, in which both ends are variable and in which both will be the starting point of other trips.

Let us first consider jobs. The first reason for needing a job (e.g., an independent professional activity with employment under one heading) is to provide an adequate income for financing consumption. Therefore, the income to be derived from a job is an important factor in choice. Of course, each job requires certain skills, and the more general these skills are, the more likely it is to find a job in a certain place. Another important factor is the level of satisfaction related to the job, which is determined by the type of work, its ancillary benefits (or lack thereof), and the status provided by it.

All of these elements are of variable importance for individual members of the employed population. Income seems to be a fairly stable element, but sometimes unequal pay is given for similar labor (when age or seniority plays a role in salary scales) or equal pay for unequal labor (when the quality of work is not reflected in salaries). In addition, the propensity to costly consumption is not the same for everybody, and fiscal arrangements can also play a role.

Satisfaction with a job is perhaps still a more personal question. Of course there are jobs that are generally found to be pleasant, but to a large extent pleasure in work is a question of the type of work being matched to one's skills. The more rare the skills are, the less likely it is that a satisfactory job can be found in a given place.

The factors influencing the choice of home are still more complicated. It is not only price, dimensions, and quality of the home itself that are important for our choices in this respect. The home is the place from which daily travel and activity patterns of all members of the household are organized. Trips to school, shops, recreation, relatives, and friends in many cases start from home. Insofar as travel has an influence on the choice of home, it is not only the journey to work that counts, but all the journeys that may be made from home. We can express these journeys as the potential accessibility to various types of destinations. The higher the potential, the more attractive is the location from the viewpoint of minimizing transport cost. These potentials are different for each member of the household, both as destinations and as travel costs and dissatisfaction; for example, children travel to school mostly on foot or by bicycle and can use cars only when traveling with older people.

The following are factors that are involved in choosing a home and a job:

- Income of all employed members of the household,
- Work satisfaction of employed members of the household,
- Price of home,
- Dimension and amenities of home,
- Environment of home,
- Travel impedance from home to potential activities for all members of the household, and
- Travel impedance between home and work for employed members of the household.

The last factor cannot be expected to play an overwhelming role in the choice of home and job, and it is therefore not astonishing that many distribution models for home-to-work journeys do not satisfactorily explain this phenomenon (*1*).

TRAVEL AS PART OF PRODUCTION PROCESS

Many productive activities imply travel. Doctors visiting patients and maintenance staff working at installations in homes or other buildings travel between two successive activities. In the course of their work, some people need to meet other people who are not in the same building, or they must attend formal meetings, visit exhibitions, go to conferences, and so on.

The cost of a journey during work time can be considered to include, apart from the actual cost of travel (fare or the cost of a car), the wage cost (including all overheads) of the person traveling. When we consider the firm to aim at an economic optimum, the decisions of whether and how a journey will be made will depend on how important the journey is to the production process and the cost of alternative ways of making the journey, calculated in the above-mentioned way. In real life, however, there are many reasons that things happen otherwise. We will try to enumerate the most important of them, but this enumeration can never be exhaustive.

A firm can have widely varying dimensions: the firm can be a single person working for him- or herself or it can be a very large organization. In the latter case, not all decisions can be made by central management, so there will be much delegation. In many cases occasional journeys will be planned by the person who will actually travel. For trips made regularly, certain rules may be set.

An enterprise or other organization does not always strive for a minimum-cost situation. As long as costs can be met and travel patterns are not clearly extravagant, management may not bother too much about efficiency.

Productivity of working time is not constant throughout the day; for example, after a period of very intensive work, people need some relaxation, and a trip between two successive activities may well serve this purpose.

Apart from relaxation, travel time—in particular on intercity trains—can also provide the opportunity to work (or read) quietly, thus using that time twice for different purposes. Netherlands railways even have "working compartments" in their intercity coaches, in which people are asked not to distract their fellow passengers by talking or other types of intrusion.

Travel during work is often considered as a fringe benefit, particularly when travel can take place in a pleasant way or to pleasant places, or both. Many workers prefer not to be clustered in the same place, and travel can give both variation and relaxation. These factors can even have an influence on the choice of job: the variation provided by travel can augment the satisfaction of work.

Apart from immaterial benefits, business travel can also provide material benefits. Some employers provide company cars that can also be used for private trips; other employers provide rail season tickets. When one of these situations exists, it will influence travel behavior as far as the choice of mode is concerned.

The same effect results from a reimbursement of car cost, including a part of the fixed annual cost of keeping the car. Workers with this type of arrangement will have an interest in making as many business trips as possible by car.

Many business journeys take place, at least partly, outside normal working hours without compensation for the time spent traveling. And, for some people, hours of work are determined not so much by office hours as by the workload: if time is lost during office hours, the remaining work may have to be done by working late in the office or at home. In these situations there are interactions between time spent for business travel and time spent for private activities.

The neat and simple model of optimization between travel cost, travel time, and work time must be made far more complicated. We have to be aware of imperfect decision processes, fringe benefits, discontinuities, and, perhaps most important, the interactions between work travel and private feelings. It is not just a problem of business calculations that determine if and how professional journeys will be made, but the travelers themselves are influencing and being influenced by decisions. Travel in the course of work has therefore more similarities with the other travel purposes than is normally expected and produces utilities and disutilities as other journey purposes do.

Institutional factors are again important, and they are relevant in two areas: reimbursement of travel cost for journeys during work, and treatment of travel time outside normal working time (payment of overtime or free time given in compensation).

INTRODUCING TIME

Allocation of Time

Until now, we have not explicitly introduced time as a factor that influences travel decisions. We have reasoned in terms of satisfaction, dissatisfaction, effort, income, and cost. Some of these are related to time.

We will now start discussing time as a separate element that influences our activity patterns and thereby our travel decisions. In general we can say that our work, consumption, noneconomic activities, and travel all cost time, and we can therefore construct a time budget similar to the income budget as follows:

$$t_w + \mathbf{t}_c\mathbf{c} + \mathbf{t}_n\mathbf{n} + \mathbf{t}_v\mathbf{v} = T$$

It seems trivial to state that T, the total time available, is 24 hr a day, but we can to a certain extent use the same time twice: it is possible to meet a friend during a meal, to read during a journey, or to knit while watching television. Nevertheless, the possibilities for such "double use" of time are limited, but we can see that those for whom time is very scarce are looking for a maximum of combinations.

Satisfaction and effort or utility and disutility can normally be considered as opposite categories. One could place time on the same side of the balance as effort: using time for a certain purpose takes away other options. On the other hand, spending time on pleasant things gives satisfaction: eating a good meal at leisure gives more satisfaction than eating it in a hurry; having to leave a theater before the play has ended is on most occasions very unfulfilling.

We can generally say that time is spent to obtain satisfaction, either directly by doing things one likes to do, or indirectly by working for an income, by producing consumption goods ourselves, or by traveling to a destination of one's choosing. There are trade-offs between doing pleasant and unpleasant things. On the one hand, we must do unpleasant things to be able to do pleasant things. On the other hand, a sequence of activities that are all of approximately the same level of pleasantness gives less satisfaction than a sequence of varying degrees of pleasantness and unpleasantness (*2*). It is as if we can appreciate pleasantness more if we can compare it with another, less pleasant, activity that is close in time. And, as stated by Scitovsky (*2*), it is not only pleasantness itself that gives satisfaction, but the change from a less pleasant to a pleasant state of activity.

For these reasons we cannot expect an organization of use of time that is based on trade-offs between individual activities. We can presume that there is a tendency to equalize the satisfaction of the last dollar spent on each activity, but this does not apply to minutes. Instead, we can expect people to organize their time in such a way that an optimal result is reached as an overall state of satisfaction.

When we use the word organize, it does not mean that the whole pattern of activities that people undertake is organized in advance. Some activities, such as going to work or school, are fixed for a long time, but others are undertaken when the idea comes up. It is an advantage in itself to be able to respond to ideas of the moment rather than having the whole day filled with planned activities.

Optimization does not mean more than choosing the best of the options available, according to our priorities at the moment of choice.

Time Budget for Transportation

We said earlier that the total time spent on all activities is not limited in the strict sense, because it is possible to combine some activities to be simultaneous. Nevertheless, we cannot circumvent the fact that our day contains only 24 hr. The more time that is spent on one activity, the less time that is left for the others.

Travel can be useful only if there is some time left for activities at the destination. Most people do not want to spend too much time on travel. All of these circumstances imply a tendency toward keeping total travel time within some limits. There is, however, no reason to assume that there will be a strict limit on total travel time per person. This time can vary according to personal activity patterns and propensity to travel, and in fact individuals can spend very different proportions of their time traveling.

Time budget analysis can provide important information that is relevant to travel behavior, but it should not be limited to time spent on travel. It should show activities at home, outdoors, and travel, and make explicit the constraints imposed by factors such as work and school hours, opening hours of shops, and showtimes of theaters and cinemas. From such an analysis, it will be able to derive what changes can occur because of a shorter working day or other shopping hours.

Traditional forecasting techniques do not generally provide linkages between individual travel purposes, to allow the time needed for trips (travel and activity) for one purpose influence decisions on other purposes. In real life, people need to accommodate all activities within the time available. It is therefore improbable that those who travel a long distance between home and work will also make long recreational trips in the evening, simply because recreational activities are also located far from their home.

Value of Time

The value of travel time has been a classical subject in discussions on passenger transportation demand and project evaluation (*3,4*). Some governments even issued directives on values of time to be used in project evaluations (*5*). The subject also has been widely contested, mainly from a political approach.

To first bring terminology in order: travel time generally has a negative value; it should therefore be considered as a cost of travel. This is in line with the place of travel time in generalized cost. It is the saving of time that is a benefit to users and, therefore, a positive value in evaluation. Faster travel, however, does not always mean that a saving in time is realized. Even when we do not assume that total travel time is constant, it can be that faster travel will enable people to take longer journeys.

Attempts to derive values of travel time from value of productive time have been made. The most straightforward reasoning for this can be given for travel during work time, but even in this case there can be doubts, as was explained earlier. For other journeys, the argument that a savings in travel time can be used to work longer hours and in that way obtain a higher income does not hold for several reasons. First, many people are not in a position to determine their working hours at will, and if they are, they might decide to spend the time saved on travel for additional leisure activities instead of for more work. The most important argument against this relationship is that it considers time and money but not effort or level of satisfaction, or both. Oort has argued (*6*) that time and effort (or exertion) cannot be isolated in analysis. Any estimate of value of time includes both the effect on total available time and an estimate of the effort or utility associated with the activities undertaken during that time.

Estimates that have been made derive values that are different from hourly income and that vary according to the ways in which time is spent (*3*). Most studies arrive at values of time as a proportion of hourly income that is variable according to the ways in which time was spent and sometimes travel purpose, but not, for example, time of day or direction of travel.

An analysis directed toward individual situations, however, will show that there are considerable differences between people, with respect to both the options they have to use their time and their evaluation of various kinds of effort. For the same person, there can be considerable differences according to the time of day and day of the week.

Value is nothing more (or less) than the expression of the importance of an economic good for a person's well-being. When we consider time as an economic good, it can have an economic value. When we want to express this value in terms of money, the money value of time is determined by both the value of time and the value of money. That is why values of time have been shown to vary with income, as income influences the (marginal) value of money.

However, the role of time in daily activity patterns does not vary with income in the first place. Household situation and employment or education are far more important factors. People who both have a full-time job and are in charge of managing a household (even a one-person household) have probably the smallest amounts of time, at least during some periods of the day. People who are retired may have difficulty in filling their days and can consider the time spent on a journey as a benefit to them, as long as the journey can be made in reasonable comfort.

In addition, a shorter journey time can make it possible to make a trip without staying away from home for the night or to combine activities in different locations in one day.

For all these reasons, we should not expect to find unique values of time, even when expressed as a proportion of household income. Values of time can be expected to vary, and any analysis in this field should be organized so as to show the extent of this variation, both between people and according to time and circumstances. Only in this way will it be possible to obtain information that can be useful for economic analysis of travel behavior and, later, for an evaluation of policy alternatives. The practice of applying standard values, also when expressed as a ratio of income, will lead to erroneous forecasts and discriminatory decisions.

SYNTHESIS

Human life can be considered to consist of a sequence of activities, some of them regular and dependent on rules, conventions, or decisions made for a long period; others dependent on decisions of the moment, perhaps made without thinking of other options. Sometimes activities can be combined during the same time (meeting a friend during a meal, reading on a train), but in most cases each activity takes up a definite period of time.

The activities can be categorized into production (in the course of employment or our own enterprise or for our own consumption), economic consumption, noneconomic activities, and travel. People can be presumed to organize their activities in such a way that they choose the most favorable combination of the options open to them. That people do not recognize all available options, do not always explicitly consider the pros and cons of each option they see, and sometimes act on impulse is not contrary to this presumption. The time and effort needed to make a choice are also part of the optimization process. For important decisions, such as taking a new job or buying a house, people tend to consider carefully all the alternatives they can think of; for unimportant ones, such as drinking a cup of tea in one place or another, or not at all, they do not spend hours making up their minds.

Travel is related to some of the other activities, and in general it is a cost of the activity, in that the income or utility can be obtained only when some time, money, and effort are spent on travel. On some occasions, however, the journey itself can provide utility, and then it can be considered as a part of economic consumption. In most cases the journey is at the cost of other activities; of course the journey will be minimized for each activity but will grow with the amount and quality of activities to be undertaken. This apparent paradox is similar to the growth of total cost of industrial production combined with diminishing costs per unit, which is the normal pattern of an advancing industrial society.

The cost of travel, measured in time, money, and effort, can be of a very different importance relative to the activity to which it gives access. It is therefore not expected that a satisfactory explanation of travel behavior can be obtained from variables relative to the transportation system only.

The study of transportation must be integrated into the study of human activities in general and cannot be pursued as an isolated subject of scientific work. Only when we gain some understanding about factors determining human activity patterns may we be able to also understand transportation decisions.

ACKNOWLEDGMENTS

This paper is based on ideas developed during many discussions with colleagues in the fields of transportation economics and travel behavior. During the process of drafting the text, which took place a considerable time before its publication, invaluable support was given by the late Leo H. Klaassen, of the Erasmus University at Rotterdam. Professor Klaassen gave a considerable amount of time to reviewing and discussing the drafts and helping to arrange the arguments and formula in an appropriate way. All those who had the pleasure of working with Professor Klaassen will remember the high quality of his contributions to the field of transportation economics and economic science in general but, more important, his extreme kindness. He will always be remembered.

REFERENCES

1. *SIGMO*, Part 2 (Report of a Study on Travel Behavior). The Netherlands Ministry of Transport, The Hague.

2. T. Scitovsky. *The Joyless Economy*. Oxford University Press, New York, N.Y., 1976.
3. A. J. Harrison and D. A. Quarmby. *The Value of Time in Transport Planning. Report of the 6th Roundtable of European Conference of Ministers of Transport*, Paris, France, 1969.
4. P. B. Goodwin. *Value of Time. Report of the 30th Roundtable of European Conference of Ministers of Transport*, Paris, France, 1976.
5. P. T. McIntosh and D. A. Quarmby. *Generalised Costs and the Estimation of Movement Costs and Benefits in Transport*. MAU-Note 179. Department of the Environment, London, England, 1970.
6. C. J. Oort. The Evaluation of Travel Time. *Journal of Transport Economics and Policy*, Vol. 3, No. 3, Sept. 1969.

Publication of this paper sponsored by Committee on Transportation Economics.

Modification of QUEWZ Model To Estimate Fuel Costs and Tailpipe Emissions

PATTABI SESHADRI, HERNÁN E. DE SOLMINIHAC, AND ROBERT HARRISON

Traffic congestion created by urban highway construction and rehabilitation activity imposes substantial cost burdens on highway users. In their attempts to quantify and predict such costs, transportation planners have customarily relied on computer models that focus principally on vehicle operating cost and time delay increases. Increasingly, however, planners are recognizing that highway users themselves can inflate system costs—particularly those costs associated with increases in tailpipe emissions. Thus, the challenge to highway planners is to model these effects so that a full range of cost effects can be used at the planning stages of a traffic engineering project. The preliminary results of a study undertaken to modify an existing work zone model—QUEWZ—to report net increases in energy and tailpipe pollution associated with various traffic-handling scenarios within construction zones are given.

Researchers have determined that net vehicle pollution tends to increase when vehicles travel at slower speeds (*1*), and recent research (*2,3*) has underscored the significance of this pollution problem in construction work zones, in which vehicle slowdowns and the concomitant congestion inevitably result from that construction activity. However, there are now strategies and technical solutions to mitigate the adverse effects of rehabilitation and reconstruction work. Because these strategies and solutions generally are more expensive, it is crucial when such options are being evaluated that the extra costs of the mitigation strategies be weighed against their social benefits, including those accruing to highway users. Increasingly, these costs must include pollution effects in construction work zones.

Work zone models were first developed using time delay as the main decision-making criterion (*4*). Currently, several manual and computerized procedures are available to evaluate the effects of work zone closures on traffic. In the United States there is a comprehensive, manual work zone evaluation procedure (*5*) that selects the most appropriate traffic control strategy for a particular maintenance task. Other manuals include a user guide on planning and scheduling (*6*), which evaluates the effects on the basis of accidents, vehicle stops and delays, vehicle operating costs, and fuel consumption. Another evaluation procedure is the 1985 *Highway Capacity Manual* (*7*), which provides estimates of work zone capacity and procedures for estimating queue lengths and delays using input-output analysis.

P. Seshadri, Louis Berger and Associates, Inc., 303 Bear Hill Road, Waltham, Mass. 02154. H. E. de Solminihac, Depto. Ingeniería de Construcción, Pontificia Universidad Católica de Chile, Casilla 306 Correo 22, Santiago, Chile. R. Harrison, Center for Transportation Research, University of Texas, 3208 Red River, Suite 200, Austin, Tex. 78705.

Some of these assessment routines are available on microcomputers. For example, additional user costs (time and vehicle operation) associated with lane closures can be predicted using Queue and User cost Evaluation at Work Zones or QUEWZ (*8*). And a number of computer programs, including DELAY (*9*), FREWAY (*10*), and FREQ10PC (*11*), are able to evaluate freeway work zone lane closures.

Another program, Computer-Assisted Reconstruction–Highway Operations and Planning, or CARHOP (*12*), can model traffic disruption caused by maintenance projects. Although several comprehensive computer programs, such as FREECON (*13*), were developed to analyze work zones on freeways, few models have been developed to analyze the arterial system with the work zone as an overall comprehensive unit. One model, Work Zone Analysis Tool for the Arterial or WZATA (*14,15*), analyzes and evaluates the system consisting of the lane closure between two signalized intersections.

Evaluating these planning tools, we decided that QUEWZ best met the objectives of our research. Accordingly, we decided to modify the model, first to report fuel as a separate user cost element, and then to report tailpipe emissions. To differentiate between the two models, the new version was termed QUEWZ-E.

QUEWZ-E MODEL

The original QUEWZ model is a comprehensive evaluation tool that estimates vehicle operating and time delay costs, reporting the former as a total cost figure. Two general configurations of work zone lane closures are incorporated into QUEWZ. The first configuration models situations in which one or more lanes are closed in one direction, whereas the traffic moving in the opposite direction is not affected. The second configuration models crossover strategy, in which all lanes in one direction are closed and two-lane, two-way traffic is maintained in the other direction. A maximum of six lanes in each direction can be handled in the model.

In modifying the existing program, we first altered the vehicle operating costs subroutine in the personal computer version of QUEWZ to report energy consumption. Specifically, operating costs were disaggregated so that fuel and oil could be reported separately from total operating costs. This required reference to, and use of, research findings published by the

Texas Research and Development Foundation (*16,17*). Figure 1 illustrates the input and output for the new QUEWZ-E program.

Although the task of reporting fuel consumption was simple, that associated with predicting emissions was complex. The remainder of the paper therefore describes both the methodology developed for this process and the sources of data used to predict emission levels. The model calculates and reports the additional emission levels by first predicting free-flow pollution and then using that figure as a base case, comparing that level with the pollution from the various traffic-handling schemes. The model then reports the difference between the predicted and actual emission levels.

WORK ZONE MOBILE SOURCE EMISSION PREDICTION

The emission prediction function in QUEWZ-E uses a four-step process. Recognizing that traffic behavior varies according to the location being modeled, the first step involves characterizing the traffic where emissions are to be evaluated. For example, if emissions from free-flowing traffic on a highway are required, the key variable will be vehicle speeds and flow. These speeds can be used with an emission model that predicts the emissions of a vehicle cruising at a given speed to determine the source strength. On the other hand, if it is necessary to compute emissions at an intersection, then information may be required on traffic signal phasing, queue lengths, delay times, acceleration and deceleration rates, or capacity.

The second step is the estimation of the source strength, which requires an emissions model to account for vehicle conditions and driving patterns existing in the zone of interest. Most emission rate analysis models (both freeway and intersection air quality models) are based on data obtained from two major studies on mobile source emissions administered by the Environmental Protection Agency (EPA), namely, the Modal Analysis Model (*18*) and the MOBILE series of models (*19–22*).

The third step in modeling mobile source pollution near a roadway uses the emission profile from Step 2 to model the dispersion of the emitted gases along, and in the vicinity of, the roadway. The dispersion of the emissions is dependent on several factors, including source strength, width of roadway, wind direction and speed, source height, and mixing height. The fourth step involves the calibration of the dispersion model using actual dispersion data collected from the site being modeled.

Because the scope of this study was limited to the determination of the source strength of work zone traffic emissions, we modeled the carbon monoxide, hydrocarbon, and nitrogen oxide emissions of passenger cars and trucks.

Traffic Analysis Model

The flow of traffic in the region of a work zone on a freeway is unique to the extent that it needs to be described by a combination of free-flowing traffic and stop-and-go traffic. When traffic volumes are not large enough to cause congestion and queueing, the traffic can be characterized entirely by the volume and speeds. When congestion occurs, additional information (such as queue lengths) is needed to characterize the traffic. Hence, a traffic model that is capable of comprehensively defining the work zone problem is required. In pursuing such a traffic model, we characterized traffic passing through a work zone in three ways, from the viewpoint of emission prediction (*23*):

- *Vehicles proceeding undelayed through the work zone*—When the capacity of the work zone is sufficiently greater than the demand, the vehicles passing through the work zone

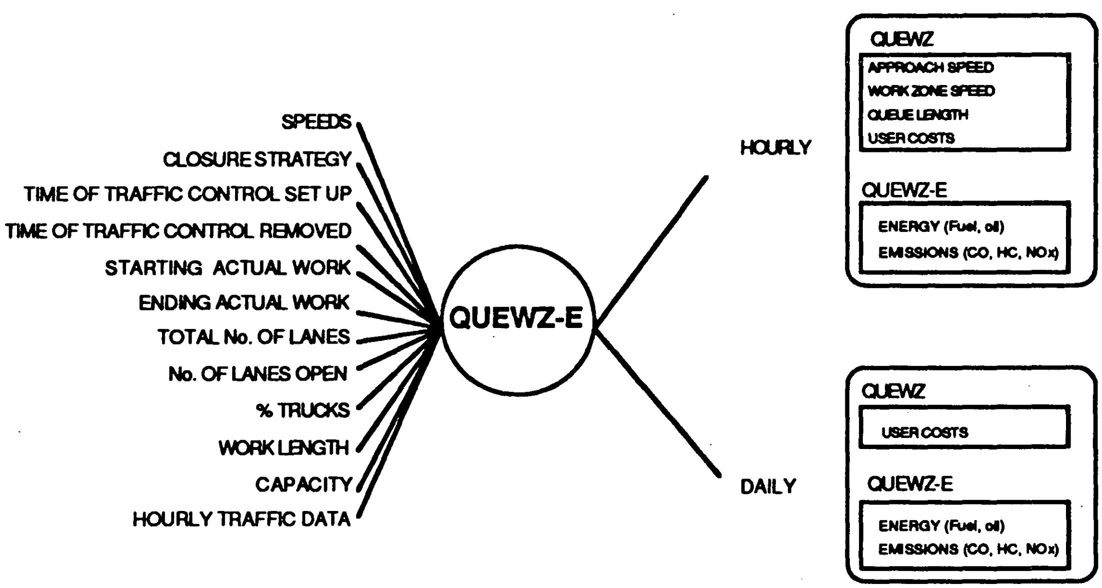


FIGURE 1 Input and output for QUEWZ-E program.

are processed without any delay. This scenario does not contribute toward excess emission levels.

- *Vehicles proceeding through the work zone at a reduced speed*—As the traffic demand at the work zone approaches the capacity of the work zone, the rate at which vehicles are processed through the work zone decreases, lowering the overall speeds of vehicles. This involves a deceleration from the approach speed to a minimum speed near the work zone, an acceleration to the work zone average speed from this minimum speed, travel at a lower average speed through the work zone, and an acceleration from the work zone speed to pre-work zone speed at the end of the work zone. The lower average speeds in the work zone might result in less pollution when compared with cases in which vehicles proceed unhindered at higher average speeds.
- *Vehicles stopping near the work zone because of queue formation*—When the traffic demand at the work zone exceeds the capacity of the work zone, queue formation takes place upstream of the work zone and involves a deceleration from the approach speed to idling at the end of the queue, short acceleration-deceleration movements (creeping motion) through the queue, acceleration to work zone speed at the beginning of the work zone, passage through the work zone at the average work zone speed, and acceleration to prework zone speed at the end of the work zone. The characteristics of this traffic behavior are illustrated in Figure 2. Because this scenario has the maximum impact in terms of excess emissions, an appropriate analysis is needed.

Excess vehicle emissions at a work zone are defined as the difference between the total emissions produced at and near the work zone minus those produced had traffic cruised unhindered through the work zone. These excess emissions can be determined as follows. The time spent by each vehicle in each mode of operation (accelerate, decelerate, cruise, queue) is first computed. The average emission rate for each mode and for each pollutant is then multiplied by the time spent in that mode to obtain the emission values. These emission values, when multiplied by the total number of vehicles in the analysis period, give the total mass of pollutants. The mass of pollutants generated if the vehicles were traveling over the affected length in the absence of the work zone is also computed. The difference between the two gives the required excess emissions of the given pollutant during the analysis period.

If the speeds at the beginning and end of the zones described in Figure 2 are known, the time spent by the vehicle in these zones can be calculated by assuming constant acceleration and deceleration rates for the vehicles. To provide information on speeds at and near the work zone, as well as the time spent

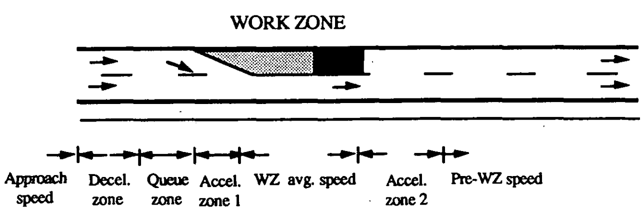


FIGURE 2 Traffic behavior near a work zone as a result of queue formation.

by vehicles in each zone, the traffic analysis model would require the following data:

1. Work zone capacity,
2. Speed-flow relationship,
3. Length of work zone,
4. Average length of queue,
5. Average vehicle speeds in queue,
6. Vehicle mix, and
7. Acceleration and deceleration rate of vehicles.

The QUEWZ work zone model (*24*) satisfies most of these data requirements, which led to its adoption over other, more complex models. The acceleration and deceleration rates of passenger cars and trucks, however, are not provided by the QUEWZ model. For the work zone emissions problem, the acceleration rates for passenger cars and trucks were assumed to be constant values of 4.5 and 1.6 ft/sec^2, respectively, and the deceleration rates were assumed to be constant values of 6.0 and 2.2 ft/sec^2, respectively, on the basis of values given previously (*25,26*).

Using the information on vehicle speeds provided by QUEWZ, the time spent by each vehicle in each mode can be determined using kinematics.

Modal Emission Rate Models

The pollution emission rates under various modes of operation (e.g., acceleration, cruise, deceleration, and idling) now need to be quantified so that the excess work zone emissions can be computed. And because the emission levels of pollutants vary widely with the mode of operation of the vehicle, modal emission rates are required to model air quality where there are wide variations in the traffic flow speeds.

In this context, the unusually high levels of pollution associated with high traffic volume at urban intersections has led highway planners to focus on modeling air quality specifically at intersections. Accordingly, various approaches have been used to obtain the modal emission rates of vehicles. One approach has been to use modal emissions from the Modal Analysis Model and to correct this using the ratio of the results from the MOBILE model for actual and base scenarios. The base scenario is for conditions used in the modal model, namely, a 1977 calendar year, a light-duty vehicle fleet, 100 percent hot stabilized operating conditions, a temperature of 75°F, and the average speed of the user-defined driving sequence. The actual scenario is for the corresponding conditions in the calendar year being modeled.

This approach is used in the IMM (*27*) and the TEXIN2 (*28*) models. The main problem associated with this strategy is that for every speed at which modal emissions are required, the MOBILE model must be run for the base and actual scenarios for that speed. This implies that the MOBILE model should be merged with the traffic analysis model, which in the present context is the QUEWZ model. If this operation is performed, the work zone model will grow cumbersome, will require more inputs, and will take a much longer time to run. Also, only a specific version of the MOBILE model can be coded into the work zone model. Frequent updates of the work zone emissions model will be required as the EPA updates the MOBILE model.

The second approach has been to use emission rates from the MOBILE model and to correct them using modal correction factors. These correction factors have been derived using limited sets of emissions data from the SDS and FTP driving cycle tests (*29*). The correction factors are usually functions of the vehicle speed and acceleration. This approach has been used in the MICRO2 (*30*) and CALINE4 (*31*) models.

The CALINE4, TEXIN2, and IMM programs were developed exclusively for modeling carbon monoxide (CO) hot spots at intersections. The equations used for modeling modal CO emissions for the purpose of work zone emission prediction will follow closely those used in MICRO2 and CALINE4. The approach used in IMM and TEXIN2 is not used for reasons stated previously.

MICRO2 is the only program among these four that models hydrocarbons (HC) and nitrogen oxide (NOx) emissions. The modal HC and NOx emission models for the work zone problem will make use of the results from the MICRO2 model.

All the modal emission rate models in the CALINE4 and MICRO2 programs were developed using data from light-duty gasoline vehicles. However, the work zone model also requires modal emission rates for diesel trucks. If we assume that the behavior of diesel vehicles in various modes of operation is similar to that of gasoline vehicles, then the modal correction factors developed for passenger cars can be applied to the composite emissions from trucks to obtain the modal emission rates for trucks.

Modal CO Emission Rates

The MOBILE4.1 model (*22*), which is the recent update of the MOBILE series of models, provides idle emission factors for the hot stabilized mode of operation. Hot stabilized idle emissions have been included in the MOBILE models to facilitate quantification of emissions resulting from idling in queues. These idle emission factors are used for the idle emission rate model.

In the development of the CALINE4 model, an analysis of the data from the California Air Resources Board (CARB) study (*32*) showed that emission rates under deceleration were relatively constant over the 16 deceleration modes of the surveillance driving sequence (SDS) driving cycle. These rates were approximately 50 percent higher than the idle emission rates. This observation was found to be consistent with the practice of decelerating by gradually releasing the accelerator during a planned deceleration. Hence, deceleration mode emissions are assumed to be 1.5 times the idle mode emissions in the work zone emissions model.

The MICRO2 model assumes that emissions in the cruise mode are constant and equal to the idle emission rates. However, as a vehicle cruises at higher speeds, the CO emissions increase. The CALINE4 model uses a cruise correction factor to the MOBILE model scenario rate emissions, which was derived as follows.

SDS emissions test data for various idling, acceleration, cruise, and deceleration segments were given artificial time weightings to provide a simulated federal test procedure (FTP) stabilized mode sequence. Data from the cruise portion of the SDS testing cycle were then analyzed to develop correlations with the SDS simulation of the FTP stabilized mode emission rates. The dependent variable was the ratio of SDS to FTP (simulated using SDS) emission rate. The independent variable was cruise speed. The following relationship was obtained from this analysis:

$$\text{cruise emission rate (g/hr)} = 16.2\,\text{mph MOBILE scenario rate} \times (\text{g/mi}) * 16.2\,(\text{mph}) * (0.494 + 0.000227 * S^2) \quad (1)$$

where S is the cruise speed of the vehicle. This result is consistent with the fact that the drag force on a vehicle cruising at a speed S is proportional to the square of the speed, S^2. Hence, as the vehicle cruises at higher speeds, the higher drag force exerts a greater load on the engine, leading to increased CO emissions. This cruise emission model is used in the work zone emissions model.

The MICRO2 and CALINE4 models develop acceleration correction factors to the composite MOBILE emission value as a function of the product of acceleration and speed (AS). The product of acceleration and speed is equivalent to power per unit mass. Therefore, the power expended by a vehicle during acceleration is proportional to AS. As power demand approaches engine capacity, vehicles tend to burn fuel less efficiently, resulting in higher carbon monoxide emissions (*30*).

The acceleration model in MICRO2 was based on data from 45 light-duty, 1975 vehicles tested in Denver on the SDS cycle (*33*). Results were analyzed using a ratio of the time rate of modal emissions and the time rate of FTP emissions. Use of this ratio allowed the direct conversion of average route speed emission factors to modal emission rates. The acceleration model developed through this analysis was

$$\text{acceleration emission rate (g/hr)} = \text{MOBILE scenario rate (g/mi)} * S\,(\text{mph}) * [0.182 - 0.00798(AS) + 0.000362 * (AS)^2] \quad (2)$$

with AS representing the product of the average acceleration and average speed for the acceleration event in units of ft^2/sec^3. This acceleration emission rate model is adopted for the work zone emissions model. Internally, the model assumes a constant AS value. For passenger cars this value is 97 ft^2/sec^3, representing an average acceleration of 4.5 ft/sec^2 from 0 to 30 mph. For each speed at which the acceleration emission rate is required, the MOBILE scenario rate needs to be determined for an average route speed equal to that speed. To circumvent this problem, an equivalent regression model was developed using MOBILE4.1 scenario rates for average route speeds from 2.5 to 65 mph. The acceleration and cruise emission rate models for CO for passenger cars are shown in Figure 3.

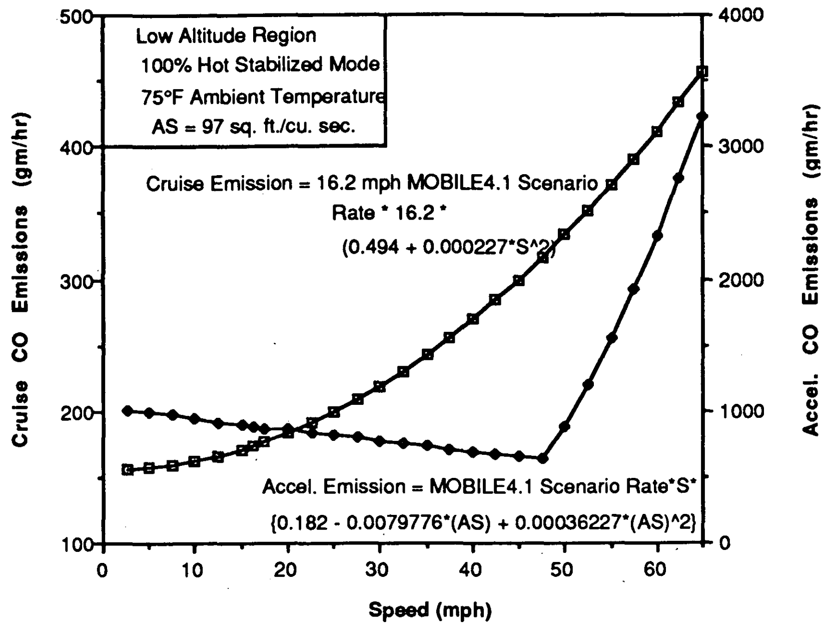


FIGURE 3 Plot of CO acceleration and cruise emission rate models for passenger cars.

Modal HC Emission Rates

As in the case for idle carbon monoxide emissions, hot stabilized idle hydrocarbon emissions can be obtained directly from the base scenario run of the MOBILE4.1 model. As in the MICRO2 model, the deceleration and cruise emission rates are assumed to be equal to the idle emission rate.

For emission of hydrocarbons under acceleration, the MICRO2 model applies correction factors to the composite MOBILE emission value as a function of *AS*. The hydrocarbon acceleration model in MICRO2, presented below, was developed in a manner similar to that described for carbon monoxide.

acceleration emission rate (g/hr)

$$= \text{MOBILE scenario rate (g/mi)} * S\text{ (mph)} * \{[0.018 + 5.266 \times 10^{-4}(AS)] + 6.1296 \times 10^{-6}(AS)^2\} \quad (3)$$

Modal NOx Emission Rate Models

As in the case for idle carbon monoxide and hydrocarbon emissions, hot stabilized idle nitrogen oxide emissions can be obtained directly from the base scenario run of the MOBILE4.1 model. The cruise emission rate is assumed to be equal to the idle emission rate, as in the MICRO2 model.

For emission of nitrogen oxides under deceleration, the MICRO2 model applies correction factors to the composite MOBILE emission value as a function of the deceleration-speed product ($-AS$). The resulting equation is presented here.

deceleration emission rate (g/hr)

$$= \text{MOBILE scenario rate (g/mi)} * S\text{ (mph)} * [0.00143 - 1.7005 \times 10^{-4}(-AS)] \quad (4)$$

The acceleration model for nitrogen oxide emissions uses results from the MICRO2 model, which applies correction factors to the composite MOBILE emission value as a function of *AS*. The resulting equation is

acceleration emission rate (g/hr)

$$= \text{MOBILE scenario rate (g/mi)} * S\text{ (mph)} * [0.00386 + 8.1446 \times 10^{-4}(AS)] \quad (5)$$

Default MOBILE4.1 Scenario Rates

The emission rate equations described in the previous sections apply modal correction factors to the composite emission rates from the MOBILE model to arrive at the modal emission rates. The composite emission rates used in the QUEWZ model were obtained from MOBILE4.1 (*22*), which is the latest update in the MOBILE series of models. To run the MOBILE4.1 model, reasonable assumptions have to be made regarding input variables that influence emissions so as to obtain a representation of the work zone problem. These assumptions constitute the default scenario used in the work zone emissions model, the main elements of which are the following:

- A low-altitude region is modeled,
- An ambient temperature of 75°F is used,
- All vehicles are assumed to be operating in a hot stabilized mode,
- No antitampering program is in effect,
- No inspection/maintenance program is in effect, and
- MOBILE4.1 default vehicle age distributions are used.

To model scenarios other than this default scenario, correction factors need to be applied to the emission rate equations used in the work zone model. Correction factors based solely on the ratio of idle emissions in the actual scenario being modeled to the idle emissions in the default scenario are used in the work zone emissions model.

QUEWZ-E Model

The methodologies and models described in the preceding sections were implemented in the QUEWZ program. The revised version of the QUEWZ model has been dubbed QUEWZ-E, indicating the additional capability of emission-value prediction. The only additional inputs needed by the program for emissions calculation are the hot stabilized idle emission rates of CO, HC, and NOx for passenger cars and trucks for the scenario being modeled.

The program gives an output of the CO, HC, and NOx emissions for each hour that the work zone is in operation. In addition, the program gives an output of the total daily excess emissions of these pollutants.

To test the model and to illustrate its use, eight sample work zone problems were analyzed. All the sample problems use an average daily traffic (ADT) of 25,000 vehicles in each direction of the freeway. The original number of lanes, the number of open lanes at the work zone, the work zone length, and the work zone activity period were the various parameters used in these sample problems. Table 1 summarizes and shows that queueing creates significant excess emissions in the work zone.

This model can be usefully applied when various work zone traffic management strategies are being evaluated. Say, for example, that a planner must choose between Problems 5 and 7. The QUEWZ-E model shows that the excess emissions caused by the closure of two lanes at the same time is more than double the excess emissions created by the closure of one lane at a time. In a similar manner, planners can compare various strategies to arrive at desirable work zone configurations and work schedules.

This model could also be applied when analyzing the work zone (and the elements affected by it) as a total system. Indeed, such a systems approach in the analysis of a work zone has been recently proposed (*34*). Given a major freeway or highway reconstruction or rehabilitation project, the analysis would take into account the agency costs, the business costs, the road user costs, the environmental costs, and costs to other parties (e.g., utility companies). The construction strategy that resulted in the lowest total system cost would then be selected. The results could also be used as leverage for construction strategies that make use of expediting techniques.

To illustrate the enormous impact imposed by work zones on freeways, we estimated the excess emissions at reconstruction and rehabilitation sites on the nation's deficient bridge infrastructure. Research conducted by Weissmann and Harrison (*35*) to predict user costs for similar purposes was used as the basis for this analysis. In their study, only bridges having high traffic levels were considered, defined by an ADT >20,000. The authors identified 524 deficient two-lane bridges, 297 deficient three-lane bridges, and 363 deficient four-lane bridges. Bridges with ADTs between 20,000 and 30,000 were assumed to be two lanes one way, with one lane closed during work. Bridges having ADTs between 30,000 and 45,000 were assumed to be three lanes one way, with two lanes closed and one lane of traffic from the bridge under construction being switched to run counterflow in the closed inside lane. Bridges with ADTs greater than 45,000 were assumed to be four lanes one way, with two lanes closed during work. The total user costs of reconstruction work on these bridges were calculated to be approximately $6 billion.

TABLE 1 Summary of Inputs to and Outputs from Test Problems

Prob. No.	No. of Lanes	No. of Open Lanes	WZ Length (mi)	Normal Capacity (vph)	Restricted WZ Capacity		Hours of Restr. Capacity		Hours of WZ Activity		Longest Queue (mi)	Total Emissions		
					Inactivity Hrs. (vph)	Activity Hrs. (vph)	Beg.	End	Beg.	End		CO (Kgs)	HC (Kgs)	NOx (Kgs)
1	2	1	1.0	4000	1800	1485	8	17	9	16	0.8	96.5	9.1	2.0
2	2	1	2.0	4000	1800	1485	8	17	9	16	0.8	101.7	11.7	2.6
3	2	1	1.0	4000	1800	1485	0	24	9	16	2.2	449.1	42.3	9.6
4	2	1	2.0	4000	1800	1485	0	24	9	16	2.2	464.5	48.5	11.1
5	3	2	1.0	6000	3600	2970	8	17	9	16	0.0	6.7	0.2	0.0
6	3	2	2.0	6000	3600	2970	8	17	9	16	0.0	6.4	0.4	0.1
7	3	1	1.0	6000	1800	1250	8	17	9	16	1.6	449.2	42.6	9.7
8	3	1	2.0	6000	1800	1250	8	17	9	16	1.6	459.1	46.9	10.8

TABLE 2 Predicted Emissions Using QUEWZ-E Model

ADT	No. of Lanes[1]	No. of Def. Bridges	CO/day/ Bridge (kgs)	HC/day/Bridge (kgs)	NOx/day/Bridge (kgs)	Total[2] CO Em. (tons)	Total[2] HC Em. (tons)	Total[2] NOx Em. (tons)
> 20,000 < 30,000	2	524	70.6	6.4	1.4	11098.32	1006.08	220.08
> 30,000 < 45,000	3[3]	297	41.8	2.0	0.4	3724.38	178.20	35.64
> 45,000	4	363	493.3	46.6	10.9	53720.37	5074.74	1187.01
Totals		1184				68543.07	6259.02	1442.73

[1] One lane closed for two-lane capacity, two lanes for four-lane capacity.
[2] Total Emissions assume a 300-day contract cycle per structure.
[3] Construction first rehabilitates two lanes, then the third. Two lanes always open to traffic. When only one lane is open, the matching bridge is opened to diverted traffic and both bridges are reduced to two-lane travel.

We then performed a similar analysis for the emission of CO, HC, and NOx. Applying QUEWZ-E to these bridge scenarios, we found that the predicted CO emissions totaled 68,543 tons, HC emissions totaled 6,259 tons, and NOx emissions totaled 1,443 tons, as shown in Table 2. Like the estimates of user costs, these emissions figures are conservative; that is, they relate only to a subset of the rural bridge population and use a truck ADT of only 14 percent, which is lower than most Interstate values. Thus, the results presented above give an indication of the magnitude of the emissions problem at work zones.

CONCLUSIONS

This paper has described the development of QUEWZ-E, a computer model used to predict excess energy consumption and excess mobile source emissions at freeway work zones. Given the characteristics of the work zone (e.g., configuration, schedules), the characteristics of traffic at the work zone (e.g., volume, percent trucks), and the emissions characteristics of vehicles in the area, the model is capable of providing the excess emission values for two vehicle types and three pollutant types. Thus, QUEWZ-E can be used for comparing work zone construction and traffic management strategies specifically in terms of air pollution, with the results from the model then used for expedited construction strategies that reduce air pollution.

There are still a few shortcomings. The model as structured currently does not take into account the diversion of traffic away from the work zone that results when long queues form. Further research is needed to quantify the nature of this traffic diversion.

Also, the lack of validated models characterizing modal emissions tends to handicap the present study. But as new data and modal emission rate models become available, the QUEWZ-E model can be easily updated to incorporate the new findings.

We should also note that the increase in truck traffic on freeways has resulted in an increase in particulate emissions. The nature of these emissions, as well as their quantification by mode of operation, needs to be explored in greater detail.

Nonetheless, the study shows that even when conservatively estimated, emission levels at work zones represent a substantial problem and cost burden. The QUEWZ-E model can be useful in identifying the construction strategy that most effectively reduces emission levels.

REFERENCES

1. P. J. Claffey. Time and Fuel Consumption for Highway User Benefit Studies. *Bulletin 276*, HRB, National Research Council, Washington, D.C., 1960.
2. M. Cameron. *Transportation Efficiency: Tackling Southern California's Air Pollution and Congestion.* Environmental Defense Fund, Regional Institute of Southern California, March 1991.
3. *The Air Pollution Transportation Linkage.* California Air Resources Board, Sacramento, 1989.
4. B. C. Butler. *"EAROMA" Work Zone Time Delay Model.* 1973.
5. *Planning and Scheduling Work Zone Traffic Control.* FHWA, U.S. Department of Transportation, 1981.
6. C. M. Abrams and J. J. Wang. *Planning and Scheduling Work Zone Traffic Control.* Implementation Package FHWA-IP-81-6. FHWA, U.S. Department of Transportation, 1981.
7. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
8. J. L. Memmott and C. L. Dudek. Queue and User Cost Evaluation of Work Zones. In *Transportation Research Record 979*, TRB, National Research Council, Washington, D.C., 1984.
9. J. Morales. Analytical Procedures for Estimating Freeway Traffic Congestion. *ITE Journal*, Vol. 57, No. 1, 1986, pp. 45–49.
10. N. Rouphail. FREWAY/SIGNAL. McTRANS Center, University of Florida, Gainesville, 1984.
11. J. Zhang, L. Leiman, and A. May. Evaluation of Operational Effects of Freeway Reconstruction Activities. In *Transportation Research Record 1232*, TRB, National Research Council, Washington, D.C., 1989, pp. 27–39.
12. J. D. Leonard and W. W. Recker. *CARHOP: An Environment for Computer Assisted Reconstruction—Highway Operations and Planning.* University of California, Irvine, 1986.
13. N. M. Rouphail. *A Model of Traffic Flow at Freeway Construction Lane Closures.* Ph.D. dissertation. Ohio State University, Columbus, 1981.
14. C. T. Joseph. *Model for the Analysis of Work Zones in Arterials.* Master's thesis. Arizona State University, Tempe, Dec. 1987.
15. C. T. Joseph, E. Radwan, and N. M. Rouphail. Work Zone Analysis Model for the Signalized Arterial. In *Transportation Research Record 1194*, TRB, National Research Council, Washington, D.C., 1988.
16. J. P. Zaniewski, B. C. Butler, G. Cunningham, G. E. Elkins,

M. Paggi, and R. Machemehl. *Vehicle Operating Costs, Fuel Consumption, and Pavement Type and Condition Factors.* Report FHWA-RD-80. Texas Research and Development Foundation, Austin, Aug. 1981.
17. W. F. McFarland, J. L. Memmott, M. L. Chui, R. Salleh, and E. Chang. *Preliminary Design for MICROBENCOST Computer Program and Updating Program.* Interim Report NCHRP-7-12. Texas Transportation Institute, Texas A&M University, College Station, April 1990.
18. P. Kunselman, H. T. McAdams, C. J. Domke, and M. E. Williams. *Automobile Exhaust Emission Modal Analysis Model.* EPA-460-/3-74-005. Office of Mobile Source Air Pollution Control, Environmental Protection Agency, Ann Arbor, Mich., 1974.
19. *User's Guide to MOBILE1 (Mobile Source Emissions Model).* EPA-400/9-78-007. Office of Air, Noise, and Radiation, Environmental Protection Agency, Washington, D.C., 1978.
20. *User's Guide to MOBILE2 (Mobile Source Emissions Model).* EPA-460/3-81-006. Office of Mobile Source Air Pollution Control, Environmental Protection Agency, Ann Arbor, Mich., 1981.
21. *User's Guide to MOBILE3 (Mobile Source Emissions Model).* EPA-460/3-84-002. Motor Vehicle Emissions Laboratory, Environmental Protection Agency, Ann Arbor, Mich., 1984.
22. *User's Guide to MOBILE4.1 (Mobile Source Emissions Model).* EPA-AA-TEB-91-01. Motor Vehicle Emissions Laboratory, Environmental Protection Agency, Ann Arbor, Mich., 1991.
23. P. Seshadri. *Work Zone Mobile Source Emission Prediction.* Master's thesis. University of Texas, Austin, Aug. 1992.
24. J. L. Memmott and C. L. Dudek. *A Model to Calculate the Road User Costs at Work Zones.* Research Report 292-1. Texas Transportation Institute, Texas A&M University, College Station, 1981.
25. W. S. Homburger (ed.). *Transportation and Traffic Engineering Handbook*, 2nd ed. Institute of Transportation Engineers. Prentice-Hall, Englewood Cliffs, N.J., 1982.
26. *A Policy on Geometric Design of Highways and Streets.* AASHTO, Washington, D.C., 1990.
27. *Carbon Monoxide Hot Spot Guidelines, Vol. V: User's Manual for the Intersection-Midblock Model.* EPA-450/3-78-037. Environmental Protection Agency, Research Triangle Park, N.C., 1978.
28. J. A. Bullin, J. J. Korpics, and M. W. Hlavinka. *User's Guide to the TEXIN2 Model—A Model for Predicting Carbon Monoxide Concentrations Near Intersections.* Research Report 283-2. Texas Transportation Institute, Texas A&M University, College Station, 1986.
29. R. D. Sculley. Vehicle Emission Rate Analysis for Carbon Monoxide Hot Spot Modeling. *Journal of Air Pollution Control Association*, Vol. 39, 1989.
30. R. G. Griffin. *MICRO2—An Air Quality Intersection Model, Final Report.* FHWA/CO/RD-83/14; PB85-219202. Colorado Department of Highways, Denver, 1983.
31. P. E. Benson. *CALINE4—A Dispersion Model for Predicting Air Pollutant Concentrations Near Roadways, Final Report.* FHWA/CA/TL-84/15. California Department of Transportation, Sacramento, 1984.
32. P. Dickey. *Lake Tahoe High Altitude Vehicle Surveillance Program.* Project T. California Air Resources Board, Dec. 1976.
33. J. A. Rutherford. *Automobile Exhaust Emission Surveillance—Analysis of the FY 1975 Program.* EPA-460/3-77-022. Environmental Protection Agency, Dec. 1977.
34. H. de Solminihac. *System Analysis for Expediting Urban Highway Construction.* Ph.D. dissertation. University of Texas, Austin, 1992.
35. J. Weissmann and R. Harrison. Impact of Turnpike Doubles and Triple 28s on the Rural Interstate Bridge Network. In *Transportation Research Record 1319*, TRB, National Research Council, Washington, D.C., 1991.

Publication of this paper sponsored by Committee on Transportation Economics.

Critique of Texas Research and Development Foundation Vehicle Operating Cost Model

PETER BEIN AND DAVID C. BIGGS

Between 1979 and 1981 the Texas Research and Development Foundation (TRDF) investigated the effects of highway design and pavement condition on vehicle operating costs (VOC) for FHWA. TRDF VOC data have been included in many highway planning and project evaluation models, including the Canadian Highway User Benefit Assessment Model, and two recent models, Highway Economic Requirements System and MicroBENCOST. The estimation methodology adopted and current applicability of VOC data developed by TRDF are critically examined. In addition, the structure and major assumptions underlying the TRDF model are described, and the TRDF data are compared with actual consumption of the VOC components and with fuel consumption predictions by ARFCOM. The TRDF model encodes highway, vehicle technology, operation, and economic conditions typical of the 1970s. Judgmental manipulation of the data base by TRDF has introduced further problems. Typical vehicles are fixed to those reflecting the fleet in 1978, and the model lacks a representative modern heavy truck. All VOC components, especially fuel consumption, have been proven to be erroneous to some degree. The model cannot adequately serve present highway investment appraisal needs. A mechanistically based substitute for the TRDF model should be developed from the components of state-of-the-art models: HDM-III, VETO, ARFCOM, and South African VOC methodology.

Between 1979 and 1981 the Texas Research and Development Foundation (TRDF) investigated the effects of highway design and pavement condition on vehicle operating costs (VOC) (*1*). The TRDF VOC data have been used in the Highway Performance Monitoring System (HPMS), and some highway departments have included the relationships into computerized models. The effects of pavement condition from the TRDF data has been included by Elkins et al. (*2*), whose aggregated equations, with only minor modifications and price indexing, have been included into the Highway Economic Requirements System (HERS). The TRDF relationships form the VOC prediction module of the Canadian Highway User Benefit Assessment Model (HUBAM).

The most current application of the TRDF data is in the MicroBENCOST software being developed under NCHRP Project 7-12. MicroBENCOST features multiple regression equations fitted to VOC tables developed by Zaniewski et al. (*1*) but modified for fuel consumption of trucks at zero grade in accordance with data collected by France (*3*). The component unit prices from 1980 have been updated using price indexes.

This paper critically examines the estimation methodology adopted, and the transferability of VOC data developed by TRDF. The TRDF data are compared with actual consumptions of the VOC components and with fuel consumption predictions by ARFCOM. The disaggregate VOC relationships, as used in MicroBENCOST in November 1991, are considered to represent the TRDF data for the present critique.

P. Bein, Planning Services Branch, Ministry of Transportation and Highways, 3B-940 Blanshard Street, Victoria, British Columbia, Canada V8W 3E6. D. C. Biggs, Sypher:Mueller International, 220 Laurier Avenue West, Suite 500, Ottawa, Ontario, Canada, K1P 5Z9.

OUTLINE AND GENERAL CRITIQUE OF TRDF MODEL

Major Assumptions

The TRDF study made a number of assumptions to obtain a workable model of VOC from available statistical and experimental data, but inaccuracies resulted in the VOC. Independent verifications in Canada and Australia confirmed the inaccuracy of fuel, tire, maintenance and repair, and depreciation predictions of VOC models on the basis of TRDF data.

Elemental Vehicle Operations

The TRDF model assumes that the following four classes of vehicle operation adequately describe all traffic operations relative to road variables in highway investment analyses: (a) running at constant speed under uniform road and traffic conditions on level tangents and on grades, with an adjustment for the effect of pavement condition; (b) changing speed between road sections with different physical road and traffic characteristics; (c) negotiating a horizontal curve; and (d) idling the engine while the vehicle is stopped.

Disaggregation

Aggregate data on vehicle resource consumption can be disaggregated to yield data on representative vehicle types. This information can be further disaggregated by road characteristics and vehicle operating conditions, using proxy methods as described below. Except for fuel consumption, the functional dependencies between consumption of vehicle operating resources and road conditions were judgmental.

Vehicle Type

Seven vehicle classes represent the highway fleet. Each vehicle is representative of its class, and all members of a vehicle population relevant to highway investment appraisal are covered. In reality, there is a variation of vehicle characteristics within each vehicle type, and the borderlines between types may be arbitrary. Actual vehicles in a traffic stream differ in characteristics and utilization and thus in the amount of resources consumed under identical road conditions.

Data Base

The collection of data in the United States was limited to truck operating costs and fuel consumption experiments for the seven vehicle classes. Operating costs for 12,489 trucks were provided by 15 intercity line-haul carriers operating primarily on Interstate highways. Truck ages and mileages were obtained from the Bureau of Census 1977 inventory, which was supplemented with historical vehicle registrations. The operating costs of light vehicles were estimated from pre-1980 data whose vehicle technology, utilization, and cost factors are now obsolete. Fuel consumption was measured on eight vehicles during idle, acceleration, deceleration, and constant-speed driving. Tests for various speeds, grades, and road surfaces were conducted at constant speeds.

The TRDF researchers had the benefit of access to data collected for the study in Brazil from which eventually the World Bank's HDM-III model resulted. Preliminary results of the Brazil project were used to estimate the effect of road roughness on VOC for FHWA.

Model Structure

MicroBENCOST calculates VOC for representative vehicles as a function of road parameters and traffic operating conditions. The operating cost components include fuel, oil, tire, maintenance and repair (MRP), and mileage-related vehicle depreciation. The highway vehicle representatives are three light vehicles, two straight trucks, and two truck combinations. Road conditions are represented by grade, curvature, and pavement surface condition.

Elemental VOC

MicroBENCOST disaggregates the VOC components by the class of vehicle operation: (a) uniform speed costs, (b) pavement condition excess costs, (c) horizontal curve excess costs, (d) speed change cycle excess costs, and (e) idling costs. Table 1 shows which VOC components are calculated for each vehicle operation class. Logically, no tires are consumed in the idling mode, but TRDF did not consider that there is additional oil consumption and depreciation on curves or added fuel consumption on pavement surfaces in poor condition.

Total VOC

For a vehicle type, each VOC component is the product of the resource consumed and the unit resource price. These component costs sum up to a vehicle operation class subtotal cost for that type of vehicle. The subtotal is then multiplied by the amount of elemental operations, which is the segment length for uniform speed, pavement condition, and horizontal curve; the number of cycles for speed change; and the number of idling hours for idling. All elemental subtotals relevant to the planning case add up to the total VOC for that vehicle type. A product of the total VOC and the annual volume of the vehicle type represents the vehicle's total VOC per year. A sum of these VOC per year over all vehicle types is the grand total VOC per year of the highway planning case.

Major Variables

Traffic Speeds

Vehicle speed is the speed of a vehicle running along a section with uniform physical and traffic characteristics. All uniform-speed VOC components are dependent on the speed of vehicle on positive, zero, and negative grades, except that depreciation cost has no models for grades other than zero. Horizontal curve excess costs of fuel, tires, and MRP also depend on vehicle speed in MicroBENCOST.

Whenever a vehicle changes its traveling speed to stop or slowdown or to resume its initial speed, it incurs additional costs. "Begin cycle speed" is the constant travel speed of a vehicle before the change in speed takes place. After the change, the new constant speed is the "end cycle speed." After the speed reduction to the slower end cycle speed, which includes a stop, the vehicle is assumed to return to the begin cycle speed. The full speed change cycle thus goes from the begin speed to the lower end speed and back to the begin speed. These variables appear only in the speed change cycle VOC submodels.

Many speed change profiles cannot be adequately represented with the begin and end cycle speeds. Returning to a

TABLE 1 VOC Components by Vehicle Operation Class

	Uniform Speed	Horizontal Curve	Speed Change	Idling	Pavement Condition
Fuel	yes	yes	yes	yes	no
Oil	yes	no	yes	yes	yes
Tires	yes	yes	yes	no	yes
MRP	yes	yes	yes	yes	yes
Depreciation	yes	no	yes	yes	yes

speed different from the begin cycle speed is usual in congested traffic and urban driving, but this type of speed profile is not covered by the TRDF method.

Road Conditions

Road conditions are represented by Grade −8 to +8 percent, horizontal curvature 1 to 30 degrees, and pavement service index (PSI)—a measure of longitudinal roughness and the independent variable in all pavement condition excess VOC submodels. Such submodel does not exist for fuel consumption in MicroBENCOST.

The TRDF relationships do not contain pavement surface variables other than PSI. A large contribution to vehicle rolling resistance arises from the road surface texture, and the effect of pavement deflection bowl that forms under a heavy wheel could also be substantial on gravel roads, surface treatments, and thin pavements (*4*). The TRDF relationships for heavy vehicles are most likely not appropriate for appraising low-volume roads.

Gross Vehicle Weight

Gross vehicle weight (GVW) includes the mass of the vehicle and its payload. MicroBENCOST uses GVW only for vehicles larger than pickup trucks and vans. For a road investment proposal involving significant changes to grades, actual average GVW on the forehaul and backhaul legs of a trip should be considered. When a higher level of aggregation is required, average round-trip GVW of trucks can be assumed. Average GVW on each leg of a trip, average round-trip GVW, and variations from the averages are not well surveyed. Time and regional and road class variability can be expected. Future predictions depend on factors such as possible changes to truck axle and GVW limits, economic conditions in the country, and truck fleet operations management.

In MicroBENCOST, truck GVW affects fuel consumption on all grades at uniform speeds, excess MRP caused by pavement surface condition, and excess fuel and depreciation caused by idling. In the latter relationships GVW is obviously used as a proxy of a more detailed description of vehicle parameters such as engine size. GVW should also affect tire consumption, but TRDF data do not show this.

Unit Resource Prices

The resources consumed by a vehicle on a road are fuel, oil, tires, maintenance and repair parts and labor, and depreciable value of vehicle. The unit prices of these resources differ between vehicle types.

Uncertainty

TRDF data do not address uncertainty of VOC estimates. Highway investment appraisals would benefit from analysis of uncertainty of the economic indicators calculated from the uncertain cost components. Road condition variables are considered deterministic in the analysis of planning cases using MicroBENCOST. This is not strictly true for aggregated analyses, such as corridor studies, in which gradients and curvatures are estimated. However, these estimates bear much less uncertainty than traffic or operating cost assumptions and can thus be regarded as deterministic. Pavement condition is a random variable, depending on pavement quality, traffic and environmental loading, maintenance effort, and rehabilitation policy, all of which change over time.

The most uncertain inputs into VOC calculations are traffic volumes and mix, typical vehicles, and characteristics of these vehicles. The MicroBENCOST VOC model does not provide any procedures for calculating the variance of total VOC. A vital piece of information for decision making is thus missing.

Transferability

User costs depend on a region's economy, vehicle technology, driver behavior, regulations, and fleet operating decisions (*5*). Since the TRDF VOC data embody the effects of the various conditions in the 1970s, updating these data—as well as transfer to regions with local conditions that are different from the average considered by the TRDF—is problematic. The TRDF fuel consumption values are based on eight vehicles chosen to be typical in 1980. Other VOC components are based on data collected in the 1960s. The TRDF VOC data are not suitable at all to examine road investment policy, program, and project analysis questions arising from current and expected developments affecting operations and VOC of light and heavy vehicles. The transferability limits are discussed by Bein (*4*).

Vehicle types in MicroBENCOST are based on 1970s vehicle fleets analyzed by Zaniewski et al. (*1*) and France (*3*). The vehicle characteristics and utilization encoded into the vehicle types and VOC relationships are thus largely obsolete. Truck combinations larger than 3-S2 are not represented.

MicroBENCOST updated the 1980 unit resource prices (*1*). However, price indexes cannot reflect relative price changes within a VOC component, such as parts and labor in the MRP. It is unreasonable to keep updating an old price over a long period, particularly if drastic changes have taken place in the economic environment.

CRITIQUE OF FUEL CONSUMPTION

The fuel consumption values are deficient and unsuitable for use in VOC models. Fuel consumption tests were carried out by TRDF for idling, acceleration, deceleration, and constant-speed driving. The latter mode received most of the experimental effort, and it tested the effects of speed, grade, surface type, and pavement condition. Measured values were transferred into sets of fuel consumption tables. No tests were carried out for large truck combinations, and results for a 3-S2 unit were assumed. The effect of curves was derived by comparing horsepower needed to traverse a curve at a constant speed with the horsepower required to climb a grade at the same speed, for which fuel consumption was measured.

Reliability of Measured Values

The fuel consumption values are based on only one test vehicle in each class, except that two identical vehicles were used for a medium car. If any of the single test vehicles had unusual fuel consumption characteristics, all future estimates for that class would be affected. For example, the two large TRDF trucks cut off fuel during engine motoring on steep grades and hard decelerations, and all trucks in those classes are assumed to behave this way.

The measured values were smoothed, but some inconsistencies remain. For example, the five-axle diesel truck traveling at 80.5 km/hr consumes less fuel on a 5.6 percent grade than on a 2.6 percent grade, and the constant speed fuel consumption rate of a four-axle truck is less at 113 km/hr than at 72 km/hr on zero grade.

The measured values tended to be lower than expected when fuel flow rates were high. Fuel consumption is difficult to measure for diesel engines because of a fuel return loop that takes unused fuel from the engine back to the fuel tank. It appears that this problem had not been overcome, but it should have been identified when the data were collected. This problem raises the question of how well other factors, such as wind, tire pressure, and engine temperature, were controlled during the tests.

Inconsistencies

Clayton (*6*) gives expressions for calculating on-road fuel consumption rates for a given GVW for each season on the basis of a large amount of data collected for trucks across Canada. The data apply to nonurban travel at about 90 km/hr on high-class, paved rural and intercity highways. The fuel rates would be slightly lower in the United States because of the milder winters. Table 2 compares TRDF fuel consumption values with average summer values for those trucks calculated using Clayton's expression. The TRDF estimates include the effect of speed fluctuation, but not the effect of the number of stops, which are few in intercity travel and will have little effect on fuel consumption. TRDF values are given for two running speeds and are 40 to 60 percent greater than typical on-road values for Canadian trucks.

For uniform speeds, observed Canadian vehicle fuel consumption is only about half of that predicted with the TRDF relationships in the HUBAM model for four vehicle types examined (*7*).

Excess Fuel Consumption per Stop or Slowdown

The method used in the TRDF study to calculate acceleration and deceleration fuel consumption and excess fuel per stop gives very poor estimates. The test vehicle was accelerated (decelerated) from rest (maximum cruise speed) to maximum cruise speed (rest), and the fuel flow was measured every 2 sec. A very simple linear function was fitted to the cumulative fuel flow during the maneuver. The fuel consumption for accelerating (decelerating) from any Speed v1 to a Speed v2 was found by subtracting the cumulative fuel flow at Speed v1 from the cumulative fuel flow at v2. Excess fuel consumption per slowdown or stop was then found in the usual way from the acceleration, deceleration, and cruise fuel consumption. The estimates are poor for two reasons:

1. The acceleration and deceleration profile and rate depend on the initial and final speeds. For example, for small decelerations at high speed the brake is often not used and thus there is little fuel penalty. During a deceleration from 110 km/hr to rest, the brake will certainly be applied and the estimated fuel penalty for that slowdown using the TRDF approach will be much greater.
2. The fuel that flows through a flow meter in a 2-sec period when the vehicle is accelerating or decelerating is not a good indication of the actual fuel being burnt in the engine, especially for diesel engines. For an accurate measurement, the vehicle must be in a stable operating condition, preferably idling, for at least 5 sec before and after a maneuver.

TABLE 2 Comparison of TRDF Truck Fuel Consumption Estimates with On-Road Values for Trucks in Canada

Variable	4 axle truck		5 axle truck	
Total vehicle mass	18100	kg	28300	kg
Observed on-road fuel rate	401	ml/km	451	ml/km
From TRDF model *:				
1. Running speed 88.5 km/h				
Constant speed fuel	395	ml/km	475	ml/km
Average speed change	16	km/h	16	km/h
Number of speed changes	1.4	cycle/km	1.4	cycle/km
Excess fuel per change	149	ml/cycle	178	ml/cycle
Overall fuel rate	604	ml/km	724	ml/km
1. Running speed 80.5 km/h				
Constant speed fuel	398	ml/km	468	ml/km
Average speed change	17	km/h	17	km/h
Number of speed changes	1.3	cycle/km	1.3	cycle/km
Excess fuel per change	136	ml/cycle	170	ml/cycle
Overall fuel rate	574	ml/km	689	ml/km

* *Assuming multi-lane road (fuel values are greater assuming a two-lane road and slightly less assuming a freeway). Ambient temperature 26°C (Celsius).*

Effect of Road Roughness

Estimates of the effect of road roughness on vehicle fuel consumption range from 0 to 30 percent (*1,8*). TRDF could find no significant relationship between fuel consumption and road roughness and has assumed zero effect. Road roughness increases rolling resistance and must therefore increase fuel consumption. Vehicle oscillations from a rough ride produce a more turbulent flow of air, and consequently a higher aerodynamic drag, compared with a smooth ride on a smooth road. The road user costs study in Brazil (*9*), from which all nonfuel costs related to roughness were adopted in the TRDF model, found a significant relationship between rolling resistance and roughness [see the work by Biggs (*10*) also for analysis of these data]. The TRDF data were analyzed to determine the proportional increase in rolling resistance caused by roughness, and by combining data into a smaller number of classes, a significant relationship was found (*10*).

Only about a third of fuel consumption at cruise speeds is caused by rolling resistance. Because fuel consumption is difficult to measure accurately, only large changes in rolling resistance can be identified by measuring fuel consumption. On the basis of the available data, the estimated effect of an increase in roughness is given in Table 3. The effect is small, but significant, and should not be ignored. Curiously, TRDF has not studied the smaller effect of road curvature but has included it into the model.

TABLE 3 Effect of Roughness on Rolling Resistance and Fuel Consumption

PSI		Change in	
Initial	Final	Resistance (%)	Fuel Consumption (%)
4	3	5	2
4	2	16	5

Comparison of TRDF and ARFCOM Fuel Estimates

Such problems with the TRDF fuel values are examined using ARFCOM, a detailed mechanistic model of vehicle fuel consumption, developed by the Australian Road Research Board (*11,12*). ARFCOM vehicle parameters can be easily changed to allow for technological improvements and various vehicle classes, fleet composition, and operating conditions. ARFCOM estimates have been checked over a wide range of vehicles to ensure that the effect of changes in vehicle parameters is reflected correctly. ARFCOM has been extensively tested on a wide range of fuel consumption data for cars and trucks in the United States (*13,14*), Australia (*15–17*), the United Kingdom (*18–20*), Canada (*6*), and the World Bank Study in Brazil (*9*). ARFCOM estimates compared well with measured values from these studies. Biggs (*21*) compared observed heavy vehicle fuel consumption in Canada with ARFCOM estimates. ARFCOM produced 2 percent error overall compared with a 50 percent error for TRDF estimates. ARFCOM also reliably estimated the effect of seasonal temperature variations on fuel consumption and the effectiveness of fuel conservation measures.

In ARFCOM acceleration and deceleration profiles and rates depend on both the initial and final speeds, as does the contribution of the change in kinetic energy to fuel consumption. The calculation of acceleration fuel consumption is based on sound theoretical principles, and estimates have been calibrated and validated using data collected over complete acceleration and deceleration maneuvers.

Table 4 compares TRDF and ARFCOM estimates for a range of constant speeds. ARFCOM estimates were made for the vehicles used in the TRDF study (Set 1) and for typical vehicle classes in the United States today (Set 2). The TRDF constant-speed fuel consumption rates are reasonably good for cars but not good for all the truck classes. In particular, fuel consumption does not increase fast enough as speed increases. The TRDF estimates for the three- and five-axle trucks show very little change with speed increases from 72 to 97 km/hr.

TRDF data grossly overestimate the effect of speed changes (Table 5). Low values of excess fuel consumption caused by a slowdown can occur because aerodynamic drag is significantly reduced and much of the change in kinetic energy during a slowdown is used to overcome rolling and aerodynamic drag. This is often the case for minor slowdowns and for vehicles with a high proportion of fuel consumption caused by aerodynamic drag (e.g., two-axle truck at high speeds).

TABLE 4 Comparison of Constant Speed Fuel Consumption Estimates

Vehicle	Source *	Constant Speed (km/h) 56.3	64.4	72.4	80.5	88.5	96.6
Medium car	TRDF	88	89	95	101	113	124
	ARFCOM-1	85	91	98	106	115	137
	ARFCOM-2	81	85	91	99	107	128
3-axle truck	TRDF	360	350	351	351	360	367
	ARFCOM-1	218	238	260	285	312	376
	ARFCOM-2	195	214	236	260	287	349
5-axle truck	TRDF	475	473	468	468	475	487
	ARFCOM-1	426	450	415	454	498	602
	ARFCOM-2	381	404	372	411	453	554

**ARFCOM-1: estimates for vehicles used in TRDF study.*
ARFCOM-2: estimates for current typical vehicles in the U.S.
Values given in milliliters per kilometer.

TABLE 5 Effect of Speed Change From 88.5 km/hr to a Minimum Speed on Fuel Consumption

Vehicle	Source	Minimum Speed (km/h) 0.0	16.1	32.2	48.3	64.4	72.4
Medium car	TRDF	68	60	52	42	27	18
	ARFCOM-1	33	28	23	19	15	12
	ARFCOM-2	27	23	19	16	13	11
3-axle truck	TRDF	469	435	390	326	224	156
	ARFCOM-1	188	138	86	43	19	18
	ARFCOM-2	201	153	99	52	22	16
5-axle truck	TRDF	636	571	492	386	258	178
	ARFCOM-1	351	254	154	71	31	59
	ARFCOM-2	376	289	185	95	39	35

**ARFCOM-1: estimates for vehicles used in TRDF study.*
ARFCOM-2: estimates for current typical vehicles in the U.S.
Values given in milliliters per kilometer.

The large errors in the TRDF estimates of the effect of speed changes could significantly overestimate the benefits of some types of road improvements and could unjustifiably favor road improvements involving reductions in speed changes.

CRITIQUE OF OTHER COMPONENTS

Figure 1, which is based on previous work (*7*), compares observed VOC with predictions by HUBAM, which uses TRDF VOC data. The observations refer to a new, urban-driven car selected by the Canadian Automobile Association (CAA). The equivalent vehicle in HUBAM is operating in free-flow traffic; consequently one would expect that it has lower fuel consumption and lower total VOC than the CAA car. The Canadian trucking industry operating cost data from Trimac Consulting Services Ltd. for a five-axle truck confirm the gross inaccuracy of TRDF fuel consumption predictions in HUBAM. It is also clear that the other VOC components are inaccurate. HUBAM maintenance cost is too low because newer cars, such as that of CAA, have lower maintenance costs than a fleet average. The industry data presented include total depreciation costs as a result of both use and passage of time, whereas HUBAM accounts only for depreciation from use. The following are some of the reasons for the discrepancies in the other cost components.

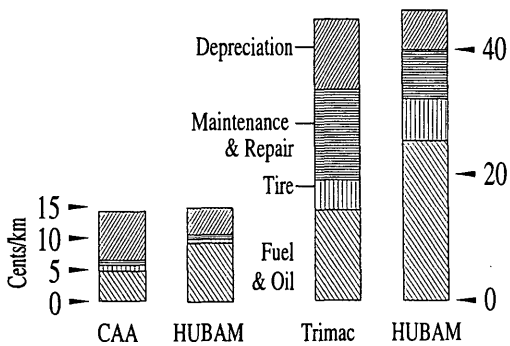


FIGURE 1 Comparison of observed VOC and HUBAM predictions. *Left*, medium car; *right*, five-axle truck.

Oil Consumption

Oil consumption is a minor VOC item and elaborate models are not warranted. Oil consumption on grades was adjusted by the ratio of the horsepower required on the grade to the horsepower required for the same speed on a level tangent section. No correction was made for oil consumption on curves. Effects of pavement roughness were adopted from the study in Brazil for lack of U.S. data. It is not clear whether the TRDF data and MicroBENCOST account for the cost of labor necessary to change engine oil.

Tire Consumption

Tire consumption is a small VOC item. MicroBENCOST expresses the consumption of a set of tires installed on all wheels of a vehicle as an equivalent percent of wear of a single tire. Tire wear was estimated by TRDF with a model (*22*), for which coefficients were selected by comparison of results with Winfrey's values corrected for greater tire cost and longer tread life. Brazilian relationships were used to determine tire cost adjustment factors for surface roughness. The U.S. and Brazilian data used reflect bias-ply tire technology, now obsolete for all vehicle types. Coefficients for the tire wear model were based on highly variable data and representative of asphalt concrete surfaces only.

Maintenance and Repair

The consumption of materials and labor necessary to maintain and repair a vehicle type is expressed as a percentage of an average MRP cost of that vehicle type. Percentage MRP costs were estimated by category (general maintenance, brakes, drive train) for light vehicles. The MRP cost categories were allocated to other trucks using trucker survey data for 3-S2. The distributed costs were then employed to calculate correction factors to net costs of brakes at constant speed on level tangents. For acceleration, grades, and curves, excess costs were calculated from a regression between horsepower and constant-speed costs. Adjustment of MRP costs for surface roughness was made using the Brazilian data. To distribute the brake cost between deceleration and holding constant speed on negative grades, it was converted to a cost per unit of work using rather limited data to calculate the cost/work coefficient. This coefficient was then multiplied by the brake work per distance in deceleration and on negative grades to obtain the excess maintenance cost of brakes.

The questions arising from the MRP component are as numerous as those from the fuel consumption component. Is the distribution of MRP cost categories in the 1970s fleet data still valid? Can MRP data for 3-S2 trucks be extrapolated to other types of trucks? Are correction factors to MRP costs and regression between horsepower and constant-speed MRP costs defendable approaches? Is the approach to brake cost analysis acceptable? Because the cost of MRP covered under factory warranty is included in the new vehicle price, is it accounted for properly? Are trailer MRP costs handled properly? The Brazilian data on surface roughness effects on MRP are likely not applicable to North America because operators adjust vehicle technology and utilization policies in response to economic and road conditions. A mechanistic approach, such as that seen previously (*23*), relating consumption and wear rates to road and traffic conditions through the dynamic forces acting on a vehicle, would be more suitable.

Mileage-Related Depreciation

Mileage-related vehicle depreciation cost is expressed as a percent of a depreciable value. The mileage-related depreciation was estimated with a survivor curve method. The use of the highest 3-percentile class of annual mileage in conjunction with the survivor curve for the entire vehicle type to determine average extreme annual mileage seems arbitrary. Given that North American trucks go through a number of life stages with different uses, the survivor curve of the entire fleet cannot possibly be a good base to estimate mileage-related depreciation. A simpler method would probably yield as good or better results.

The age and accumulated mileage of vehicles were compiled from the 1977 census, and the number of registrations corresponding to the census data was obtained from 1945–1977 statistics. These data are not representative of newer technology and use of vehicles. The estimates were updated using relative adjustment factors for the range of operating speeds produced, but are the assumed distributions of vehicle depreciation costs to speeds, speed changes, and idling reasonable? Depreciation expenses were not distributed to grades and horizontal curves, but the excess time consumed in speed changes relative to constant speed was considered in the updating.

Brazil data were used to adjust the updated estimates of depreciation for different roughness conditions. The data cannot possibly reflect accurately the effect of pavement roughness on depreciation of vehicles. Operators adjust vehicle maintenance and scrapping policies in response to economic and road conditions, which are quite different between the two countries.

Strictly speaking, the depreciable value of a vehicle should be reduced by that portion of new vehicle price that is added by manufacturers to cover the cost of factory warranties. Trailers undergo uses that are different from those for truck tractors or straight truck units, but this fact is not accounted for in depreciation cost estimations. Given that vehicle depreciation and MRP costs are interdependent, the estimation methods for the two components of operating costs are deficient in MicroBENCOST. A better approach, the optimal life method, was recently implemented in South Africa (*24*).

CONCLUSIONS

The TRDF model of VOC and any aggregated relationships derived from the data, such as those in HERS, as well as updates incorporated into MicroBENCOST—although the models of choice in the U.S. highway policy, planning, and project evaluation—all have a number of deficiencies. The deficiencies arise mainly because the model has a statistical rather than mechanistic or other causal foundation. The data encode highway, vehicle technology, and operating, and economic conditions typical of the 1970s, which are not adequate to examine questions arising today in highway transportation planning. The judgmental manipulation of the data base by TRDF has introduced further problems.

A representative of a modern heavy truck is missing in the TRDF model. The typical vehicles are fixed and cannot be altered by the user by changing vehicle characteristics and utilization parameters—in contrast to the mechanistic models.

Without exception, all VOC components are inaccurate at least for one vehicle operating class: running at uniform speed. Fuel consumption has been proven erroneous for all operating classes. The speed change VOC estimation method does not address the real conditions of impeded traffic flows in congestion and urban driving. A similar conclusion was reached independently by the HERS study (*25*).

Of the vital highway decision variables, longitudinal roughness is excluded from fuel consumption relationships, whereas pavement type and surface texture do not appear at all in the TRDF model. Surface texture alone has at least as great an effect on VOC as does roughness.

VOC are the major user cost in maintenance, rehabilitation and upgrading evaluations. A large part of the transportation improvement budgets in the United States is directed toward solving urban traffic and highway congestion problems. The lack of a better VOC model in the United States to serve these needs must be remedied. The HERS study reached a similar conclusion.

A mechanistically based substitute for the TRDF model should be developed as soon as possible. The development work should draw on the best elements from HDM-III (*9*), VETO (*23*), and ARFCOM (*11*). It should also implement the optimal life method for calculating vehicle depreciation.

ACKNOWLEDGMENTS

Analysis presented in this paper comes from the author's research sponsored by the Saskatchewan Highways and Transportation (1987–1988), the former Roads and Transportation Association of Canada (1988–1989), the Canadian National Highway Policy Study Committee (1989), and the British Columbia Ministry of Transportation and Highways (1990 to the present).

REFERENCES

1. J. P. Zaniewski, B. C. Butler Jr., G. Cunningham, G. E. Elkins, M. Paggi, and R. Machemehl. *Vehicle Operating Costs, Fuel Consumption and Pavement Type and Condition Factors.* PB82-238676. Texas Research Foundation; FHWA, U.S. Department of Transportation, 1982.
2. G. E. Elkins et al. *Estimating Vehicle Performance Measures.* FHWA, U.S. Department of Transportation, July 1987.
3. C. France. *Fuel Economy of Heavy Duty Vehicles.* Environmental Protection Agency, Ann Arbor, Mich., Sept. 1976.
4. P. Bein. Evaluation of VOC Models. *Road and Transport Research*, Australian Road Research Board, Ltd., Vermont South (submitted) (1993).
5. A. Chesher and R. Harrison. *Vehicle Operating Costs: Evidence From Developing Countries.* Highway Design and Maintenance Standards Series, Johns Hopkins University Press, Baltimore and London, 1987.
6. A. Clayton. *Heavy Vehicle Fuel Consumption in Canada.* Transport Canada Report TP5781E. Energy Planning Directorate, Transport Canada, 1984.
7. P. Bein, J. B. Cox, and N. D. Lea. *Adapting the Brazil/UNDP Cost Research to Canadian Conditions.* Final Report. Roads and Transportation Association of Canada, Ottawa, 1989.
8. F. R. Ross. *The Effect of Pavement Rougnness on Vehicle Fuel Consumption.* Research Unit, Materials Section, Division of Highway and Transportation Facilities, Wisconsin Department of Transportation, Madison, 1981.
9. T. Watanatada, K. Tsunokawa, W. D. O. Paterson, A. Bhandari, and C. G. Harral. *The Highway Design and Maintenance Standards Model (HDM-III).* Highway Design and Maintenance Standards Series, Johns Hopkins University Press, Baltimore and London, 1987.
10. D. C. Biggs. *Estimating Fuel Consumption of Light to Heavy Vehicles.* Internal Report AIR 454-1. Australian Road Research Board, Victoria, 1987, pp. 31–34.
11. D. C. Biggs. *Vehicle Fuel Consumption Estimation Program, ARFCOM—User Guide.* Technical Manual ATM 28. Australian Road Research Board, Victoria, 1988.
12. D. C. Biggs. *ARFCOM—Models for Estimating Light to Heavy Vehicle Fuel Consumption.* Research Report ARR 152. Australian Road Research Board, Victoria, 1988.
13. R. E. Knight. *Correlation of Truck Tire Rolling Resistance as Derived from Fuel Economy and Laboratory Tests.* SAE Technical Paper 821266. Society of Automotive Engineers, Warrendale, Pa., 1982.
14. D. J. Schuring and J. S. Redfield. *Effect of Tire Rolling Loss on Fuel Consumption of Trucks.* SAE Technical Paper 821267. Society of Automotive Engineers, Warrendale, Pa., 1982.
15. E. Khatib. *TABESAM—Truck and Bus Energy Simulation Model.* Report T85/68. Department of Mechanical and Industrial Engineering, University of Melbourne, Australia, 1986.
16. E. Khatib. *Development and Verification of Heavy Vehicle Fuel Consumption Test Procedures.* SAE-Australia, Parkville, Victoria, 1987.
17. D. C. Biggs and G. J. Murphy. *Collection of Fuel Consumption Data at Isolated Intersections.* Internal Report AIR 428-1. Australian Road Research Board, Victoria, 1986.
18. M. A. Renouf. *Prediction of the Fuel Consumption of Heavy Goods Vehicles by Computer Simulation.* Report SR 453. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
19. M. A. Renouf. *An Analysis of the Fuel Consumption of Commercial Vehicles by Computer Simulation.* Report LR 973. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
20. I. C. P. Simmons. *Fuel Consumption of Commercial Vehicles: Instrumentation and Analysis of Results.* Report SR 508. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
21. D. C. Briggs. Comparison of Observed Heavy Vehicle Fuel Consumption in Canada with Estimates Produced by ARFCOM. *Australian Road Research*, Vol. 20, No. 3, 1990, pp. 11–16.
22. W. Barreire, T. E. Smith, and H. C. Hodges. *Transportation Cost Study: Tire Wear Test.* Forest Service, U.S. Department of Agriculture, 1974.
23. U. Hammarström and B. Karlsson. VETO—A Computer Program for Calculation of Transport Costs as a Function of Road Standards (in Swedish with English summary and figure captions). *VTI Meddelande 501.* Swedish Road and Traffic Research Institute, Linköping, 1987.
24. H. W. du Plessis and I. C. Schutte. *Road Roughness Effects on Vehicle Operating Costs: Southern Africa Relations for Use in Economic Analyses and in Road Management Systems.* Project Report 88/070/3. Division of Roads and Transport Technology, CSIR, Pretoria, South Africa, March 1991.
25. Jack Faucet Associates. *The Highway Economic Requirements System.* Technical Report. Highway Needs and Investment Branch, FHWA, U.S. Department of Transportation, July 1991.

The authors are solely responsible for the opinions expressed in the paper.

Publication of this paper sponsored by Committee on Transportation Economics.

Understanding the Competing Short-Run Objectives of Peak Period Road Pricing

DAVID BERNSTEIN AND JÉRÔME MULLER

The interest in peak period road pricing has grown considerably in recent years both in the United States and abroad. This increase in interest is usually attributed to worsening congestion and improved electronic toll collection technologies. However, there may be a third reason as well: peak period pricing can be used to generate revenues. This use of peak period road pricing is explored and compared with programs that are designed to minimize social cost. Using some simple examples, it is shown that it is possible to increase toll revenues but at a significant cost to society. In addition, it is shown that most of the revenues and costs can be attributed to the length and end of the toll period.

Peak period pricing has been an accepted part of life in the United States for many years. For example, most long distance telephone companies charge higher prices during the day than they do during the evening and night. In addition, many public transit systems (e.g., the Washington, D.C., Metro) also charge higher prices during the peak period. However, in spite of the urgings of many economists, peak period pricing is not yet in widespread use on U.S. highways.

There are, of course, many reasons for this. In the United States, some of these reasons became evident during the 1970s when UMTA, with the help of the Urban Institute, offered to assist several cities in establishing programs that would demonstrate that peak period pricing could be used to bring congestion levels down to the "socially optimal" level (which is a specific type of peak period pricing that is usually referred to as "congestion pricing"). The response was considerably less than expected (*1,2*). Although three cities did agree to further study (Madison, Berkeley, and Honolulu), all of the preliminary studies ended without requests for further funding because of public opposition. Congestion pricing was perceived as being unfair, discriminatory, regressive, coercive, and antibusiness.

Yet, in spite of these past failures, the United States and other countries are again beginning to consider peak period road pricing. In the United States, the recent Intermodal Surface Transportation Efficiency Act (ISTEA) has allocated funding for up to five pilot congestion pricing programs. In the rest of the world, several such programs are now either under way or in the planning stages. For example, Singapore has had a program in place since 1975 (*3–6*); Bergen, Norway, has had a program in place since 1986; and Hong Kong tested a program from 1983 to 1985 and is now considering a full-scale implementation (*7–12*). The Netherlands had planned on having a full-scale program in place by 1996 and is now considering a somewhat scaled-down program instead (*13*); Cambridge, England, intends to initiate a trial program by 1993 or 1994 and a full-scale program by 1997 (*14*); and Oslo and Trondheim, Norway, and Stockholm and Gothenberg, Sweden, are all considering programs of one kind or another.

Several reasons are usually given for this renewed interest. First, congestion is now much worse than it has ever been in the past. As two-income families have become more prevalent and suburban rings have grown in both absolute population and area, the number of commuters using private automobiles in the United States has increased dramatically. In fact, from 1960 to 1980 the number of people driving to work nearly doubled (at a time when the number of people in the work force increased by only 50 percent). In addition, there has been a continued increase in the amount of truck transport, and hence in highway truck miles, both inside and outside of urban areas. The result is that more than 55 percent of urban freeway travel during the peak period takes place during congested conditions (*15*) and that more than 11 percent of the total vehicle miles of travel takes place during recurring congestion (*16*). Hence, although congestion pricing has been unpopular in the past, it is needed more than ever before. Second, toll collection technology [see for example, Bernstein and Kanaan (*17*)] has now advanced to the point where congestion pricing can be implemented.

However, in our casual conversations with policy makers who are now considering peak period pricing, it has become apparent that there is another reason for this increase in interest. Many of them seem to view peak period pricing as a mechanism for increasing revenues (which should not, strictly speaking, be referred to as congestion pricing). In addition, there may also be increased support for road pricing simply as a means of reclaiming road space for pedestrians [see Goodwin (*18*)].

Not surprisingly, the existing literature does not consider these aspects of peak period road pricing. Instead, it focuses on the inefficiencies that are inherent in (high-volume) roadway travel (*19–28*). Thus, given the possibility that peak period road pricing may be used as a revenue generation mechanism, it is important to consider the impacts of such policies. This paper represents a first step in such an investigation.

We begin by describing the specific setting we consider throughout the remainder of the paper and by discussing the various competing objectives of peak period road pricing. We then consider the impacts of pursuing these objectives, first within the context of tolls that are in place throughout the entire peak period and then within the context of time-varying tolls. We conclude with a discussion of future avenues that still need to be pursued. We should point out in advance that we do not evaluate any specific road pricing programs in this paper. Studies of this kind can be found elsewhere (*29–32*).

Massachusetts Institute of Technology, Cambridge, Mass. 02139.

COMPETING OBJECTIVES

Road pricing can have several effects on traveler behavior in the short run. In particular, it can result in changes in route, changes in departure time, changes in mode, changes in the total number of trips taken, and changes in the origin and destinations of those trips. As a result, it can be used to achieve a variety of different ends. (It is for this reason that we generally often use the term peak period road pricing rather than congestion pricing in this paper. We use the latter term only to describe programs that are designed to increase social welfare.) For example, one can imagine policy makers making well-reasoned arguments about why road pricing should be used to

- Reduce social costs (time, money, etc.),
- Reduce out-of-pocket transportation costs,
- Reduce travel time,
- Reduce the number of vehicle miles,
- Increase toll revenues,
- Increase transit revenues, or
- Increase total (toll and transit) revenues.

Furthermore, these objectives could be either general (e.g., reduce the total number of vehicle miles or reduce out-of-pocket transportation costs) or very specific (e.g., reduce the total number of vehicle-miles on the Gotham Expressway or reduce out-of-pocket transportation costs for people earning less than $20,000/year).

In general, the various objectives of road pricing need not conflict, yet in some cases they may be entirely contradictory. For example, it may be possible to both increase toll revenues and reduce social costs. On the other hand, it may not be possible to both reduce out-of-pocket transportation costs and increase total revenues. In either case, it is important that we have some way of comparing these different objectives so that policy makers can make informed decisions.

For the purposes of this paper, we will use a relatively simple technique to conduct the comparison. In particular, we will consider only two different types of policies, one aimed at minimizing social cost (which in the examples that follow is simply the total travel cost minus the toll revenue) and the other aimed at maximizing toll revenues (given certain constraints). For each of these policies we will calculate and compare a variety of different impact measures using some simple examples.

BASIC MODEL: SRD EQUILIBRIUM

We will work with a model of commuter behavior on a simplified transportation network composed of a set of nonoverlapping (i.e., with no arcs in common) paths (P) and a single origin-destination pair (O-D pair). Commuters traveling between this single O-D pair choose both a path (or route), $p \in P$, and a departure time, $t \in [0, T]$.

Although we discuss commuters as if they are discrete entities, we actually work with departure rates. In particular, for each path ($p \in P$) the departure rate on p at time t is denoted by $r_p(t)$, and each possible pattern of departures is described by the vectors $r(t) = (r_p(t) : p \in P)$ and $r = (r(t) : t \in [0, T])$. The total number of commuters between the single O-D pair is denoted by N. Thus, the set of feasible path flow vectors is given by

$$H = \left\{ r : \sum_{p \in P} \int_0^T r_p(t)\,dt = N \right\} \tag{1}$$

The cost on each path (C_p) is a function of the total travel time on the path and the arrival time at the destination (plus any tolls). We assume that the time needed to traverse path p when entered at time t can be modeled as a deterministic queueing process in which

$$D_p(t) = d_p + \frac{1}{s_p} x_p(t + d_p) \tag{2}$$

where

d_p = (fixed) travel time on path p,
s_p = service rate of the queue on path p, and
$x_p(t + d_p)$ = number of vehicles in the queue on path p for a vehicle that enters the path at time t (and hence reaches the queue at time $t + d_p$).

Further, because travelers may arrive early or late, we introduce an asymmetric schedule cost given by

$$\Phi_p(t) = \begin{cases} \beta[t^* - (t + D_p(t))] & \text{if } [t + D_p(t)] < t^* \\ 0 & \text{if } [t + D_p(t)] = t^* \\ \gamma[(t + D_p(t)) - t^*] & \text{if } [t + D_p(t)] > t^* \end{cases} \tag{3}$$

where

t^* = desired arrival time,
β = dollar penalty associated with early arrival, and
γ = dollar penalty associated with late arrival.

Note that we do not include a window of "equally acceptable" arrival times to simplify the analysis that follows. Also note that we will often use the notation F_p to denote the free-flow travel cost on path p (i.e., $F_p = \alpha d_p$).

Using the preceding definitions, we may now define the generalized travel cost as follows:

$$C_p(t) = \alpha D_p(t) + \Phi_p(t) + \tau_p(t) \qquad \forall\, p \in P \tag{4}$$

where α is the value of travel time and $\tau_p(t)$ is the toll on path p at time t (if any). Further, by appropriately defining the min operator [see Friesz et al. (*33*) for details], and letting

$$\mu_p(r) = \min\{C_p(t) : t \in [0, T]\} \tag{5}$$

and

$$\mu(r) = \min\{\mu_p(r) : p \in P\} \tag{6}$$

we can define an equilibrium as follows: A departure rate pattern $r \in H$ is a simultaneous route and departure-time choice equilibrium (SRD equilibrium) if and only if (iff) r satisfies the following condition for all $p \in P_w$, and t:

$$r_p(t) > 0 \Rightarrow c_p(t) = \mu_w(r) \tag{7}$$

That is, we assume that the only flow patterns that can persist are those in which all used path and departure time choices have minimum cost.

Although it can be quite difficult to solve for SRD equilibria [see Bernstein et al. (34)], for this simple model earlier works (23,35) present an analytic solution. For the case when $\tau(t) = 0$ (i.e., the no-toll equilibrium), they demonstrate that the equilibrium can be characterized as follows:

$$t_q = t^* - \left(\frac{\gamma}{\beta + \gamma}\right)\frac{N}{S} - d \tag{8}$$

$$t_{q'} = t^* + \left(\frac{\beta}{\beta + \gamma}\right)\frac{N}{S} - d \tag{9}$$

$$t_q = t^* - \left(\frac{\beta\gamma}{\alpha(\beta + \gamma)}\right)\frac{N}{S} - d \tag{10}$$

$$r(t) = \begin{cases} s + \dfrac{\beta_s}{\alpha - \beta} & \text{if } t \in [t_q, \tilde{t}) \\ -\dfrac{\gamma_s}{\alpha + \gamma} & \text{if } t \in [\tilde{t}, t_{q'}] \end{cases} \tag{11}$$

$$C = \left(\frac{\beta\gamma}{\beta + \gamma}\right)\frac{N}{s} \tag{12}$$

where

t_q = time that the queue begins to form,
$t_{q'}$ = time the queue falls to zero, and
$\tilde{t}$ = departure time that leads to an arrival at t^*.

Patn subscripts were omitted for the sake of clarity.

COMPARING OBJECTIVES: PERMANENT TOLLS

With this as background we now turn to the question at hand. That is, we now consider the differential impacts of the socially optimal and revenue maximizing tolls. We begin with a very simple case for which we need only consider the effects of the revenue maximizing toll. This enables us to easily illustrate the magnitude of these effects. Specifically, in this section we consider the case in which there is a toll in place throughout the entire period (i.e., a permanent toll), and we assume that this toll does not vary with the flow level [for a discussion of the use of step tolls see Bernstein and Smith (36) and for the case of tolls that vary continuously with flows see Dafermos (26)]. We also limit ourselves to a simple example in which commuters traveling between a single origin and a single destination choose both their departure time and whether to use the single highway (which can be tolled) or the single local road (which cannot be tolled). Finally, we assume that the system reaches an equilibrium in which both paths are used.

Using the results above and letting $\eta_p = \gamma/[s_p(\beta + \gamma)]$, we can write the equilibrium cost (as a function of the number of users) on the highway as

$$C_h(N_h) = F_h + \eta_h N_h \tag{13}$$

and for the local (free) road

$$C_l(N_l) = F_l + \eta_l N_l \tag{14}$$

where

N_h = number of users on the highway,
N_l = number of users on the local road, and
$N = N_h + N_l$.

Thus, in this case, the equilibrium condition is given by

$$\underbrace{F_h + \eta_h N_h + \tau}_{C_h} = \underbrace{F_l + \eta_l(N - N_h)}_{C_l} \tag{15}$$

which implies that $N_h = (F_l - F_h - \tau + \eta_l N)/(\eta_h + \eta_l)$.

Because we know that the socially optimal toll is zero in this case, we can turn directly to determining the revenue maximizing. The toll revenue maximization problem is given by

$$\max_{0 \le \tau \le (F_l - F_h + \eta_l N)} N_h \cdot \tau \tag{16}$$

Substituting in for the equilibrium value of N_h yields the following equivalent problem:

$$\max_{0 \le \tau \le (F_l - F_h + \eta_l N)} \frac{F_l - F_h - \tau + \eta_l N}{\eta_h + \eta_l} \cdot \tau \tag{17}$$

Hence, the revenue maximizing toll is given by

$$\tau_{tr} = \frac{(F_l - F_h + \eta_l N)}{2} \tag{18}$$

and

$$N_h^{tr} = \frac{(F_l + F_h + \eta_l N)}{2\eta_h + 2\eta_l} \tag{19}$$

with the toll revenues given by $R_{tr} = \tau_{tr} N_h^{tr}$.

We can now gain some insight into the effects of the revenue maximizing toll by considering a numerical example. Using the values of α, β, and γ estimated elsewhere (37) (i.e., $\alpha = 6.40$, $\beta = 3.90$, and $\gamma = 15.21$), and setting $N = 5{,}000$, $s_h = 6{,}000$, $s_l = 4{,}000$, $d_h = 0.333$, and $d_l = 0.50$, we find that the equilibrium cost when there is no toll is \$4.11 with 3,825 commuters using the highway. With the revenue maximizing toll of \$2.47 in place, the equilibrium cost increases to \$5.56 with only 1,913 commuters using the highway. This means that the revenue maximizing toll results in a 35 percent increase in commuting cost (assuming that toll revenues are not redistributed) and toll revenues of \$4,725. Each dollar increase in toll revenues therefore costs society \$1.53. Hence, although toll agencies may be tempted to adopt the revenue maximizing toll, such a policy has a very high social cost.

COMPARING OBJECTIVES: TIME-VARYING TOLLS

Given the significant effects of the revenue maximizing toll, it is important to take a closer look at what causes those effects. To do so, we now consider an example with a single path (i.e., the highway from above) and explore the influence of the "timing" of the toll.

To remove any toll mechanisms from consideration that we think are socially, technically, or politically unacceptable, we place several restrictions on these tolls. First, we consider only a step toll (i.e., there is a peak toll and an off-peak toll), the value of which is τ within the toll period $[t^+, t^-]$ and zero outside. We further restrict our investigations to toll schemes for which some commuters exit the bottleneck before the toll period (in equilibrium). Finally, we will require that $t^+ < t^* < t^-$ (i.e., the period during which the step toll is levied) is the period when congestion is at its worst under the no-toll situation and that the bottleneck remains used at capacity throughout the "rush" period (i.e., during the interval $[t_q, t_{q'}]$. (Note that, in equilibrium, a large number of commuters arrive at the bottleneck some time before the step toll is lifted. We will refer to these commuters as the "bulk" group.)

Socially Optimal Step Toll

The socially optimal step toll has been discussed at length elsewhere (*35*). Hence, we will present only a summary of their results. In particular, the socially optimal step toll can be characterized as follows:

$$\tau_{sc} = \frac{\beta\gamma}{\beta + \gamma}\frac{N_h}{2s_h} \quad (20)$$

$$t_q = t^* - \frac{\gamma}{\beta + \gamma}\frac{N_h}{s_h} + \frac{\gamma - \alpha}{(\beta + \gamma)(\alpha + \gamma)}\tau_{sc} \quad (21)$$

$$t^+ = t_q + \tau_{sc}/\beta \quad (22)$$

$$t^- = t_q + \frac{N_h}{s_h} - \frac{2\tau_{sc}}{\alpha + \gamma} \quad (23)$$

so that $R_{sc} = \tau_{sc}s_h(t^- - t^+)$ and $C_{sc} = \beta(t^* - t_q) - R_{sc}/N_h$.

Revenue Maximizing Step Toll

We now turn to the derivation of the revenue maximizing step toll to gain some insight into why the "timing" of the toll is important. Any given toll can lead to one of two possible traffic patterns. In the first, drivers pass through the bottleneck before, during, and after the toll period, whereas in the second, drivers pass through the bottleneck before and during the toll period only. We will denote the set of tolls that result in the first pattern by $\mathcal{S}_1$ and those that result in the second pattern by $\mathcal{S}_2$ and consider each separately below.

Traffic Pattern 1

We begin our derivation of the revenue maximizing toll in $\mathcal{S}_1$ by observing that this toll will never result in departures after the bulk group. With this in mind, it is relatively easy to derive the revenue maximizing toll. In particular, recall that the size of the bulk departure in equilibrium must be $(2s\tau)/(\alpha + \gamma)$ (*35*). Hence, it must be the case that the number of people exiting the queue between t^- and t_q is given by

$$s(t^- - t_q) = N_h - \frac{2s\tau}{\alpha + \gamma} \quad (24)$$

and hence that

$$(t^* - t_q) + (t^- - t^*) + \frac{2\tau^s}{\alpha + \gamma} = \frac{N_h}{s_h} \quad (25)$$

which further implies that

$$C_h = F_h + \beta\underbrace{\left[\frac{N_h}{s_h} - (t^- - t^*) - \frac{2\tau^s}{\alpha + \gamma}\right]}_{t^* - t_q} \quad (26)$$

Also, observe that several conditions must be satisfied for the toll to be in $\mathcal{S}_1$. First, commuters should not be deterred from passing through the toll booth right after the toll period begins. This implies that

$$\frac{\beta}{\alpha}(t^+ - t_q) - \frac{\tau}{\alpha} > 0 \quad (27)$$

Second, commuters should keep passing through the bottleneck until the very end of the toll period. This implies that

$$\frac{\beta}{\alpha}(t^* - t_q) - \frac{\tau}{\alpha} - \frac{\gamma}{\alpha}(t^- - t^*) > 0 \quad (28)$$

Finally, commuters should have no incentive to pass through the bottleneck after the bulk group. This implies that

$$\beta(t^* - t_q) < \gamma(t_{q'} - t^*) \quad (29)$$

Now, using Equation 25 and Equations 27 through 29 and letting $q = \beta(\gamma - \alpha)/[(\gamma + \alpha)(\gamma + \beta)]$, it can be shown that the revenue maximizing toll in $\mathcal{S}_1$ must satisfy the following conditions:

$$\frac{N_h}{s_h} > (t^- - t^+) + \frac{\tau}{\beta}\left(1 + \frac{2\beta}{\alpha + \gamma}\right) \quad (30)$$

$$\frac{\beta}{\beta + \gamma}\frac{N_h}{s_h} > (t^- - t^*) + \frac{\tau}{\gamma}(1 + q) \quad (31)$$

With these results, it is now relatively easy to show that the revenue maximizing toll within $\mathcal{S}_1$ must satisfy Equations 30 and 31 as equalities.

First, observe that C_h is independent of t^+. Thus N_h is independent of t^+. As a result, toll revenues are maximized when t^+ is set at the lower bound given in Equation 30.

Now, observe that, in general, $R = \tau s_h(t^- - t^+)$. Hence, for this subset of toll schemes we have

$$R = N_h\tau - \frac{s_h}{\beta}\left(1 + \frac{2\beta}{\alpha + \gamma}\right)\tau^2 \tag{32}$$

Further, it follows from Equation 26 and the equilibrium condition that N_h is an increasing function of t^-. Hence, for any given toll scheme with t^+ defined by Equation 30, toll revenues are maximized by setting t^- at its upper bound in Equation 31.

It thus follows that

$$(t^- - t^*) = \frac{\beta}{\beta + \gamma}\frac{N_h}{s_h} - \frac{\tau}{\beta}(1 + q) \tag{33}$$

and since $(t^* - t^+) = \gamma/\beta(t^- - t^*)$ in equilibrium, it must be the case that

$$t^* - t^+ = \frac{\gamma}{\beta + \gamma}\frac{N_h}{s_h} - \frac{\tau}{\beta}(1 + q) \tag{34}$$

To maximize toll revenues, we simply set

$$\tau_{tr}^{(1)} = \frac{1}{1 + q}\frac{\beta\gamma}{\beta + \gamma}\frac{N}{2s_h} \tag{35}$$

and

$$(t^* - t^+)_{tr} = \frac{\gamma}{\beta}(t^- - t^*)_{tr} = \frac{\gamma}{\gamma + \beta}\frac{N}{2s_h} \tag{36}$$

which means that

$$C_{tr}^{(1)} = F_h + \left(1 - \frac{q}{2(1 + q)}\right)\frac{\beta\gamma}{\beta + \gamma}\frac{N}{s_h} \tag{37}$$

and

$$R_{tr}^{(1)} = \frac{1}{1 + q}\frac{\beta\gamma}{\beta + \gamma}\frac{N^2}{4s_h} \tag{38}$$

Traffic Pattern 2

We now turn our attention to finding optimal toll schemes within the subset $\mathcal{S}_2$. In this case (i.e., when drivers exit the bottleneck before and during the toll period only), it is relatively easy to show that the equilibrium conditions imply that

$$t^* - t_q = \frac{\gamma}{\beta + \gamma}\frac{N_h}{s_h} + \frac{\tau}{\beta + \gamma} \tag{39}$$

$$t_{q'} - t^* = \frac{\beta}{\beta + \gamma}\frac{N_h}{s_h} - \frac{\tau}{\beta + \gamma} \tag{40}$$

$$C_h = F_h + \frac{\beta\gamma}{\beta + \gamma}\frac{N_h}{s_h} + \frac{\beta\tau}{\beta + \gamma} \tag{41}$$

In addition, several conditions must be satisfied by tolls in $\mathcal{S}_2$. First, commuters should not be deterred from passing through the toll booth right after the toll period begins. This implies that

$$t^* - t^+ < \frac{\gamma}{\beta + \gamma}\frac{N_h}{s_h} - \frac{\gamma}{\beta}\frac{\tau}{\beta + \gamma} \tag{42}$$

Second, no commuter should pass through the bottleneck after t^-. This implies that

$$t^- - t^* > t_{q'} - t^* \tag{43}$$

Hence, given that $t^* - t^+$, $t^+ - t_q$, and $t_{q'} - t^*$ must be positive, it follows that $\tau < \beta\ (N_h/s_h)$.

It is now easy to show that the revenue maximizing toll must satisfy Equation 42 as an equality. In particular, observe that toll revenues are given by

$$R = s_h\left[\underbrace{\left(\frac{\beta}{\beta + \gamma}\frac{N_h}{s_h} - \frac{\tau}{\beta + \gamma}\right)}_{t_{q'} - t^*} + (t^* - t^+)\right]\tau \tag{44}$$

Also, observe once again that C_h does not depend on t^+ and hence that N_h is independent of t^+. As a result, toll revenues are maximized only when t^+ is set at its lower bound (as given by Equation 42).

It then follows that

$$R = \tau\left(N - \frac{s_h}{\beta}\tau\right) \tag{45}$$

To maximize this expression, we set

$$\tau_{tr}^{(2)} = \beta\frac{N}{2s_h} \tag{46}$$

and

$$(t^* - t^+)_{tr} = \frac{\gamma}{\beta + \gamma}\frac{N_h}{2s_h} \tag{47}$$

where $(t^- - t^*)$ must be chosen in such a way that $(t^- - t^*)_{tr} > \beta/\beta + \gamma\ N_h/2s_h$. Hence, for this toll mechanism, we have

$$C_{tr}^{(2)} = F_h + \left(1 + \frac{\beta}{2\gamma}\right) \frac{\beta\gamma}{\beta + \gamma} \frac{N_h}{s_h} \qquad (48)$$

and

$$R_{tr}^{(2)} = \beta \frac{N^2}{4s_h} \qquad (49)$$

Comparison

When the toll must be set in such a way that commuters choose to exit the bottleneck before, during, and after the toll period we find that (using the same values of the parameters as above) switching to the revenue maximizing scheme leads to a 0.7 percent increase in toll revenues, an 8 percent drop in the level of the step toll, and an increase in the length of the toll period (as compared with the socially optimal toll). On the other hand, when the toll can be set in such a way that nobody departs after the end of the toll period, we find that switching to the revenue maximizing scheme leads to a 37 percent increase in toll revenues, a 25 percent increase in the level of the step toll, and an increase in the length of the toll period.

Hence, it seems that the additional revenues (and additional costs) generated by the revenue maximizing permanent toll (as compared with the socially optimal permanent toll) are, in large part, a result of the fact that commuters cannot depart after the end of the toll period. In situations in which commuters must be allowed to depart after the toll ends, there is much less difference between the revenue maximizing and socially optimal tolls.

CONCLUSION

Although these results are in no way intended to be conclusive, we believe they yield some interesting insights. Most important, they provide evidence that the different objectives of peak period pricing do "compete" in some situations.

In general, it appears that toll revenues can be increased substantially by changing toll policies. In fact, by experimenting with various values of the above parameters, we have found that the revenue maximizing toll results in an increase in toll revenues of between 20 and 50 percent over the socially optimal toll. However, this increase in revenues comes at great cost to commuters—each additional dollar of toll revenue costs society between $2.00 and $4.00.

The most substantial revenue gains can be realized by imposing a toll that cannot be avoided (i.e., either a toll that lasts throughout the entire day or a toll that ends so late in the day that nobody is inclined to depart after the toll period ends). Unfortunately, this is exactly the type of toll that most increases the cost to society. On the other hand, when the toll is set in such a way that commuters will choose to depart after the toll period ends, society does not suffer much if the revenue maximizing toll is imposed. Not surprisingly, however, very little additional revenue is generated in such situations.

In short, these models seem to provide some evidence that the length and end of the toll period, and not the value of the toll, has the biggest impact on both revenues and social costs. Hence, it is the length and end of the toll period that policy makers may find most tempting to change. Yet it is probably the value of the toll that the public will react to most strongly, at least in the short run. In practice this could result in a very interesting political dynamic.

Of course, much more work needs to be done before we can have any real confidence in these results. Most important, these results need to be extended to more general networks. We are hopeful that recent advances in dynamic network equilibrium modeling [see, for example, previously published works (*33,34*)] will make this possible. In addition, longer-run decisions, such as the decision to travel and the choice of destination, should also be included, as should nonwork trips. The results of these extensions will be reported in subsequent papers.

ACKNOWLEDGMENT

This research was funded, in part, by a grant from the Mitsui Corporation.

REFERENCES

1. T. Higgins. Road Pricing: Managing the Risks. *Policy Analysis*, Vol. 3, 1977, pp. 579–582.
2. T. Higgins. Road-Pricing Attempts in the United States. *Transportation Research*, Vol. 20A, 1986, pp. 145–150.
3. P. L. Watson and E. P. Holland. Congestion Pricing—The Example of Singapore. *Finance and Development*, Vol. 13, 1976, pp. 20–24.
4. E. P. Holland and P. L. Watson. Traffic Restraint in Singapore. *Traffic Engineering and Control*, Vol. 27, 1978, pp. 14–22.
5. S. F. Williams. Getting Downtown: Relief of Highway Congestion Through Pricing. *Regulation*, March/April 1981, pp. 45–50.
6. A. T. Armstrong-Wright. Road Pricing and User Restraint: Opportunities and Constraints in Developing Countries. *Transportation Research*, Vol. 20A, 1986, pp. 123–127.
7. J. A. L. Dawson and F. N. Brown. Electronic Road Pricing in Hong Kong: A Fair Way to Go? *Traffic Engineering and Control*, Vol. 26, 1985, pp. 522–529.
8. I. Catling and B. J. Harbord. Electronic Road Pricing in Hong Kong: The Technology. *Traffic Engineering and Control*, Vol. 26, 1985, pp. 608–617.
9. B. Harrison. Electronic Road Pricing in Hong Kong: Estimating and Evaluating the Effects. *Traffic Engineering and Control*, Vol. 27, 1986, pp. 13–19.
10. J. A. L. Dawson. Electronic Road Pricing in Hong Kong: Conclusion. *Traffic Engineering and Control*, Vol. 27, 1986, pp. 79–83.
11. J. A. L. Dawson and I. Catling. Electronic Road Pricing in Hong Kong. *Transportation Research*, Vol. 20A, 1986, pp. 129–134.
12. I. Catling and G. Roth. Electronic Road Pricing in Hong Kong: An Opportunity for Road Privatization. In *Transportation Research Record 1107*, TRB, National Research Council, Washington, D.C., 1987, pp. 51–55.
13. H. J. Stoelhorst and A. J. Zandbergen. The Development of a Road Pricing System in the Netherlands. *Traffic Engineering and Control*, Vol. 31, 1990, pp. 66–71.
14. G. Hughes and S. Ison. The Cambridge Congestion Metering Scheme. *Transport*, May/June 1992, pp. 7–9.
15. *The Status of the Nation's Highways: Conditions and Performance*. U.S. Department of Transportation, June 1985.
16. J. A. Lindley. Urban Freeway Congestion: Quantification of the

Problem and Effectiveness of Potential Solutions. *ITE Journal*, Jan. 1987, pp. 27–32.
17. D. Bernstein and A. Kanaan. Automatic Vehicle Identification: Technologies and Functionalities. *IVHS Journal* (forthcoming).
18. P. B. Goodwin. The "Rule of Three": A Possible Solution to the Political Problem of Competing Objectives for Road Pricing. *Traffic Engineering and Control*, Vol. 29, 1989, pp. 495–497.
19. M. Beckmann, C. McGuire, and C. Winsten. *Studies in the Economics of Transportation*. Yale University Press, New Haven, Conn., 1956.
20. A. A. Walters. The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica*, Vol. 29, 1961, pp. 676–699.
21. W. Vickrey. Pricing in Urban and Suburban Transport. *American Economic Review*, Vol. 53, 1963, pp. 452–465.
22. W. Vickrey. Optimization of Traffic and Facilities. *Journal of Transport Economics and Policy*, Vol. 1, 1967, pp. 123–136.
23. W. Vickrey. Congestion Charges and Welfare. *Journal of Transport Economics and Policy*, Vol. 2, 1968, pp. 107–118.
24. C. H. Sharp. Congestion and Welfare: An Examination of the Case for a Congestion Tax. *Economic Journal*, Vol. 76, 1966, pp. 806–817.
25. C. H. Sharp. "Congestion and Welfare" Reconsidered. *Journal of Transport Economics and Policy*, Vol. 2, 1968, pp. 119–125.
26. S. C. Dafermos. Optimal Resource Allocation and Toll Patterns in User Optimized Transportation Networks. *Journal of Transport Economics and Policy*, Vol. 5, 1971, pp. 1–17.
27. R. Layard. The Distributional Effects of Congestion Taxes. *Economica*, Vol. 44, 1977, pp. 297–304.
28. J. V. Henderson. Road Congestion: A Reconsideration of Pricing Theory. *Journal of Urban Economics*, Vol. 1, 1974, pp. 346–365.
29. M. B. Johnson. On the Economics of Road Congestion. *Econometrica*, Vol. 32, 1964, pp. 137–150.
30. T. E. Keeler and K. A. Small. Optimal Peak-Load Pricing, Investment, and Service Levels on Urban Expressways. *Journal of Political Economy*, Vol. 85, 1977, pp. 1–25.
31. A. E. Boardman and L. B. Lave. Highway Congestion and Congestion Tolls. *Journal of Urban Economics*, Vol. 4, 1977, pp. 340–359.
32. P. A. Viton. Equilibrium and Short-Run Marginal-Cost Pricing of a Transport Facility. *Journal of Transport Economics and Policy*, Vol. 14, 1980, pp. 185–203.
33. T. L. Friesz, D. Bernstein, T. E. Smith, R. L. Tobin, and B.-W. Wie. A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem. *Operations Research*, Vol. 41, 1993, pp. 179–191.
34. D. Bernstein, T. L. Friesz, R. L. Tobin, and B.-W. Wie. A Variational Control Formulation of the Simultaneous Route and Departure-Time Choice Equilibrium Problem. *Proc., 12th International Symposium on the Theory of Traffic Flow and Transportation*, 1993, in press.
35. R. Arnott, A. de Palma, and R. Lindsey. Economics of a Bottleneck. *Journal of Urban Economics*, Vol. 27, 1990, pp. 111–130.
36. D. Bernstein and T. E. Smith. *Equilibria for Networks with Lower Semicontinuous Costs: With an Application to Congestion Pricing*. MIT Working Paper. Massachusetts Institute of Technology, Cambridge, 1992.
37. K. Small. The Scheduling of Consumer Activities: Work Trips. *American Economic Review*, Vol. 72, 1982, pp. 467–479.

Publication of this paper sponsored by Committee on Transportation Economics.

Manufacturing Logistics for the 21st Century

MARK A. TURNQUIST

Eight major external influences in global trade, manufacturing methods, and public policy are identified that are likely to have profound effects on logistics as the 21st century begins. These forces are discussed and general conclusions are reached on likely results in the logistics system. Finally, a set of ideas that could help the freight transportation sector adapt to the changing logistics system is described.

What are the changes in manufacturing and the global economy that will force fundamental changes in logistics as we enter the next century? What changes will result in logistics operations? How will this affect transportation providers? How does public policy with respect to transportation affect these changes? Lacking a functional crystal ball, we are forced to speculate on the answers to these questions, but by looking carefully at both global economic changes and changes in manufacturing, we can make some educated guesses about changes in logistics that are likely to be coming over the next 10 years or so. By thinking about these changes now, transportation providers can begin to position themselves to take advantage of new opportunities that will arise. Further, by thinking about the interplay between private logistics decisions and public policy with respect to transportation, we can gain additional important insights.

From the perspective of the logistics system, changes in the global economy, in manufacturing methods, or in public policy are all external influences. Furthermore, they are of sufficient magnitude that they can force fundamental reorganization of logistics operations. The purpose of this paper is to identify important trends that will affect logistics operations in the future and to suggest some potential responses within the freight transportation system.

WHAT ARE THE MAJOR DRIVING FORCES?

At least eight major forces will influence the structure and function of logistics systems over the coming years and lead to dramatically different expectations for the freight transportation sector:

1. The growing importance of international trade and the emergence of large multinational trading blocs,
2. Changes in the nature of production and assembly operations in manufacturing,
3. Efforts by manufacturers to reduce the number of suppliers they deal with and to emphasize long-term relationships with the remaining supplier base,
4. Continuing emphasis by manufacturers on reducing overall logistics costs and improving service quality,
5. Increasing pressure on manufacturers to take responsibility for recycling their products after use as a part of worldwide environmental consciousness and efforts to reduce solid waste disposal problems,
6. Rapid increases in the scope and capacity of data networks for moving and organizing information,
7. Increasing levels of highway congestion in and around the urban areas of the United States, and
8. Changing perceptions and policies of the federal government concerning transportation systems in the United States.

In the following subsections, the nature of each of these influences will be discussed.

Growth in International Trade and Trading Blocs

Figure 1 shows U.S. merchandise imports and exports during the period between 1980 and 1990 (*1*). This clearly illustrates the growing importance of international trade in our economy. The recent breakup of the Soviet Union and the Warsaw Pact has removed another major set of barriers to globalization of the world economy and will result in even more rapid increases in U.S. trade volume in the years ahead.

The European community is forming an integrated trading bloc, and this has enormous implications for manufacturers, both in and outside of Europe (particularly in North America and Japan). A recent report by Andersen Consulting for the Council of Logistics Management (*2*) emphasizes movements toward both "integration" and "rationalization" in European logistics operations. Integration efforts are aimed at combining what previously have been separate national production and distribution systems into a coherent pan-European system. Rationalization efforts are aimed at cost reduction through elimination of duplicate or redundant facilities. Opportunities for rationalization are being created by efforts toward greater integration.

The recently negotiated North American Free Trade Agreement involving Canada, Mexico, and the United States is likely to have comparable significance, creating a free-trade zone spanning North America. The result of this agreement is likely to be a dramatic increase in the magnitude of north-south flow of goods, including raw materials, work-in-process, and finished goods. The predominant historic pattern for goods

School of Civil and Environmental Engineering, 309 Hollister Hall, Cornell University, Ithaca, N.Y. 14853-3501.

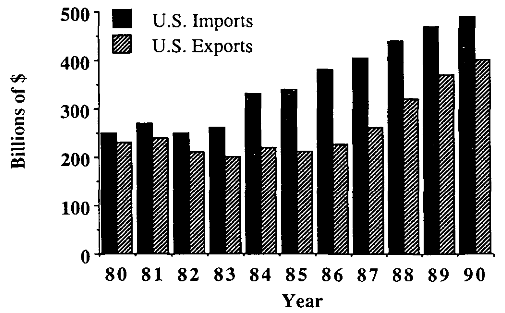


FIGURE 1 U.S. imports and exports, 1980–1990.

movement in both the United States and Canada has been east-west, and the increasing north-south flow is likely to require changing operations practices, investments in new capacity for both manufacturing and transportation, and changes in facility location decisions.

All of these changes will create both challenges and opportunities for manufacturing logistics and will reverberate through the freight transportation industries. The manufacturing process will become more decentralized, drawing on materials, resources, and labor in a wider variety of locations and depending on the logistics systems to bring all the pieces together at the right place at the right time. At the same time, markets for products (and hence distribution channels) are widening, but in a way that may require a variety of "regionally customized" products.

A simple, but illustrative, example of this regional customization issue is the manufacture of electrical appliances for European use. With the removal of tariffs among the European Community countries, the potential market for a manufacturer of appliances in France, for example, is now much larger. However, there are still several different standards for electrical plugs, and thus a slightly different model must be produced for each of those markets.

This variation in standards may affect the logistics system, because one strategy for dealing with these variations is to postpone the attachment of the electrical cord until the appliance reaches a distribution center within each country and at that point attach the correct cord for that country's use. This changes the logistics requirements because the cords and plugs, which may be produced by an outside supplier, now must be shipped in a pattern very different from that used in the past. In effect, the distribution center has become an assembly location, and that changes the way the logistics system must operate.

Changing Production Methods and Technology

Manufactured products are becoming increasingly complex, and the customers who buy those products have increasing expectations for product quality and reliability. As described in the previous section, manufacturers are also doing more and more "niche marketing"—producing several different variations on a product aimed at slightly different sets of consumers. The ever-growing emphasis on total quality management as a focus on meeting customer desires will also force companies to change their manufacturing methods. Among the likely changes in manufacturing as we enter the next century are the following:

- Greater use of smaller (cellular) production facilities capable of responding quickly to local market demands;
- Increased adoption of flexible production methods capable of making a wide variety of customized products, each in relatively low volumes;
- Shorter and shorter life cycles for products, which also will put pressure on manufacturing facilities to be more flexible and easily adaptable to production of new products.

These ideas (and others) have been proposed in the "Manufacturing 21" study sponsored by eight major Japanese companies and several universities in that country (*3*).

An early indication of the implementation of these ideas may be seen in Nissan's goal of five "anys": any volume, anywhere, any time, anything, anybody (*4*). Translated, Nissan wants to be able to make any one of its models, at any of its plants, anywhere in the world, at any volume level, and at any time demanded by local market conditions. Moreover, they want to do so with a production system capable of being run by anybody.

Another illustration of the ideas is the production of bicycles in Japan by a subsidiary of the electronics giant, Matsushita (*5*). In a small factory in Kobuku, in western Japan, 20 employees and a computer are ready to produce any of 11,231,862 variations on 18 models of racing, road, and mountain bikes. Production starts with a customer order faxed from a local retail store, and the bike is made to fit for a specific customer, with delivery in 2 weeks.

The implication of these examples is that in the future there are likely to be more production and assembly locations, each producing a wider variety of products in smaller volumes, and doing so on demand to meet customer orders rapidly. For transportation of materials and products, this means smaller lot sizes, more frequent orders, a more dispersed set of origins and destinations, and tighter standards for on-time delivery.

Changing Relationships with Suppliers

Many U.S. manufacturers are making major changes in their relationships with suppliers, and this includes suppliers of transportation services as well as suppliers of raw materials and component parts. The objective is to reduce the number of suppliers and to establish a longer-term cooperative relationship, or partnership, with the remaining supplier base.

An excellent example of this sort of change in the transportation services area is provided by Reynolds Metals Company (*6*). By reorganizing their logistics operations, they reduced the set of trucking carriers they dealt with from 200 to 14. Each of the remaining "core carriers" agreed to tailor specific services to meet Reynolds' needs and, in return, received a larger share of Reynolds' shipments.

Other examples of manufacturers that have made major changes in their relationship with transportation suppliers are Xerox (*7*), DuPont (*8*), and Olin (*9*). These companies are trend setters, and their lead is likely to be followed by many other manufacturing companies over the next few years. These changing relationships affect both the logistics operations of the manufacturing company and the operations of the transportation provider.

Emphasis on Reducing Logistics Costs

Statistics cited by Foster (*10*) indicate that in 1991, U.S. companies spent $655 billion on logistics, amounting to 11.6 percent of the entire gross domestic product (GDP). This is about the same amount as is spent on health care and twice what is spent on defense. Expressed as a percentage of GDP, logistics costs have been reduced from their high point (1981), when they reached 17.9 percent of GDP, largely as a result of deregulation in the transportation industries. However, further reductions are going to require more systematic analyses and structural changes in the logistics systems of most companies.

Systematic study of the full logistics chain has proven beneficial for several companies. Two excellent examples are the efforts at Reynolds Metals Company, cited previously, and at General Motors, documented by Blumenfeld et al. (*11*). The work at General Motors, in particular, has illustrated the close connections between production decisions (and costs) and logistics decisions (and costs).

Movements toward just-in-time (JIT) deliveries by many companies in the last 3 to 5 years have clearly reduced inventory carrying costs and are one illustration of attempts to tie production decisions and logistics decisions more closely together. However, careful study of the relationships between production and logistics decisions involves consideration of much broader issues than JIT deliveries of materials.

One of these broader issues is production of components and subassemblies in plants scattered literally around the world. Many major U.S. manufacturers are (or are at least considering) locating some production facilities outside the United States, primarily in a search for inexpensive labor. This lengthens the links in the logistics chain and increases logistics costs. Pressures to contain the total logistics bill for a company must be understood against the backdrop of a complex web of interrelated production and logistics decisions.

Responsibility for Product Recycling

Historically, manufacturers have worried only about putting their products in the hands of their customers. The logistics chain has been a one-way movement. However, the growing environmental consciousness in the major industrialized societies of the world is forcing some reconsideration of this assumption. Recycling and reprocessing of used products is growing in importance and is likely to become much more important by the early 21st century.

Some manufacturing industries, such as aluminum processing, already have made major changes in this regard. Others, such as manufacturers of plastics and lead-acid storage batteries, are beginning to use recycled materials more extensively. A wider variety of manufacturers is likely to follow over the next several years.

The resulting reverse flow in the logistics chain has implications for both the manufacturers and transportation providers. On the plus side, it may create backhaul opportunities that could lead to more efficient use of transportation resources. However, it also creates the need to integrate reprocessing facilities into the overall logistics system and to manage material flow patterns among an even wider set of origins and destinations.

Improved Information Management Capabilities

The processing speed, storage capacity, and networking ability of computer hardware have increased at a phenomenal rate in the last decade. These changes have provided opportunities for data exchange, processing, and organization in real time (or near real time) that only a few years ago would have been beyond our imaginations. Over the coming decade, we will see even more dramatic improvements, particularly in computer networking and distributed data base management, and these changes will open up even broader possibilities for managing the freight transportation system differently and more effectively.

Dertouzos (*12*) and Cerf (*13*) have outlined a variety of possible changes, ranging from installation of computer networks, which would become an "interstate data highway system" transmitting data at rates exceeding 2,400 million bits/sec, to "knowbots"—programs that could be launched into the network to look automatically for a variety of types of information relevant to a particular request, organize what is found, and return the results to the initiator.

Even in the relatively short run, increasing use of electronic data interchange will change the character of many logistics transactions and operations. For example, the widespread use of electronic point-of-sale terminals in retail outlets has allowed retailers to have virtually real-time information on sales and stock levels of products in their stores. Many retailers have used this information to their advantage by reducing in-store inventories, putting pressure on manufacturers for more frequent and smaller deliveries to retail outlets, and in some cases reorganizing their own inventory operations (*14*). All of the changes in information management that are used to reduce inventories in the overall logistics chain result in changes in shipment sizes, frequency, and composition that have consequences for both the manufacturer and the transportation service provider.

In the longer run, we can look forward to having much wider knowledge of shipment options, status of current shipments, equipment availability, and so forth easily accessible to shippers, carriers, and receivers. This should lead to increased responsiveness of the transportation system to demands placed on it and to improved utilization of resources in the system.

Increasing Highway Congestion

Rao et al. (*15*) have recently argued that highway congestion and JIT operations are both growing rapidly and are probably on a collision course. The smaller and more frequent shipments, shorter lead times, and precise scheduling called for by JIT can be severely impeded by travel times that are rising and becoming more uncertain as traffic congestion grows.

Extrapolation of current trends is certainly risky, but forecasts based on such extrapolation indicate that (a) total vehicle delay on urban freeways is expected to increase over 400 percent by 2005 (*16*) and (b) the intercity speed of an average railroad freight car will exceed that of an average tractor-trailer by 2010 (*17*).

Rao et al. (*15*) describe several potential strategies for alleviating some of the effects of congestion on JIT operations, including use of off-peak deliveries, computer-assisted routing and scheduling of movements, and consolidation strategies by either vendors or transportation companies. The last two of these strategies are dependent on better information organization and transmission and thus relate directly to the issues discussed in the previous subsection.

Changing Federal Transportation Policy

The Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) presents a new vision of the nation's transportation system—a concept in which the various modes form an integrated, closely coordinated system that provides "seamless" multimodal service for both passengers and freight. The individual modes are seen as different facets of a unified system that provides services appropriate to the needs and demands of the shipper or person being served.

This change in focus, from one dominated by differentiating the modes on the basis of technology to one emphasizing the services they can provide together, represents a fundamental shift in public policy at the federal level. This change will be played out in a variety of ways over the coming years, but the emphasis on improving intermodal connections surely will lead to new service offerings and should open additional opportunities for improving both service quality and resource utilization by transportation providers.

HOW WILL LOGISTICS CHANGE?

Bianco (*18*) has outlined a general model of logistics networks, including suppliers, parts production facilities, assembly facilities, distribution centers, warehouses, and markets. Figure 2, reproduced from his work, shows the basic connections. In this network diagram the logistics chain involves several stages of movement, connections between stages may bypass other stages, and at each stage there may be many locations (and potentially many different organizations) involved. Although Figure 2 shows the various stages with equal numbers of nodes (locations), we should not infer from that that the numbers of locations are really equal at the various stages. For the purposes of this paper, we can use the representation of the logistics network in Figure 2 as the basis for projecting potential changes in the logistics system that might be driven by the major forces described in the previous section. This perspective leads us to focus on changes in

1. The number and locations of facilities for suppliers, parts production, assembly, distribution centers, and warehouses;
2. The size or location of markets; and
3. The connections among the various stages of the process.

Figure 3 (again in an idealized sense) represents the general nature of changes in the logistics system under the influence of the driving forces described above. At least five major structural changes are likely.

First, the efforts of manufacturers to reduce their supplier base will "squeeze" the top (supplier) stage of the network. A smaller set of core suppliers will each be providing a wider range of raw materials or producing a wider range of parts and components. These suppliers will be asked to ship in mixed loads to meet JIT delivery schedules, and many manufacturers will expect their transportation providers to use multistop "milk runs" to collect shipments from several different suppliers for a single delivery to the destination plant.

Second, increasing globalization of the economy will "stretch" the bottom of the network, creating broader markets. However, those markets may have notable regional differences, requiring many customized variations in products.

Third, the competitive pressures to reduce logistics costs will compress the network from top to bottom, resulting in elimination of layers in the hierarchy of distribution inventories. This is shown in Figure 3 as the elimination of the "warehouse" stage. However, this top-to-bottom compression does not necessarily mean that the logistics links will become shorter. In fact, as production becomes more global the individual links are likely to become longer.

Fourth, the distinctions among the production, assembly, and distribution center stages will become less distinct. This change is a result of the combination of pressures to reduce logistics costs, the changing nature of production technology, and the desire to be more responsive to rapidly changing conditions in a wide range of markets.

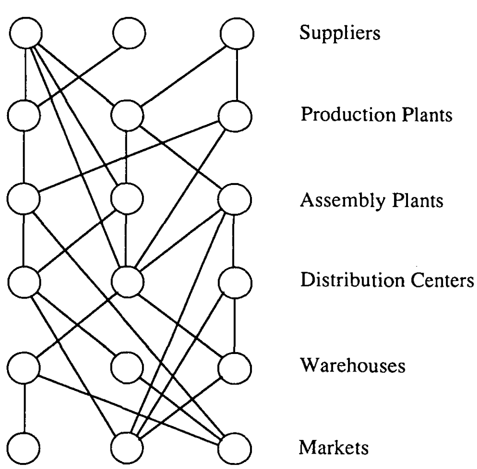


FIGURE 2 Representation of a general logistics network.

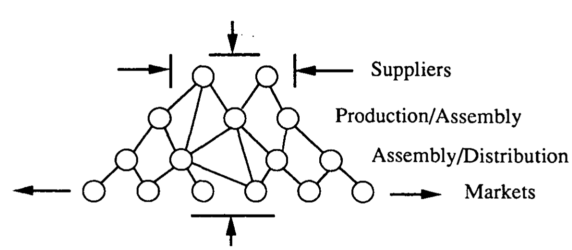


FIGURE 3 Response of a logistics network to external forces.

Finally, there will be greater flow of materials in the "reverse" direction, as a result of recycling used products. This also may create a new set of locations in the overall network, corresponding to reprocessing facilities in which materials from used products can be separated and prepared for reuse.

Perry (*19*) has argued that there are seven elements of strategic importance in adapting today's logistics systems to meet the future: asset productivity, horizontal management, selective risk (tailored, customer-driven service standards), postponement of resource commitments in the face of uncertainty, substitution of information for other resources (vehicles, materials, and labor), integrated planning, and system flexibility. His description of these seven ideas will not be repeated here, but the following section does illustrate, in a somewhat different way, a similar set of ideas about how we need to focus attention on the transportation elements of the logistics system.

NEW VIEW OF TRANSPORTATION SYSTEM OPERATIONS

Adaptation of transportation operations to the changing requirements placed on the logistics system requires new goals and objectives. Three critical goals for transportation operations are

- Coordination,
- Responsiveness, and
- Resource utilization.

These three goals do not make up an exhaustive list—certainly cost-effectiveness, safety, and reliability continue to be important, for example—but particular attention needs to be focused on these three. They point to total quality management for the transportation system—an emphasis on managing the performance of the whole system with constant attention to the customer. This focus on services provided is also the logical result of the public policy perspective expressed in ISTEA and reflects the influence of that force on changing the way the transportation system functions.

In an operational sense, coordination is the goal of bringing all the required pieces together in the right place at the right time. In a container port, for example, this means having gantry cranes, trucks (tractor and chassis) or railcars, and people available on the dock when the ship is ready to be unloaded, as well as having the required information for each container (contents, customs clearance, destination, tariff information, etc.).

The goal of improved system responsiveness stems from the fact that the system must be able to meet changing demands quickly and effectively. As markets shift from one place to another or commodity price changes cause changes in supplier locations, the system must be able to change delivery schedules or capacity requirements to meet new needs.

One way to meet the responsiveness goal is to provide excess capacity—to ensure that enough equipment and facilities will always be able to accommodate the demand that arises. But this leads to poor resource utilization and increased total cost, in contradiction to the third goal. Because responsiveness and utilization are conflicting goals, it is better to think about advancing the "responsiveness utilization" frontier, as shown in Figure 4. From a given operating point (combination of responsiveness and utilization), such as Point A in the graph, we want to move to a point at which improvement occurs in one or both objectives without a sacrifice in either.

Improvements in coordination, responsiveness, and utilization are intrinsically linked to information flow—the ability to provide more punctual and accurate information about system status and projected future events. The system needs an ability to complete operational transactions in a punctual, efficient, and accurate fashion, with "instantaneous" dissemination of pertinent information to all parties involved. Furthermore, improved quality of information offers the opportunity for effective real-time planning based on real-time data about present status and anticipated near-term demands.

Figure 5 shows a way of thinking about the related flows of goods, physical resources, and information, as we strive to achieve improvements in coordination, responsiveness, and resource utilization. Each of the three elements moves in its own "network"—goods move from vehicle to dock to vehicle on their way from origin to destination; trucks or railcars move from shipment to shipment as they are used and reused; and information moves from computer to computer (or person to person) through its own channels. However, these three networks can be treated as "layers" of a larger network, with connections among the layers that are necessary for processing steps. In the port example above, the container unloading process cannot begin until the ship with the containers, the crane, and the trucks or railcars are all present together, and the information about this shipment has also arrived, been assembled, and distributed to the necessary people.

The representation of connected layers of goods flow, resources flow, and information flow, as shown in Figure 5, emphasizes one of the major sources of delay in the transportation system. Delays occur when the layers are not tightly connected, and one or more of the elements required for a processing step are not present when needed. This delay lowers resource utilization, because those resources already assembled must wait for the missing elements, and also reduces overall service quality.

Figure 5 is also important because it places concern for information flow on an equal footing with concern for flows of goods and physical resources. As Perry (*19*) has suggested, the cost of information as a resource is falling relative to the costs of the other major resources (vehicles, labor, fuel, etc.)

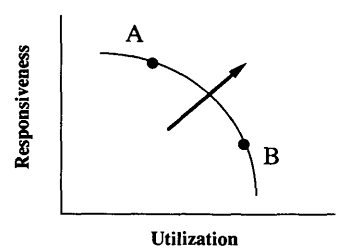


FIGURE 4 Responsiveness-utilization trade-off.

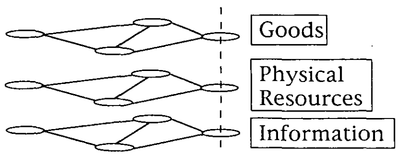


FIGURE 5 Connected layers in an overall network.

used in the logistics function. As rational managers, we should be trying to use more information and less of the resources it can replace. This will be vital as we enter the 21st century.

IMPLICATIONS FOR TRANSPORTATION SERVICE PROVIDERS

If the changes identified in this paper are occurring or accelerating, what should transportation providers do to preserve their existing business and create new opportunities for themselves? First and foremost, they must be aware that their customers' businesses are changing and that their logistics are likely to be changing as well. To reduce overall logistics costs and improve performance, the customers are likely to want to deal with fewer carriers, and the carriers they choose will have to understand their business. The implication is clear: transportation providers must get to know their customers better.

As manufacturing customers begin to rely on sources of materials and components from around the world, their inbound logistics problems get more difficult, and they are likely to be looking for carriers who can help them effectively manage the whole inbound supply chain. This may require joint ventures with other transportation companies so that together services over broader geographic areas and over multiple modes can be provided also. Similar pressures on expanding markets for finished products will create opportunities for similar joint ventures in the distribution end of the logistics chain.

Customers are likely to want more frequent shipments of goods, in smaller lot sizes, and with greater mixtures of commodities in each shipment. This is likely to mean greater use of containerization for domestic as well as international movements, and transportation providers need to be prepared to respond to that need.

Working more closely with customers, engaging in joint ventures with other carriers, and providing more complex services all will require mastery of the information flow that accompanies shipments. Transportation providers will need to exchange data with their customers, partners in joint ventures, and agencies such as Customs. Gaining the ability to do this effectively will require investments in both information technology and the people who will need to learn to use the technology efficiently.

FINAL NOTE

It has often been said that the only thing true about forecasts is that they will be wrong. Peering into the future is tricky business, but failing to plan for the future is almost certainly worse than planning for an uncertain future. It is my hope that the picture of the future sketched out here, dim and clouded though it may be, will shed at least a little light on someone's planning efforts.

REFERENCES

1. *Business America.* May 6, 1991, p. 3.
2. *Reconfiguring European Logistics Systems.* Andersen Consulting, Inc., 1993.
3. J. H. Sheridan. A Look at the 21st Century. *Industry Week*, Nov. 20, 1989, pp. 38–39.
4. B. S. Moskal. Nissan Heads "Anywhere." *Industry Week*, Jan. 6, 1992, p. 19.
5. S. Moffat. Japan's New Personalized Production. *Fortune*, Oct. 22, 1990.
6. E. W. Moore, Jr., J. M. Warmke, and L. R. Gorban. The Indispensable Role of Management Science in Centralizing Freight Operations at Reynolds Metals Company. *Interfaces*, Vol. 21, No. 1, 1991, pp. 107–129.
7. J. A. Cooke. How Xerox Cut Millions from Its International Freight Bill. *Traffic Management*, Vol. 30, No. 6, 1991, pp. 67–69.
8. J. Gordon. In Search of Perfection. *Distribution*, Vol. 89, No. 9, 1990, pp. 41–43.
9. T. A. Foster and J. V. Barks. The Right Chemistry for Single Sourcing. *Distribution*, Vol. 89, No. 9, 1990, pp. 44–52.
10. T. A. Foster. In a Difficult Year, Logistics Shines. *Distribution*, Vol. 91, No. 7, 1992, pp. 6–10.
11. D. E. Blumenfeld, L. D. Burns, C. F. Daganzo, M. C. Frick, and R. W. Hall. Reducing Logistics Costs at General Motors. *Interfaces*, Vol. 17, No. 1, 1987, pp. 26–47.
12. M. L. Dertouzos. Communications, Computers and Networks. *Scientific American*, Sept. 1991, pp. 62–69.
13. V. G. Cerf. Networks. *Scientific American*, Sept. 1991, pp. 72–81.
14. A. Mercer. The Consequences for Manufacturers of Changes in Retail Distribution. *European Journal of Operational Research*, Vol. 64, 1993, pp. 457–461.
15. K. Rao, W. Grenoble, and R. Young. Traffic Congestion and JIT. *Journal of Business Logistics*, Vol. 12, No. 1, 1991, pp. 105–121.
16. M. Jacks. Urban Congestion—A Local Problem Requiring a National Response. *ITE Journal*, Feb. 1989, pp. 21–23.
17. J. Hagen and W. K. Smith. Transportation—Some Practical Advice for the 1990's. Presented at Annual Meeting of the Council of Logistics Management, Boston, Mass., Oct. 1988.
18. L. Bianco. Mathematical Models in Logistic System Design. In *Freight Transport Planning and Logistics* (L. Bianco and A. LaBella, eds.), Springer-Verlag, Berlin, 1988, pp. 210–257.
19. J. H. Perry. Emerging Economic and Technological Futures: Implications for Design and Management of Logistics Systems in the 1990's. *Journal of Business Logistics*, Vol. 12, No. 2, 1991, pp. 1–16.

Publication of this paper sponsored by Committee on Transportation Economics.

PART 3

Socioeconomic Impacts

Analyzing Effects of Highway Rehabilitation on Businesses

HERNÁN E. DE SOLMINIHAC AND ROBERT HARRISON

Urban highway rehabilitation projects create problems not only for state highway agencies and motorists but also for commercial premises. For businesses dependent on highway traffic, these projects can potentially disrupt the flow of customers that provide essential business revenue. Notwithstanding the fact that, following the construction, businesses invariably report improved sales from the increased highway traffic, construction projects can be so disruptive as to cause some companies to fail. To reduce these negative effects, highway planners are increasingly expediting construction projects in a variety of ways. An attempt was made to determine whether expediting—an approach that makes use of a systems analysis—in fact provides benefits to adjoining businesses.

The objectives of this study were to develop, first, a set of fundamental questions relating to the effects of urban construction and then a methodology that could be used to answer them—that is, a methodology that could effectively quantify the economic impact of an expedited highway rehabilitation program on a business. We then applied the methodology to an expedited rehabilitation project currently under way on the Southwest Freeway (US-59) in Houston, Texas.

PROPOSED METHODOLOGY

Using information from a literature review (*1,2*) and from the Southwest Freeway case study, we posed and then attempted to answer five basic questions regarding the effect of road construction on business activity. These questions included the following:

1. Do road construction activities significantly affect the sales of the abutting business?
2. Do road construction activities of a major corridor significantly affect the economy of the city in which those activities take place?
3. Do road construction activities affect some businesses more than others?
4. Does the use of a public relations officer fully dedicated to an important construction job reduce user and business effects?
5. Does phased construction minimize negative business effects?

To study the effect of construction activities on the abutting businesses, we used two approaches. The first included analyzing historical sales data of the businesses located in the area of the construction activities. In the second approach, we interviewed the owners of businesses located along the road being rehabilitated.

PROJECT DESCRIPTION

Transportation improvements in the Southwest Freeway Corridor are an important aspect of the Houston regional mobility plan, which consists of a comprehensive network of freeways, major thoroughfares, and transit improvements for the Houston metropolitan area. To provide better mobility for motorists and transit users in the southwest area, the Texas Department of Transportation (TxDOT) and the Metropolitan Transit Authority of Harris County have joined with FHWA and FTA to complete the Southwest Freeway project.

The construction was begun in August 1989 and was scheduled for completion in December 1992. It involves the reconstruction of 11.6 mi (divided into four segments) of the busiest roadway in Houston (and in Texas), one having an average daily traffic volume exceeding 250,000 vehicles per day on some sections. The $200 million project includes freeway reconstruction and widening, as well as construction of a high-occupancy-vehicle lane with associated park-and-ride lots and transit center.

New frontage roads will add as many as two lanes (four lanes total) in some areas, and three lanes (five lanes total) at the major street intersections. The number of main lanes will basically be doubled, from 6 to 12. The high-occupancy-vehicle lane will be constructed in the center of the freeway.

Once completed, the project will provide added freeway capacity, a transitway, improved frontage roads, better intersections, and improved driveways; in addition, users will benefit from better freeway signing, drainage, and lighting.

This massive project has been designated an "expedited" project, and TxDOT has phased the construction plans accordingly. In keeping with such efforts, a full-time project public relations (PR) officer has been employed. Contacting all affected businesses, this officer has attempted to address specific business problems relating to the construction; in addition, the PR effort has worked to ensure the use of all possible media to inform highway users of the various phases of the project and the various traffic management schemes used.

SALES ANALYSIS

In analyzing the effect of Southwest Freeway construction on business sales in the Houston area, we selected a methodology

H. E. de Solminihac, Dept. Ingeniería de Construcción, Pontificia Universidad Católica de Chile, Casilla 306 Correo 22, Santiago, Chile. R. Harrison, Center for Transportation Research, The University of Texas at Austin, 3208 Red River, Suite 200, Austin, Tex. 78705.

that involved collecting sales data from the research division of the comptroller of public accounts of the state of Texas. This information, classified by trade industry and zip code, has been collected by the state since early 1984.

The first step in this methodology was to select the city of Houston, the Southwest Freeway project (US-59), and the IH-10 freeway as a control section. The latter was selected for its proximity and geometrical similarity to US-59. The information requested in these reports included the following retail categories: building materials (BUILD), general merchandise (MERCH), food store (FOOD), automotive (AUTO), clothing (CLOTH), home furnishings (HOME), restaurants (REST), drug stores (DRUG), liquor stores (LIQUOR), miscellaneous retail (MISC), and total retail trade (TOTAL).

The second step categorized the zip code information according to three areas: Houston, IH-10, and US-59. We then developed 11 sales-versus-time plots, including total retail gross sales (Figure 1) and the other 10 categories already mentioned in the methodology (*3*). Each figure had two plots: one for the city of Houston and the other for IH-10 and US-59.

In general, Houston total retail sales exhibit cyclical behavior, with the first three quarters showing low sales and the fourth reporting much higher sales. In addition, the figure shows that the economic activity, after decreasing from 1986 to 1988, starts increasing again during 1988. On the other hand, the US-59 corridor shows a behavior similar to that for the Houston area; but the recession that appears in Houston from 1986 to 1988 is less pronounced in this corridor. Finally, the control section shows behavior similar to that for the

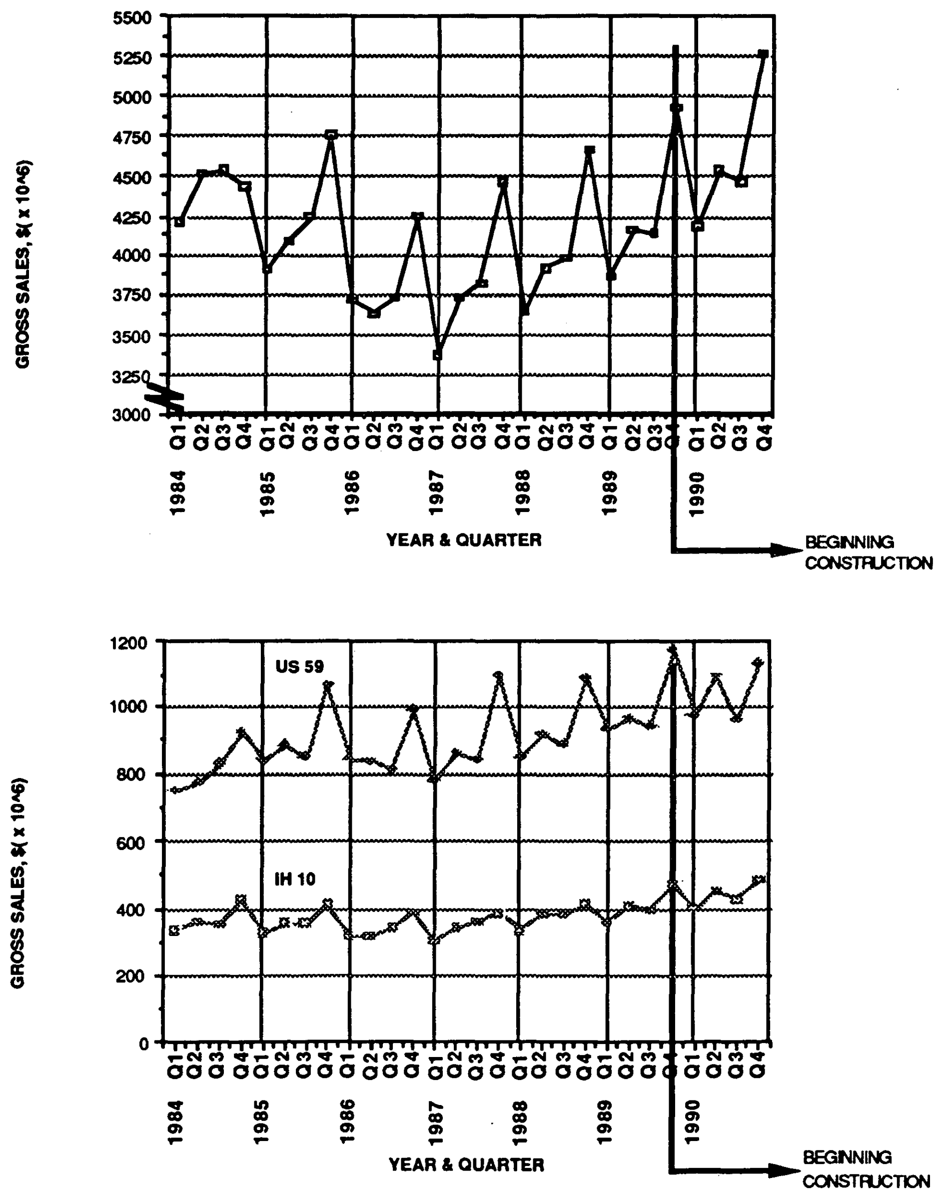


FIGURE 1 Historical sales for total retail: *top*, for Houston; *bottom*, for IH-10 and US-59.

US-59 corridor, but with less pronounced fourth-quarter peaks.

Thus, there is little evidence to suggest that the expedited project on US-59 has had a detrimental total retail sales impact beyond the immediate vicinity of the project.

Methodology

We next looked at an early statistical analysis of the sales data obtained from the comptroller's office to study the impact of the Southwest Freeway project on retail business sales in the project area (*3*). Although this analysis was performed using quarterly information, the results showed that the number of businesses reporting in the fourth quarter was much greater than the number reporting in any one of the other three quarters. The comptroller's office, contacted about this question, explained that some businesses report their sales figures for the year in the last quarter, and these inflate figures in the fourth quarter. Consequently we decided to undertake a new analysis. For this we aggregated the data by year, assuming that all businesses that reported in the first three quarters also reported their sales in the fourth quarter.

The first step was to aggregate the sales data by quarter for each retail category of interest (total of 10 categories) and for each area of interest (total of 3 areas). The next step was to develop a regression equation for each case (total of 33 equations), including the information on the first 5 years (1984 through 1988). The information for 1989 was excluded because the construction started at the end of the third quarter and, hence, included information from both before and after construction began. Next, the sales for 1990 were predicted using the regression equations developed in the previous step; the confidence interval for that prediction, using a confidence level of 90 percent, was also calculated.

We next compared the predicted value with the actual value for that year. For this comparison, we assumed that if the actual value fell between the limits of the confidence interval, the expected value would be statistically similar to the actual value. In addition, if the actual value was greater than the upper limit, the actual value would be statistically greater than the expected value. At the same time, a similar analysis was considered for the lower limits. To include the economic effects in the analysis, IH-10 was introduced as a control section. Thus, to isolate the effects of the construction on US-59, a relative analysis was performed using IH-10 as a base case.

Results

For the city of Houston, 80 percent of the retail categories have actual sales greater than prediced on the basis of historical trends, whereas only 20 percent have no statistical difference between the predicted and actual values. For US-59, 60 percent of the retail categories have actual sales greater than predicted, whereas 30 percent have no statistical difference between the two values, and 10 percent of them have actual sales lower than predicted. Considering the sum of all retail sales, the actual value is statistically greater than the predicted sales for the Southwest Freeway. For IH-10, all retail categories have actual sales greater than predicted. Table 1 gives a summary of these three tables, along with the net effect of the construction on US-59.

As Table 1 indicates, four retail groups—general merchandise, food stores, automotive outlets, and home furnishings—have been adversely affected by the construction on US-59. This is to be expected, given customer mobility, the wide range of consumer choices available, and the difficulty in accessing the businesses on US-59. However, once construction is complete, such businesses generally benefit from the increased traffic passing their business sites. Agencies seeking to minimize these adverse effects should thus work closely with these specific business types, targeting them for special planning considerations.

ANALYSIS OF THE BUSINESS SURVEY

Introduction

Business premises, when adjacent to highway construction work, must weigh the prospect of short-term costs (sometimes claimed to be severe) against possible long-term financial benefits. This trade-off is a critical feature of a mitigation policy and, accordingly, must be evaluated by the planning agency. But there exists in the literature very little information about such effects; moreover, the TxDOT team discovered little about the specific impact of US-59 work on Houston businesses. Despite their considerable effort to inform businesses and to keep them advised of developments, TxDOT did not uncover much hard evidence about economic effects. We therefore decided to conduct a comprehensive business survey that could quantify some of these effects on the Southwest Freeway project; for this, a random sample of the businesses

TABLE 1 Summary of Analysis Results

Retail Trade Categories	Houston	IH-10	US 59	Net Effect using IH-10 as Reference
Building Materials	Similar	Greater	Greater	No effect
General Merchandise	Greater	Greater	Similar	Negative
Food Stores	Greater	Greater	Similar	Negative
Automotive	Greater	Greater	Similar	Negative
Clothing	Greater	Greater	Greater	No effect
Home Furnishings	Similar	Greater	Lower	Negative
Restaurants	Greater	Greater	Greater	No effect
Drug Stores	Greater	Greater	Greater	No effect
Liquor Stores	Greater	Greater	Greater	No effect
Miscellaneous	Greater	Greater	Greater	No effect
TOTAL	Greater	Greater	Similar	Negative

located on the US-59 corridor was selected as a way of representing the views of all businesses adversely affected (*3*).

Business Survey Methodology

There are several ways to obtain the opinions of business managers. One is to mail the survey, requesting that they be completed and returned for analysis. Another way involves direct interview. For this study, we decided to interview business owners directly where possible. When the business owner or manager was unavailable, we left the survey (along with a self-addressed envelope) to be filled out and returned by mail.

The first step was to develop a draft questionnaire. With feedback from two professionals (one a business expert, the other an expert in survey design), the questionnaire was modified and then pilot tested at three different sites in Austin; a modified version of the questionnaire was then prepared and customized to suit the Southwest Freeway project.

At first, a 2-day visit was planned to test the questionnaire on 10 businesses that were contacted and informed in advance of the surveyor's intent. This 2-day visit was organized as follows: for the first day, as many of the businesses as possible were visited up until 4:00 p.m.; after that, the research team went to the base hotel to make phones calls using the users' survey previously discussed; the next day the study team went back to some of the businesses interviewed on the first day to conduct new interviews. At that point, it was clear that some form of sampling was necessary to characterize the large number of businesses located along the 12-mi section. Accordingly, the team estimated the population of active businesses by driving along the frontage road, canvassing and categorizing each business passed.

The next step was to analyze the initial results and modification of the questionnaire. Four additional 2-day visits were made to obtain a sample approximating 20 percent of the total number in the construction area. The last two steps in the methodology included the final analysis of the results and the establishment of the main conclusion of the study.

Identification of the Sample of Businesses Interviewed

Of the 337 businesses counted in the drive-by survey, about 118 were contacted (35 percent). Of all those contacted, 74 (63 percent) responded either by completing the survey immediately or by mailing it to the return address. About half the businesses surveyed filled out the survey immediately.

Table 2 compares the answers received (sample) with the population of the businesses on the Southwest Freeway. Our intent was to survey the 10 different retail categories used for the sales analysis, described in the previous section, and to collect as much data on services as resources would allow. A sample of more than 20 percent was collected for each of the retail categories.

Analysis of the Results

The following analyzes the answers given to the questions asked in the Southwest Freeway survey. The survey form had four parts: (a) general, (b) during construction, (c) after construction, and (d) final comments.

General

About 72 percent said they rented, whereas about 28 percent said they owned their facility. This issue is especially important for those highway projects involving the acquisition of

TABLE 2 Classification of Businesses Existing and Surveyed Along the Southwest Freeway from Shepherd to Beltway 8

Category	Existing	Surveyed	Surv./Exist.
1. Building materials	6	3	50.0 %
2. General merchandise	57	19	33.3 %
3. Food stores	4	2	50.0 %
4. Automotive retails	67	18	26.9 %
5. Clothing stores	20	8	40.0 %
6. Home furnishings	16	7	43.8 %
7. Restaurants	33	8	24.2 %
8. Drug stores	1	1	100.0 %
9. Liquor stores	1	1	100.0 %
10. Miscellaneous stores	30	0	0.0 %
Retail-Subtotal	**235**	**67**	**28.5 %**
1. Travel	3	0	0.0 %
2. Health	36	1	2.8 %
3. Hotel	12	3	25.0 %
4. Insurance	6	0	0.0 %
5. Governmental	2	0	0.0 %
6. Office buildings	29	2	7.0 %
7. Business services	9	0	0.0 %
8. Malls	4	1	25.0 %
9. Churches	1	0	0.0 %
Service-Subtotal	**102**	**7**	**6.9 %**
TOTAL	337	74	22.0 %

right-of-way. About 17 percent of the businesses responded that they had been at that location less than 2 years—that is, they started operating after the beginning of the construction. About 22 percent had been there from 2 to 6 years. About 24 percent had been there from 6 to 10 years, and 37 percent had been there 10 years or more.

Around 23 percent of the individuals were informed by a letter sent to them by TxDOT, whereas 19 percent were notified in person by a TxDOT representative. The media played an important role in the construction coverage: 36 percent of the businesses knew about the construction from the media (24 percent from Houston newspapers, the rest from watching television). These responses answered affirmatively the question related to the importance of a public relations officer. About 76 percent of the businesses were informed about the construction directly or indirectly by TxDOT personnel. About 16 percent considered the communication to be very effective, around 20 percent said it was good, 24 percent said it was normal, and 40 percent said it was poor. Again, these responses tended to confirm the presumed importance of a public relations officer in a mitigation strategy.

During Construction

About 49 percent said that they were affected considerably by the construction. About 32 percent said that they were somewhat affected, about 8 percent said that the effect was minor, and 11 percent said that the construction did not affect them at all. These responses helped to confirm the notion that some businesses are affected more than others. Table 3 summarizes the responses to this question. Almost 47 percent of the respondents considered the economy in Houston a major problem for their business. Some even blamed their poor business performance on the economy and not on the construction. About 34 percent of the respondents thought that the construction imposed the only effect on their business. This question had a wide range of answers. Some respondents said that the construction started 2 years ago, whereas others said that it lasted for 6 months. These discrepancies derive from the fact that construction activity was ongoing on the main lanes even after the frontage road had been completed.

Table 4 shows the responses to how sales were affected during construction. Most of the respondents gave percentages for this answer, like "30 percent down" or "I estimate a 20 percent decrease in sales." Some refused to answer this question because they did not want to reveal their sales records. These responses confirmed that some businesses are affected more than others. Six percent reported improved sales owing to the construction, about 60 percent of the businesses reported that sales were down by less than 20 percent, and only 12 percent reported a sales drop greater than 40 percent.

Only 24 percent reported that they reduced the number of employees. Answers to this question included "Did not affect much," "Employees arrive late," and "Laid off two employees." In general, it can be said that the abutting business employee rate did not significantly change as a result of the construction.

The question seeking to identify what TxDOT had done to ease adverse effects evoked a range of answers. These answers were categorized according to the following: communicating, expediting, keeping access open, directing traffic, working at night, putting up signs, making things worse, and nothing (see Table 5).

About 27 percent of the business owners who answered this question said TxDOT communicated with them. (They noted that this communication was appreciated and needed.) Conversely, 55 percent felt that the highway agency did nothing to ease their inconvenience. About 50 percent of the businesses thought that TxDOT was doing something to mitigate the impact on their businesses, whereas the other 50 percent thought that they were doing nothing.

Businesses were then requested to comment on how they mitigated the negative effects of construction. Table 6, which gives answers to this question, indicates that businesses usually advertised, redirected customers, and tried to encourage customer visits with sales and free pickup and delivery; they also requested that TxDOT keep the feeders and business entrances opened.

TABLE 4 Effect on Sales During the Construction Period

Effect	Number	Percentage
Improved	4	6.1 %
No effect	11	16.7 %
Dropped by 10%	8	12.1 %
Dropped by 10% down to 20%	15	22.7 %
Dropped by 20% down to 30%	13	19.7 %
Dropped by 30% down to 40%	7	10.6 %
Greater than 40% drop	8	12.1 %
Total	66	100.0 %

TABLE 3 Other Internal or External Factors, Besides Construction, That Could Affect Business Activities

Factors	Number	Percentage
Economy	30	46.9 %
Business Strategy	4	6.3 %
Seasonal Variation	4	6.3 %
Competition	2	3.1 %
Media	1	1.6 %
Persian Gulf conflict	1	1.6%
No other Factors	22	34.4 %
Total	64	100.0 %

TABLE 5 Highway Agency Actions To Make Things Easier for Businesses During Construction

Action	Number	Percentage
Communication	17	26.6 %
Expedited	3	4.7 %
Kept access open	2	3.1 %
Directed traffic	2	3.1 %
Worked at night	2	3.1 %
Put up signs	2	3.1 %
Made things worse	1	1.6 %
Did nothing	35	54.7 %
Total	64	100.0 %

TABLE 6 Business Mitigation Strategies

Action	Number	Percentage
Advertise	8	12.5 %
Redirect customers	8	12.5 %
Erect signs for entrances and exits to property	7	10.9 %
Use alternate routes in and out of property	5	7.8 %
Have sales	4	6.2 %
Free pick up & delivery	3	4.7 %
Left home earlier	3	4.7 %
Keep in touch with Highway Department	2	3.1 %
Others	8	12.5 %
Nothing	16	25.0 %
Total	64	100.0%

TABLE 7 Construction Impact on Businesses After Completion of Feeder

Construction Impact	Number	%
No improvement	16	33.3
Business steadily and dramatically picked up	14	29.2
Improved slightly	11	22.9
Construction still going	4	8.3
Helped initially till exits were closed	2	4.2
Business back to normal	1	2.1
Total	48	100.0

After Construction

We next asked what happened after the feeder road construction was completed, and answers are given in Table 7. About one-third of the business owners interviewed noticed no improvement. Others said the construction was still there and the feeder was not yet finished. About 29 percent said that their business improved steadily and dramatically. Only 2 percent characterized their business as being "back to normal." The increase in sales varied from 5 to 30 percent (Table 8). These responses tended to confirm that phased construction reduces business effects. Our survey suggests that the phases undertaken by TxDOT proved to be an effective strategy for mitigating the impact on businesses. By adopting a strategy in which initial operations targeted the lanes directly in front of the businesses, TxDOT allowed these businesses to start receiving the benefits of the rehabilitation before the end of the project.

Final Comments

We asked business owners whether they would consider selling temporary access rights to the contractor and closing during an expedited construction period. Most of the businesses expressed the fear that, if they closed down, customers would seek retail products and services at other locations; thus, about 84 percent said that they would not consider closing down. Only 9 percent said that they would consider closing. The results of the business survey analysis, divided into seven different topics, are summarized in Table 9.

CONCLUSIONS OF THE HOUSTON BUSINESS SURVEY

We now return to the five questions posed earlier in this paper. Road construction can clearly affect sales of abutting businesses; common sense, anecdotal evidence, and the small number of research publications conclusively demonstrate this. This study shows, however, that highway agencies can adopt a range of policies, from construction techniques to closely working with adjoining businesses, to mitigate these effects. In this regard the US-59 project was a great success. Second, construction activities in a major corridor of a large city have a negative effect on a selection of abutting businesses but

TABLE 8 Effect on Sales After Completion of Frontage Road

Effect	Number	Percentage
No positive effect	17	34 %
Sales up 30%	4	8 %
Sales up 25%	3	6 %
Sales up 20%	1	2 %
Sales up 15%	4	8 %
Sales up 10%	5	10 %
Sales up 5%	6	12 %
Back to Normal	2	4 %
Still Dropping	6	12 %
Construction still ongoing	2	4 %
Total	50	100 %

TABLE 9 Summary of Business Survey Results

Topic	Results
Starting of the business	• 17% of the businesses started after the construction began
Information about the construction	• 42% directly from the TxDOT (letters in person) • 36% indirectly from the TxDOT (newspaper and TV) • 10% others • 12% nobody
Effectiveness of the communication	• 60% good- normal • 40% poor
Effect of the construction	• 19% of the businesses said that the construction had little or no effect on their business
Effect on sales during the construction	• 6% improved • 17% no effect • 77% drop
Additional mitigation strategies	• 49% of the businesses said that TxDOT could not do anything differently or do not know
Effect on sales after the feeders were opened	• 34% no effect • 62% positive effect

have no overall regional effects. Construction effects are highly selective, which links to the third basic question posed in the introduction. In particular, general merchandise (−28 percent sales), food stores (−37 percent sales), automotive outlets (−32 percent sales), and home furnishings (−17 percent sales) were found to be particularly vulnerable. Clearly, those projects having a dedicated public relations person should focus attention on these retail categories, which should help reduce adverse business effects. Researchers also believe that the carefully phased construction plans prepared by the agency helped to reduce the negative effects on businesses.

Of course, there is always a problem with drawing general rules from a single (although admittedly large) project. More research needs to be undertaken to develop fully transferable results. However, lessons learned from this study were later applied to a traditional nonexpedited bridge project on I-30 near Dallas, where considerable business complaints had been received. Target businesses were counseled, changes were made to some construction phasing, and traffic management was undertaken. Applying some lessons learned from US-59 led to a more harmonious relationship between the contractor and the abutting business community.

It is likely that an increasing number of urban projects will need to be planned on a systems basis, in which user and business effects are explicitly recognized at every stage in the process. This study of an expedited project shows that this approach can be extremely beneficial to abutting businesses and that many of the inconveniences feared by businesses with respect to urban construction activities can be reduced to tolerable levels.

However, as with any human enterprise, there is a large degree of intrinsic variability that can never be adequately accounted for in a construction project. It is likely, for example, that some businesses will always fail through either mismanagement or insufficient capital, no matter how solicitous the highway department's construction program. Moreover, some highway planners and some public relations officers are more effective than others, although all may adopt an expedited approach. Yet the Southwest Freeway project, undertaken in one of the densest commercial areas in the United States, shows that an expedited approach has the potential to minimize the negative effects that urban highway construction sometimes imposes on adjoining businesses.

REFERENCES

1. J. L. Buffington, W. F. McFarland, J. C. Memmott, and K. N. Womack. *Estimated Impact of Widening U.S. Highway 80 (Marshall Avenue) in Longview, Texas.* Draft Technical Report 1, Research Study 01990. Texas Transportation Institute, Texas A&M University, College Station, Sept. 1987.
2. *Highway Reconstruction and Repair Impact Study.* Legislative Report. Wisconsin Department of Transportation, Nov. 1989.
3. H. de Solminihac. *System Analysis for Expediting Urban Highway Construction.* Ph.D. dissertation. The University of Texas at Austin, May 1992.

Publication of this paper sponsored by Committee on Social and Economic Factors of Transportation.

Economic Impact of Highway Bypasses

S. JOHANN ANDERSEN, HANI S. MAHMASSANI, REIJO HELAAKOSKI, MARK A. EURITT, C. MICHAEL WALTON, AND ROBERT HARRISON

The economic effects of bypasses are addressed from several methodological perspectives. A data base for bypassed cities in the state of Texas was developed. A control city was chosen for each bypassed city in the sample to control for the effect of the bypass. Econometric models were developed to relate retail sales, gasoline sales, restaurant sales, and service receipts to the pertinent characteristics of the area. The models showed that a bypass generally brought a small, but statistically significant, decrease to business volumes in bypassed cities. Cluster analysis, a multivariate statistical procedure, was used to explore a possible grouping of cities that can help predict the economic impact of bypasses. Cluster analysis emphasized the importance of the economic base of a city, as captured in the geographic regions in which the city is located. Inclusion of the regional cluster variables in the econometric models improved the specification and predictive abilities of the models. Beyond formal analysis, individual case studies showed that local communities might not necessarily perceive bypasses as negative but as one of many factors contributing to the overall performance of a city in a rural setting.

Highway bypasses have long provided a practical approach to improving transportation levels of service through small cities in primarily rural settings by rerouting through traffic around such cities.

Road investment in highway bypass construction normally produces benefits for road users in the form of reduced journey times and vehicle operating costs and an improvement in safety. It reduces environmental nuisance from traffic to residents and pedestrians along the bypassed roads. In addition, highway transportation projects such as bypass construction normally produce local economic impacts of the following nature: (a) the creation of jobs and subsidy revenue from facility planning, construction, and operations; (b) the indirect impact of increased production because of reduced transportation costs and delays; and (c) the indirect impact of all of the above on nonusers because of the multiplier effect.

The construction of bypasses, however, has not always met with unanimous approval. Communities have feared that their economies would be adversely affected by the highway bypass construction. Business interests in the bypassed cities have generally resisted efforts to build bypasses in the belief that large numbers of customers would be diverted from the business district, thereby impairing the community's economic health. In voicing their concern, communities have raised the following questions: Does the economy of the bypassed city suffer from these new highways? Are retail sales harmed by bypassing? What specific types of businesses are harmed, if any? What are the temporary effects and what are the long-term economic effects? For the community as a whole, what is the net effect of the highway bypass on economic activity?

Several methodological perspectives are used in this paper to address the economic effects of bypasses. A case study analysis of several Texas cities provides insights that form a basis for more formal analyses with greater applicability. A data base is established, containing data on pertinent variables for both bypassed cities and control cities. Control cities are introduced to control for the effect of the bypass. Econometric models are used to identify economic effects of highway bypasses on business activities by examining both highway-related and non-highway-related factors. Cluster analysis, a multivariate statistical procedure, is used to explore the possible existence of an underlying structure within the bypassed cities. The results of the clustering process are then used to improve the specification of the econometric models.

BACKGROUND REVIEW

Historically, transportation has been a vital component in almost every aspect of economic development. The traditional view in the literature has been that the improvement of the transportation infrastructure is a necessary precursor to economic development in a region. Some researchers have found a significant relationship between highways and economic growth. However, a summary of economic impact studies made in the 1970s and 1980s generally concludes that many other factors besides highway improvements come into play to affect regional growth (*1*). In well-integrated economic systems, the effects of transportation improvements are complex and difficult to predict.

Several highway bypass studies have explored the economic effects on small communities (*2–5*). These effects can take many forms, such as a drop or increase in retail sales, employment, or personal income. Highway bypasses have been reported to have seriously affected highway-oriented businesses (i.e., those providing fuel, food, and accommodations for travelers). To remain competitive, service stations and restaurants often have successfully adjusted their merchandise and their methods of operation to attract local trade. In a review of several Texas highway bypass studies, Skorpa et al. (*6*) found it difficult to draw a relationship between highway bypass construction and changes in local business volumes. In almost all cases, the non-traffic-oriented businesses had experienced increases in annual gross sales, whereas many traffic-serving businesses, such as service stations and motels, showed large decreases.

S. J. Andersen and R. Harrison, Center for Transportation Research, The University of Texas at Austin, Suite 200, 3208 Red River, Austin, Tex. 78705. H. S. Mahmassani, M. A. Euritt, and C. M. Walton, Department of Civil Engineering, The University of Texas at Austin, Austin, Tex. 78712. R. Helaakoski, Suunnittekolmio Oy, Vetehisenkuja 3, 00530, Helsinki, Finland.

The economic impact of a bypass on a city in a rural setting must also be seen against the background of ongoing non-highway-related economic and social changes. The continuing centralization of trade and economic and social relationships has diminished the importance of many small towns (*7*). The ability of a community to retain its residents is largely dependent on its economic base (*8*). For instance, agricultural communities have experienced a nearly steady loss of population over the past two decades as a result of increased mechanization and a shift to corporate-owned farms.

It is evident that the effects of bypass construction are neither conclusive nor uniform across locations. Many factors influence the economy in a given area, severely complicating one's ability to establish simple cause-and-effect relationships and limiting the ability of any one methodological approach to uncover reliable results. Several methodological perspectives are often necessary to obtain robust results and form conclusions with a reasonable level of confidence. In this study, results from econometric modeling, cluster analysis, and exploratory data analysis are combined with the findings of more qualitative case studies to obtain insights into the economic effects of bypasses as well as into the factors likely to influence the direction and magnitude of these effects.

DEFINING THE SAMPLE

Highway bypasses in Texas were inventoried and categorized to identify those bypasses relevant to the study's objectives. Highway bypasses in Texas can be categorized according to highway characteristics, geographical location, population characteristics, and year of construction. The segment of a new highway intended to reroute through traffic around a central business district constitutes, for this study, the working definition of a bypass. A sample of 23 Texas cities, with bypasses conforming to this definition, was obtained (*9,10*). Interstate bypasses are excluded from this analysis, because the Interstate system is largely in place and future bypass construction will involve mainly state and U.S. highways. It is also postulated that the characteristics of the road users on the Interstate system are different from those using other highways.

It was decided to compare changes in the economies of the bypassed city with the changes (over the same period) in selected control areas. A control city was selected for each bypassed city so that both cities would ideally share the following common characteristics: highway district, proximity to a larger city, economic base, magnitude and trend of retail sales, population size category and growth trend, and highway network characteristics.

DATA BASE

A data base was established by assembling data on pertinent variables for all cities to form the basis for further analysis. Distinction is made between dependent and explanatory variables.

Several measures were used as dependent variables intended to capture the changes in business activity. Typically, total retail sales is used as a short-term indicator reflecting the economic viability of a city. Service receipts are also used as a short-term indicator, representing a sector providing a variety of services for individuals, businesses, and government establishments. Also, it is expected that a bypass should specifically affect highway-oriented businesses. To this effect, data for gasoline sales, restaurant sales, and hotel/motel receipts are pertinent. However, data on hotel receipts were available only when there were more than three hotel/motel establishments in the city, which was rarely the case with this sample of small cities.

The explanatory variables in the data base reflect the demographic, geographic, economic, and highway characteristics of each city. Influential explanatory variables cited in earlier studies include the population of the bypassed city, the distance to a comparably sized city or a larger city, the distance between the old and the new routes, and the number of state and U.S. highways (*11*). Explanatory variables, together with the dependent variables, are given in Table 1.

Summary descriptive statistics were calculated for the variables in the data base to examine the overall similarity of bypassed and control cities (see Table 2). Several tests were performed (*9*), including tests of the differences in the means of the key variables between the bypassed cities and the control cities before the opening of the bypass. The differences between bypassed and control groups are minor and insignificant except for the variable ADT-TOTAL (average daily traffic on incoming highways). Higher traffic volumes in the bypassed cities were, presumably, one of the reasons to construct the bypass in the first place. Descriptive statistics confirmed the similarities of bypassed and control cities in the period preceding the opening of a bypass.

The pertinence of variables and their importance in establishing a relationship between highway bypasses and business activity are explored further in the rest of the paper by following several methodological approaches.

CASE STUDY METHODOLOGY

The case studies involve six cities in Texas (Navasota, Grapeland, Taylor, Alvord, Bowie, and Littlefield) with different population and economic characteristics. The study methodology included the following: a review of each city's history and economy; the tracking of changes in the spatial distribution of highway-oriented businesses in the city; and a site visit including interviews with local business people. The interviews focused on the following topics: economic viability of the city, effect of the bypass on businesses, adjustment to the bypass, opinions regarding the desirability of the bypass, downtown improvement programs, land use changes, and traffic characteristics. A full account of the case studies is presented by Helaakoski et al. (*9*) and Andersen et al. (*10*).

The site visits and interviews helped elucidate much of the inner functioning of small cities in rural areas. The subjective input provided by local people during the visits shed some useful light on the perceived effects of bypasses. The key findings are summarized as follows.

1. In general the bypass is not perceived by local residents to have had a devastating impact on any of the communities that were visited. The case studies do not suggest a strong

TABLE 1 Variables in the Data Base

Variable Name	Description
Dependent Variables	
SALES	Total Retail Sales within city, in 1987 dollars
GAS STATION SALES	Gas Station Sales within city, in 1987 dollars
RESTAURANT SALES	Restaurant Sales within city, in 1987 dollars
SERVICE RECEIPTS	Value of Services provided to city, in 1987 dollars
Explanatory Variables	
POPULATION	within city boundaries
DLARGER	distance in miles to a city of larger size
METRO-AREA	=1, if the city is located in a metropolitan area, otherwise 0
INCOME	average personal income per capita in the county in 1987 dollars
GROWTH1	the growth rate per capita of real GNP in the USA during the period between year t and t-1
GROWTH5	the growth rate per capita of real GNP in the USA during the period between year t and t-5
US	number of incoming US highways to the city
STATE	number of incoming state highways to the city
YEAR'[a]	overall trend in gas station sales
ADT-TOTAL	average daily traffic volumes on incoming highways
LENGTHOLD	length of the old bypassed route in miles
LENGTHNEW	length of the bypass in miles
DISTANCE	the average distance in miles between a bypass and a bypassed route
C1 through C24	city specific dummy variables for bypassed cities
C101 through C124	city specific dummy variables for control cities
CLASS	classification of the bypass (US highway=1, state =0)
ADT-BYPASS	Average daily traffic volumes on the bypass
ACCESS	access type for the bypass (=1, if a bypass has limited access and grade separation, otherwise 0)

[a] $YEAR' = -3074 + 111.5*YEAR - 0.8283*(YEAR)^2$

relationship between a bypass and economic growth. Other factors, such as fluctuation in the agriculture or oil business, continuing urbanization trends, and establishment of large discount stores within the market area, have a much stronger effect on local businesses.

2. Local business and political leaders can exert a strong influence on a local community and businesses and their evolution after bypass opening.

3. Spatial changes are often confined to increased activity toward and at the point at which another highway intersects the bypass. Few establishments were found at the split between the bypass and the bypassed route.

4. The removal of a portion of through traffic from the downtown streets, especially heavy vehicles, is seen in a positive light. Improved safety and cleaner air are perceived as the most important benefits.

5. Downtown businesses have typically experienced a drop in sales after the opening of the bypass. However, this drop was in many cases temporary, as business owners restructured their stores or reoriented their businesses. Many gas stations have closed on the bypassed route, corresponding to general declining trends as a result of industry restructuring nationally.

The case study findings agree to a large extent with those from previous case studies (*5*) and others referenced by Helaakoski et al. (*9*) and Andersen et al. (*10*).

ECONOMETRIC MODELING

One of the purposes of this research is to develop a qualitative predictor of business activity that would capture the effect of the underlying determinants of such activity and allow formal testing of hypotheses pertaining to the relative effects of various such determinants. Multivariate regression models are developed to explain the following measures of business activity: total retail sales, gas station sales, restaurant sales, and service receipts in small cities. The models for each of the four dependent variables have the following usual linear form:

$$Y_{it} = bX_{it} + e_{it}$$

TABLE 2 Means, Standard Deviations, and Medians for Dependent and Key Explanatory Variables Before Bypass Was Opened

	BYPASSED CITIES			CONTROL CITIES		
Variables	Mean	Std Dev	Median	Mean	Std Dev	Median
Total Retail Sales / Person	6,783	2,165	6,249	6,494	1,837	6,003
Gas Station Sales / Person	576	199	532	587	188	561
Restaurant Sales / Person	269	110	254	268	98	246
Service Receipts / Person	494	202	459	500	282	466
Income / Person	5,264	1,549	5,323	4,890	1,353	4,934
Population	6,981	3,974	6,142	6,088	3,812	5,459
Distance to Larger City	26	12	24	29	11	29
Number of Highways	4.1	1.1	4	3.8	1.1	4
ADT, all incoming highways	13,630	5,660	13,490	10,220	4,440	9,040

where

Y_{it} = measure of business activity (total retail sales, gas station sales, restaurant sales, or service receipts) for a city in year t;
X_{it} = vector of explanatory variables for city i in year t;
b = vector of parameters to be estimated; and
e_{it} = error term of the usual type, with mean 0 and constant variance.

The vector of explanatory variables X_{it} consists of the kind of variables included in the data set and shown in Table 1. It may also include city-specific binary variables that capture city-related differences in culture, such as base of economy and geography, that change very slowly over time and that are not captured well by the other explanatory variables in the model. The vectors of parameters b were estimated using ordinary least squares.

Total Retail Sales Model

Retailing is generally the most important component of the local business infrastructure in small cities. The specification and associated parameter estimates of the model developed to explain retail sales are as follows (t-statistics are reported in parentheses):

$$\begin{aligned}\text{SALES} = & -14{,}495 + 5.561\text{POPULATION} \\ & (-5.99) \qquad (22.84) \\ & + 0.576\text{INCOME} + 3{,}027\text{LARGERCITY} \\ & (1.41) \qquad (1.81) \\ & + 1.305\text{ADT-TOTAL} - 12{,}402\text{ACCESS} \\ & (9.76) \qquad (-4.91) \\ & + 31{,}470\text{C22} - 44{,}747\text{C23} + 15{,}186\text{C101} \\ & (5.88) \qquad (-7.09) \qquad (3.21)\end{aligned}$$

The most significant variable is POPULATION, as expected, since more residents generate more sales. A simple regression analysis performed for retail sales shows that a relatively high 74.4 percent (R^2) of total variation can be explained by the POPULATION variable alone. Theoretically, INCOME is considered as one of the most important explanatory variables. This variable is less significant than perhaps expected. However, in this specification INCOME is taken as the average income per capita over a whole county and may not entirely reflect buying power in a small city within the county.

The distance between a given city and the nearest city of equal or larger size (DLARGER in Table 1) is expected to exert a positive effect on business volumes because the further away the larger city, the less pull there is for residents to shop away from their own city. This attribute was specified as a binary indicator variable to reflect the finding that the distance to a larger city has a positive effect on retail sales only if such a city is situated at least 20 mi away. If a larger city is very close, it is easy for shoppers to drive a few miles and thereby reach a greater variety of shops. In this model LARGERCITY is a binary indicator variable equal to 1 if the distance to a larger city is 20 mi or more; 0 otherwise.

Two traffic-related attributes are included in the model: ADT-TOTAL (average daily traffic volumes on incoming highways) and the bypass variable, ACCESS. The estimated value of the coefficient of this attribute shows that a bypass has a significantly negative effect on total retail sales in cases in which the geometric characteristics of the facility provide for limited access from adjoining property. The estimated coefficient of the ACCESS variable indicates that the decrease in sales is on average about 20 percent per city in the cases in which access is limited on the bypass. This value was obtained by applying the estimated model, using the sample mean for each variable and the estimated coefficients to calculate the corresponding value of the dependent variable (*9*). Ten cities in the sample have bypasses with limited access.

To decrease autocorrelation and to control for intercity differences, the most significant dummy variables were included in the model. City-specific dummy variables C22 (Silsbee), C23 (Edinburg), and C101 (Clarksville) were found to be statistically significant. With these dummy variables, it was possible to determine part of the intercity differences not captured by other explanatory variables in the model.

Highway-Oriented Business

Gas Station Sales

The estimated parameters and corresponding t-statistics of the selected model specification are as follows:

$$\begin{aligned}\text{GAS STATION SALES} = & -4{,}390 + 4.596\text{YEAR}' \\ & (-5.74) \qquad (4.69) \\ & + 0.438\text{POPULATION} + 0.205\text{INCOME} \\ & (15.05) \qquad (4.14) \\ & + 182\text{HIGHWAYS} + 0.0544\text{ADT-TOTAL} \\ & (2.37) \qquad (2.70) \\ & - 0.131\text{ADT-BYPASS} + 2{,}344\text{C13} \\ & (-2.98) \qquad (3.96)\end{aligned}$$

As expected, POPULATION and INCOME are significant attributes. The nature of gasoline station sales also explains the significance of two highway-related variables in the model: the number of incoming highways (HIGHWAYS) and the traffic volumes on these highways (ADT-TOTAL). Higher traffic volumes will definitely cause a higher volume of gasoline sales.

An overall gas station sales trend (YEAR′) was included as an explanatory variable and proved to be statistically significant. It captures exogenous influences unrelated to bypass effects or other local factors. The trend was calibrated for the control cities to avoid contamination by bypass effects. For the cities studied, gas station sales reached a peak in the late 1960s. A reversal of the trend is caused mainly by more energy-efficient automobiles.

The final attribute included in the model is traffic volume on the bypass, ADT-BYPASS. It is statistically significant and indicates that a highway bypass has a negative effect on overall gasoline station sales in the sampled cities. This variable indicates, in principle, that the more traffic that is diverted to the bypass from the bypassed route, the lower the gas sales that can be expected in the city. It can be estimated, on the basis of the mean values of the variables and the corresponding estimated coefficients, that a highway bypass causes on the average about a 15 percent decrease in gasoline station sales in a small city.

Restaurant Sales

The model for restaurant sales is as follows:

$$\begin{aligned}\text{RESTAURANT SALES} &= \underset{(-9.68)}{-1{,}827} + \underset{(17.33)}{0.366\text{POPULATION}} \\ &+ \underset{(1.94)}{0.062\text{INCOME}} - \underset{(-1.77)}{296\text{METRO-AREA}} \\ &+ \underset{(8.57)}{0.016\text{ADT-TOTAL}} - \underset{(-2.42)}{0.0674\text{ADT-BYPASS}} \\ &- \underset{(-3.54)}{1{,}704\text{C23}} - \underset{(-2.75)}{1{,}022\text{C112}} + \underset{(4.51)}{1{,}745\text{C113}}\end{aligned}$$

Again, POPULATION is the most significant variable. Also significant are ADT-TOTAL and INCOME. The only new variable is METRO-AREA, which appears to have a significant negative effect. This binary variable is equal to 1 if the city is located in a metropolitan area; 0 otherwise. Apparently, a greater variety of restaurants in a nearby large metropolitan area reduces sales in small cities for reasons similar to the retail sales findings.

The highway bypass–related variable, ADT-BYPASS, is found to have a significant negative effect on restaurant sales. Additional calculation based on the mean values of the variables and the corresponding estimated coefficients suggests that a highway bypass is associated on the average with a 10 to 15 percent decrease in restaurant sales in a small city.

Service Receipts

In developing the model to explain service receipts, the variables were transformed by taking their natural logarithms (ln-ln), eliminating heteroskedasticity and giving a significantly better R^2 value. The model is stated hereafter.

$$\begin{aligned}\ln(\text{SERVICE RECEIPTS}) &= \underset{(-12.08)}{-8.78} + \underset{(7.03)}{0.0243\text{YEAR}} \\ &+ \underset{(15.99)}{1.022\ln(\text{POPULATION})} + \underset{(4.35)}{0.388\ln(\text{INCOME})} \\ &+ \underset{(2.26)}{0.00303\ln(\text{NEARBYCITY})} \\ &+ \underset{(4.36)}{4.36\ln(\text{ADT-TOTAL})} - \underset{(-1.54)}{0.116\text{ACCESS}} \\ &- \underset{(-4.34)}{0.671\text{C6}} + \underset{(3.55)}{0.545\text{C118}} - \underset{(-5.08)}{0.870\text{C119}}\end{aligned}$$

As expected, POPULATION is again the most significant variable, although this time population alone explains only 48.8 percent of the variation in service receipts. INCOME level also has a significant influence on service receipts. In this model NEARBYCITY is introduced as a new variable, taking the value of 0 if the distance to the nearest larger city is less than 20 mi, otherwise, it is set to the value of the variable DLARGER, the distance in miles from a bypassed or control city to the larger city. The new variable NEARBYCITY means that the geographical location of a city leads to higher service receipts in a small city only if a larger city is more than 20 mi away. With longer distances, the positive effect still increases gradually. Furthermore, service receipts also apparently are a traffic-related phenomenon, as can be seen from the significance of the ADT-TOTAL variable. Also, an overall increasing trend captured by the linear variable YEAR was found to be significant.

The last traffic characteristic found to have a significant effect on service receipts is the variable ACCESS. This variable indicates that a bypass has a significantly negative effect on service receipts in cases in which the geometric characteristics of the facility provide for only limited access.

In addition to the model development presented above, the hypothesis was examined that cities with larger populations have a somewhat better chance of adjusting to economic changes that may be induced by the bypass (*9*). It was concluded that the negative effect of a highway bypass on total retail sales and highway-oriented sales has about the same significance for cities of less or more than 6,000 inhabitants. In small cities, a highway bypass does not have a significant negative effect on service receipts, whereas large cities are found to suffer losses because of a bypass. Finally, it was found that the econometric models developed can be used in predictions with fairly reasonable accuracy (*9*).

CLUSTER ANALYSIS

Cluster analysis, a multivariate statistical procedure, involves the grouping of entities that are similar to one another. This problem is frequently stated as one of finding the "natural groups." Cluster analysis may be used as a tool to explore and reveal structure and relations in the data. Measures of similarity or distance between entities are computed. Different heuristic clustering methods can then be used to obtain the various groupings (*12*). Cluster analysis is used in this study to identify some underlying structure within the set of bypassed cities. This is done by comparing clusters formed for bypassed cities with those formed for the control cities. It is also used to define variables that may improve the specification and predictive ability of models similar to those discussed in the previous section.

Cluster Analysis Procedure

The complete linkage method, part of the family of hierarchical clustering methods, is used for this analysis. At each stage in this method, after clusters p and q have been merged, the similarity between the new cluster (labeled t) and some other cluster r is determined as follows:

$$s_{tr} = \max(s_{pr}, s_{qr})$$

The quantity s_{tr} is the distance between the most distant members of clusters t and r. If clusters were merged, every entity in the resulting cluster would be no farther than s_{tr} from every other entity in the cluster. The value of s_{tr} is the diameter of the smallest sphere that can enclose the cluster resulting from the merger of clusters t and r. The method is called complete linkage because all entities in a cluster are linked to each other at some maximum distance or minimum similarity.

Cluster analysis is performed separately for both the bypass set and the control set. In both cases the choice set consists of 23 cities. The explanatory variables given in Table 1 are utilized as variables for the clustering procedure, with the following additions: (a) variables representing prebypass growth and (b) the economic region each city is located in. These regions are the Plains, Metroplex, East Texas, Gulf Coast, Central Corridor, and the Border (*13*). Their locations are shown in Figure 1. The cluster variables as input to the cluster procedure are summarized as follows: population, growth in population, income, growth in income, distance to a larger city, total incoming traffic, growth in traffic, economic regions of Texas (Plains, East Texas, Border, Metroplex, Gulf Coast, and Central Texas), and access control.

Cluster Analysis Results

Results from the cluster analyses for both groups (bypass and control groups) appear to be very similar. Three clear clusters emerge for both groups. The geographical variable is the most important clustering variable. These clusters clearly represent the geographical regions of Texas. One cluster represents the Plains (or West Texas), which has a predominant agricultural economic base; another represents the oil- and petroleum-based economic region of the Gulf Coast region; whereas the third cluster represents the Metroplex and Central Corridor economic regions, with high-tech and other manufacturing forming the basis of the economy together with federal and state government and higher education. The East Texas and Border regions are represented by only one city each in the sample. The city of East Texas clusters with the Metroplex and Central Corridor group, whereas the border city stands on its own.

Inferences can be drawn about the effect of the bypass on a small city by comparing retail sales trends of bypassed and control cities. Both groups can be characterized by the retail sales trend for the prebypass period. The control cities were chosen on the basic premise of having the same retail sales trend for the period before the bypass was opened. The respective trends before bypass construction are therefore similar. The prebypass trend is extended for the postbypass period to yield the projected trend. Actual data points for the postbypass trend are then compared with the projected trend,

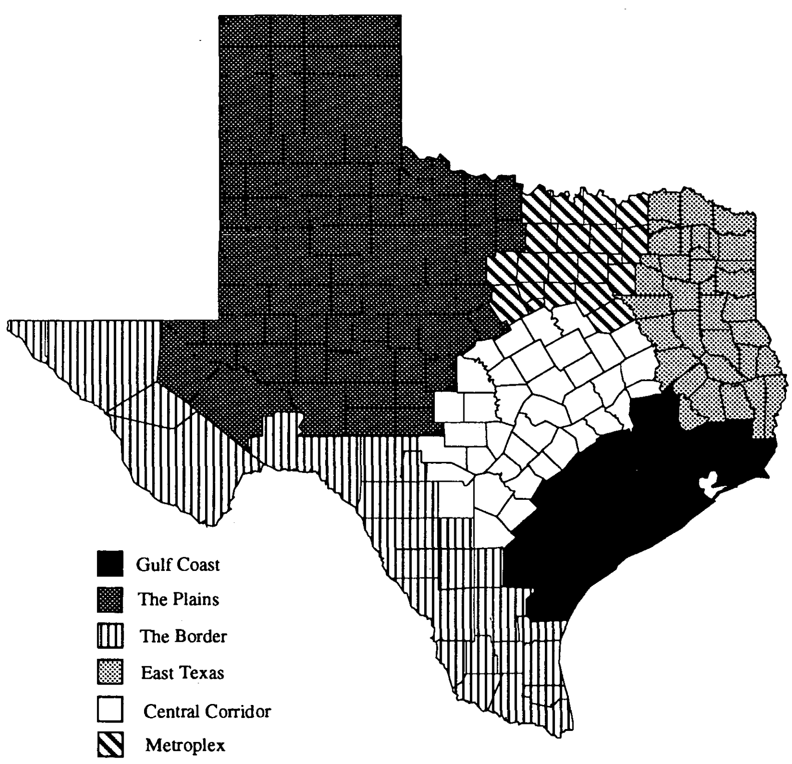


FIGURE 1 Economic regions of Texas.

and the difference is determined. The differences for the two groups are summarized by informal descriptive statistics [see Andersen et al. (*10*) for a full account of the analysis]. Comparisons are drawn between the bypass group and the control group as a whole. Also, corresponding clusters between the sets are compared. This analysis indicates that in all cases the differences for the bypassed groups are lower than the differences for the control groups, suggesting that the bypass has a small but negative effect on the sales volumes of a small city.

The cluster does not detect any new phenomena that might have been generated by the introduction of a bypass that will drastically change the characteristics of a city. The similarity between the clustering of the bypassed and control groups suggests that the control cities were, in most cases, well chosen. A comparison of retail sales trends between similar clusters of the two sets suggests that a bypass has a slight negative impact on sales volumes. It appears that business performance is intrinsically tied to the area's particular economic base. This question was explored further by incorporating results from the cluster analysis into the previously described econometric models.

FINAL MODEL DEVELOPMENT

Results of the cluster analysis and econometric modeling are now combined by introducing indicators for the economic regions as explanatory variables in the econometric model. By including these variables, instead of the city-specific variables, the applicability of the econometric models becomes more general. This effort also shows the consistency between various approaches and ties together the results of the overall research effort.

The economic regions were introduced individually into the models as well as in subsets as suggested by the cluster analysis. These variables are binary indicator variables equal to 1 if a city falls in a specific region or group; 0 otherwise. The effect of an economic region is not necessarily additive, because location in a particular region may influence the effect of other factors on retail sales, so various interaction variables were also introduced and tested, as discussed hereafter. All additional variables are defined in Table 3. Final regression models for the four business categories are compared with the initial models and presented in Table 4. A brief discussion of each category follows.

Total Retail Sales Model

To test whether the type of access control on the bypass affected retail sales differently in the various economic regions, separate coefficients for the ACCESS variable were estimated for each economic region by including appropriate interaction terms in the specification. An *F*-test of the restriction that all the coefficients of these interaction terms are equal did not lead to rejection of the null hypothesis at the 5 percent level of significance. Similarly, interaction with various subsets of economic regions (corresponding to the clusters obtained earlier) did not support the existence of differential effects of the bypass across economic regions or groups. The coefficient for ACCESS is negative and statistically significant, suggesting that a bypass has a negative impact on retail sales in all regions.

POPULATION was also interacted with the geographic region, resulting in statistically significant coefficients for POPPLAINS (a variable equal to the city's POPULATION if the city is in the PLAINS region, 0 otherwise) and POPGCM (similarly defined for the GULF, CENTRAL, and MPLEX regions taken as a group). This indicates that population size contributes differently to retail sales in different regions, possibly reflecting differences in purchasing power in the various regions. The restriction that the parameters for POPPLAINS and POPGCM are equal was rejected, indicating that purchasing power in the PLAINS region appears to be significantly lower.

Note that the INCOME variable is not statistically significant and thus is not included in the final model specification. This is likely because of the correlation between this variable and the regional variables, which capture varying income levels across the various regions.

By replacing the city-specific dummy variables with regionwide variables (without losing much explanatory power),

TABLE 3 Variables Added to Data Base

Variable Name	Description
Explanatory Variables	
ACCREGION	interaction variable between ACCESS and any economic region (=1, if a city has limited access and falls within a specific geographic region, =0 otherwise)
POPGMC	interaction between POPULATION and the GULF, CENTRAL, and MPLEX-regions as a group (=POPULATION, if city falls within any of these economic regions, 0 otherwise)
POPPLAINS	interaction between POPULATION and the PLAINS-region (=POPULATION, if city falls within PLAINS, 0 otherwise)
POPGC	interaction between POPULATION and the GULF and CENTRAL-regions as a group (=POPULATION, if city falls within either of these two regions, 0 otherwise)
POPEAST	interaction between POPULATION and the EAST-region (=POPULATION, if city falls within EAST, 0 otherwise)
ADTGC	interaction between ADTTOT and the GULF and CENTRAL-regions as a group (=ADTTOT, if city falls within either of the two regions, 0 otherwise)

TABLE 4 Summary of Initial and Final Models

VARIABLES	TOTAL RETAIL SALES INITIAL	TOTAL RETAIL SALES FINAL	GAS STATION SALES INITIAL	GAS STATION SALES FINAL	RESTAURANT SALES INITIAL	RESTAURANT SALES FINAL	SERVICE RECEIPTS INITIAL	SERVICE RECEIPTS FINAL
INTERCEPT	-14495 (-5.99)	-8437 (-4.613)	-4390 (-5.74)	-4632 (-5.947)	-1827 (-9.68)	-19657 (-10.91)	-8.78 (-12.08)	-10.80 (-13.949)
POPULATION	5.561 (22.84)	4.440 (17.665)	0.438 (15.05)	0.478 (22.470)	0.336 (17.33)	0.355 (18.898)		
INCOME	0.576 (1.41)		0.205 (4.14)	0.229 (4.778)	0.062 (1.94)	0.151 (4.429)		
ADT-TOTAL	1.305 (9.76)	1.349 (8.768)	0.054 (2.70)		0.106 (8.57)	0.061 (5.101)		
ADT-BYPASS			-0.131 (-2.98)	-0..67 (-1.822)	-0.067 (-2.42)			
ACCESS	-12402 (-4.91)	-15760 (-5.50)				-439 (-2.246)	-0.116 (-1.54)	
YEAR'			4.596 (4.96)	4.656 (4.740)				
HIGHWAYS			182 (2.37)	265 (3.427)				
METRO-AREA					-296 (-1.77)	-575 (-3.649)		
LNPOP							1.022 (7.03)	1.191 (18.590)
LNINCOME							0.388 (4.35)	0.671 (6.213)
LNNEARBY							0.00303 (2.06)	0.003 (2.181)
LNADTTOT							0.315 (4.36)	0.154 (1.999)
YEAR							0.0243 (7.03)	0.019 (5.121)
NEARBYCITY								
C6							-0.671 (-4.34)	
C118							0.545 (3.55)	
C119							-0.870 (-5.08)	
C22	31470 (5.88)							
C23	-44747 (-7.09)				-1704 (-3.54)			
C101	15186 (3.21)							
C13			2344 (3.96)					
C112					-1022 (-2.75)			
C113					1745 (4.51)			
POPGMC		1.517 (6.161)						
POPPLAINS		1.227 (4.639)				-0.040 (-1.885)		-0.014 (-1.915)
POPGC						0.077 (4.739)		
POPEAST								0.039 (3.265)
ADTGC				0.036 (-1.822)				
F	248	303	106	119	193	226	215	228
C.V.	23.2	26.1	34.4	34.9	37.1	36.7	4.1	4.5
Adjusted R^2	0.883	0.849	0.739	0.725	0.855	0.855	0.881	0.855
Durbin-Watson	1.167	0.918	1.407	1.394	1.060	1.035	1.248	1.165

Note: Every cell contains a corresponding estimated coefficient, with the t-statistic in parentheses

the specification of the final model is conceptually improved and more generally applicable.

Highway-Oriented Business

Gas Station Sales

An interactive effect was captured between ADTTOT and the GULF and CENTRAL regions, combined in the interaction variable ADTGC. This variable is equal to the total ADT of the city if it falls within either the GULF or the CENTRAL regions, 0 otherwise. The statistically significant positive coefficient for this variable indicates the significant positive effect of daily traffic on gas stations in these areas.

The bypass-related variable, ADT-BYPASS, is statistically significant and negative, clearly indicating the negative effect of the bypass. By applying various restrictions, no differential effect of this variable was observed across regions.

Restaurant Sales

Different behavior across geographic regions is also captured in the model for restaurant sales. Separate coefficients for POPULATION for the PLAINS region (POPPLAINS, as defined before), and for the CENTRAL and GULF regions as a group (POPGC, defined similarly for the CENTRAL and GULF regions as a group), were significant. Several restrictions were applied to test various hypotheses regarding interaction with the geographic variable. The null hypothesis that the parameters for POPULATION in the GULF region and POPULATION in the CENTRAL region are equal was not rejected. The restricted model containing the grouping of the two (POPGC) was obtained. On the other hand, the null hypothesis that the parameters for POPPLAINS and POPGC are equal was rejected, and both variables are retained in the final specification. The POPPLAINS variable is negative, showing that for the same population, there is a lesser tendency to support restaurants in the Plains region.

Service Receipts

In estimating the model for service receipts, the geographic region played a less significant role than in the other models. Intuitively, this can be expected, since the number of services offered (and the corresponding number of service receipts) is probably more dependent on local characteristics than on regional characteristics. However, by interacting POPULATION with PLAINS (POPPLAINS) and also with EAST (POPEAST, defined similarly to POPPLAINS, but for the EAST region), some differential effects are observed. The significant positive contribution in the EAST region shows that more services are offered for a specific population compared with what is offered in other regions. This is probably because of the expansion, growth, and diversification of this region's economy (*12*). The significant negative interaction between POPULATION and PLAINS indicates that fewer services are offered in this predominantly agricultural region.

With this specification, the bypass-related variable is not significant. The only explanatory variable with a negative coefficient is POPPLAINS (apart from the effect captured in the intercept). This is an indication that bypass construction does not necessarily have a negative impact on service receipts; in this case the decreasing population in a mostly agricultural area appears to be the predominant negative factor.

Throughout the final modeling process, the importance of the geographic region is evident. The diversity and size of the state of Texas are thus better captured in these models. Also, in all the models except the one for service receipts, the impact of the bypass remained statistically significant and negative. In all models, the city-specific dummy variables were replaced by variables related to the economic regions without any significant loss in explanatory power. The final models are thus more general and improve the specification.

CONCLUSIONS

The economic impact of highway bypasses on small cities in a rural setting is not uniform across cities and in most cases appears to be rather minor. The way in which a social and business community responds to a highway bypass is complex and involves the interaction of several factors. Several approaches were used in this study to address this issue.

Econometric models showed that a bypass generally brought a small, but statistically significant, decrease to business volumes in bypassed cities. These models were developed to relate total retail sales, gasoline service receipts, restaurant sales, and service receipts to the pertinent characteristics of the area. Cluster analysis highlighted the importance of the economic base of a city, as captured in the geographic regions. Cities in rural settings in Texas were clustered according to their characteristics, resulting in similar clusters for both the bypassed and the control cities. Inclusion of the regional cluster variables into the econometric models improved the specification of the models.

Beyond the formal models, individual case studies show that local communities might not necessarily perceive bypasses as negative. Rather, the construction of a bypass is seen as one of many factors contributing to the overall economic performance of a city in a rural setting. The initial decreases in certain types of sales were often counteracted by reorientation of local stores. Political and business leadership in a given area seems to play an important role in the evolution of the city after bypass opening.

ACKNOWLEDGMENT

Financial support for this study was provided by the Texas Department of Transportation.

REFERENCES

1. G. E. Weisbrod and J. Beckwith. Measuring Economic Development Benefits for Highway Decision Making: The Wisconsin Case. Presented at 69th Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1990.
2. A. M. Mackie. Effect of Bypasses on Town Development and Land Use. *Planning and Transportation Research*, Vol. P239, London, 1983.
3. R. D. Pashek and E. M. Horwood. Community Consequences of Highway Improvement. In *Highway Research Record 96*, HRB, National Research Council, Washington, D.C., 1965.
4. *A Literature Review of Urban Bypass Studies*. Iowa Department of Transportation, 1989.
5. J. L. Buffington. *An Economic Impact Study of Interstate Highway 35E on Waxahachie, Texas*. Research Report 4-6. Texas Transportation Institute, March 1966.
6. L. Skorpa, C. M. Walton, and J. Huddleston. *Transportation Impact Research: A Review of Previous Studies and a Recommended Methodology for the Study of Rural Communities*. Council of Advanced Transportation Studies. University of Texas at Austin, March 1974.
7. G. V. Fuguitt, D. L. Brown, and C. L. Beale. *Rural and Small Town America*. Russell Sage Foundation, New York, 1984.
8. L. Garkovich. Population and Community in Rural America. *Contributions in Sociology*. No. 84, Greenwood Press, New York, 1989.
9. R. Helaakoski, H. S. Mahmassani, C. M. Walton, M. E. Euritt, R. Harrison, and S. J. Andersen. *Economic Effects of Highway Bypasses on Business Activities in Small Cities*. CTR Research Report 1247-1. Center for Transportation Research, University of Texas at Austin, Oct. 1992.
10. S. J. Andersen, H. S. Mahmassani, C. M. Walton, M. E. Euritt, and R. Harrison. *Traffic and Spatial Impacts and the Classification of Small Highway Bypassed Cities*. CTR Research Report 1247-2. Center for Transportation Research, University of Texas at Austin, Oct. 1992.
11. J. Buffington and D. Burke. *Employment and Income Impacts of Highway Expenditures on Bypass, Loop and Radial Highway Improvements*. Research Report 1066-F. Texas Transportation Institute, College Station, 1989.
12. M. S. Aldenderfer and R. K. Blashfield. *Cluster Analysis*. Sage Publications, Beverley Hills, Calif., 1984.
13. *Texas Almanac 1990–1991*. The Dallas Morning News, Dallas, Tex., 1989.

The views and findings expressed and described in this paper are the authors' and do not necessarily reflect those of the Texas Department of Transportation or the Center for Transportation Research at the University of Texas.

Publication of this paper sponsored by Committee on Social and Economic Factors of Transportation.

Running on Empty: Travel Patterns of Extremely Poor People in Los Angeles

ALEX MEYERHOFF, MARTINE MICOZZI, AND PETER ROWEN

Although previous studies have focused on the travel needs and patterns of various social groups—including commuters, the elderly, the disabled, and women—few have specifically examined the travel patterns of public assistance recipients. The lack of information on such a large and growing segment of the population prompted a group of graduate students from the University of California at Los Angeles' Department of Urban Planning to initiate a study of the transportation needs of the very poor in Los Angeles county. Various subgroups of the population were examined, including the homeless, those living in shelters, and recipients of public assistance. Many of the latter depended on General Relief (GR), the form of public assistance obtained as a last resort, mandated by the state of California. For the very poor, accessibility to transportation is a major factor influencing the capacity to satisfy even the most rudimentary needs, such as food, shelter, employment, and medical care. The poor struggle daily to reach trip destinations that are necessary to sustain life or maintain eligibility for public assistance. This is particularly problematic in Los Angeles, where the geographic expanse of the region adds to the already formidable barriers that the poor must overcome. Moreover, the county, unlike other heavily populated counties in California, allocates few transportation subsidies to defray transportation costs to crucial destinations. Even the few transportation subsidies that are granted are allocated randomly by case workers. Dragged from one bureaucracy to another, with inadequate resources to defray the necessary transportation costs, many of the poor find themselves literally running on empty.

For the very poor, accessibility to transportation often constrains one's ability to satisfy even the most rudimentary needs, such as food, shelter, and medical care. Many must spend much of their day simply traveling to and from shelters, clinics, missions, and social service agencies that are highly decentralized. The urgency of meeting these immediate basic needs impedes the very poor from seeking and securing long-term employment.

In addition, to maintain eligibility for public assistance, employable recipients must complete assigned county work projects. Frequently, however, the very poor lack the resources to pay for the cost of transportation to their assigned job sites. Without a means of transportation to assigned job sites, the very poor often lose their eligibility for continued public assistance benefits. Many in this predicament find themselves homeless, destitute, and literally running on empty.

It is anticipated that in the current recession, this dire situation will worsen. In 1992 Los Angeles county submitted a proposed budget that fell $2.2 billion short of that necessary to maintain the existing level of services for the poor. Since the survey, the county has cut the General Relief (GR) monthly allowance from $341 to $296. At a time of record unemployment, such reductions in public assistance will only serve to push those already on the verge of homelessness out into the streets.

The purpose of this paper is to analyze and quantify the travel patterns and mobility constraints of very poor people in Los Angeles county. To this end, the research team designed a survey that was administered to more than 200 respondents. The sample consisted of homeless individuals and persons living in shelters or single-room occupancy dwellings, many of whom depend on GR benefits. Respondents were selected from nine geographic locations throughout Los Angeles county.

GENERAL RELIEF

Approximately half of the 203 survey respondents depend on monthly GR benefits. For this reason, it is necessary to provide a brief description and explanation of the state's GR program and its effect on the mobility of the very poor.

GR, the form of public assistance obtained as a last resort, is administered by the state of California but funded entirely by the counties. The purpose of the GR program is to ensure that the indigent population not receiving state or private support is provided with "minimum assistance."

To qualify for GR benefits, an applicant must

- Be a resident of Los Angeles County,
- Be a citizen of the United States or a lawful alien,
- Have a net monthly income of less than $341,
- Own less than $500 (excluding a motor vehicle with a market value of less than $1,500),
- Have less than $34,000 in real property value, and
- Apply for unemployment insurance, accept assignments to Los Angeles county workfare projects, register with the State Employment Development Department, and actively seek work.

The California Welfare and Institutions Code—under which the state's GR program falls—stipulates that "minimum assistance" include, at the very minimum, allocations for food, housing, utilities, medical care, and transportation. The state mandates the minimum requirements and oversees county actions, but the onus of administering the program, including all decisions on benefit amounts, is left to the counties.

Los Angeles county has relegated transportation to a "special need" category rather than incorporate it as a component

A. Meyerhoff, 1226 Franklin Avenue, Apartment Four, Los Angeles, Calif. 90404. M. Micozzi, 1308 Oronoco, Alexandria, Va. 22314. P. Rowen, 10737 Wellworth Avenue, Los Angeles, Calif. 90024.

of minimum subsistence. Thus, the $341 (now $296) monthly allowance allocated to qualified recipients in Los Angeles does not factor in the cost of transportation. Yet, for recipients to remain eligible for continued GR benefits, they must perform up to 17 days of assigned county work projects each month and complete 24 job searches within an 8-week period. It is an arduous challenge for these recipients to arrive at all their necessary destinations without funding to defray the cost of travel. Without a monthly transportation allowance, many GR recipients are forced to forego food or shelter in order to pay for transportation. Even those who make such tradeoffs are sometimes unable to pay the cost of transportation. Thus, despite the provisions mandating transportation allocations, few GR recipients receive the full allotment.

The lack of mobility causes some GR recipients to miss social service appointments, employment interviews, or county workfare assignments, which results in the suspension of their benefits. It is easy to see why so many GR recipients in Los Angeles county find themselves caught in this vicious cycle.

Recent Los Angeles County Statistics

As of May 1993 more than 105,000 persons were receiving GR in Los Angeles county. This marks a dramatic increase over the 50,000 recipients in June 1990. In addition, welfare cases in Los Angeles county stand at a record 1.5 million, a 40 percent increase over the 1990 figure (*1*). At present 1 in 10 people in Los Angeles county depends on some form of public assistance. Of the total number of GR recipients in Los Angeles county, 65 percent are homeless and 85 percent are male.

Demographics of Homeless Population in Los Angeles County

Given that many of the survey respondents were homeless, it is important to examine the homeless as a major subset of the very poor. In 1985 a survey by the National Coalition for the Homeless placed the homeless population in the United States at close to 2.5 million, or approximately the population of metropolitan Boston.

Today, 177,000 people are homeless in Los Angeles county (*2*). Almost 40 percent of the homeless population lives on Skid Row, a section of downtown Los Angeles. The largest increases in the homeless population are among women, youth, and families with children (*3*, p. 59). Some 60 percent of these people suffer from health problems that render them unemployable.

The characteristics of homeless people have changed dramatically. The stereotypical homeless person—once a white, middle-aged, alcoholic male—has given way to a new group of homeless, who are younger than their predecessors and more likely to be single women, single-parent households, families, the unemployed, victims of domestic violence, legal and undocumented immigrants, and the mentally ill (*4*, p.1). For an overview of basic characteristics of survey respondents of this study, see Table 1.

METHODOLOGY

The goal of this study was to gain quantitative data on the travel patterns and transportation needs of the very poor in Los Angeles county. A survey was designed to examine the general characteristics and background of the very poor, including information about monthly income and expenditures, living arrangements, and travel patterns.

Defining, locating, and accessing the sample population for the survey proved to be one of the greatest challenges of the study. As a consequence of the inability to isolate an exclusive sample of GR recipients from the county Department of Public and Social Services (DPSS) office, the research team opted to conduct personal interviews with patrons at nine homeless shelters and residential social service agencies located in the Los Angeles metropolitan region.

The county was divided into five major regions for the purposes of analysis: San Fernando Valley (North Hollywood), Downtown, East Los Angeles, South Bay (San Pedro), and West Los Angeles (Santa Monica). This sampling procedure resulted in a survey sample that was more diverse than if the research team had been able to obtain an exclusive sample of GR recipients from the DPSS. Aside from GR recipients, the sample includes the homeless, those on other forms of public assistance such as Aid to Families with De-

TABLE 1 Basic Characteristics of Survey Respondents

CHARACTERISTIC	UCLA SURVEY, 1992	DPSS SURVEY, 1987
Average Age	38	39
Ethnicity	38% African American 30% White 24% Hispanic 8% Other	56% African American 19% White 17% Hispanic 8% Other
Birth Place	23% California 59% Other States 18% Other Countries	24.7% California 31.2% Other States 10.1% Other Countries
Years Residing in L.A.	33% Resident > 4 Yrs.	83% Resident > 4 Yrs.
Native Language	77% English 20% Spanish	No Data
Marital Status	55% Single 26% Divorced 11% Married	44% Single 28.5% Divorced 3.2% Married
Average Yrs. Schooling	11.5	No Data
Gender	83% Male 17% Female	66% Male 34% Female

pendent Children, and other very poor individuals not receiving public assistance.

The survey contained questions pertaining to travel destinations and expenses, trip frequencies, and mobility constraints as well as information on living arrangements, public assistance, work status, monthly income, and demographics.

SURVEY ANALYSIS

As noted, the sample consisted of 203 respondents. Among those surveyed, the primary form of public assistance was GR (45.6 percent), although one in four persons reported receiving some other form of welfare. Results indicated that on average, respondents made about four trips per day. Walking was the primary mode of transportation, accounting for 60 percent of all trips. The second most frequently used mode was transit, with an average monthly expenditure of $35. On the basis of a medium monthly income of $360, close to 10 percent of the respondent's income was spent on transit.

Although it may be more economical to buy a monthly bus pass than to pay fares trip by trip, a $42 disbursement is a heavy burden for someone with a monthly income of $360.

The vast majority of respondents indicated that they were unemployed (84 percent), and more than 70 percent reported that they considered finding employment to be a very important priority.

Demographic Profile of Survey Population

Respondents ranged in age from 16 to 78 years. Males constituted 83 percent of the survey sample; only 17 percent were female. The ethnic composition of the survey population is described in Figure 1 ($N = 186$).

The typical survey respondent is a single, unemployed, African-American male, 38 years old, with a high school education and an average monthly income of $442. In general, respondents had achieved a relatively high level of education, with an average of 11.5 years of schooling. More than three-quarters of the respondents were born in the United States and speak English as their native language; 14 percent were born in Mexico. Spanish is the first language for 20 percent of the respondents. Approximately one-quarter of those surveyed were born in California; 67 percent had lived in Los Angeles for less than 4 years, and more than 33 percent, for more than 4 years.

Marital Status and Current Living Arrangements

Three-quarters of the respondents are single, and only 11 percent are married. Approximately 64 percent of the respondents reported that they live alone. More than half of the respondents indicated that they had lived in the same location for less than 1 month before their interview. Only 38 percent of the respondents reported regularly paying rent. Among those who pay rent regularly, nearly half indicated that they pay on either a weekly or a daily basis. The data suggest that more than three-quarters of the respondents have unstable living arrangements, a characteristic typical of the homeless population.

The median monthly rent paid by respondents is $158. Since their average monthly income is only $360, the average monthly income left for food, utilities, transportation, clothing, and personal items is just $202 per month. It is difficult to imagine surviving, let alone thriving, on so little money.

Public Assistance and Income

Less than half of the respondents (43 percent) were receiving GR assistance at the time the survey was conducted.

Employment Status and Work History

By conservative estimates, more than half a million Californians are currently unemployed. These conditions have taken an unusually harsh toll on Los Angeles county residents, as evidenced by the large numbers of unemployed people in the survey sample. Over three-quarters of the sample (86 percent) are currently out of work (Figure 2).

Half of the unemployed survey respondents indicated that they stopped working as a result of layoffs or the completion of seasonal work. More than 80 percent of those sampled had participated in job searches a week before the interview date, and more than half of those job-seekers had looked for employment more than twice a week. Approximately 25 percent of those currently employed are searching for a new job. The percentage of employed males and females is evenly distributed. Figure 2 illustrates that less than 5 percent of the survey population indicated that they work every day, slightly over 5 percent work a few days a week, 2.3 percent work a few days a month, and less than 1 percent work on an irregular basis.

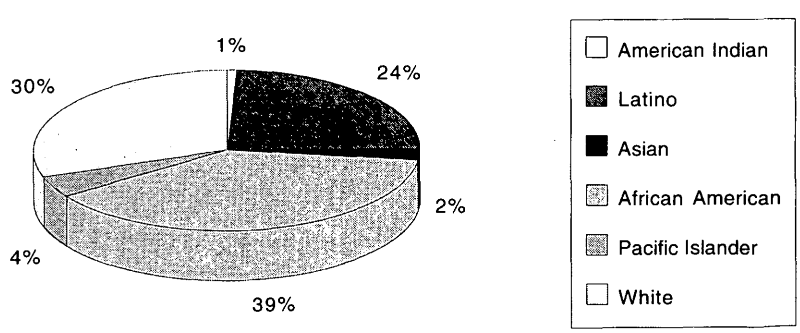


FIGURE 1 Ethnic composition of survey respondents.

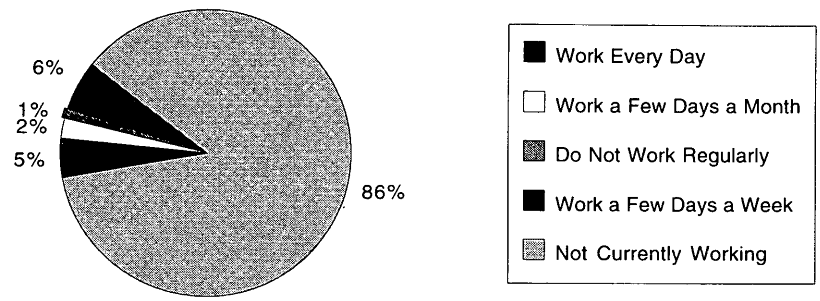


FIGURE 2 Employment characteristics of survey respondents.

Trip Diary

Transportation questions were posed to respondents in a trip diary format that attempted to assess how and where they traveled the day before the administration of the survey. The results revealed that respondents made an average of 4.1 trips during the day before the survey. Figure 3 illustrates that more than half of the respondents made between two and four trips the day before the survey. Fewer than 10 percent of the respondents reported having made between 8 and 10 trips.

Trip Destinations

Trip destinations were grouped into seven categories. The highest percentage of trips were taken for the purpose of finding shelter. The next most frequent were personal trips, followed by social and recreational, food-related, and job-related trips. Approximately 17 percent of the trips were reportedly made to specific locations such as Downtown or Santa Monica. Of these, approximately 40 percent were on the West Side and the rest were divided between Downtown, East Los Angeles, South Bay, San Pedro, the Valley, and South-Central Los Angeles (Figure 4).

Trip Modes

Figure 5 illustrates the distribution of trips by mode. Clearly, walking is the primary mode of transportation used by the poor and homeless, as nearly 60 percent of the total trips were made on foot. Transit was the second most frequently used mode, followed by riding as a passenger in a vehicle, driving, and bicycling.

Income significantly influenced mode choice. The range of monthly income reported by respondents from all sources varied from $0 to $3,750, with a median of $360. The transportation modes most affected by income were automobile use and public transit. To determine how income influenced the respondents' travel modes, car ownership was cross-tabulated with income. Given a median monthly income of only $360, it is not surprising that just 11 percent of the respondents reported owning an operable automobile. Overall, the findings indicated that as income increases, so does car-ownership.

Given that almost 90 percent of the respondents did not own a car, it was important to examine how income affected the use of other travel modes. As stated, walking was the primary mode of transportation used by respondents. In addition, an analysis of the average number of walking trips by income categories concluded that income had an inverse relationship with walking as a mode of transportation; as income increased, the percentage of trips made on foot decreased (Figure 5). No clear patterns emerged regarding other modes of transportation.

Although walking is generally the primary mode of transportation among the poor, national policy fails to recognize it as a legitimate form of transportation. A comparison between the average daily trips in the survey sample and the findings of the 1983–1984 National Personal Transportation Study (NPTS) revealed a major difference in that the NPTS study excluded trips made on foot. The NPTS found that individuals with an income of less than $10,000 made an av-

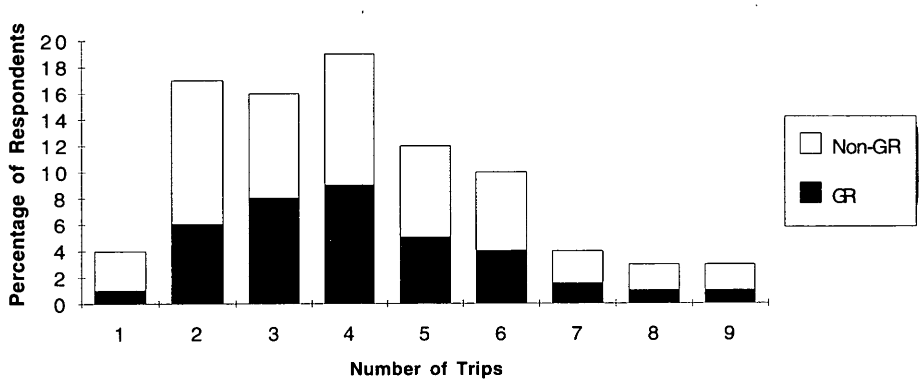


FIGURE 3 Number of trips made by respondents.

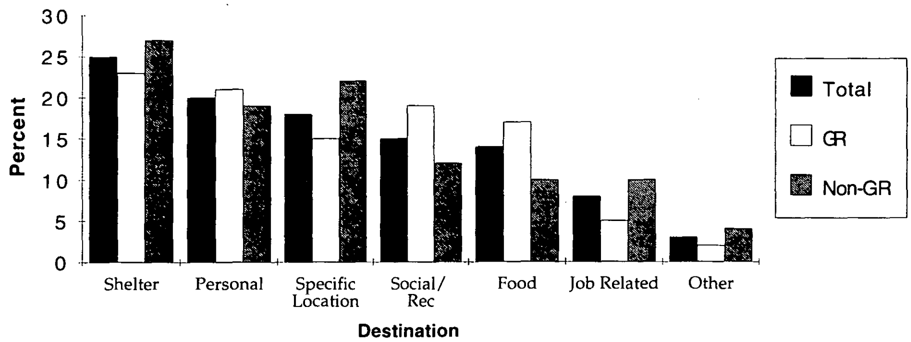


FIGURE 4 Trip diary information.

erage number of 2 daily trips, whereas this survey sample revealed a higher daily average number of trips, 4.1. When walking trips were excluded from these data, the figure for the average daily trips decreased by nearly 60 percent to 2.47.

Employment and Mode Choice

Employment status was shown to correlate with mode choice. Data on transportation modes used by the employed poor to access their jobs are displayed in the following table:

Mode	*Users (percent)*
Bus	43.0
Walk	32.1
Car	14.1
Carpool	3.6
Other	7.2
Total	100.0

These findings indicate that more than three-quarters of the employed poor are either transit-dependent or must walk to their jobs. Another interesting finding is that the percentage of car ownership of the employed poor (11 percent) is slightly higher than that of the sample on the whole.

Expenditures on transit varied according to the individual's employment status. Half of the employed respondents spent at least $35/month on transit, but only about 11 percent of the unemployed spent the same amount. The employed had higher transit usage than the unemployed. One interpretation is that increased transit usage and mobility are correlated to increased financial resources.

Transit Use

A number of questions about the respondents' transit use were asked to determine the transportation characteristics of very-low-income individuals and the factors that may influence these patterns. First, the number of respondents who used transit as well as the frequency of their transit trips was determined. Next, the monthly transit expenditures of the sample during the previous month was calculated. Finally, the number of transit tokens received by GR recipients from the DPSS was used to document the level of transportation support offered by the county welfare program.

Almost 77 percent of the respondents indicated that they had used public transit during the previous month, with higher usage among women than men. Most of those responding made more than 10 trips. The average monthly transit expenditure was $35, accounting for an average of 30 bus trips made over the month. As these figures suggest, the poor are clearly transit-dependent, and many devote a large part of their monthly income to transportation.

As noted in Figure 6, more than 60 percent of the respondents spent between $13 and $60/month on bus travel. In some cases, the total amount paid exceeded $42, the cost of a monthly bus pass offering unlimited rides. Although it is often more economical to buy a monthly bus pass, many respondents indicated that they could not allocate such a large portion of their monthly income to transit, all at once in a single expenditure. Senior citizens and the disabled qualify to purchase a discounted monthly bus pass for only $4, but most of the poor surveyed in this study did not qualify for this special benefit.

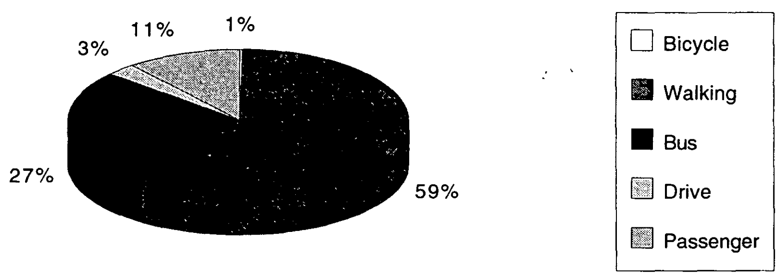


FIGURE 5 Trip destination by mode.

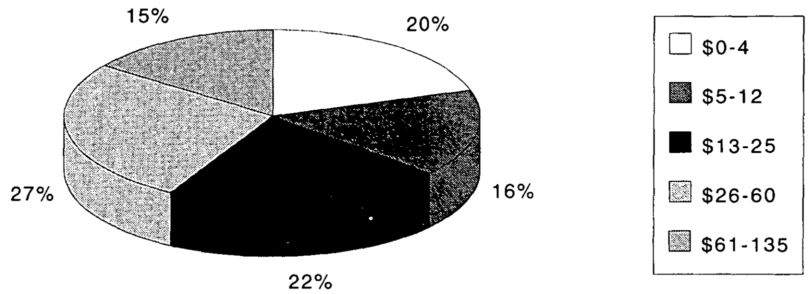


FIGURE 6 Number of transit trips last month.

Despite the stated policy of the DPSS that GR recipients receive transit subsidies when they need them, in actuality, more than 47 percent of the eligible GR recipients in the sample had not received a single bus token from the county. Moreover, of the GR recipients who had received bus tokens from the county, more than half had received only one or two tokens. For many people, lack of access to transit may contribute to patterns of failure.

There was a wide variation in transit use among the survey locations. Figure 7 demonstrates high percentages of transit use by respondents at all survey locations with the exception of East Los Angeles. The interview sites in East Los Angeles are within walking distance of services in Downtown Los Angeles. The low transit use by respondents at the East Los Angeles sites may be because these centers attracted a large number of senior citizens who had access to automobiles and to dial-a-ride taxis, shuttles, and other paratransit services for the elderly.

In contrast with the findings in East Los Angeles, the San Fernando Valley and Downtown locations had the highest percentages (greater than 78 percent) of respondents who used transit. In Santa Monica, approximately 50 percent of the respondents indicated that they had used transit at least 25 times in the last months compared with less than 10 percent of the East Los Angeles respondents. The high levels of ridership in Santa Monica and Downtown Los Angeles are consistent with the respective areas' high levels of transit service (Santa Monica) and high density (Downtown).

The influence of income on transit usage was also explored. Table 2 illustrates that the highest number of respondents who used transit were members of the income category $351–$500. The two most commonly cited trip purposes were related to accessing employment and social services.

Assessing Mobility Needs of the Poor

To ascertain whether the mobility needs of the sample were being met, interviewers asked respondents if there were any places they wanted to go but couldn't get to. An overwhelming 67 percent of the respondents indicated that there were such places. The results are shown in Figure 8. Interestingly, more than a quarter of the mobility problems were related to accessing employment. Survey results indicate that nearly 40 percent of the employed respondents relied on public transit to get to work.

The analysis indicated that gender did not significantly affect mobility. Men and women experience mobility problems equally.

Analysis of Frequent Destinations

Respondents were asked if there were places where they go regularly. The responses to this question were coded and grouped into six different categories: food and shopping, shelter, personal, social and recreational, job, and other. Food and shopping consisted of trips to such locations as grocery stores, restaurants, soup kitchens, and miscellaneous shopping destinations. The shelter category includes permanent and temporary shelters in addition to some outdoor locations, such as areas under bridges where one may obtain temporary shelter. The personal category includes more than 25 types of destination, the most prevalent being the DPSS office, City Hall, storage facilities, child care centers, and lawyers' offices. Regular visits to places such as libraries, movie theaters, beaches, and shopping centers were coded as social and recreational destinations. Work destinations include visits to the

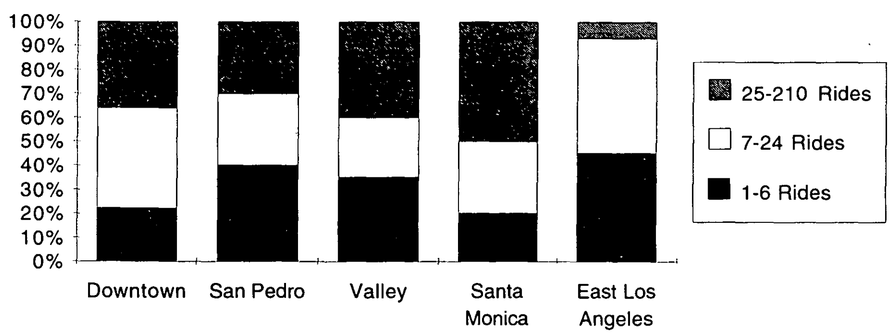


FIGURE 7 Transit trips by location.

TABLE 2 Influence of Income on Transit Use

INCOME	$0-200	$201 -350	$351- 500	$501- 900	$901+	TOTAL
% THAT USE TRANSIT	72	81	88	66	80	77
% THAT DIDN'T USE TRANSIT	28	19	12	34	20	23
TOTAL	100%	100%	100%	100%	100%	100%

employment offices, employment sites, or other locations associated with employment searches. Miscellaneous locations such as streets, gas stations, train stations, and boathouses were amalgamated under "other."

On the basis of their responses, 43.5 percent of respondents made regular trips to social and recreational destinations. By comparison, 10 percent of the respondents indicated that they made regular trips to shelters. The percentages of those who regularly make personal, food/shopping, and employment-related trips were 42, 32.2, and 20.1 percent, respectively.

Respondents were also asked how many trips a week they visited each destination mentioned. The largest percentage of trips were made for social and recreational purposes, followed by trips made for personal needs, including visits to social service agencies. It is interesting to note that only 5 percent of the trips were made for the purpose of seeking shelter.

Summary

The very-low-income population is heavily dependent on public assistance. The primary form of public assistance received by respondents was GR (45.6 percent), and one in four persons reported receiving some other form of public assistance. The vast majority of the respondents were unemployed (84 percent), and when queried, more than 70 percent of the respondents reported that it is very important for them to find employment.

Walking was the primary mode of transportation among the respondents. Nearly 60 percent of the daily trips of the respondents were made on foot, and on the average each respondent made about four trips a day. The next most often used mode was transit. The average monthly transit expenditure was $35. Close to 10 percent of the respondents' median monthly income of $360 was spent on transit. Given that the cost of a monthly bus pass in Los Angeles county is $42, it is often impossible for the respondents to buy monthly bus passes, even though it may be more economical to do so. Results of our analysis showed that 47 percent of the GR recipients in the sample had not received a single token from the county. Overall, only 11 percent of the survey respondents possessed bus tokens, and they indicated that the amount and frequency of this allocation was left solely to the discretion of the county worker.

From these results, it is not surprising that 67 percent of the survey respondents indicated that there were places they wanted to go to but could not get to. More than one in four respondents indicated that among such places were social service agencies, and some had become ineligible for GR because of missed appointments. It is apparent that the GR elibility is contingent on a considerable degree of mobility, but it is impossible to attain that mobility without monetary assistance in the form of a transit allowance. Despite the DPSS's policy, the agency has failed to provide this assistance to a group of people with no other means of sustaining themselves.

CONCLUSIONS

The purpose of this study has been to document and evaluate the travel needs of GR recipients. Survey findings indicate that there are deficiencies between the transportation that is

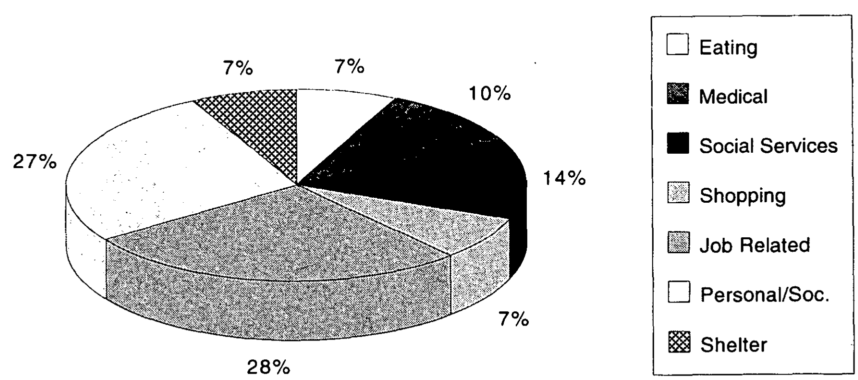


FIGURE 8 Places respondents could not get to.

needed and the transportation that is provided. Possibly the most telling statistic from the survey results is that 75 percent of the respondents who indicated using transit for the month before the survey spent an average of $35, a figure only $7 less than the cost of a monthly bus pass for the Southern California Rapid Transit District (RTD, which in 1993 became part of the Los Angeles County Metropolitan Transportation Authority). This finding is important because even without a formal transportation allowance each month, most people deemed transportation so critical to their own needs that they were willing to allocate a portion of their already extremely limited budget toward fulfilling their travel needs. Furthermore, the percentage of people spending $35 on travel is higher than the percentage of people deemed employable by DPSS standards. Though many people have transit needs related to employment, many people make trips for other purposes. Most people made use of transit, whether for employment or other purposes. It then becomes important to recognize that transportation is considered a vital need for most individuals and that a transportation allocation for all GR recipients in the form of a cash supplement to the existing GR check is the most appropriate response.

Furthermore, enormous amounts of time and labor resources are currently being spent in determining transportation benefits. Since there is really no way for case workers to singlehandedly make equitable decisions on the benefits that individuals should receive, it seems most appropriate to streamline and standardize the entire system. In sum, providing an allowance for transportation complies with the requirements as mandated in the Health and Welfare Code and for the critical travel needs of all GR recipients; it could also save the DPSS valuable time and money at a time when it is in need of 3,000 additional staff workers.

The survey results reinforce the idea that there is a significant link between mobility and the ability to secure many of life's basic needs. The study suggests that non-employment–related transportation is a basic and legitimate need for all people, including those receiving public assistance. Recognizing this, the dollar value of the current DPSS transportation allowance should be reevaluated to better reflect the basic travel needs of those currently receiving public assistance.

The evidence also suggests that the present administration of the GR transportation allowance program should also be reevaluated. Administration of the program is time-consuming and overly bureaucratic; often many individuals do not receive their maximum entitlement. Moreover, procedural constraints often make it difficult for case workers to make equitable decisions about benefits that individuals should receive. Given these concerns, it seems appropriate to streamline the system in such a way as to maximize benefits while reducing administrative costs.

REFERENCES

1. *Los Angeles Times*. May 17, 1992.
2. *Los Angeles Times*. February 23, 1992.
3. R. Schwartz. The End of the Line: The Homeless and the Transportation Industry. *WTS National*, Vol. 10, No. 5, Sept.–Oct. 1991.
4. A. Rahimian. *Migration and Mobility of the Urban Homeless*. Dec. 1990.

Publication of this paper sponsored by Committee on Social and Economic Factors of Transportation.

PART 4

Management and Productivity

On the Response Consistency of Questionnaire Surveys of State Department of Transportation Management

J. L. HULSEY, P. A. KOUSHKI, AND C. D. VAUGHAN

The approach to decision making in state departments of transportation (SDOTs) is in the process of change because of several important and complex factors. The results of two recent surveys addressed to the secretaries/commissioners and regional directors of SDOTs are reported. The result of the analysis of response consistency of these two groups of management officials to a set of questions dealing with resource allocation decisions is also reported. The officials' views on the impact of political change in DOT upper management on resource allocation decisions and implementation also are presented and shortcomings associated with current management system programs are identified. The results of chi-square test and correlation analysis performed on the responses to the two questionnaire surveys showed a significant degree of consistency in responses of these two groups of upper SDOT management officials to the questionnaire surveys.

This paper presents the results of a research study aimed at determining the response consistency of questionnaire surveys of the upper managements in the state departments of transportation (SDOTs).

Examples of questionnaire surveys in the transportation industry abound. Surveys have been used to analyze park-and-ride lot use (*1*); determine ridematching system effectiveness (*2*); identify commuter behavior (*3*); examine the effect of variable work hour programs on ridesharing (*4*); study commuter attitudes (*5*); measure transit route service elasticity (*6*); identify management structure and decision process (*7*); evaluate manager attitudes concerning job change (*8*); examine the effects of disseminating service information on bus ridership (*9*); perform transit market research (*10*); compare travel behavior (*11*) and travel demand management markets (*12*); and analyze changes in travel patterns (*13*).

A significant level of effort has also been directed toward improvements in surveying techniques. Examples of such studies include comparison of telephone and door-to-door surveys (*10*), testing of alternative administrative procedures and survey instruments (*14*), application and modifications of home interview travel surveys (*15*), and design of mail and telephone surveys (*16*).

However, research has not considered the variation in responses when different levels of management hierarchy in an organization respond to survey inquiries. In a recent research study aimed at the examination and evaluation of management information system programs and decision processes in SDOTs nationwide (*17*), self-administered surveys of upper managements were undertaken. In addition, telephone and personal interviews of SDOT officials supplemented the questionnaire surveys. Results of these personal interviews indicated that a potential variation in responses may be expected if individuals from different levels of the management decision-making hierarchy are surveyed. Quite clearly, the impact of a survey response is of paramount importance on findings, conclusions, and recommendations of a given study. It was to the quantification of this important point that this study was addressed.

The specific objectives of the study were twofold: First, to present the findings of two surveys and second, to test the following hypothesis—H_0: depending on the hierarchical level of the management decision maker responding to a questionnaire survey, no significant degree of variation in responses may be expected.

J. L. Hulsey and C. D. Vaughan, Department of Civil Engineering, University of Alaska Fairbanks, 263 Duckering Building, Fairbanks, Alaska 99775. P. A. Koushki, Civil Engineering Department, College of Engineering and Petroleum, Kuwait University, P.O. Box 5969, Safat, Kuwait 13060.

DATA

A structured, yet brief, questionnaire was designed to determine SDOT decision-making patterns, identify the tools and computer programs utilized, and evaluate the shortcomings of these programs. The questionnaire was mailed to the secretary/commissioner of every state DOT. A total of 43 states (86 percent) responded.

The questionnaire was later modified to incorporate a number of additional questions concerning the identification of state DOT boards/commissions and their impact on the decision process, implementation, and stability. The questions addressed in the first questionnaire survey were also included in the second. This questionnaire was mailed to a randomly generated list of regional directors in state DOTs approximately 6 months after the first survey. The random list of regional directors was generated from the organizational charts of SDOTs (*18*). A total of 48 states (96 percent) responded to the second survey.

Data were coded and compiled for statistical analysis using the mainframe computer at the University of Alaska Fairbanks. Statistical Analysis Software (SAS) was used to process the data.

ANALYSIS OF SURVEY RESULTS

The following is a frequency distribution analysis of responses to each common question in both surveys.

Question 1. With respect to source (budget) allocation, is the process of decision making in your state DOT centralized (decisions made at the top), decentralized (regional/district levels), combination and/or other?

All responses of secretary/commissioners (first survey) fell within two categories. Seventeen states (39.5 percent) indicated a "centralized" decision-making process, whereas the remaining 26 (60.5 percent) selected the "combination" category, reflecting district/regional level participation in decisions concerning the allocation of resources. The responses of regional directors (second survey) were 28 (59.6 percent) "centralized," 16 (34.0 percent) "combination," and 3 (6.4 percent) "decentralized." One director did not respond to this question.

Question 2. In your opinion, is the process of decision making in your SDOT smooth and logical?

In the first survey, no secretary/commissioner responded "no" to this question. However, 21 (48.8 percent) responded "to some extent," and the remaining 22 (51.2 percent) believed that the decision process within their organization was smooth and logical. In the survey of regional directors, the responses were 28 (58.3 percent) "yes," 18 (37.5 percent) "to some extent," and 2 (4.2 percent) "no."

Question 3. Is there any room for improvement in the decision-making process?

Twenty-three secretaries/commissioners (53.5 percent) responded "yes," and another 19 (44.2 percent) indicated "to some extent." One state commissioner (2.3 percent) responded that there was no room for improvement in the budgetary decision process. The responses of the regional directors were 27 (57.4 percent) "yes," 19 (41.7 percent) "to some extent," and 1 (2.1 percent) "no." One director did not respond to this question.

Question 4. Does the organizational structure of your state DOT support/complement the department's decision-making process?

The responses in the first survey were 1 (2.3 percent) "no," 14 (32.6 percent) "to some extent," and 28 (65.1 percent) "yes." In the second survey, the responses were 37 (77.1 percent), 10 (20.8 percent), and 1 (2.1 percent), "yes," "to some extent," and "no," respectively.

Question 5. Please indicate the number of upper-management positions in SDOT that are political appointees.

A wide range of responses was given to this question. In the first survey, the number of political appointees ranged from 0 to 74, for a mean of 12.3 positions and a standard deviation of 12.0. In the second survey, the responses to this question ranged from 0 to 22, for a mean of 7.7 and a standard deviation of 6.1.

Question 6. How does the change in the top-level DOT administrators (political appointees) affect the decision process?

Twenty-two percent of the respondents in the first survey indicated that the process of decision making was "very much" affected by change of the top-level administrators, and 63.4 percent felt it was affected "to some extent." The remaining 14.6 percent indicated that the change of top management had no impact on the decision process. The figures for the second survey were 14, 72, and 14 percent for the "very much," "to some extent," and "no" impact categories, respectively.

Question 7. How does the change of top-level DOT administrators affect the implementation of decisions?

Ten of the respondents (23.2 percent) in the first survey indicated that the implementation of decisions was "not at all" affected by management change, whereas 27 (63 percent) responded that the management change affected decision implementation "to some extent." Four secretaries/commissioners believed that the implementation of DOT decisions was "very much" affected by these management changes. Two respondents did not answer this question. In the second survey, 8 (18.2 percent) regional directors stated that the management change had "no" effect on decision implementation, whereas 33 (75 percent) believed that the implementation of decisions was affected by the change in top administrators "to some extent." Three directors (6.8 percent) responded that decision implementation was "very much" affected by the change in political appointees.

The question "How can the decision process and implementation be improved under the changing management environment?" brought recommendations of important measures from respondents in both surveys:

- Top-level positions should be civil service or under an executive contract.
- Multiyear project commitments should be exempt from changes except for some predefined reasons.
- There should be well-documented information systems that would not have to be resold to each new administration.
- Transition would be smoother if only transportation professionals were appointed.
- There needs to be good communication of objectives to all involved employees. There should be no hidden agendas.
- The basis for decisions should be documented and should be based on clearly stated objectives.
- The establishment of a long-term plan, supported by detail and a "need justification," may be the best hedge against impulse or political expedients.
- A unified management information system should be established.

In response to a question about the names of the management system software programs utilized, a total of 107 programs were reported by the responding state DOTs. Seventy-four of the software programs (69.2 percent) were individually developed "in-house" programs. The remaining 33 (30.8 percent) included a variety of commercial and specially commissioned software programs. The software programs were utilized for the management of pavement, maintenance, highway, and bridge systems.

Numerous criticisms were directed toward the current management system software programs by the respondents of both surveys. The DOT secretaries/commissioners, for example, frequently identified such shortcomings as different data base requirements, the inability to interact, lack of integration, vast data requirements, PC interact limitation, limited analysis capability, lack of common network reference ability, and lack of predictive capability.

Correlation analyses were performed to examine degrees of associations among the various questionnaire responses. The Pearson correlation coefficients, along with the probability of the null hypothesis that a given coefficient of correlation (r_{xy}) is equal to 0 (H_0: $r_{xy} = 0$), were computed. Results indicated that the degree of association between a number of response variables was rather significant. For example, in the first questionnaire survey, as the number of upper-management political appointees increased, the state DOT decision process tended to include a "combination" of district, regional, and, of course, central headquarters management into the decision-making process for resource allocation ($r_{xy} = 0.3668$; $p = 0.03$).

When the state DOT decision process was reported to be smooth and logical, there was less reported room for improvement ($r_{xy} = -0.4492$; $p = 0.002$), a higher compatibility with the organizational structure ($r_{xy} = 0.3680$; $p = 0.01$), less impact of management change on the decision process ($r_{xy} = 0.2815$; $p = 0.05$), and even less impact on decision implementation ($r_{xy} = -0.3513$; $p = 0.02$). The analysis of correlation also indicated that when a state DOT organizational structure supports the agency's decision-making process, there is less room for improvement in the decision process ($r_{xy} = -0.3781$; $p = 0.02$).

RESPONSE CONSISTENCY OF SURVEYS

The examination of the response consistency of the two upper SDOT management levels (secretaries/commissioners versus regional directors) involved the construction of contingency tables based on the observed frequency of responses to each question. The chi-square test of significance was then used to test the null hypothesis (H_0). H_0 assumes that response frequencies of the two questionnaires are independent of each other (*19*). In other words, no significant difference in response to a given question will be found to exist regardless of the responder's hierarchical management level.

The value of chi-square for a 2 × 3 contingency table may be computed from Equation 1:

$$\chi^2 = \frac{N}{N_A}\left[\frac{a_1^2}{N_1} + \frac{a_2^2}{N_2} + \frac{a_3^2}{N_3}\right] + \frac{N}{N_B}\left[\frac{b_1^2}{N_1} + \frac{b_2^2}{N_2} + \frac{b_3^2}{N_3}\right] - N \quad (1)$$

where

$a_1, \ldots, a_3, b_1, \ldots, b_3$ = category response frequency of a given question in the first and second surveys, respectively;

N_A and N_B = sample sizes of Survey 1 and Survey 2, respectively;

N_1, N_2, and N_3 = sum of question response frequencies in a given category for both surveys; and

$N = N_A + N_B = N_1 + N_2 + N_3$.

Table 1 gives the contingency tables for common questions in the two surveys.

The result of chi-square computations and the test of the null hypothesis (H_0) are presented in Table 2. With the exception of the first question, the null hypothesis could not be rejected for all other questions at the 95 percent significance level ($\alpha = 0.05$). This indicates that no significant difference was found to exist between the responses of the two groups

TABLE 1 Contingency Tables for Common Questions in Two Surveys

QUESTION NUMBER	CATEGORY RESPONSE FREQUENCY		
	I: yes/very much	II: to some extent	III: no
Q. 1:			
Survey I	17	26	0
Survey II	28	16	3
Q. 2:			
Survey I	22	21	0
Survey II	28	18	2
Q. 3:			
Survey I	23	19	1
Survey II	27	19	1
Q. 4:			
Survey I	28	14	1
Survey II	37	10	1
Q. 6[a]:			
Survey I	10	27	6
Survey II	6	32	6
Q. 7:			
Survey I	4	27	10
Survey II	3	33	8

[a] Question 5 involved the test of significance between means.

TABLE 2 Result of Chi-Square Test of Significance

Question Number	Computed Value of X^2	Test of Hypothesis H_o: No Sig. Diff. in Response Freq.
One	$X^2 = 7.9 > 5.99$ [a]	H_o: is rejected
Two	$X^2 = 2.7 < 5.99$	H_o: not rejected
Three	$X^2 = 0.14 < 5.99$	H_o: not rejected
Four	$X^2 = 1.6 < 5.99$	H_o: not rejected
Six	$X^2 = 1.4 < 5.99$	H_o: not rejected
Seven	$X^2 = 0.9 < 5.99$	H_o: not rejected

[a] Chi-square value, df = 2, $\alpha = 0.05$ (95% significance level)

of SDOT management levels to various categories of questions in the two surveys.

A correlation analysis was also performed to examine the degree of association between category response frequencies of the two questionnaires. Correlation coefficients for the three categories of responses ("yes/very much," "to some extent," and "no") were $r_{xy} = 0.951$, $r_{xy} = 0.793$, and $r_{xy} = 0.932$, respectively.

CONCLUSIONS

Findings of two study surveys of upper managements of state departments of transportation have clarified a number of issues and questions. First, the impact of political change in DOT upper management on resource allocation decisions and decision implementation seems to be significant. Second, a number of shortcomings associated with current management system software programs were identified by the upper DOT management.

Finally, statistical analyses of responses of two different groups of upper DOT management officials to a set of questions in two independent surveys demonstrated a significant degree of consistency in the responses of these two groups of decision makers. This finding is of paramount importance because a significant number of critical decisions in the transportation industry, such as those dealing with resource allocation, service operations, alternative maintenance strategies, and plan evaluations, are based on the result of surveys of transportation management officials and the user public.

ACKNOWLEDGMENTS

The authors would like to thank the University of Alaska Fairbanks for facilities and space to carry on the work. Thanks are also due to the Regional Transportation Center TransNow for funding some of this work. Thanks are due to Bobbi Chouinard for typing this manuscript.

REFERENCES

1. J. Al-Kazily. Analysis of Park-and-Ride Lot Use in the Sacramento Region. In *Transportation Research Record 1321*, TRB, National Research Council, Washington, D.C., 1991, pp. 1–6.
2. S. Beroldo. Ridematching System Effectiveness: A Coast-to-Coast Perspective. In *Transportation Research Record 1321*, TRB, National Research Council, Washington, D.C., 1991, pp. 7–12.
3. P. Tsuchida and L. Wilshusen. Effects of the 1989 Loma Prieta Earthquake on Commuter Behavior in Santa Cruz County, California. In *Transportation Research Record 1321*, TRB, National Research Council, Washington, D.C., 1991, pp. 26–33.
4. A. M. Freas and S. M. Anderson. Effects of Variable Work Hour Programs on Ridesharing and Organizational Effectiveness: A Case Study, Ventura County. In *Transportation Research Record 1321*, TRB, National Research Council, Washington, D.C., 1991, pp. 51–56.
5. C. D. Angell and J. M. Ercolano. Southwestern Connecticut Commuter Transportation Study: An Analysis of Commuter Attitudes and Practices on Connecticut's Gold Coast. In *Transportation Research Record 1321*, TRB, National Research Council, Washington, D.C., 1991, pp. 66–72.
6. E. J. Miller and D. F. Crowley. Panel Survey Approach To Measuring Transit Route Service Elasticity of Demand. In *Transportation Research Record 1209*, TRB, National Research Council, Washington, D.C., 1989, pp. 26–31.
7. J. L. Hulsey, P. A. Koushki, and L. Raad. *Revisiting Management Information Systems for the Allocation of Funds for Highway Activities/Projects/Maintenance*. Final Research Report, TNW 92-05. Civil Engineering Department, University of Alaska Fairbanks, 1992.
8. C. White and S. Edner. Manager Retention and Job Change in the Transit Industry: A Survey of Manager Attitudes. In *Transportation Research Record 1144*, TRB, National Research Council, Washington, D.C. 1987, pp. 7–16.
9. C. K. Capo and D. J. Messmer. Effects of Disseminating Service Information and Fee Coupons on Bus Ridership. In *Transportation Research Record 1144*, TRB, National Research Council, Washington, D.C., 1987, pp. 34–40.
10. R. A. Hitlin, F. Spielberg, E. Barber, and S. J. Andrle. A Comparison of Telephone and Door-to-Door Survey Results for Transit Market Research. In *Transportation Research Record 1144*, TRB, National Research Council, Washington, D.C., 1987, pp. 87–91.
11. L. Wesemann, P. Dove, and N. Roadh. Comparison of Travel Behavior Before and After the Opening of HOV Lanes in a Suburban Travel Corridor. In *Transportation Research Record 1212*, TRB, National Research Council, Washington, D.C., 1989, pp. 41–52.
12. C. P. Flynn and L. J. Glazer. Ten Cities' Strategies for Transportation Demand Management. In *Transportation Research Record 1212*, TRB, National Research Council, Washington, D.C., 1989, pp. 11–23.
13. B. B. Norris and G. A. Shunk. The 1984 Home Interview Survey in the Dallas-Fort Worth Area: Changes in Travel Patterns, 1964–1984. In *Transportation Research Record 1134*, TRB, National Research Council, Washington, D.C., 1987, pp. 1–9.
14. I. M. Sheskin and P. R. Stopher. Pilot Testing of Alternative Administrative Procedures and Survey Instruments. In *Transportation Research Record 886*, TRB, National Research Council, Washington, D.C., 1982, pp. 8–22.

15. P. R. Stopher. Small-Sample Home Interview Travel Surveys: Application and Suggested Modifications. In *Transportation Research Record 886*, TRB, National Research Council, Washington, D.C., 1982, pp. 41–47.
16. P. R. Stopher and I. M. Sheskin. Toward Improved Collection of 24-Hour Travel Records. In *Transportation Research Record 891*, TRB, National Research Council, Washington, D.C., 1982, pp. 10–17.
17. P. A. Koushki, J. L. Hulsey, and L. Raad. State Departments of Transportation: Decision Process and Management Information Systems. *Journal of Management Information System Quarterly* (submitted).
18. *Organizational Charts of State Highway and Transportation Departments*. AASHTO, Washington, D.C., 1990.
19. R. E. Walpole and R. H. Myers. *Probability and Statistics for Engineers and Scientists* (3rd ed.). Macmillan Publishing Co., New York, 1985.

Publication of this paper sponsored by Committee on Management and Productivity.

Performance Measurement: Producing Results at the Oregon Department of Transportation

LAURA WIPPER

In July 1989, a pilot project was developed to implement Performance Measurement at the Oregon Department of Transportation (ODOT). This program quantifies measures of efficiency and effectiveness for work crews and the department as a whole and equates these data on a common scale. Performance Measurement represents a change in philosophy. Rather than monitoring individual activities, the program focuses on results. Four key factors in the accomplishment of results are tracked and the outcomes are communicated on a regular basis. Efficiency measures gauge the volume of production and the cost, whereas effectiveness measures track quality and customer satisfaction. This new focus has seen increasing success as the 27 ODOT highway division work crews (approximately one-tenth of the total work force) participating in the pilot steadily improved productivity, culminating in savings amounting to $1.8 million over the last 6 months of 1991. The success of the pilot has not only led to full implementation of the program at ODOT but also has caught the eye of Oregon's executive department, which mandated the program for all state agencies.

State government in Oregon has evolved over the past century by adding commissions, boards, agencies, and, in turn, program upon program for what seemed important reasons at the time. Those reasons can become lost over time and needs can disappear, and yet activities and costs of programs often remain. Without a mechanism for ongoing evaluation, these factors can build inefficiencies along with a lack of effectiveness and accountability because of the absence of a clear mission, purpose, and focus.

The Oregon Department of Transportation (ODOT) had no readily visible signs of this malaise, yet in reality suffered from some of these symptoms. In 1988, the new state highway engineer, Don Forbes, asked the following questions, which did not always have answers at the time. How much does it cost to maintain the average lane mile? How accurate are construction contract estimates? Does the transportation planning process lead to accomplishment of department goals? What is the public's perception of the department? The search began for a method to provide answers to these questions and others and to quantify the efficiency and effectiveness of the highway division.

A program was chosen that had been developed at the Oregon State University Productivity Center by James L. Riggs and Glenn Felix. The program, called Performance Measurement, establishes measures of efficiency and effectiveness. The purpose is to improve performance by providing a tool to quantify and communicate results. It also provides data on which to base decisions for optimizing efficiency and effectiveness.

With the strong support of Forbes, now the ODOT director, Performance Measurement is currently in full implementation throughout ODOT. To date, more than 170 work groups have defined measures and are gathering data and receiving regular feedback on their overall performance.

Office of Productivity Services, Oregon Department of Transportation, 135 Transportation Building, Salem, Oreg. 97310.

WHY MEASURE PERFORMANCE?

A well-managed organization, be it public or private, needs to have a clear purpose and goals and objectives, base decisions upon data, provide regular feedback, and have some form of recognition for above-the-norm performance. The general state of our nation's economy suggests that many U.S. companies do not enjoy this type of management even under the powerful motivation of profits. Government agencies, too, suffer a similar lack. Initiatives for tax overhaul indicate that the public has lost confidence in government to operate efficiently and effectively. Over the decades, as layers of programs build up, a governmental organization can lose its focus without a regular, data-driven evaluative process in place. Performance Measurement provides that evaluative process for ODOT.

Performance Measurement clarifies the overall mission of ODOT and the purpose of its branches, sections, and units. It provides direction by presenting data against a backdrop of historical averages and historical potentials or goals. Presented in a matrix, seemingly disparate information can be converted to a common scale, which allows evaluation of the interaction between efficiency and effectiveness. This enables managers and staff alike to base decisions on data and to evaluate strategies for improvements to achieve the optimum balance between improved efficiency and effectiveness. This feedback is provided on a regular basis to help managers manage better at the program level and to communicate to those involved what is going well and what needs more attention. Because the focus is on programs and work groups, not individuals, teamwork is improved at all staff levels. The simple act of performance measurement alone usually prompts improvements because what is measured is what will surely get done.

KEY ELEMENTS OF PERFORMANCE MEASUREMENT

Results, Not Activities

Results are the point at which a product or service is delivered; activities are the actions that lead to delivery of that product or service. In the past, most forms of measurement at ODOT placed greater emphasis on forecasting and tracking activities—work load measurement. ODOT now places emphasis on results. Activity-based measurement only reinforces the accomplishment of activities; results-based measurement reinforces the accomplishment of results.

Group-Based Measures, Not Individual Measures

A key part of the process to develop performance measures is the involvement of the work group. Work groups are taught the concept of Performance Measurement and then facilitated in development of measures for their unit. In many cases, the individual members of the groups have minimal awareness of all functions of the group, so the discussion fosters a better awareness of the work group's priorities. Managers have reported improved work group cohesiveness following such discussion. Measuring the performance of individuals can be divisive, whereas measuring group-based results causes the members of the group to work better together to produce better results.

Performance Measures, Not Work Load Measures

Where work load measures capture just the number of activities, performance measures gauge results. When only activities are counted, desired results may not be produced because the focus is limited to the activities. This limited focus does not provide an environment to culture improvement strategies, whereas measurement of results does provide such an environment. As improvement strategies surface, they can be evaluated via the performance matrix.

Work Groups, Not Individuals, Develop Measures

The process of implementing Performance Measurement begins with a management team that develops broad guidelines. The work group then develops performance measures on the basis of their intimate knowledge of what they do and what they believe to be important. This ensures more accurate measures because the people who are actually doing the work know best what is being done.

Efficiency and Effectiveness Measures, Not Just Amount Done

Performance Measurement looks at both efficiency and effectiveness. Efficiency means doing the right things with the best use of resources. Effectiveness means doing the right things well and customer satisfaction with the product or service.

This program tailors measurement of quality to the product, service, and customer because quality holds different meanings for different people. For example, timeliness, accuracy, and availability of services equal quality for the driver and vehicle licensing functions of ODOT. Pavement condition and bridge sufficiency ratings are measures of effectiveness for not only highway maintenance, construction, and design, but also the department as a whole.

Credibility in State Government, Not Distrust of the Unknown

ODOT's goals, and those of other government agencies, and information about how well they are being achieved can be conveyed to the public via Performance Measurement. Budgets can be based on program performance and presented more effectively to the legislature because efficiency and effectiveness are demonstrated. This can also create a new role for government, which has not habitually played a proactive role in communicating exactly what it is trying to accomplish and how well it is doing.

The Visual Element

The performance matrix, a complex-appearing document, is actually how Performance Measurement keeps things simple. Once understood, the performance matrix will tell the user at a glance whether an entire organization's performance in key areas identified is improving or declining.

PERFORMANCE MATRIX

The matrix is not as complex as it initially appears. In fact, it can be understood in less than 30 min. In the sample matrix in Figure 1, Row A identifies emphasis areas of efficiency and effectiveness. Efficiency measures look at production volume and cost. Effectiveness measures look at such factors as timeliness, accuracy, and conformance to standards. A mandatory effectiveness measure is customer satisfaction, which is the customer's perception of products and services provided. Safety and work life quality are two more areas that should be included.

Row B identifies more specific key measures of performance important to the organization in each emphasis area. In the first column of Figure 1, the key measure is transactions per FTE (full-time equivalency). Row C contains the actual results achieved over the reporting period for each of the measures. In this sample matrix, the actual average transactions per FTE was 130.

Row D shows the potential results targeted to be achieved (in other words, a goal for each measure). Potential is based on either a historical best or an absolute goal such as 100 percent customer satisfaction or zero errors. The 10 is the level achieved when the goal is reached. In the example, the potential for transactions per FTE is 200.

A	Emphasis Areas		EFFICIENCY		EFFECTIVENESS				
			LABOR	COST	QUALITY		PERCEPTION	WORK FORCE	
B	Key Measures of Performance		Transactions Per FTE	Cost Per Transaction	Percent Delivered On Time	Percent Of Work Corrected	Percent Satisfied Customers	Work Life Quality Index	Safety
C	Actual Results		130	$2.30	90%	12%	80%	−10	0.11
D	Potential	10	200	$1.70	100%	0%	100%	100	0
		9	190	$1.75	98%	1%	98%	90	0.01
		8	180	$1.80	96%	2%	96%	80	0.02
		7	170	$1.85	94%	3%	94%	70	0.03
		6	160	$1.90	92%	4%	92%	60	0.04
		5	150	$1.95	90%	5%	90%	50	0.05
		4	140	$2.00	88%	6%	88%	40	0.06
		3	130	$2.05	86%	7%	86%	30	0.07
		2	120	$2.10	84%	8%	84%	20	0.08
		1	110	$2.15	82%	9%	82%	10	0.09
E	Baseline	0	100	$2.20	80%	10%	80%	0	0.1
		−1	90	$2.25	78%	11%	78%	−10	0.11
		−2	80	$2.30	76%	12%	76%	−20	0.12
		−3	70	$2.35	74%	13%	74%	−30	0.13
		−4	60	$2.40	72%	14%	72%	−40	0.14
		−5	50	$2.45	70%	15%	70%	−50	0.15
F	Level Achieved		3	−2	5	−2	0	−1	−1
G	Relative Weight		25	15	15	10	20	10	5
H	Earned Value		75	−30	75	−20	0	−10	−5

Performance Index: 85

FIGURE 1 Performance matrix.

Row E lists baseline results or average, standard, or regularly expected performance based on historical averages. The 0 is the level achieved when average results are achieved. In this illustration, baseline for transactions per FTE is 100.

Because neither exactly average nor potential results are always achieved, a range of performance is also identified. Since performance, when measured, is more likely to be above than below average, ODOT's format contains 10 levels above baseline and only 5 below. The range between each level is determined by dividing the difference between baseline data and potential by 10. For transactions per FTE, 200 (potential) minus 100 (baseline) divided by 10 equals a range of 10 per level. This same range is taken in the opposite direction in the negative levels.

Row F is where the level achieved based on the actual results is shown. These levels are the common scale that can compare the interrelationships between measures that would otherwise be incomparable. The level achieved is reflected here because it is multiplied by the relative weight shown in Row G.

Relative weight in Row G is a method of weighting or prioritizing the key performance measures. By convention, all the relative weights in a matrix total 100. The assignment of relative weights is determined by the work groups once their measures have been developed. This process is somewhat arbitrary, but the measure of greatest importance is the measure with the greatest relative weight. Conversely, the measure with the lowest relative weight is the measure of lowest importance. In Figure 1, the labor efficiency measure, transactions per FTE, has the greatest weight, so it is of highest importance. The measure with the least weight and of lowest importance is a workforce measure, safety.

Row H shows the earned value of each measure, which is the result of multiplying the level achieved in each measure by its relative weight. For example, Level 3 was achieved in the transactions per FTE measure in Figure 1, which has a relative weight of 25, which equals an earned value of 75.

The performance index at the bottom of the matrix is the sum of the earned values for all measures contained in the matrix. This one number indicates overall how well an organization or work group juggled its priorities. A total of 0 means that the overall performance was average. A positive number means some degree of overall above-average performance. A negative number means some degree of overall below-average work. Because the relative weights must total 100, achieving potential in all measures would equal a performance index of 1,000; achieving level −5 in all measures would equal a performance index of −500, thus giving some relativity to the positive or negative degree of overall performance. A performance index of 85 in the sample matrix indicates slightly above-average effort.

Various levels of achievement attained in each of the key measures contribute to an overall indicator. These measures can be evaluated individually to determine whether performance was below average in any specific area. When performance is below average in more than one area, the relative weights and the earned values can be examined to focus improvement strategies. In Figure 1, equal negative levels were achieved in two measures, cost per transaction and percent of work corrected. Cost per transaction would be the area of highest priority to improve because of its higher relative weight and greater negative earned value.

Analysis of the matrix in Figure 1 might reveal a work force working overtime to deliver increased products/services with a greater percentage on time. The negative side is a tired staff making more errors and working less safely. Increased timeliness counterbalanced by decreased accuracy accounts for average customer satisfaction.

IMPLEMENTATION PROCESS

Implementation begins with a steering committee consisting ideally of all senior managers or, at a minimum, the agency head, the budget officer, information services manager, personnel manager, and a performance coordinator. This group is taught the concept of Performance Measurement before going on to develop guidelines and performance measures that are very broad in scope.

The midlevel management team participates in the same workshops as Performance Measurement progresses to the next level in the agency. This group develops measures that are still broad in scope, yet specific to that level in the agency while conforming to the guidelines and measures developed by the steering committee. The measures continue to get more specific as work groups learn Performance Measurement. Through workshops, they go on to develop their measures within the steering committee's guidelines.

At each level, the group decides what is important to measure within agency guidelines. This hierarchal approach allows data from all over the agency to feed into agency-level performance measures. For example, one motor vehicles division quality measure tracks timeliness, which is a measure of the percentage of transactions meeting service levels in 12 different service areas. The work groups then develop a measure to track the timeliness of the specific service offered by the group.

RESULTS OF PERFORMANCE MEASUREMENT

Figure 2 plots the overall average performance indexes of nine pilot work groups since July 1989. The overall average indexes fluctuate a great deal, but the trend is upward in improved performance, a shown by the regression line in Figure 2.

A recent comparison of the performance of seven additional groups with the performance of the original nine work groups shows virtually the same trend when compared on the same graph in Figure 3. Whereas this 6-month comparison (July through December 1991) might give an impression of declining performance, the expectation is for an overall improvement in performance similar to that in the original nine groups. When the performance of the original nine groups is viewed

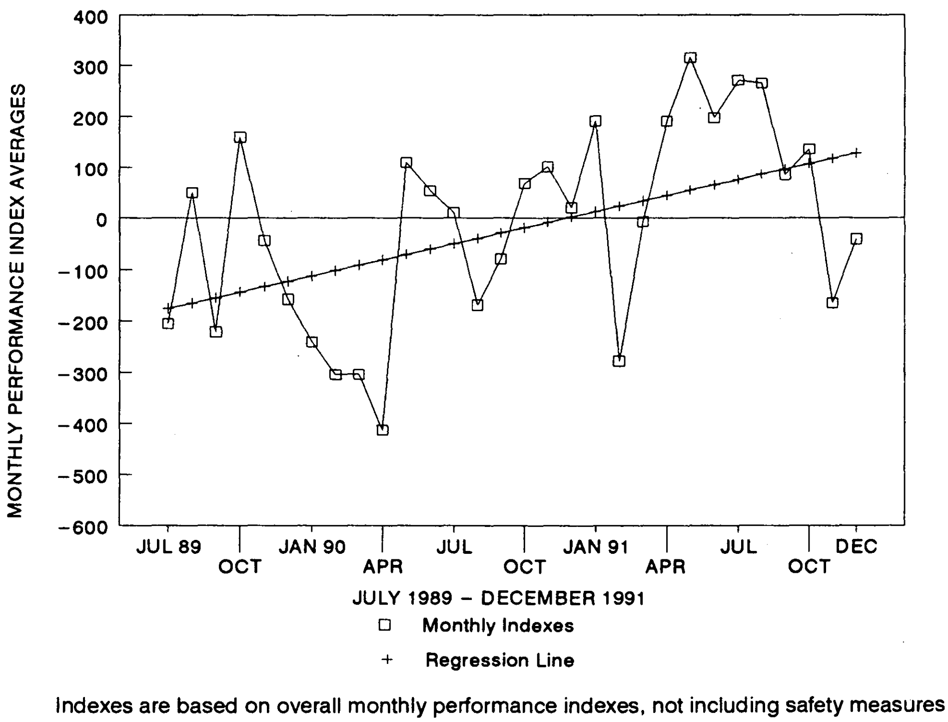


FIGURE 2 Performance index averages.

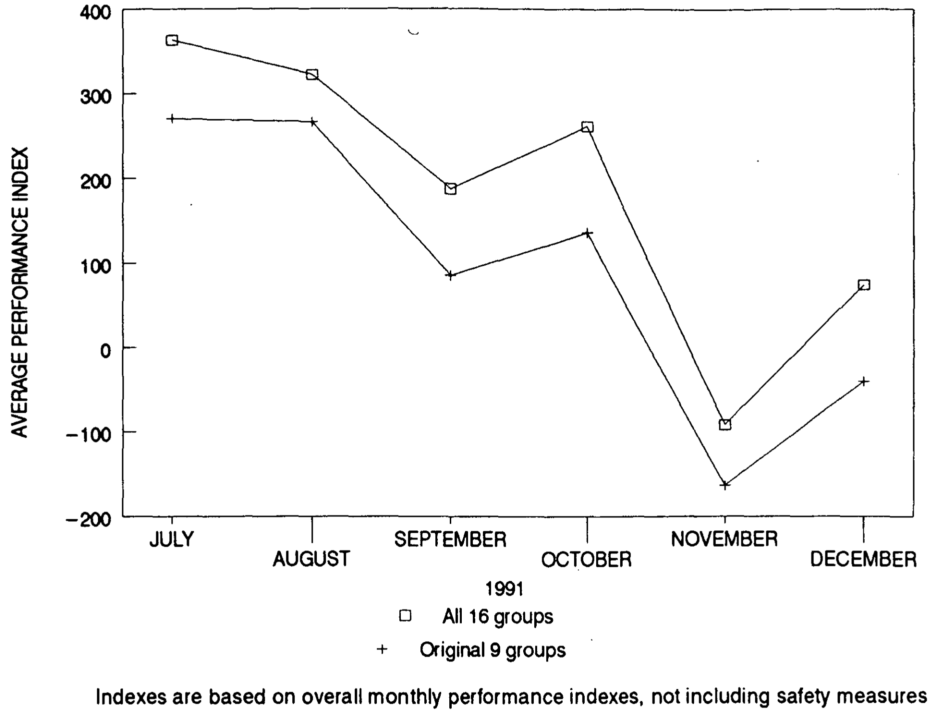


FIGURE 3 Average performance index (July–December 1991).

over a longer period, it fluctuates, but the long-term trend is upward (Figure 2).

Figure 4 shows dramatic improvements in the cumulative 6-month average indexes for the same nine pilot groups. The trend is again clearly upward but even more evident when comparing the same period, July through December, in 1989, 1990, and 1991.

Figures 5 and 6 show performance trends for two specific areas: efficiency and effectiveness. Efficiency measures look at volume of output and cost, whereas effectiveness measures quantify product/service quality and customer satisfaction. The average total earned values of the measures of effectiveness hover around baseline performance in Figure 5. Figure 6 shows a familiar trend. Improved performance causes the graph of average total earned value for efficiency measures to be virtually the mirror image of the overall average performance indexes shown in Figure 2. The slight downward trend in average total earned value for effectiveness measures over the last 6 months of 1991 warrants further analysis to maintain the optimum balance between efficiency and effectiveness.

Figure 7 isolates a single measure of effectiveness, customer satisfaction, for the nine original work groups. Although efficiency has improved, customer satisfaction has hovered so close to baseline performance that it is essentially a straight line. One group, weighmasters, who weigh and inspect trucks and subsequently cite truck drivers when the truck is out of compliance, expected very negative results from a customer survey. This group was pleasantly surprised by the initial results and went on to increase the satisfaction rate by 50 percent in the next 6 months.

LESSONS LEARNED

During the pilot phase of Performance Measurement, we learned that the program could beneficially affect results. In addition to seeing improved performance, four key lessons were learned to better implement the program department-wide.

1. An automated reporting process must be in place before agencywide implementation begins. Without automation, data gathering can become extremely labor intensive, making it difficult to produce timely reports. Once the measures have been developed and data gathering has begun, work groups are anxious to receive regular feedback. Confidence in the program can be lost if this part of the program is not performed.
2. Union representatives must be involved at every step of both a pilot and full implementation to learn the concept, the process, the reasoning behind steering committee guidelines and, above all, to realize that performance measures are based on results produced by a group and are not individually focused.
3. A communication and decision-making process must precede agencywide implementation. The steering committee must decide such things as the level of the agency responsible for review of the measures, baselines, and potentials; the frequency of review; and the criteria to be used to determine baselines.
4. All levels of management must be actively involved in the Performance Measurement process and kept informed.

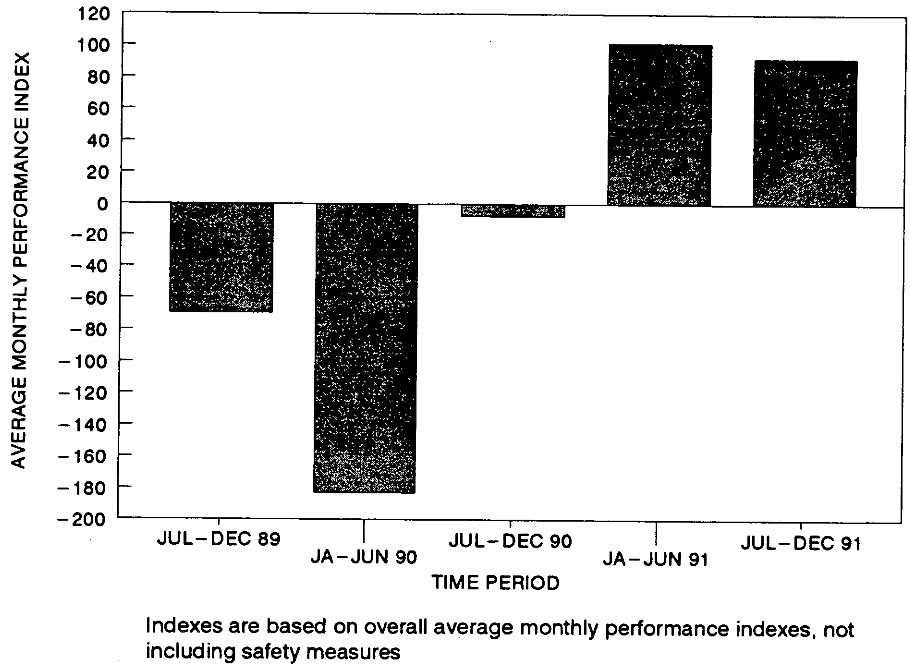


FIGURE 4 **Six-month performance index averages.**

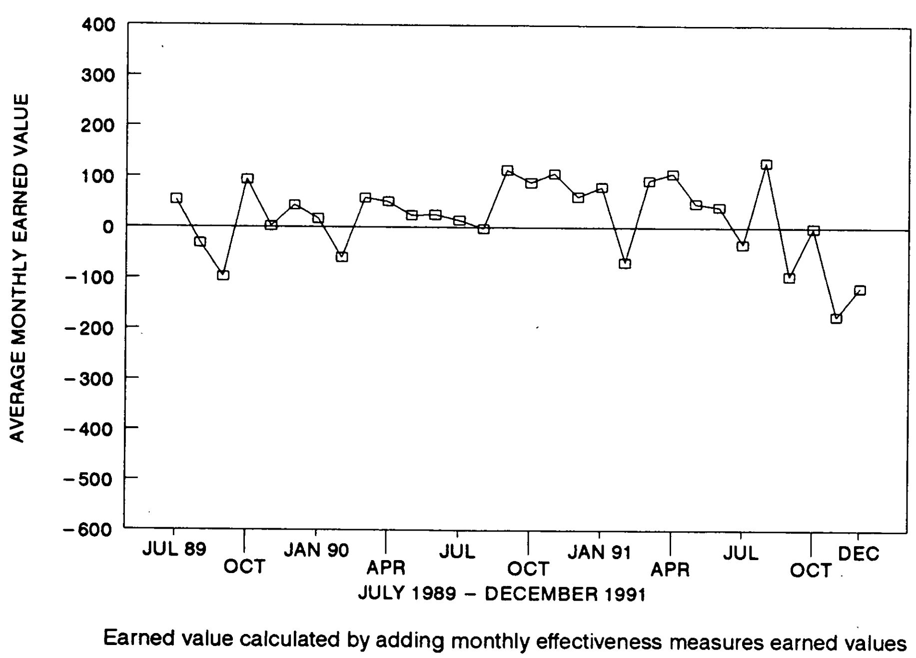
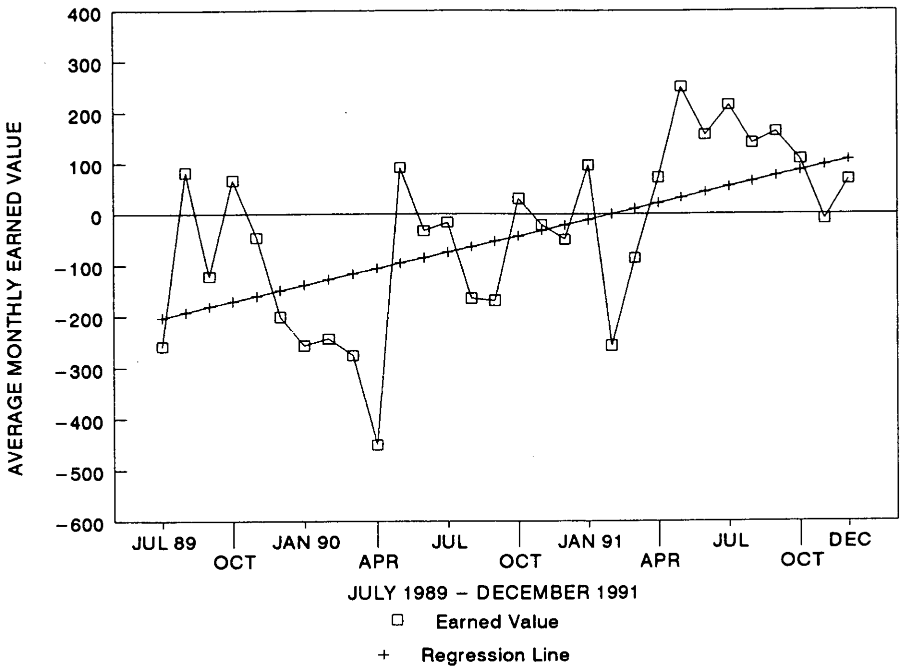
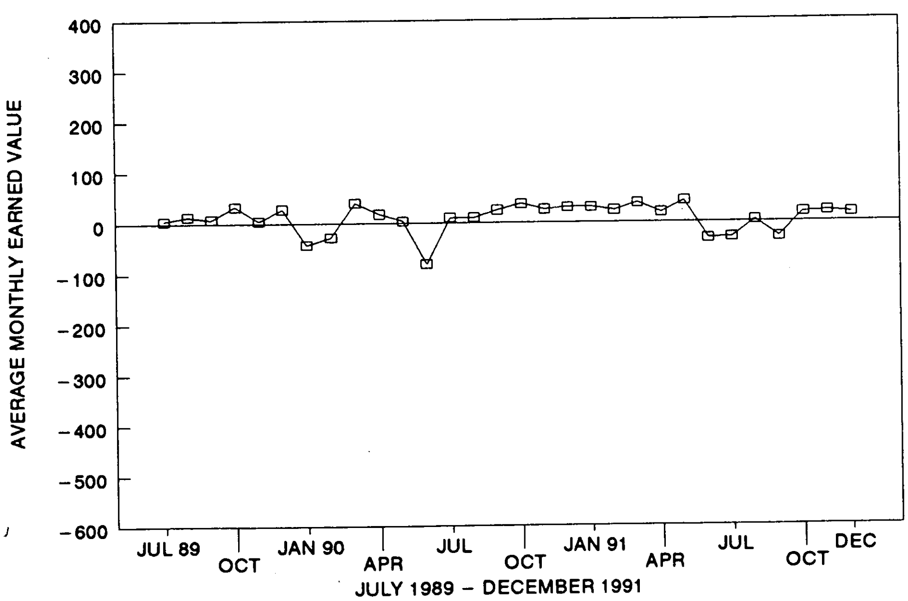


FIGURE 5 **Average monthly effectiveness earned value.**



Earned value calculated by adding labor efficiency measures earned values and materials measure earned value

FIGURE 6 **Average efficiency earned value.**



Earned value is customer satisfaction measure earned value for each group

FIGURE 7 **Average customer satisfaction earned value.**

In addition, senior management must understand, support, champion, and promote the program.

ODOT has been quick to incorporate these improvements into the program to streamline implementation as it continues through the agency.

CONCLUSIONS

From July through December 1991, 27 pilot work groups, amounting to 7 percent of the work force or 350 FTE, saved ODOT $1.8 million through improved efficiency and effectiveness. In addition, if success can be measured by what others imitate, Performance Measurement at ODOT can be considered a resounding success. What began as a pilot program within ODOT has become a full-scale initiative throughout state government in Oregon. The Oregon Executive Department recognized the value of the Performance Measurement program and mandated it for all state agencies. ODOT was instrumental in the success of this initiative by teaching representatives from over 115 state agencies the program concept and implementation. ODOT expertise assisted countless agency management teams with development of performance measures.

Oregon was awarded the "E for Effort Award" by *Financial World* magazine as a result of the magazine's annual evaluation of state government. The award is given to honor a state that has taken a leadership role in dealing with present issues facing state government. In the annual rankings by *Financial World*, Oregon has moved from 34th in 1990 to 17th in 1991 to 6th in 1992. The state's "trailblazing work in performance measurements" was the primary reason cited by the magazine for Oregon's movement into the top 10.

Beginning in August 1992, all agency budgets will be presented in the context of agency-level performance measures, thus providing a consistent platform to communicate the efficiency and effectiveness of programs throughout state government. Other agencies plan to do so, but to date, ODOT is the only agency implementing the program at all levels in the organization. We look forward to improved management of transportation programs via Performance Measurement.

Publication of this paper sponsored by Committee on Management and Productivity.

Customer Satisfaction: The Next Frontier

STEVEN SILKUNAS

In an introduction to customer satisfaction and its measurement, key concepts, conceptual frameworks, and outlines for implementation are derived from business research literature. An agenda for public transportation based on customer-driven expectations and requirements is proposed. Quality, market research, and measurements are reviewed as management tools with appropriateness in transportation.

To even the most casual of observers, American business has shifted its emphasis toward a return to basics. Econometrics, modeling, and strategic planning all have their place, but all are several steps away from the fundamentals. Financial wizardry and bottom-line obsessions have, of late, been called to task. The vogue words are quality and value. Improvement is the call of the day.

Interestingly enough, even quality and value alone or together miss the mark. Redenbaugh, Chairman of Action Technologies, Inc., has some insightful comments:

> Customers do not buy quality. They buy satisfaction. The two are categorically different.
> McDonald's . . . has continually increased its market share and eliminated competitors.
> Does McDonald's sell quality? It has won no culinary or nutritional awards that I know of. McDonald's sells customer satisfaction, and it manages this religiously. It obsessively measures, monitors, and promotes customer satisfaction.
> . . . If we confuse quality with customer satisfaction, we fall into big trouble. Satisfaction is what customers pay for, over and over again. It is this recurrence that produces profitability. Companies that forget about customer satisfaction may disappear. (*1*, p. 11)

In some respects it is surprising that *Business Month* printed Redenbaugh's remarks. Customer satisfaction has neither glitz nor glamour. It is basic. Be that as it may, a recent *Business Week* feature story was on "King Customer." On its front cover *Business Week* admonished its readers to "forget market share. Stop worrying about your competitors. The companies that are succeeding now put their customers first" (*2*).

The thoughts expressed in *Business Month* and *Business Week* were also expressed in the *Harvard Business Review*. Levitt points out that

> In the end—really, at the outset—every activity or purpose . . . is about getting and keeping customers. No matter what fashion or idea comes or goes, the one absolutely essential and therefore inescapable thing that must be done, and to which attention therefore always returns, is marketing. That is why marketing is always getting rediscovered. . . . When attention slackens, as attention invariably does in all things, things go bad. When bad enough or not good enough, management returns to first principles—getting and keeping customers. (*3*,p.8)

Southeastern Pennsylvania Transportation Authority, 841 Chestnut Street, Philadelphia, Pa. 19107-4484.

These thoughts lead to the thrust of this paper: the state of customer satisfaction in public transportation. Does public transportation perform as well, better, or poorer than its counterparts in the private sector? Do people who opt to ride buses and trains do so because public transportation is the most desirable alternative or because it is the least objectionable? In a study National Family Opinion performed for The Conference Board, public transportation had lackluster performance: 13.2 percent rated it a "good" value, 52.1 percent rated it an "average" value, and 34.8 percent rated it a "poor" value (*4*).

Table 1, which shows selected products and services, places mass transit in perspective.

THE CASE FOR MONITORING CUSTOMER SATISFACTION

Complacency and customer satisfaction are diametrically opposed. In a survey of the literature, two types of organizations pursue customer satisfaction with unrelenting zeal. The first type of organization is run by the founder or carries on the spirit of the founder. L. L. Bean and Disney are good examples of this type of company.

The second type of organization has found itself "up against the wall." Foreign competition, adverse public sentiment, or the threat of insolvency has pushed a number of firms to pursue customer satisfaction as a means to achieve a competitive advantage. Motorola and Xerox, winners of the Baldrige award, and Florida Power and Light, the winner of Japan's Deming Award, are good examples of this second type of firm.

Few good examples of a commitment to customer satisfaction exist in the public sector, although the number is increasing each year. Madison, Wisconsin, and Baltimore, Maryland, are often cited as models for emulation. Clearly "PATHursday," a program in which Port Authority Trans Hudson (PATH) managers go out to meet customers, is one good example in the transit industry. This program has been successfully copied by the Southeastern Pennsylvania Transportation Authority (SEPTA) and the New York City Transit Authority. The Long Island Rail Road was the first to measure customer satisfaction through self-directed questionnaires on a large-scale, systematic basis. A few other properties have also done major work in this regard, most notably MTA, Baltimore, and SEPTA. Boston, Chicago, Philadelphia, Vancouver, Seattle, and Portland have engaged market research firms to use telephone surveys on a smaller scale. Public transportation, however, generally follows in the shadow of the private sector, which has long used market research for developing its business ventures. Admittedly, some of this

TABLE 1 Consumer Rating of Selected Products and Services (Ranked by Composite Score)

Product	Good	Average	Poor	Composite*
Poultry	40.4	52.0	7.6	66.4
Video Rentals	39.3	52.7	8.0	65.7
Electricity	28.8	52.2	19.0	54.9
Cars (foreign)	22.3	54.1	23.6	49.4
U.S. Postage	21.2	50.5	28.4	46.5
Fast Foods	12.1	59.1	28.8	41.7
Air Fares	10.9	57.6	31.5	39.7
Mass Transit	**13.2**	**52.1**	**34.8**	**39.2**
Sports Events	7.7	50.4	41.9	32.9
Movie Tickets	8.9	40.8	50.4	29.3
Lawyers' Fees	4.8	40.6	54.6	25.1
Hospital Charges	5.1	33.5	61.4	21.9

*Composite = Good + 1/2 Average

Source: The Conference Board/National Family Opinion, Inc. February, 1990

is perception. But, as any phenomenologist can attest, perception is reality.

Why does public transportation perform at a mediocre level (according to the Conference Board study) in customer satisfaction? One can postulate a simple explanation: the customer is often only a secondary driving force. In public transportation, the emphasis is on "riders" or "passengers"; "customer" is often a foreign, or at least unfamiliar, word. This was the case when public transportation was still in private hands; public takeovers have, if anything, made the situation worse. A customer today is much more difficult to define: it is both the person who boards a vehicle and the agencies that oversee the activity, together with the elected officials that provide the funding.

Since fare box revenue accounts for less than half the income at many properties, consumers are subjugated to other interests. To wit, accountability tends to follow the source of funds. Stated another way, transit providers court the funding sources; they are typically at best indifferent to the consumer. (Privatization in this conceptualization can be viewed as a basic move toward customer accountability and satisfaction.)

The case for customer satisfaction in public transportation is straightforward: when public transit is not attuned to customer needs, the long-term support for and growth of transit is in jeopardy.

KEY CONCEPTS FROM CUSTOMER SATISFACTION RESEARCH

• The goal of a customer-driven organization is to find out what its customers want and to deliver the product right the first time. Although this is a simple concept, in execution it is far from simple.

• A customer-driven company is customer focused. This tautology is more than just words—it expresses the raison d'être. The contribution of the *Excellence* series by Peters, Austin, and Waterman is that it centered attention on the customer.

• A customer-driven company addresses the dynamics of permanence and change. The expectations of today may be the expectations of tomorrow. Then again, they might not be. This is the role of market research. Good market research identifies customer needs and perceptions. Shoddy market research on the other hand can result in a double-pronged disaster: it can add dimensions that are undesired and it can neglect dimensions that are critical (*5*). Far more devastating than poor research is the attitude that "I know what people want," which leads to the pious apothegm of "pride going before the fall."

• Customer-driven firms are fueled by enthusiasm. The word enthusiasm here is used in a narrow, classical sense of being "filled with the spirit." In both the manufacturing and the service sectors, dedicated employees, well trained and empowered to act, can be the critical difference between success and failure. Quality circles are one way of integrating this into the fabric of the organization; and there are others. What is important is that both the organization and the employees assimilate the change. Lip service and quick fixes fail; employees can tell when the organization lacks commitment.

• Customer-driven firms have a firm concept of quality derived from the customers' specifications. As a philosophic inquiry, the study of quality is extensive. The Augustine maxim

"How can you be satisfied with a good when you know there's a better?" is at one end of the spectrum. Pirsig's *Zen and the Art of Motorcycle Maintenance* is a more sophisticated undertaking. In the marketplace, though, robust quality should be the goal:

> Of course, customers do not give a hang about a factory's record for staying "in spec" or minimizing scrap. For customers, the proof of a product's quality is in its performance when wrapped, overloaded, dropped, and splashed. Then, too many products display temperamental behavior and annoying or even dangerous performance degradations. We all prefer copiers whose copies are clear under low power; we all prefer cars designed to steer safely and predictably, even on roads that are wet or bumpy, in cross winds, or with tires that are slightly under or overinflated. We say these products are robust. They gain steadfast customer loyalty. (*6*)

Although the foregoing concepts are positive, customer satisfaction would scarcely be a significant issue if perfectly consummated consumer transactions were the norm. The fact of the matter is that many marketplace transactions go awry. This subject was addressed in the landmark study *Consumer Complaint Handling in America* on the basis of research conducted between 1974 and 1979 and published in 1979. This effort was initiated by the United States Department of Consumer Affairs and conducted by the Technical Assistance Research Programs Institute (TARP), a Washington, D.C., consultancy.

Follow-up research, done in 1984 and 1985, showed the Sisyphean nature of customer satisfaction: as business has gained expertise in the handling of consumer problems (*7*), the American consumer has become more sophisticated and effective in dealing with problems encountered in the marketplace (*8*).

In general, TARP studies indicate that sooner or later a customer will experience a problem. Although research indicates that good service does not guarantee that the customer will return, the experience of poor service makes a repurchase less likely (*9*). As indicated in Table 2, which profiles various market segments, many customers do not complain, and of those who complain, many are dissatisfied with the end response.

Dissatisfaction represents a major impediment to repeat business. Table 3 indicates how a slipup in customer service affects repeat business. Not only does dissatisfaction result in the potential loss of a customer, the word-of-mouth retelling of experience has far-reaching consequences. In TARP research it was found that, in small purchases, customers who had a satisfactory experience told five people. At the same time, customers who were dissatisfied told 10 people (*9*).

Although these data can be seen as a cause for corporate distress, the very opposite may be the case. Complaints can be used as a marketing opportunity. Complainants who see their problems resolved by attentive, responsive companies often forgive past transgressions. Indeed, responsiveness is often rewarded with loyalty (*7*).

Noncomplainants complicate the satisfaction problem. In general, it has been found that the nonarticulated complaints are usually the easiest ones to resolve. With little effort, it is possible to remedy the situation and encourage repeat patronage (*7*).

The TARP conclusion is noteworthy:

> Given the high costs of marketing, it may be less expensive to resolve the problem of an existing customer than to win a new customer. Repeat sales are especially important in highly competitive industries. The data, therefore, suggest that it may be in business' self-interest to solicit complaints. (*7*, p. 44)

Noteworthy in this regard is that few transit properties offer a service guarantee. Southern California Rapid Transit District (SCRTD) launched its service warranty program. SCRTD's program was generous and administered at the operator level. SCRTD lacked specific means of gauging lost revenue, but its estimate of revenue loss was considerably less than what had been projected.

Two other concepts can also be found in the literature. One deals with turning complaint handling units into strategic busi-

TABLE 2 Unarticulated Complaints

Type of Service	Percentage of Unarticulated Complaints	Percentage of Customers Who Voice Complaints And Are Dissatisfied With Response
Financial Services	39	26
Auto Repair Services	26	69
Telecommunications	45	29
Car Rental	55	39
Direct Marketing	62	31
Utilities	45	45

Source: TARP Industry Specific Data

TABLE 3 Likelihood of Customer Repurchase After Dissatisfaction

	Major Complaints (Over $100 Losses)	Minor Complaints ($1-$5 Losses)
Non Complainants	9%	37%
Complainants		
Complaints Not Resolved	19%	46%
Complaints Resolved	54%	70%
Complaints Resolved Quickly	82%	95%

Source: TARP/National Consumer Survey

ness units with quality as a focus. Another key concept is that of churning. Churning is seldom easily detectable and refers to the erosion of the customer base. Typically, marketing efforts (e.g., special pricing and promotions) can bring in new customers, but the gross effect is dissipated when existing customers become dissatisfied and leave. If the indicators (revenues, transfers, etc.) are positive and on the increase, this critical area can go undetected.

CONCEPTUAL MODELS IN CUSTOMER SATISFACTION

Any number of conceptual frameworks can be applied to customer satisfaction. In general, all can be reduced to the quantitative expression

Satisfaction = expectations − performance

In practice, performance can be equated with perception of the performance. This basic expression has been adopted throughout the customer satisfaction measurement field, with modifications made by users and practitioners. A model based on a customer-driven organization is shown in Figure 1. This model, of course, presumes that market research has already identified desires and expectations as customer needs.

OUTLINE FOR IMPLEMENTING CUSTOMER SATISFACTION AS COMPETITIVE STRATEGY

Customer satisfaction is a call to self-examination, which is as it should be. At the same time, there is the human tendency to slough off criticism. The experts in the field seemingly know better than the customer. Herein lies the critical danger:

> It's tempting to forego analysis because you assume you know what customers expect. But assumptions don't make effective customer service strategies. Inward-looking companies that are guided by industry norms and their own past practices end up with inappropriate strategies, lower market shares, and anemic profits. Time after time, studies have shown large differences between the ways that customers define service and rank the importance of different service activities and the ways suppliers do. (*10*)

Who, then, should take the lead in the measurement of customer satisfaction? There are a number of possibilities providing that three criteria are met: first, the measurer must have independence; second, the measurer must be knowledgeable about customer satisfaction; and third, the measurer must have the appropriate research tools. The latter two are self-explanatory. Independence, however, is another story. The assessment of customer satisfaction is such that, without independence, there would be a tendency to be less than totally frank. Unless there are compelling reasons to do otherwise, the measurement of customer satisfaction should be outside the operations and marketing groups. A task force assigned this duty is one possibility. An internal audit group is another. An outside consultant experienced in customer satisfaction has distinct advantages.

The following in brief outline form is derived from a presentation of L. Crosby (at AMA/ASQC First Customer Satisfaction Conference, Chicago, 1989) delineating the process:

1. Establish the research question; begin exploratory research.
2. Define standards for performance relative to customer expectations.
3. Develop standards of comparison (a) over time and (b) against the competition.
4. Define the population to measure (current customers, former customers, noncustomers).
5. Select the sampling mode (telephone, self-administered questionnaire, intercept).
6. Select sampling methodology, procedure, and sample sizes.
7. Determine where to measure (customers' locations, provider's location).
8. Determine time frame (intermittent, periodic, continuous).
9. Get organizational buy-in.

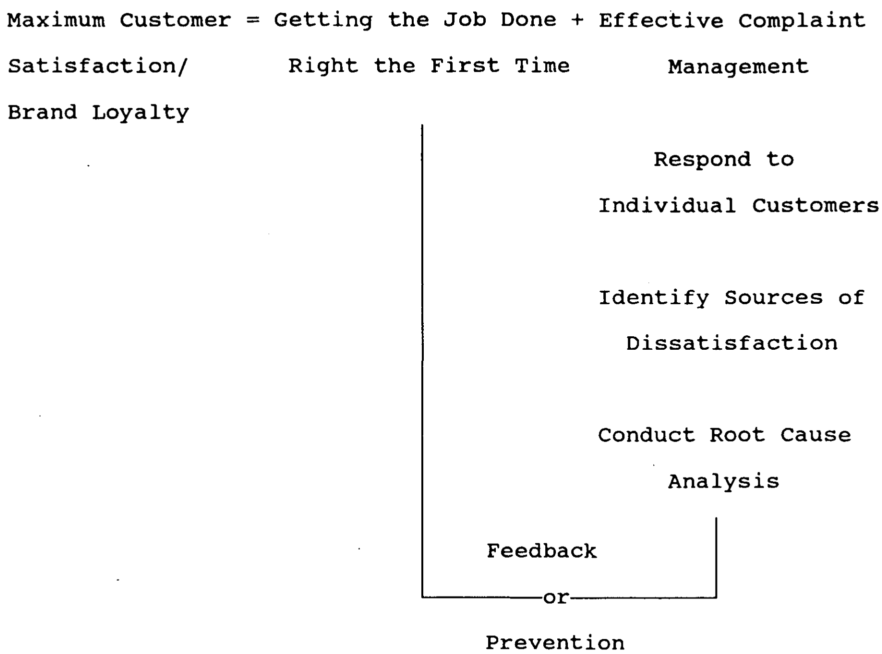


FIGURE 1 Maximum customer satisfaction (source: TARP).

THE NEXT FRONTIER: AN AGENDA

This paper has conspicuously avoided references to mass transportation literature. This was intentional: few examples exist. Programs derived from the *Excellence* series are a start. Brogan et al. (*11*) document a move in the right direction. The sessions on total quality management cosponsored by the TRB Committees on Management and Productivity and Transit Management and Performance at the 1992 TRB Annual Meeting provide another example. With several notable exceptions, though, few articles of substance have been disseminated. Although due consideration to employee involvement and buy-in is critical, more important, is a responsiveness to the customer (*11*).

The agenda for the 1990s and beyond should include the following:

1. Transportation services should be based on market research. The tendency of transportation professionals is to rely on models and "professional judgment." Market research tools are available today to form a sound basis for decision making. Considerable inroads have been made in this area. PATHursday was originally started as a public relations effort but was taken to the next step and integrated into operations and the strategic planning process. With respect to this, L. M. Rocha of PATH noted that

> by linking market research with the strategic planning process we were able to reduce the stigma associated with negative findings. Previously, market research findings that ran counter to management's intuition or that showed the need to improve certain areas were resisted as reflecting badly on management's past decision. (*12*,p.4)

2. Service standards should be based on customer demands rather than on industry standards. Headways, loading standards, and cleanliness, to name a few areas, need to be viewed from the customer viewpoint. The customer should decide whether a bus is clean; reliance on a standard of running buses through washers falls short. MDBFs and passenger miles fail to relate to a customer's direct experience and lexicon.
3. Customers should be treated as such, rather than impersonalized into fares or passengers. It is very difficult to treat living, breathing customers poorly; a remote, third person is another story altogether. Priority should be given to users of the services over funding agencies. Satisfied users can be prime movers in the funding process. The political process is ultimately sensitive to the needs of its constituents.
4. Customer satisfaction should be qualitatively defined and assiduously measured. Further, this should be done repeatedly (quarterly, monthly) and at the most basic (route and trip) levels.

This represents a start.

REFERENCES

1. R. G. Redenbaugh. Speak Out. *Business Month*, June 1990.
2. *Business Week*, March 12, 1990.
3. T. Levitt. From the Editor: Making Sense. *Harvard Business Review*, Sept.–Oct. 1989.
4. *Special Consumer Survey Report*. The Conference Board, Feb. 1990.
5. D. A. Garvin. Competing on the Eight Dimensions of Quality. *Harvard Business Review*, Nov.–Dec. 1987.
6. G. Taguchi and D. Clausing. Robust Quality. *Harvard Business Review*, Jan.–Feb. 1990, pp. 65–66.
7. Technical Assistance Research Programs Institute. *Consumer Complaint Handling in America: An Update Study; Part II*. United States Office of Consumer Affairs, March 1986.
8. Technical Assistance Research Programs Institute. *Consumer Complaint Handling in America: An Update Study; Part I*. United States Office of Consumer Affairs, Sept. 1985.
9. G. Bargatze. *Leveraging Customer Feedback To Achieve Mea-*

sureable Impact. Technical Assistance Research Programs Institute, 1990.
10. W. H. Davidow and B. Uttal. Service Companies: Focus or Falter. *Harvard Business Review*, July–Aug. 1989, pp. 83–84.
11. R. Brogan, H. Stamm, and J. Hamm. Drive for Excellence or How To Increase Transit Ridership by Turning Your Organization Upside Down and Loving It. Presented at 69th Annual Meeting of the Transportation Research Board, Washington, D.C., 1990.
12. L. M. Rocha. Designing Market Research for Customer Service. *Passenger Transport*, July 2, 1990.

Publication of this paper sponsored by Committee on Management and Productivity.

Evolution of Fare Policy: A Product of Modern Transit Management

ROBERT E. PAASWELL AND DARWIN G. STUART

When a transit agency sees costs going up and ridership going down, can fares save the day? A new fare structure, imposed by the Chicago Transit Authority (CTA), suggests that fares can be used to stimulate both ridership and revenues. Fare policy analysis as an integral part of strategic management planning is outlined. The adoption of a series of basic fare policies by the CTA Board is reviewed. The recent 1990 fare restructuring, as described, directly responds to these policies, with respect to market segmentation, reflection of service quality, and deep discounting concepts. Revenue and ridership effects and associated fare elasticities, associated with four CTA fare changes that were implemented during the 1980s, are also analyzed. Fare elasticities derived from state preference surveys, regarding a range of alternative fare structures, are presented. The broader context for fare policy analysis, as it has evolved at CTA, is also outlined, in terms of sustained demand for basic service, emerging new markets, accommodating escalating costs, and meeting associated revenue needs.

Can a product priced at less than half of its value have a future? Public transit, faced with a sustained escalation of operating costs, stagnating public acceptance, and the potential reduction of federal subsidies must address this question. Public transit agencies, often wedded to management procedures that worked so well in the 1950s and 1960s, must now integrate aspects of modern management into their planning and operation. With present-day competition for the traveler coming from many sources, public transit operators must learn new ways of not only being sensitive to their markets but of satisfying and increasing those markets through well-thought-out, corporate-like strategies.

This paper follows the development of one such strategy—a market-oriented fare policy—developed at the Chicago Transit Authority (CTA). Fare policy itself is seen as a significant component of emerging agency market orientation.

That market orientation and resultant fare policy have evolved from a strategic planning process that has set an overall context for service delivery in terms of an eroding financial climate and CTA's ability to change within that climate.

The strategic planning process, and subsequent management actions taken, including the fare restructuring, can be summarized as follows:

1. *Strategic management planning.* The CTA strategic plan identified the capital and operating constraints that will limit agency actions in the short term and serve as guidelines for agency change in the long term. The plan targets changing ridership patterns and identifies markets from which CTA must draw. Knowing the most likely limits and sources of capital funds for upkeep and growth, and knowing the constraints on non-farebox-generated operating funds, the CTA can better understand the limits of its ability to deliver service. The agency must take full advantage of knowledge of its markets to ensure that it can match the demand for existing and new service with an organizational structure to provide adequate and appropriate supply.
2. *Program-based budgeting.* To be sure that the organization follows a strategic management approach, the annual operating budget must be targeted to reflect agency goals and objectives, ensuring that programs of highest priority get adequate funding. In addition, the budget must identify revenue targets to be met. The budget, rather than being merely a control document, becomes an active planning document reflecting the strategic plan.
3. *Fare policy.* The strategic management plan and its annual documentation—the program-based budget—articulate revenue goals and ridership targets. Fare policy, in this context, becomes a dynamic and aggressive marketing tool.

The next section of this paper addresses the conditions under which these management approaches were applied.

PREPARING FOR THE FUTURE

The CTA is a complex agency, with over 100 years of history. It now serves 1,900,000 trips per day and operates 24 hr every day of the year, with more than 13,000 employees and a combined operating and capital budget approaching 1 billion dollars.

CTA, like other transit companies, has been confronted for the last two decades with rising costs of service and an eroding passenger base.

Quite often, to balance this divergence of increasing costs and declining revenues, transit agencies attempt to increase their income by a simple fare increase applied to and imposed on an existing base of riders.

Today's transit agencies, with newer business tools at their disposal than in the past, can now begin to manage this divergence in more sophisticated ways. Assuming a market orientation, strategic planning can examine various cost-revenue subsidy scenarios and establish ways in which a revenue generation policy—especially fare setting—can be used as a marketing device.

R. E. Paaswell, University Transportation Research Center/City College, Building Y-220, New York, N.Y. 10031. D. G. Stuart, Strategic Planning, Chicago Transit Authority, Merchandise Mart Plaza, Room 704, P.O. Box 3555, Chicago, Ill. 60654.

TRADITIONAL TRANSIT MARKETS

Competition with the private automobile is cited by transit operators as the single greatest challenge to urban transit service. In our larger metropolitan areas a major sustained demand for relatively high service levels of transit will undoubtedly continue, linked closely with those high-congestion, peak-hour travel corridors where high-capacity, exclusive right-of-way transit modes continue to hold clear service advantages. Sustained demand will be linked also to those transit dependents who do not have automobiles available.

However, it is problematical whether these sustained demands will experience marginal increases or decreases over the coming years. For example, many larger urban areas are experiencing a gradual decline in bus ridership levels, particularly for cross-town and nonwork travel. As bus fares have risen, some discretionary travel has been foregone or accommodated in other ways. Regaining these trips then becomes a potential marketing target.

Population and employment change within both larger and smaller urban areas will continue to be a key to future transit demand prospects. The extent to which that change occurs within central cities, either in relation to renewal of older residential areas, at higher densities, or the renewal and growth of downtown areas and environs, or both, will be especially important. The prospects for growth in rail rapid transit ridership are promising for CTA because of Central Area employment growth and a resurgence of inner-city housing.

In light of uncertainty regarding noncentral area transit travel markets, major challenges face the transit industry in terms of targeting specific market segments where transit potentials appear stronger. These market segments may be tied either to specialized transit needs (demand side) or particular types of transit service (supply side). In the Chicago region, for example, market segments on the demand side that show potential for growth include reverse commuting to suburban jobs, and off-peak travel for elderly, disabled, and student populations.

On the supply side, additional service potential appears to exist for premium express bus service, particularly to the Central Area, and for a Central Area distributor system to facilitate the present pattern of growth and to avoid the threat of downtown street gridlock. Privately provided transit services also have established a particular supply side role and market share.

EMERGING MARKETS

The pattern of transit ridership in Chicago is not dissimilar to that in other very large urban areas. Bus ridership is stable, or declining, whereas rail ridership, especially in certain emerging urban "hot" areas is gradually increasing. There is no question that the significant change in the economic and demographic profile of the region has had, and will continue to have, a significant impact on bus and rail ridership. CTA's current ridership profile is given in Table 1.

Table 1 also notes some demographic characteristics of those who do not ride CTA. Because of the current nature of its service, CTA serves city residents primarily. However, those riders are not overly skewed, as in many urban areas, to lower-income levels but are distributed throughout all income categories. A total of 12 percent of those who use CTA have household incomes greater than $50,000, compared with 14 percent of non-CTA users.

TABLE 1 Profile of Current CTA Ridership Market (*1*)

	Riders (%)	Non-Riders (%)
Live in City	81	50
Live in Suburbs	19	50
Have Drivers License	64	86
No Car in Household	27	6
Employed Full Time	52	56
Student	12	4
Occupation: White Collar	71	68
Occupation: Blue Collar	23	28
Median Income ($1,000)	24.2	26.8
- Less Than $20,000	39	31
- More Than $50,000	12	14
Reasons For Using CTA (%)		
* No Car Available	36	
* Can't Drive	11	
* Convenience	34	

When looking at why CTA is selected for travel, car availability and convenience are major reasons for 81 percent of CTA riders. Cost is a principal reason for only 9 percent of riders. But for those to whom cost is important, a greater percentage (10 percent) cite cost when they use the system every day, as compared with those (5 percent) who are infrequent users. Those attributes of service that are important, coupled with frequency of use by rider and purpose of trip, become critical design parameters for the development of a fare structure that can be used to provide market incentives to existing and new users.

In projecting new markets for which CTA might provide service, two major sets of information must be assessed: (a) ridership trends by year and by route or line and (b) significant changes in regional demographics. Two examples developed from such data give evidence of emerging markets that are different in character from the more traditional transit markets noted briefly in the previous section:

1. *Higher-income urban workforce.* An older segment of the elevated rail system serves a series of revitalized higher-income, young professional neighborhoods in the north and near northwest parts of the city. These more affluent riders (many owners or leasers of automobiles) use the Ravenwood elevated rail line to commute to work. Over the last 10 years line ridership has increased 32.5 percent, compared with an overall rail system increase during that period of 3.3 percent. Three stations more than doubled daily riders in that same period. Because these new riders are "users of the city," there is also significant new demand for improved weekend service. In summary, this is a new market of moderate- and higher-income riders, who ride for convenience and even for status, and for whom the cost of the trip is far less important than convenience and quality.

2. *Airport corridor.* CTA extended its northwest rapid transit line to reach O'Hare Airport in 1984. Three additional stations were opened at suburban-like sites. Of these three additional stations, two have major parking capacity, and all three provide kiss-and-ride access. All have excellent feeder or connecting bus service. Since opening, the extension has

added 17,000 daily riders to the system. Of all trips taken on the extension, 21 percent are for reverse commuting, responding to the rapid growth of jobs in the vicinity of the airport. O'Hare station itself is unique in ridership. It remains busy on weekends and at peak holiday periods, including summer months. Many riders are first-time CTA riders. In summary, this extension serves riders for unique purposes: journeys to airport, reverse commutes, and park-and-ride facilities at convenient station parking areas for former automobile travelers.

These examples show that emerging markets must be evaluated so that service attributes can be tailored to needs and services can be priced according to these attributes.

A more general evaluation of emerging markets to be incorporated in strategic system planning includes

- Redistributed journeys to work
  - Reverse commuting
  - Airport vicinity job markets
- Traditional journeys to work, new markets
  - More affluent young urban professionals
  - Suburban commuters at core-oriented park-and-ride locations
- Nonwork journeys
  - Midday shoppers and evening recreation
  - Expanded Loop (central city) travel
- Nontraditional markets
  - O'Hare Airport travelers (rapid rail line now in service)
  - Midway Airport travelers (rapid rail line to have opened in 1992)

The ability to attract and serve these markets is a function of the agency's ability to dedicate resources toward tailoring services to the needs of riders. Because contained costs make it difficult to satisfy even today's markets, methods to reallocate key budget items to high-priority areas must be derived. Strategic planning and program-based budgeting, described briefly later, are used for establishing agency priorities and allocating resources consistent with meeting the needs and expectations of current and emerging markets.

OPERATING BUDGET CONSTRAINTS

CTA, by law, must generate at least 50 percent of its operating budget from system-based revenue. One approach to budgeting would be to look at ridership trends and fares and calculate the maximum allowable operating budget on the basis of anticipated fares. A more strategic approach might be to develop an operating budget that responds to strategic agency priorities and to then establish the level of fare revenue needed to support that budget. The latter was the approach taken that resulted in a new fare structure.

The operating budget of CTA has been under pressure since 1981, when all systems in the region faced a major fiscal crisis. From a period of severe agency cutbacks in 1981–1982 until the present, operating costs have gradually increased so that daily demands for service could be met at least at a satisfactory level.

With public pressure to keep operating cost growth at less than inflation, with regional pressure to divert some operating funds to capital programs, with a strong unwillingness to enact general fare increases and general public pressure to sustain service at existing levels, much ingenuity must be applied to generate a budget that permits operation of a safe system. Of interest is that CTA performed as well or better than its peers during the past decade, using operating cost per passenger (both bus and rail) as an index. But inflationary growth in expenses, even with a constrained budget that defers many necessary agency programs, still outpaces growth in available revenues. New approaches to management, primarily using a longer-range strategic plan as a focus, must be brought to bear to simultaneously address market needs, anticipated budget levels, and the need for growth of system-generated revenue.

APPLICABLE MANAGEMENT TOOLS

The previous section illustrated the financial dilemma facing a large transit property. Operating costs are rising greater than the available revenues. The basic revenue source, the fare, does not necessarily reflect the nature of the service obtained. The system is aging and needs replacement. And new markets are emerging that must either be served well or be lost to competing modes.

The development of a strategic plan creates an agencywide awareness of these problems and challenges the agency to develop effective strategies to confront problems and serve its markets.

In this section the strategic planning effort at CTA is reviewed briefly. An operational partner of the plan, a program-based budget, was also an important supporting management tool.

STRATEGIC MANAGEMENT PLAN

The strategic management plan has become, for most complex agencies, a major document for setting priorities and establishing programs as they evolve.

The strategic management plan examines current conditions of operation, organization, and financial structure and projects alternative futures on the basis of overall system condition, the rate at which system improvements can be made, future funding sources other than fare-box generated revenue, ridership trends, and agency organization. For these various alternatives, revenue (fare-box) needs can be established. The purpose of the strategic planning process was to raise the following question: Should future budgets and revenue needs be addressed by simple, across-the-board periodic fare increases, with a likely negative impact on ridership, or can new fare structures be identified that can actually serve as market tools, increase revenues, and retain or increase ridership? The latter was the direction that CTA followed.

Fare policies have emerged as a major theme of the strategic management plan of the CTA. Within the broader set of financial constraints, cost containment needs, search for new non–fare box revenue sources, capital infrastructure replacement issues, productivity and organizational efficiency needs, and related issues, fare policy—and a stable relationship between passenger-generated revenues and operating costs—has become a vital element.

In 1988 the CTA Board adopted a fare policy, composed of 14 elements that in part recognized the importance of regular, relatively small incremental fare increases. The policies offer a basic foundation for fare revenues to keep pace with inflationary growth in operating costs while minimizing the ridership loss otherwise associated with fare increases.

To further explore this critical nature of fare policy, three scenarios for future population, employment, and transit demand growth in the CTA service area were examined as part of the strategic plan. The financial implications of these scenarios, in terms of associated service levels and operating costs, indicated that under any of the scenarios the ability to have passenger-related revenues keep pace with operating cost growth is the key to financial stability.

The need to explore fare restructuring that is sensitive to time-of-day and quality-of-service pricing (two variables found important to riders in previous CTA surveys) and to address specific market segments was accentuated in the strategic planning process. The long-term fare improvement program also included ongoing efforts to upgrade fare collection equipment technology, which would then provide flexibility in the pricing of prepayment instruments and differential fare structures. This is consistent with the long-term objective of a cash-free system. A new fare system then is responsive to the following:

- Market segments and the demand for service;
- Improved agency operations through integration of modern fare collection technology; and
- Budget requirements for meeting, in a responsible and planned way, an adequate share of operating costs.

A clear product of the strategic management planning process is to have an integrated agency-wide decision on market targets, methods of achieving those targets through various price schedules, and, finally, methods of providing organizational support for those price schedules via appropriate operational changes.

REVENUE AND RIDERSHIP: A CRITICAL MIXTURE

During the 1980s, the typical experience of urban transit operators was the inexorable need for periodic fare increases to match rising operating costs, coupled with resultant ridership decreases that in turn may have led to service cuts or additional fare increases, or both. Breaking this unfortunate and dangerous downward spiral has been one of the major challenges of modern transit management.

This section documents some of these typical revenue/ridership trends by examining the characteristics and impacts of four fare increases implemented by CTA in the 1980s. These increases variously addressed the pricing of cash fares, passes, transfers, and rail-bus differentials, with associated expected and unexpected effects on ridership by mode and fare payment type. This experience, coupled with the results of a 1987 stated preference survey of other fare structure options in Chicago, provides the basis for an overall better understanding of fare structure options as the transit industry enters the 1990s.

CTA FARE CHANGES DURING THE 1980s

Table 2 summarizes the four CTA fare increases that were achieved during the 1980s: in 1981 (twice), 1986, and 1988.

Different pricing strategies were embraced by each of these increases:

1. In January 1981 the cash fare was increased by one-third, whereas the pass price was increased substantially less, providing a major incentive for a shift to pass purchase and use.
2. In July 1981 when it was realized that the earlier fare increase was not generating required revenue gains, an across-the-board increase of 12 to 14 percent for both cash fares and passes was enacted.
3. In February 1986 three fare changes were made:
   - A 10¢ rail surcharge was added to reflect the higher quality of service offered by rail as compared with bus;
   - The transfer price was increased from 10¢ to 25¢, the first such increase in many years; and
   - Transfer regulations were tightened to allow only two rides per transfer, excluding the route of issuance (except for seniors during off-peak hours of service).
4. In January 1988, across-the-board fare parity was pursued, with the bus fare raised 10¢ to equal the rail fare and the pass price increased by 7 percent. This fare change resulted in the smallest (8 percent) average increase of any of four fare revisions.

CTA RIDERSHIP SHIFTS ASSOCIATED WITH FARE CHANGES

Table 3 summarizes the ridership shifts that occurred as a result of each of the four 1980s fare increases. For 1986 and

TABLE 2 Summary of CTA Fare Increases During 1980s

	Price ($)			Price Change (+%)			
	Bus Cash	Rail Cash	Monthly Pass	Bus Cash	Rail Cash	Monthly Pass	Average
1980	.60	.60	30	--	--	--	--
January 1981	.80	.80	35	33	33	16	30
July 1981	.90	.90	40	12	12	14	12
February 1986	.90	1.00	46	--	11	15	18
January 1988	1.00	1.00	50	11	--	7	8

A dash (--) is used to indicate unavailable data.

TABLE 3 Ridership Shifts Associated with CTA Fare Increases

	Annual Unlinked Trips (000)[a]			Ridership Change (%)			
	Bus Cash	Rail Cash	Pass	Bus Cash	Rail Cash	Pass	System Total
1980	341	99	85	--	--	--	--
January 1981[b]	195	59	46	-16	-12	+50	-5
July 1981[b]	110	34	52	- 4	+2	-20	-7
February 1986	304	107	164	-16	-15	+22	-5
January 1988	286	103	183	0	+6	-10	-3

[a]Compares year after against year before fare increase
[b]Ridership data are semi-annual; first 6 months 1981 vs. 1980 and second 6 months 1981 vs. 1980

A dash (--) is used to indicate unavailable data.

1988, the annual ridership before the fare increase was compared with that in the year after the fare increase. For the two closely spaced 1981 fare increases, the 6 months before the 6 months after each fare change were examined. In general, both expected and somewhat unexpected ridership shifts occurred. Analysis of these ridership shifts shows the following:

1. After January 1981 a 50 percent increase in pass sales and use was, as expected, accompanied by a significant decline in bus and rail cash fares, with bus affected more seriously than rail.
2. The continuing shifts experienced just 6 months later with the July 1981 fare increase may have, in fact, also included some continuing stabilization in response to the initial January fare increase, as well as direct response to the July 1981 fare increase itself.
3. After July 1981 the slight increase in rail cash fare ridership may have reflected a recognition that, because of its higher service level, rail was "worth" more. The unexpected drop in pass use may have reflected the increased relative value of the transfer, whose price continued at only 10¢, with some pass users moving back to using transfers. This drop of 20 percent was found to be only temporary in subsequent months, as pass users gradually returned.
4. In January 1986 pass use was back to its June 1981 level. An additional 22 percent increase in pass use occurred, attributed to the combined effects of increased transfer price (10¢ to 25¢) and tightened transfer regulations limiting their utility.
5. After January 1988 an overall ridership loss of -2.1 percent was projected, although an actual loss of -2.5 percent occurred. Effects were particularly high in reduced fare categories, which experienced a 25 percent (10¢ out of a prior 40¢) fare increase on bus, leading to ridership losses of approximately 12 percent, which was not unexpected.

DERIVATION OF FARE ELASTICITIES

The average fare and annual/semiannual ridership changes summarized in Tables 2 and 3 can be used to derive modal fare elasticities, as summarized in Table 4. These are calculated using the "shrinkage factor" elasticity formula:

$$\text{fare elasticity} = \frac{(R_2 - R_1)/R_1}{(P_2 - P_1)/P_1} \tag{1}$$

where R_1 and R_2 represent before and after ridership levels, and P_1 and P_2 represent before and after prices, respectively.

Table 4 shows considerable variation in fare elasticities across the four fare revisions of the 1980s, with average bus, rail, and system elasticities that are consistent with the literature: a rail fare elasticity (-0.14) about a third that of bus (-0.38), yielding a systemwide elasticity of about -0.33.

The following are year-by-year highlights:

1. For January through June 1981, with the highest of any of the fare increases examined, resulting elasticities were surprisingly low. The key here is the increased discount offered for passes, which encouraged and achieved a major switch from cash to pass, allowing a substantial number of riders to switch fare payment method rather than eliminate or avoid transit trips altogether. Additional discretionary or induced trips also may have been made by new pass purchasers. Note particularly that the net effect on rail was to yield any elasticity close to zero.
2. Although the July 1981 fare revision yielded the next-to-lowest average fare increase, it correspondingly had the most sensitive resulting systemwide fare elasticity, with a particularly high impact on buses. This elasticity could have been caused by continued reaction to the January 1981 fare increase and also may have been clouded by other factors, such as the economic recession then under way.
3. The average 1981 fare elasticities (across both fare increases) are more representative of the overall elasticity experience during the 1980s, suggesting that a 6-month time interval for gauging full ridership impact may be too short.
4. The elasticities derived for the February 1986 fare increase were complicated by the related effects of tightened transfer regulations, which had an unclear financial component. This worked against a clear distinction between bus and

TABLE 4 Estimated CTA Fare Elasticities During 1980s

	Average Fare Increase (+%)	Fare Elasticity Bus	Rail	System
January 1981	30	-.20	- .03	-.17
July 1981	12	-.66	- .16	-.59
February 1986	18	-.33	- .28	-.27
January 1988	8	-.31	+.22	-.29
AVERAGE		-.38	-.14[a]	-.29

[a]January 1988 not included

rail elasticities, which are relatively close, although the systemwide elasticity is reasonable compared with the experience elsewhere. The lack of clear differentiation between bus and rail elasticities is likely caused by both increased transfer pricing and tightened regulation. The 15 percent increase in pass pricing also offered a neutral alternative compared with higher cash fares.

5. The differences in bus and rail elasticities in 1988 must be interpreted carefully. The net increase in rail ridership was largely attributed to the return to rail from bus of former pre-1986 rail passengers who had reluctantly switched to bus to avoid the 1986 10¢ rail surcharge. A cross elasticity between bus and rail was therefore in effect, as it undoubtedly was in 1986.

RESPONSE OF CTA RIDERS TO ALTERNATIVE FARE STRUCTURES

On the basis of 1987 stated preference surveys of CTA travelers, market segment analyses revealed a relative preference for peak versus off-peak fare structures, as well as for differential pricing for radial central business districts versus local neighborhood trips. Limited sensitivity to transit pricing based on the length of the trip was also found. A series of market-segmented elasticities for peak and off-peak travel were derived, after adjusting stated preference survey results to reinterpret individual mode choice preferences (automobile versus transit), as aggregate fare elasticities.

Table 5 summarizes these fare elasticities. The average elasticity of −0.33 for these different peak/off-peak pricing variations is consistent with the average bus-rail system elasticity of −0.33 shown in Table 4, so that together the two tables offer a broad range of fare elasticities for basic fare structure options. A review of this analysis shows that

- Overall, off-peak elasticities are more than twice peak-hour elasticities. Off-peak elasticities tend to hold at this higher level, between −0.36 and −0.49 for other market segmentations, including travel to the Chicago Central Area versus local neighborhood travel.
- Radial (to and from the Central Area) travel markets have a lower fare elasticity than local markets, particularly during the peak hour, where relatively low elasticities of −0.11 to −0.13 were observed. This most likely reflects the greater difficulty and congestion associated with peak-hour radial travel by automobile. The primary work purpose associated with such travel may also be a factor, with an assumed greater ability to absorb fare increases supported by employment income.
- Journeys of less than 2 mi also displayed peak versus off-peak fare increase sensitivities, as well as a higher level of sensitivity to fare changes than journeys greater than 2 mi. However, partly because of a probable lack of experience with distance-based fare structures, survey respondents did not display any further sensitivity to fare increases scaled on a per-mile basis for longer trips.
- Shorter trips within the Chicago Central Area also displayed off-peak elasticities higher than peak-hour elasticities. In particular, peak-hour elasticities for these shorter trips are significantly higher (double or more) than those for longer radial trips.

TABLE 5 Fare Elasticities Derived from 1987 Stated Preference Survey Data

	Peak	Off-Peak
Radial, to Central Area	-.11 to -.13	-.36 to -.39
Local Neighborhood	-.19 to -.24	-.41 to -.44
< 2 Miles	-.29	-.49
Within Central Area	-.26	-.39
OVERALL	-.19	-.44
AVERAGE ALL DAY	-.33	

NOTE: The ranges quoted above for Radial and Local market segments correspond to a further disaggregation into inner and outer zones

EVOLUTION OF A FARE POLICY

It was noted earlier that during the last decade CTA transit fares have been adjusted primarily in response to balancing the growing demands of an inflating operating budget with subsidies available from public funding and secondarily with recognition of quality differences between bus and rail service. Between 1986 and 1988 a rail surcharge was implemented to reflect the fact that there are differences between bus and rail service to which riders will respond. Lower rail fare elasticities were a practical basis for this fare differential but were only incompletely understood. As noted previously, consultant studies were undertaken, using stated preference surveys, to further examine the market sensitivity of a full range of alternate fare structures.

As the organization became more aware of market needs, a range of consumer issues associated with fares were addressed:

- Elderly and handicapped fare level mandates;
- Low-income riders, affordable fares—inability to purchase monthly passes;
- High-income riders—willingness to pay more for increased quality of service;
- Sustained declines in ridership in what are perceived to be unsafe areas; and
- Ridership growth in growing middle- and upper-income neighborhoods.

Recognizing that there must be a balance between meeting base service budget needs and maintaining or even improving ridership, the CTA Board adopted a series of "CTA fare policies." These 14 points (see Table 6) deal with issues of equity and quality as well as operational issues related to the collection and handling of money. Such problems were targeted to be the basis for further discussion as future fare structures are implemented. It is believed that CTA is the first transit property to adopt such policies independent of a system's particular fare increase. With an emphasis on prepaid fare instruments, improved technology, and value for service, such policies give strong direction to evolving fare structures that meet market needs and are consistent with an emerging strategic management plan.

TABLE 6 CTA Fare Policies

Policy Number	Policy
1	While pursuing all avenues of cost containment and innovation in service provision, CTA must regularly adjust its pricing and fare structure to reflect changes in the overall costs of service. Small affordable changes and structural readjustment are preferable over large one-step increases.
2	CTA fares should be structured to reflect the quality of service delivered.
3	CTA fares should be structured to reflect the relative costs of service delivered.
4	The CTA fare structure should afford maximum convenience to its customers, in terms of reasonableness, understandability, and acceptability.
5	Changes in CTA's fare level and structure should be designed to increase passenger revenues, attract new customers wherever possible, and minimize any associated losses in ridership, with an equitable distribution of financial impacts on existing and future riders.
6	The fares for elderly and handicapped customers during off-peak hours will not exceed half the fares charged other passengers during peak hours.
7	Special fares may be established for some customers or some types of service, to increase or facilitate ridership.
8	The availability and diversity of prepaid fare instruments should be increased, matching ridership markets and market segments.
9	Changes in CTA's fare structure should provide for an orderly, timely, and cost-effective implementation.
10	Exact fare should be required for boarding both bus and rail services.
11	CTA's fare structure should minimize opportunities for fare cheating, in order to protect its revenue base and the interests of all fare-paying passengers.
12	Cash-handling and change-making by operating personnel and ticket agents should be minimized.
13	Where cost/benefit ratios and improvements in passenger convenience are favorable, increased use of automated fare collection equipment and ticket/token-vending and coin-changing equipment should be made.
14	Regional fare coordination between CTA, Metra, and Pace should be achieved.

SETTING PRICE OF TRANSIT

Among the fare restructuring options considered in Chicago was the "deep discount" concept. This type of option recently had been implemented in different forms in Milwaukee and Denver and several smaller urban areas. Under this concept, one form of convenient prepayment of a multiple-ride fare instrument is discounted from the base cash fare, on the order of 20 to 25 percent. This is balanced against a correspondingly higher increase in the base cash fares usually so that the overall need for higher passenger revenues, at whatever level is indicated, can be achieved.

Increased revenues associated with the increased cash fare would be offset somewhat by a modest ridership loss among cash fare riders, reflecting fare elasticities associated with a price increase. However, some (perhaps most) of those price-sensitive riders could switch to the discounted prepayment instrument, softening that ridership loss. Experience has shown that the attractiveness of this prepayment instrument can be sufficient, in fact, to induce additional discretionary transit travel, to the extent that this ridership growth may itself offset the ridership loss associated with the cash fare price increase. It appears entirely possible to therefore achieve a gain in both fare-box revenue and passenger ridership through this deep discount pricing strategy.

TABLE 7 1990 CTA Fare Structure

	1989	1990
Full Fare ($)		
bus-peak	1.00	1.25
bus-off-peak	1.00	1.00
rail-peak	1.00	1.25
all-modes		
weekday (5-day)	none	45.00
monthly pass		
everyday monthly pass	50.00	60.00
tokens	.95	.90
transfer	.25	.25
Reduced Fare ($)		
bus-peak	.50	.45
bus-off-peak	.50	.40
rail-peak	.50	.45
rail-off-peak	.50	.45
everyday monthly pass	25.00	25.00
tokens	.50	.40
transfer	.15	.15

The fare structure selected by CTA was a clear product of the management principles discussed above. The characteristics of the adopted structure are given in Table 7. The base price of travel is set at $1.25. The differential in quality between bus and rail is shown in the off-peak price of $1.00 for bus, whereas rail remains at $1.25. Equity is addressed for daily riders, who are given a variety of discounts from token prices to weekday passes. This is the most ambitious fare change undertaken by CTA—but one well founded in good management principles.

The strategic management planning process showed clearly the nature of fiscal constraints that will affect the operating budget for years to come. Since half of the operating budget comes from the fare box, a judicious approach to fare policy is warranted. This means that ridership, and the markets that riders represent, must be understood and that service must be tailored as much as possible to meeting market targets. Market surveys also have shown important rider sensitivity to various fare structure options. The end result was to select and institute a fare policy that would help sustain or improve ridership while meeting revenue needs.

REFERENCE

1. *Household Survey of the CTA Service Area.* Market Opinion Research and Midwest System Sciences, Inc. Aug. 1988.

Publication of this paper sponsored by Committee on Management and Productivity.

PART 5

Education and Training

Enhancing the Future Pool of Civil Engineers

JOHN M. MASON, JR., JOSEPH P. TARRIS, EMAN ZAKI, AND MICHAEL S. BRONZINI

Recent demographic trends and employment facts indicate that the traditional major source for the supply of civil engineers will decline. In the civil engineering profession, ethnic minorities and women are significantly underrepresented. Additionally, although overall college enrollments held steady during the 1980s, the number of students graduating from high school has declined from its peak in the mid-1970s. To confound the situation, it is a national expectation that the fastest-growing sectors of the work force will be the areas requiring the highest skill levels. It is anticipated that the civil engineering profession will find itself competing for high-quality, competent individuals who will simultaneously be considering the appeal of other careers. To meet the changing needs of the civil engineering work force, an NCHRP study was undertaken in which a conceptual model (herein referred to as "ARC") was developed. The ARC model contains three comprehensive and interrelated strategies: (a) to heighten the awareness of the civil engineering profession, (b) to improve the retention rate of the existing pool of potential civil engineering candidates, and (c) to enhance the curriculum of pre-college and college programs. The findings of the initial research to identify existing programs and practices to promote the civil engineering profession and the results of 17 diverse focus group sessions are also presented.

The concept of supply and demand applies to human resources as well as to traditional consumer goods. ASCE reports that the supply of civil engineers graduating with bachelor's degrees each year decreased from 10,547 in 1981 to 7,688 in 1989 (*1*). At the same time, the demand for civil engineering professionals is seen to be on the rise. Several of the predicted growth rates for engineering and civil engineering are as follows:

- The Hudson Institute predicts a rate of growth of 41 percent between the years 1985 and 2000 for engineering, architecture, and surveying (*2*).
- TRB estimates that future demand for civil engineers will increase at an annual rate of 5 percent, with 60 percent of the increase exclusively a result of growth and the remaining 40 percent a result of death and retirement (*3*).
- The Bureau of Labor Statistics (BLS) projects that the employment of civil engineers will increase 17.9 percent between 1988 and the year 2000 (*4*).

If the supply of civil engineers continues to decrease, the required demand will need to be met in an even more highly competitive market.

Some argue that the shortage issue is really one of providing adequate and commensurate compensation. Success to overcome the financial arguments requires long-term, continuous efforts. The broad-based employment of civil engineers (private versus public sector), however, does not lend itself to very timely or convenient solutions. Nonetheless, near-term strategies for recruitment and retention must consider the emerging demographic changes in the future work force, which will severely curtail the size of the traditional pool of entering students. Additionally, students currently enrolling in civil engineering score lower on SAT/ACT examinations than students pursuing other engineering fields, which introduces a student quality dimension into the shortage situation (*5*).

Immigration currently adds approximately 750 to 800 civil engineers to the pool annually (*3*). This population influx is relatively stable but is always subject to modifications in U.S. immigration policy. The largest supply to this pool, approximately 90 percent, is from the university system. Currently, regarding the outflow, approximately 20 percent of graduate civil engineers elect to pursue a career path other than civil engineering (*5*).

The makeup of the U.S. work force is changing with respect to an aging population, low birth rate, population shifts, and increased educational and per-capita income levels (*6*). The ramification of all these changes is that the size of the traditional college-age population (18- to 24-year-olds) will decrease approximately 25 percent from current levels by the year 1996. This drop is predicted to reduce college enrollments by 12 to 16 percent (*7*). Additionally, approximately 75 percent of the undergraduate degrees conferred annually are obtained by white males. However, data indicate that this group, although currently the clear majority, may be a minority in the 21st century (*8*).

Table 1 presents a comparison of the number of civil engineering degrees conferred for 1980 and 1989 in total for women and for ethnic minority groups (*1*). The percentage of women conferred civil engineering degrees has increased during this period from 9 percent of all civil engineering degrees to 14 percent. The percentage of African-Americans has remained constant, and those for Hispanic-Americans and Asian-Americans have shown modest increases. However, women and minorities remain underrepresented relative to their numbers in the population.

Although the need for civil engineers will continue in the future, the reduced levels of college-age individuals in the

J. M. Mason, Jr., J. P. Tarris, and E. Zaki, Pennsylvania Transportation Institute, Pennsylvania State University, 201 Research Office Building, University Park, Pa. 16802. M. S. Bronzini, Center for Transportation Analysis, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tenn. 37831.

TABLE 1 Summary of Civil Engineering Degrees Conferred—1980 and 1989 (*1*)

Group	1980	1989	Percent Change
All	10,346	7,688	-25.7
Women	931	1,036	11.3
African-Americans	158	170	7.6
Hispanic-Americans	257	274	6.6
Asian-Americans	290	353	21.7
Native-Americans	9	28	211.0
Foreign	1,309	864	-34.0

population combined with the changes in the demographic makeup of this group will make it difficult to maintain an adequate number of quality civil engineers to satisfy reported demands. Efforts to encourage more students, particularly women and minorities, to pursue civil engineering as a profession may be categorized as either supply side or demand side. The supply side involves increasing the number of students interested in civil engineering and prepared to enter the profession. The demand side involves the improvement of the "product," in which the product includes both enhancing engineering employment conditions and a responsive engineering curricula.

This paper is based on the findings of the initial efforts of NCHRP Project 20-24(3), entitled "Expanding the Civil Engineering Pool."

NCHRP STUDY

Project Objective

The overall scope of the NCHRP study includes the identification, development, and testing of specific actions that will increase the quality and diversity of civil engineers available to the transportation profession. Particular attention and emphasis have been placed on the implications of changing demographics on the future work force.

Research Approach

The research approach comprises three phases; a brief description of the primary research activities follows.

Survey of Practices and Programs To Increase Pool of Civil Engineers

Many professional associations and resource organizations were contacted for relevant information and sample documents and programs. Letters of request for samples and descriptions of pertinent materials were sent to 224 civil engineering departments and 114 civil engineering technology departments on record with the ASCE and 50 state highway agencies/departments of transportation (DOTs) in the United States. The final survey response is shown in Table 2. The findings are summarized in the following sections of this paper; specific details are documented in the NCHRP project reports.

Market Research Study

The principal means of field data collection in this study was a market research study designed to determine how people make career choice decisions and what underlying attitudes and perceptions they have about civil engineering and other professions. A method of qualitative research known as the "depth group" or "focus group" was used. The study was conducted by the Brand Consulting Group, Southfield, Michigan, coordinated by the Pennsylvania State University study team and with the assistance of the study consultants located at the field data collection sites.

Seventeen focus group sessions were conducted in five locations (Table 3). Selected groups consisted entirely of African-American or Hispanic-American subjects, whereas the balance were ethnically mixed (although predominantly Caucasian). All groups contained a mix of male and female participants. All sessions were viewed by at least one study team principal, and all were recorded on both audio- and videotape.

Documentation of Phase I and Phase II Efforts

A comprehensive research report was prepared that contained the detailed findings of the Phase I and Phase II efforts. The report fully documents the results and products of the overall project. The research report is divided into two parts: Part 1 summarizes the overall research activities, and Part 2 presents a synthesis of current practices to increase specific labor supplies and contains seven detailed appendixes related to state

TABLE 2 Summary of Survey Response

	Number of Departments/Agencies Contacted	Number of Departments/Agencies Responding
Civil Engineering Departments	224	83
Civil Engineering Technology Departments	114	30
State Transportation/ Highway Agencies	50	44
Resource Organizations	9	5
Professional Associations	16	10

TABLE 3 Focus Group Design

	State College, PA	Pittsburgh, PA	Lafayette, IN	Austin, TX	Los Angeles, CA
Junior High School Students			■	■	
High School Students		■*		■	■
Parents			■	■**	■*
Junior High School/High School Teachers	■				
Junior High School/High School Counselors					■
College Students	■		■*	■	■**
College Faculty			■		■
Civil Engineers		■			

*AA--African-American

**HA--Hispanic-American

agency, academic institution, and professional practices to increase interest in civil engineering and closely aligned technical areas. A full copy of the market research study and recommended action plans are also provided in a set of appendixes.

A second executive summary style report cites the activities of the overall project, concisely presents the results of the Phase I and Phase II efforts, discusses the strategies for enhancing the civil engineering pipeline, and presents a list of recommended action plans to be considered for further development and eventual implementation.

Phase III Activities

Phase III of this research is under way as of this writing. The goal of the overall project is to produce a user's manual of selected techniques to be used by transportation agencies, educational institutions, national organizations, and others interested in increasing the pool of civil engineers. Phase III includes the development, testing, and refinement of selected action plans identified at the conclusion of the Phase II efforts. A specific set of recommended, highly selective strategies has been assembled on the basis of the recorded findings of Phase I, the results of the Phase II evaluation procedures described above, and the articulation of the concerns, issues, and suggestions of the project team, consultants, and the NCHRP panel.

CHANGING DEMOGRAPHICS

General Trends

The predicted changes in the composition of the work force into the next century are documented in several research studies and papers. In 1987 the Hudson Institute published *Workforce 2000* (*2*). This study examined the changes occurring in the economy and the work force and projected the jobs that this work force will perform between the years 1985 and 2000. Five demographic trends were reported by the study:

1. The population and work force will grow more slowly than at any time since the 1930s.
2. The average age of the population and work force will rise, and the pool of young workers entering the labor market will shrink.
3. More women will enter the work force. The report predicts that approximately two-thirds of the new entrants into the labor force between the years 1985 and 2000 will be women.
4. Minorities will represent a larger share of new entrants into the labor force. The report predicts that nonwhites will make up 29 percent of the new entrants into the labor force between the years 1985 and 2000. This percentage is twice the current minority share of the work force.
5. Immigration will account for the largest share of the increase in the population and the work force since World War I. The report predicts that the majority of the new workers entering the labor pool from immigration will be located in the South and West.

The combined effect of these demographic trends is an expected shift in the makeup of the work force between the years 1985 and 2000. Women, nonwhites, and immigrants are expected to contribute 85 percent of the net additions to the work force. Currently, these groups constitute approximately one-half of the work force.

The Office of Technology Assessment has reported on two additional demographic trends and their impact on the scientific and engineering work force (*7*). These demographic trends are (a) the decline in the number of 18- to 24-year-olds (the traditional college-age population) from a peak of

30 million in 1982 to approximately 24 million in 1995, after which this number is expected to rise; and (b) the increased representation of ethnic minorities in the 18- to 24-year-old group, from 20 percent to 27 percent in 1998.

The Hudson Institute study predicts not only a change in the supply of new entrants to the work force but also a change in the demand for skills. Specifically, the new jobs to be created will demand far greater skill levels. More than one-half of new jobs will require education beyond high school, and approximately one-third of these new jobs will be filled by college graduates. Furthermore, the Hudson Institute reports that if occupations are rated and categorized according to skill level (comprising mathematics, language, and reasoning skills), 27 percent of the current (1985) jobs are within the three highest categories. However, the study predicts that 41 percent of the new jobs created by the year 2000 will require a skill level within these three highest skill categories. [Engineering had a skill rating of 5.1. Only the natural sciences (skill rating, 5.7) and law (skill rating, 5.2) were rated higher.]

Effects of Changing Demographics on Engineering and Civil Engineering

Historically, white males have been the principal recipient of the engineering degrees awarded, representing 90 percent of all practicing engineers (*9*). Currently, excluding degrees awarded to foreign nationals, white males obtain 70 to 75 percent of the degrees awarded annually (*10,11*). Women, African-Americans, and Hispanic-Americans are underrepresented groups; that is, their representation in the engineering profession is less than their representation in the general population. Women constitute 47 percent of the population; however, the proportion of undergraduate degrees in engineering awarded to women grew from 1 to only 15 percent during the years between 1970 and 1985 (*10*). This percentage has remained relatively constant over the past several years and is well below the representation of women in the working population. Women earned 15.3 percent of the B.S. engineering degrees awarded in 1989 (*12*).

Over the past 10 to 15 years, third-year (junior-year) civil engineering enrollments have been characterized by the following:

- Enrollments peaking between 1977 and 1981;
- A long downward trend until 1989, when enrollments were 27 percent lower than their peak years;
- White males composing about 70 percent of civil engineering enrollments;
- Female enrollments peaking in the beginning of the 1980s, similar to the general trend, with women currently representing 16.6 percent of total enrollments;
- Ethnic groups representing a very small percentage of undergraduate enrollments;
- Hispanic-Americans representing the highest enrollment among ethnic groups, with increasing representation;
- African-American enrollments peaking in the beginning of the 1980s and then starting to decline;
- Increasing Asian-American enrollment; and
- Already very small and still decreasing Native-American enrollment.

Regarding retention characteristics, the following can be stated:

- In general, retention rates are promising at the junior undergraduate level.
- 94 percent of junior-year civil engineering students graduate.
- 85 percent of female junior-year civil engineering students graduate.
- Retention rates among the ethnic minority civil engineering students are lower and stand at about 65 percent.

The general belief that foreign nationals are accounting for an ever-increasing share of civil engineering students is not completely true. The research found that

- Foreign national representation is very high at graduate levels only. At the undergraduate level, representation has dropped to below 10 percent.
- In 1988 foreign nationals constituted 43 percent of all master's-level enrollments in civil engineering and 65.7 percent of all doctorate enrollments.
- Retention rates among foreign nationals are high.

Studies by the Hudson Institute, the U.S. Office of Technology Assessment, and the Western Interstate Commission for Higher Education indicate the following by the end of the century:

- The number of high school graduates will continue to decline until about 1995 and then begin to rise.
- The western and south/south-central regions of the country will experience an increase in the number of high school graduates, whereas the northeastern and north-central regions will experience a decrease.
- Women and ethnic minority groups, who have a history of underrepresentation in the engineering profession, are expected to represent 62 percent of the net additions to the work force.
- White males, who have traditionally filled the civil engineering ranks, will represent only 15 percent of the net new workers entering the work force.
- The fastest growing sectors of the work force are those sectors requiring the highest skill level. These sectors include lawyers, natural scientists, and engineers.

A decrease in the number of civil engineering graduates is likely based on the preceding discussion. However, shortages are difficult to predict. Alexander (*13*) has examined previous impending shortages in civil engineering during the 1950s, 1960s, and 1970s. He determined that in each of the periods following the shortage prediction, the average salary for civil engineers decreased relative to the average worker, demonstrating an oversupply, not a shortage. For 1990 the College Placement Council's September *Salary Survey* reports that the average starting salary for civil engineering graduates, both males and females, was the lowest of all engineering specialties (*14*, p. 1). Another study analyzed the BLS predictions made for several engineering occupations in 1960 and 1965 for the years 1975 and 1980 (*15*). The analysis determined that the BLS overestimated, by 20 to 55 percent, the requirements for aeronautical, civil, and mechanical engineers.

In summary, current knowledge reveals that

- The fastest growing sectors of the work force are those sectors requiring the highest skill levels. These sectors include lawyers, natural scientists, and engineers.
- White males, who have traditionally filled the engineering and civil engineering ranks, will represent only 15 percent of the net new workers entering the work force through the year 2000.
- Women and ethnic minority groups, who have continued to be underrepresented in the engineering profession, are expected to represent 62 percent of the net additions to the work force between 1985 and 2000.
- Long-term engineering work force predictions are debatable.
- Occupational mobility has served to balance short-term labor shortages.

SUMMARY OF DATA GATHERING—PHASE I

The first phase of the project concentrated on collecting and documenting information on current practices, perceptions, and attitudes in the overall career choice decision-making process. Significant efforts were placed on identifying, synthesizing, and evaluating related literature, products, and programs related to civil engineering, engineering in general, and the mathematics and science disciplines.

Three specific activities were conducted in the Phase I research efforts:

1. Documentation of practices used by undergraduate civil engineering/engineering technology programs and state transportation agencies to increase interest in civil engineering and civil engineering–related careers.
2. Documentation of practices used by other professional disciplines to address their labor shortages.
3. Identification of the attitudes and expectations of various constituencies affecting civil engineering career choice decisions (via intensive "focus" group interviews).

A complete presentation of the documentation of selective and related practices and programs is beyond the scope of this paper. Particular attention and emphasis of the research were placed on the implications of the changing demographics in the U.S. work force throughout the data-gathering phase. Each inquiry attempted to identify specific considerations being given to underrepresented groups. A common assessment among the many sources was a fundamental need to heighten the awareness of the public to the attributes of the civil engineering profession. Efforts to better inform prospective students and their parents, teachers, and counselors are seen as having the broader benefit of educating the public regarding the role of the civil engineer in society.

Practices To Increase Interest in Civil Engineering

Federal and State Transportation Agencies

Several initiatives by federal and state transportation agencies were identified to increase interest in civil engineering with an emphasis on the participation of minorities and women. On-the-job training, minority engineering scholarships, and summer intern programs are representative of transportation agency programs. Several state transportation agencies reported the use of "adopt-a-school" programs to encourage employees to volunteer their services for mentoring, speaking, guidance, and judging at science fair activities. The Transportation and Civil Engineering (TRAC) careers center concept would strengthen the existing initiatives used by the various states (*16*). The TRAC program would provide quality information that could be readily implemented through the programs already identified. Such a national outreach program would permit the state agencies to use their time disseminating consistent, reliable, and effective materials rather than using their resources to produce new material.

The survey of civil engineering/engineering technology departments revealed that few schools are very active in recruiting students in civil engineering. The responses indicated that the college of engineering or the entire educational institution, or both, generally provides precollege career information and undergraduate information.

Professional and Industrial Organizations

ASCE In 1990 ASCE had a marketing plan prepared for career guidance. The marketing plan provides information about the development and implementation of an educational career guidance campaign. The plan includes specific strategies for attracting young people to civil engineering, with an emphasis on women and minority students. The recommended action plan for ASCE includes

- Development of programs at the branch and section levels of ASCE to promote science and mathematics at all grade levels;
- Promotion of interest in civil engineering by involving the ASCE membership in in-class presentations and demonstrations;
- Encouragement and education of high school science and mathematics teachers through involvement in science fairs and mathematics competitions and with information about the benefits and rewards of a civil engineering career;
- Provision of tours of construction sites and related civil engineering facilities;
- Continuation and expansion of advertising initiatives at the national level to increase the awareness of civil engineering;
- Coordination of ASCE activities with other efforts to promote and expand science and mathematics awareness;
- Increased participation and encouragement of underrepresented groups to pursue the civil engineering career; and
- Encouragement of undergraduate degrees holders to pursue graduate study and consider faculty careers.

ITE ITE also has been active in recruiting and retaining civil engineering students in the traffic and transportation engineering discipline. Its activities include fellowship programs, support of student chapter activities, committee activities, investigation of professional career opportunities in transportation engineering, and preparation of informational materials in the form of printed material and videotape.

Career Choice Decisions

Another task in the data-gathering process was the conduct of focus group sessions to determine

- Attitudes and expectations of students, teachers, and guidance counselors;
- Degree of awareness of civil engineering as a discipline and career alternative; and
- Perception of what a civil engineer does.

A summary of the key findings of a selective literature review and the major results of the focus group sessions follow. The final NCHRP reports contain a complete copy of the Brand Consulting Group findings on the focus group sessions.

The focus group activities proved to be a highly effective means of data collection, in that they produced observations that are very difficult to capture through questionnaires or other quantitative methods. The information provided in this report is based on the results of 17 intensive group sessions.

Attitudes Toward Math and Science

- Teaches most directly affect subject preferences.
- Teachers' enthusiasm and qualifications motivate the student.
- Math classes, at all levels, pay very little attention to practical applications.
- Teachers would use materials featuring technical applications if they were prepackaged and integrated into the existing curriculum.
- Parents want their children to do well in all classes. Many parents do not feel qualified to help their children in math and science.
- Secondary school guidance counselors do not appear to greatly influence student attitudes toward math and science.

Attitudes Toward Career Planning

- Elementary school students are stimulated more by outside influences than by what specifically happens at school.
- Students and parents both indicate that junior high school is too early for specific career planning but definitely appropriate for career awareness.
- In high school there is increased recognition that career options must be seriously considered. Students and parents seek specific career information; students, however, respond with mixed enthusiasm regarding programs specifically oriented toward career awareness.
- Summer programs and other types of enrichment experiences (exposing target groups to technical careers), however, are received enthusiastically and are considered highly successful.
- There is little or no exposure to career-related material in the classroom.
- Teachers are interested in using career-related materials, but these materials can potentially overcrowd a curriculum, so using them effectively requires great motivation and dedication on the part of the teacher.
- Students who do well in math and science are attracted or directed toward careers in engineering.

Attitudes Toward Careers

- Students (and virtually everyone else) have very little information about specific careers. This is especially true of engineering and particularly true of civil engineering. The term "civil engineering" means almost nothing to those who are uninformed.
- Individuals relate better to career terms such as transportation engineering, structural engineering, construction, and environmental engineering.
- Individuals indicated that the civil engineering field would not be challenging because most techniques for building roads and bridges are well known. (Similar misconceptions were noted for other civil engineering specialty areas.)
- Professional licensing tended to confer distinction on the civil engineering field as a career choice.

Attitudes of Women and Minorities

- Women and minority reactions to the civil engineering profession were not markedly different from those of the other groups already identified.
- Minority students reported favorable reactions to math/science/engineering enrichment programs.
- College-bound minority students recognize the need for financial aid and are highly motivated to seek and qualify for such aid. The remuneration potential among engineering careers was of relatively greater concern to minority students, although this seems to be related more to economic status than to ethnicity.
- Parents appeared somewhat troubled by recruiting efforts directed toward their children; several cautioned against painting a false picture of the relative desirability of engineering or civil engineering careers.
- Students as role models were mentioned often, and with greater emphasis by the minority group members.
- College students were mentioned as good role models for communicating with secondary school students.
- Women and minority engineering students are concerned about the continuing "chilly climate" for women and minorities in the workplace.
- Many women and minority group members felt that they would have to start their own firms to get around a "glass ceiling" that blocks their advancement in larger organizations.
- Civil engineering was noted as a field in which entrepreneurship is highly possible, but few students are aware of this.

RECOMMENDED PLAN OF IMPLEMENTABLE ACTIONS—PHASE II

Phase I resulted in the identification of several fundamental obstacles facing the civil engineering profession in general and the future pipeline of entrants:

1. An image problem exists for the civil engineering profession.
2. Institutional barriers exist that contribute to the increasing attrition rates among high school and college students.
3. Changes are necessary in precollege and college curricula if enhancements are to be made in mitigating the previous two items.

ARC Model

To address these obstacles effectively, the research effort developed a conceptual framework consisting of three interrelated long-range market strategies:

1. Heighten the *awareness* of technology, engineering, and civil engineering.
2. Increase the *retention* of the existing pool of future civil engineers.
3. Modify the existing *curriculum* from kindergarten through college.

These key strategies serve as a focus such that future efforts will recognize the developmental stages of the pool of potential engineers. Table 4 gives the primary components of the recommended conceptual framework, which is referred to as the "ARC" (awareness, retention, curriculum) model. The target audiences for the candidate action plans include students and their adult influencers (teachers, counselors, parents, and practicing professionals).

The ARC model has the following characteristics:

- It is a continuum model that builds on previous experiences.
- The action plans, although not independent of each other, are implemented differently at each developmental stage.
- To achieve a long-term effectiveness, the model as a whole must be implemented.
- The action plans for the three strategies are broad at the early developmental stages and narrow to more specific actions as higher developmental stages are reached.
- Overlap and crossover benefits occur among the three strategies.

Recommended Action Plans

The ARC model argues that to successfully enhance the civil engineering pipeline, future marketing strategies must consider awareness, retention, and curriculum issues in the development of individual action plans. Specific objectives must likewise be defined at each developmental stage of the model for the pipeline constituents.

As one proceeds through the developmental stages, a selected objective becomes more narrowly defined (from global engineering to civil engineering to the specialties of civil engineering). Similarly, as one proceeds from the awareness strategy to the retention strategy to the curriculum strategy, the target audience also narrows.

Awareness strategies target both the potential pool and the influencers (parents, teachers, counselors, and practicing professionals). The retention strategies tend to be more personalized. Curriculum strategies intimately affect the students (and teachers).

Fifteen candidate action plans were developed to essentially address specific issues at each defined developmental stage. The goal of each recommended action plan is cited in Table 5. If the ARC model is fully implemented, the awareness, retention, and curriculum strategies applied at the precollege developmental stages should reduce the need for the college actions.

The NCHRP reports contain the detailed individual action plans recommended for further development in Phase III of this NCHRP project.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The research completed in this project leads to eight principal conclusions:

1. The poorly defined image of civil engineering is a serious impediment to recruiting new entrants to the profession.
2. Additional efforts to enhance the civil engineering pool are warranted.

TABLE 4 ARC Model

Market Strategies	DEVELOPMENTAL STAGES				
	Pre-College (grade level)			College (year of study)	
	K-6	7-8-9	10-11-12	1-2	3-4
AWARENESS	The Environment ⇒	The Engineer ⇒	The Civil Engineer	Introduction to Civil Engineering Disciplines ⇒	Summer/ Co-op Employment
RETENTION	Field Trips ⇒	Role Models ⇒	Peer Mentors	Clustering ⇒	Professional Mentors
CURRICULUM	Technology Applications ⇒	Mathematics and	Science Emphasis	Introduction to Design ⇒	Project Design

TABLE 5 Goals of Candidate Action Plans

Market Strategies	Pre-College			University	
	Elementary (Grades K-6)	Junior High (Grades 7-8-9)	Senior High (Grades 10-11-12)	Engineering/Non-Engineering Students (Freshman/Sophomore)	Civil Engineering Students (Junior/Senior)
AWARENESS	**ARC-1** Discuss technology and global engineering.	**ARC-4** Present engineering as a career alternative with an introduction to civil engineering.	**ARC-7** Provide civil engineering career information.	**ARC-10** Provide engineering career information to: • Civil engineering students • Students in other majors • Students at other institutions	**ARC-13** Provide information regarding the disciplines within civil engineering.
RETENTION	**ARC-2** Build confidence in mathematics and science.	**ARC-5** Maintain confidence in mathematics and science.	**ARC-8** Promote and retain technical interest.	**ARC-11** Retain and graduate civil engineering students.	**ARC-14** Retain and graduate engineering students.
CURRICULUM	**ARC-3** Integrate early relevancy of mathematics and science.	**ARC-6** Encourage further pursuit of mathematics and science.	**ARC-9** Complete mathematics and science requirements.	**ARC-12** Integrate early exposure to engineering/civil engineering disciplines.	**ARC-15** Integrate design projects (collaborative learning) into the curriculum.

3. Enhancement efforts should go forward as a coordinated program with the themes of awareness, retention, and curriculum.
4. Enhancement efforts should address all of the educational development stages of the civil engineer, from preschool through college.
5. Enhancement efforts should target the adults who influence career choice decisions as well as students.
6. Enhancement efforts should effectively use existing programs.
7. Personal involvement of civil and transportation engineers is crucial to the success of efforts to enhance the civil engineering pool.
8. Additional market research is needed.

Recommendations

The basic recommendation resulting from the research is to adopt the ARC model and further develop and implement the candidate action plans. Table 6 shows the ARC matrix of candidate action plans and highlights those that are recommended for testing in Phase III of the research:

1. Develop and test a coordinated set of awareness actions, across all student developmental levels.
2. Develop and test a coordinated set of awareness, retention, and curriculum actions at the high school level.
3. Develop and test retention actions for civil engineering students.

TABLE 6 Actions Recommended for Phase III

Market Strategies	Pre-College			University	
	Elementary (Grades K-6)	Junior High (Grades 7-8-9)	Senior High (Grades 10-11-12)	Civil Engineering Students	Freshman/Sophomore Engineering Students
AWARENESS	Discuss technology and global engineering.	Present engineering as a career alternative with an introduction to civil engineering.	Provide civil engineering career information.	Provide civil engineering career information to: • Other engineering majors • Other institutions	Provide engineering career information to: • Other majors • Other institutions
RETENTION	Build confidence in mathematics and science.	Maintain confidence in mathematics and science.	Promote and retain technical interest.	Retain and graduate civil engineering students.	Retain and graduate engineering students.
CURRICULUM	Integrate early relevancy of mathematics and science.	Encourage further pursuit of mathematics and science.	Complete mathematics and science requirements.	Integrate early exposure to civil engineering disciplines.	Integrate early exposure to engineering disciplines.

4. Formulate a complete marketing strategy for civil engineering careers.

5. Disseminate the findings of this research to a broad audience of civil engineering, engineering, and education decision makers.

6. Encourage AASHTO to spearhead the recommended national implementation effort.

ACKNOWLEDGMENTS

This paper is based on the reports that were prepared by the Pennsylvania Transportation Institute, Pennsylvania State University, as partial fulfillment of an NCHRP project. The Brand Consulting Group, Southfield, Michigan, was the subcontractor that performed the market research study.

REFERENCES

1. *Civil Engineering Enrollment & Degree Data—1989.* ASCE, New York, N.Y., Aug. 1990.
2. W. B. Johnston and A. E. Packer. *Workforce 2000: Work and Workers for the 21st Century.* Hudson Institute, Inc., Indianapolis, Ind., June 1987.
3. *Special Report 207: Transportation Professionals—Future Needs and Opportunities.* TRB, National Research Council, Washington, D.C., 1985.
4. *Press Release USDL 89-485.* Bureau of Labor Statistics, U.S. Department of Labor, 1988.
5. ITE Technical Council Committee 2-32. Attracting Students to a Professional Career in Transportation Engineering. *ITE Journal*, Vol. 60, No. 1, Jan. 1990, pp. 42–48.
6. L. C. Liburdi. *Special Report 210: Education and Training Needs of Women in Transportation.* TRB, National Research Council, Washington, D.C., 1984, pp. 184–191.
7. U.S. Congress. Office of Technology Assessment. *Demographic Trends and the Scientific and Engineering Work Force.* 99th Cong., 2d sess., 1986. Transmitted to the Task Force on Science Policy of the House Committee on Science and Technology, Dec. 1986; U.S. Government Printing Office, 1987.
8. J. R. Brown. The Engineering Student Pipeline. *Engineering Education*, Vol. 78, No. 3, May 1988, pp. 733–734.
9. D. S. Gedney. Educating 21st Century Highway Engineers—Industry Perspective of Highway Engineer Needs. *Proc., Conference on Engineering 21st Century Highways*, San Francisco, April 1988, pp. 216–221.
10. Women in Engineering. *Engineering Manpower Bulletin*, No. 99, Feb. 1990.
11. Minorities in Engineering. *Engineering Manpower Bulletin*, No. 97, Oct. 1989.
12. R. A. Ellis. Engineering and Engineering Technology Degrees, 1989. *Engineering Education*, Vol. 80, No. 3, April 1990, pp. 391–422.
13. J. A. Alexander. Professionalism and Marketing the Profession of Civil Engineering. *Journal of Professional Issues in Engineering Education and Practice*, Vol. 117, No. 1, Jan. 1991, pp. 10–20.
14. Female Engineering Graduates Received Highest Salary Offers in 1990. *Engineering Education News*, Vol. 17, No. 5, American Society for Engineering Education, Washington, D.C., Dec. 1990.
15. *Labor Market Conditions for Engineers: Is There a Shortage?* National Research Council, Washington, D.C., 1984, pp. 89–92.
16. *"TRAC" Careers Center, A Non-Profit Organization.* Draft Prospectus. AASHTO, Washington, D.C., April, 1990.

Publication of this paper sponsored by Committee on Transportation Education and Training.