Texas Department of Highways and Public Transportation, Austin, Feb. 1976.
3. J.L. Memmott and J.L. Buffington. Predicting Traffic Volume Growth Rates Resulting from Changes in Highway Capacity and Land Development. Texas Transportation Institute, Texas A&M Univ., College Station, Res. Rept. 225-23, Jan. 1981.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Vehicle Life-Cycle Costing with Probabilistic Part Replacement and Repair Options

GEORGE C. JACKSON AND T. H. MAZE

The purchase of transportation parts and equipment is a complex task that requires more than just the simple comparison of prices submitted by potential vendors. Ideally, the financial implications of a purchase should be analyzed over the entire life cycle of the item. The theory of how Markov chains were used by a purchasing analyst to solve the problem of whether to equip a new fleet of semitrailers with radial or conventional tires is described. Markov chains are used to develop total tire costs over the life cycle of the semitrailer for both types of tires. Although the methodology is demonstrated by using a truck tire example, the methodology is equally applicable to analyze the life-cycle costs of other transportation equipment.

The use of a life-cycle costing methodology to determine the most cost-efficient type of tire to purchase with new truck trailers is described in this paper. Although the methodology is applied to a truck trailer costing problem, it is equally applicable to the purchase of bus tires, maintenance truck tires, garbage truck tires, school bus tires, and so on, or to other cases where vehicle parts have probabilistic replacement or repair options or both.

Choosing the right tire for a new fleet of trailers is an important problem for many fleet managers. Tires will normally account for 10 to 15 percent of the purchase price of the trailer and can contribute significantly to maintenance costs over the life of the trailer. Trailer tires also play an important role in customer service by influencing the incidence of late deliveries caused by tire problems. Thus, the choice of tires to be specified on new trailer fleets is an important decision, i.e., one that requires careful analysis.

The choice is generally between steel-belted radials and conventional belted tires. The initial cost of radial tires is substantially greater than the cost of conventional tires. However, radials have been found to wear longer and to be able to be recapped more times, thus reducing the total tire cost over the life of the trailer. The question that remains is which type of tire to buy: radial or conventional.

The methodology developed to answer the question of which type of tire to purchase uses a Markov chain. The Markov chain is employed to develop estimates of total tire costs over the expected life of a truck trailer. By using these cost estimates, the present worths are calculated for both steel-belted radial tires and conventional belted tires and the minimum cost tire type is selected.

ANALYTICAL APPROACH

Markov chains are used as a mathematical means of forecasting the probability of a particular item transcending from one state to another during a time period. For example, after a period of wear, a new tire can either be recapped or, if no longer recappable, scrapped. Thus, there is a probability that a new tire will either transcend to a state where it is recapped or to a state where it is scrapped. If the tire is recapped, after another period of wear it again faces the possibilities of being recapped or scrapped. Every time wear causes the tire to reach the end of its safe tread, the tire can transcend into one of two states (scrap or recap). Markov chains are used to quantify the probability of an item transcending from one state to the next. The probability of transcending from one state to any other state is defined in a Markov chain by a transition matrix (1).

To demonstrate a transition matrix, suppose that there are m states. For example, the states could be new, scrap, first recap, second recap, and so on. Let the probability of transcending from one state to another be represented by p, where p_{12} represents the probability of transcending from state 1 to state 2. For example, p_{12} could represent the probability of transcending from the first recapping to the second recapping. Of course, the probability of impossible transitions would be zero. For example, if p_{32} represented the probability of transcending from the third recapping to the second, then p_{32} would equal zero. An example of a transition matrix with m states is shown below:

$$\underline{P} = \begin{array}{c|ccccc} \text{States} & \underline{0} & \underline{1} & \underline{2} & \underline{3}\ \cdots & \underline{m} \\ \hline 0 & p_{00} & p_{01} & p_{02} & p_{03} & p_{0m} \\ 1 & p_{10} & p_{11} & p_{12} & p_{13} & p_{1m} \\ 2 & p_{20} & p_{21} & p_{22} & p_{23} & p_{2m} \\ 3 & p_{30} & p_{31} & p_{32} & p_{33} & p_{3m} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ m & p_{m0} & p_{m1} & p_{m2} & p_{m3} & p_{mm} \end{array}$$

To demonstrate the use of the transition matrix for forecasting tire states, suppose that a new tire has a probability of 0.6 of being able to be recapped successfully. The transition from the state new to the state first recap is defined by the transition probability whose value is 0.6. At some point, the tire must be discarded and replaced by a new tire. Returning to the example cited above, the probability of an unsuccessful recap would be 0.4, in which case a new tire would be purchased. This would mean that the probability of remaining in the state new would be 0.4. Once a tire is scrapped, it cannot be recapped. Thus it is found that one of the rows of the matrix of transition probabilities must have all elements associated with recapping equal to zero and the element that describes the acquisition of a new tire must be equal to one.

The transition matrix permits the determination of the probability of transcending from one state to another in one step. However, the transition matrix can be used to determine the probability of transcending from one state to another in some finite number of steps n. If the probability of going from state i to all other states (k) in one step is p_{ik}, and the probability of going from all other states (k) to state j in one step is p_{kj}, then the probability of going from i to j in two steps is equal to

$$\sum_{k=0}^{m} p_{ik}\, p_{kj}$$

This is the same as multiplying the ith row by the jth column. This value is also the p_{ij} element of the squared transition matrix ($\underline{P}$). The probability of transcending from any state i to any state j in two steps can be similarly determined by examining the corresponding element in the squared transition matrix. Similarly, the probability of transcending from the ith state to the jth state in n steps is the corresponding element in $\underline{P}^n$.

The life of a tire (actually a tire casing) can be described by the transition matrix $\underline{P}$. The length of time between transition steps defined by $\underline{P}$ is equal to the average life of the tread. Let the state vector $\underline{\Pi}$ be the percentage allocation of the total set of fleet tires in each cycle of the tire casing life at the start of the process. For example, at the start of the process, all tires would be in the state new. At various steps in the Markov process (periods of time in the future), tires would be in several states: some would be new, some at first recap, some at second recap, and so on. The state vector describes the percentage of the total number of tires in each of these states at a point in time. The status of the fleet's tire casings may be found at period n from the equation

$$\underline{\Pi}_n = \Pi_o\, \underline{P}^n \quad (1)$$

where $\underline{\Pi}_o$ is the state of the fleet's tires at the beginning of the process (all tires new).

To find the actual number of tires in each state at any point in time, one needs only to multiply the state vector $\underline{\Pi}$ by the number of tires in the fleet or fleets. That is,

$$\underline{S}_n = S_o\, \underline{\Pi}_n \quad (2)$$

where $\underline{S}$ is defined as the vector that describes the number of tires in each state in the tire life cycle.

To find the costs associated with the tire fleet at each transition, the costs of purchasing and recapping a tire ($\underline{C}$) are multiplied by the number of tires in each state ($\underline{S}$); thus,

$$D_n = \underline{C}\, S_n \quad (3)$$

where D_n is the dollar cost of the fleet's tires at time period n.

The final step in the analytical process is to compare the present worths of the cost streams (D_N) generated by different types of tires.

It is reasonable to expect that the matrix $\underline{P}$ will change with each different type of tire. It is also reasonable to expect that the cost vector $\underline{C}$ will be dependent on the tire under examination.

EXAMPLE OF COMPARING TIRES BY USING MARKOV CHAINS

In order to demonstrate the analytical approach described above, consider the problem of the selection of the type of tire to be ordered with the acquisition of 50 new trailers. There are 8 tires on each of these trailers; thus the tire selection decision encompasses 400 tires. (Note: The data used in this example were collected while Jackson was a distribution analyst for Anchor Hocking, Inc. The values used in the example were generated from Anchor Hocking truck fleet records and other sources.)

There were two types of tires under consideration for the fleet: steel-belted radials and conventional belted tires. The purchase price of a steel-belted radial is $220; recapping costs $55. The purchase price of a conventional belted tire is $150; recapping costs $45. The average tread lives of steel-belted radials and conventional belted tires are approximately 130 000 and 80 000 miles, respectively. These tread lives are based on a relatively normal mix of highway and city miles over normal road surfaces by a medium-grade tire. These tread lives also apply to the life of a retread.

The data given in the table below present the probabilities of a tire casing holding up for subsequent retreadings:

	Probability	
Retread	Conventional	Radials
1st	0.625	0.85
2nd	0.35	0.60
3rd	0.10	0.30
4th	0.00	0.10

Radials exhibit the desirable property of a higher probability for retreading at each retread transition. Radials also exhibit the potential for more successful retreads than do conventional tires. Retreading has a distinct advantage because the cost is substantially less than the purchase price of a new tire.

Because the tires would be delivered with the new trailers, the state vector $\underline{\Pi}_o$ was initialized with all tires in the new state. The values within the transition matrix $\underline{P}$ are the probabilities of being retreaded or replaced by a new tire and were developed from the data shown in the table above. The transition matrix for radials ($\underline{P}_R$) is shown in the table below:

		Retread			
Phase	New	1st	2nd	3rd	4th
New	0.15	0.85	0	0	0
Retread					
1st	0.40	0	0.60	0	0
2nd	0.65	0	0	0.35	0
3rd	0.90	0	0	0	0.10
4th	1.00	0	0	0	0

To demonstrate how to interpret the table above, the first element in the first column (0.15) is the probability of a new tire not withstanding retreading after 130 000 miles of wear, which is 15 percent. The first element in the second column (0.85) is the probability of the new tire taking a recap after 130 000 miles of wear. The second row indicates that, given that the tire is on its first retread, there is a 40 percent probability that a new replacement will be required and a 60 percent probability that the current tire can be retreaded for a second time. The 4th retread row shows that there is no chance of a fifth retread; that is, a new replacement tire must be purchased. The zero cells in the matrix reflect assumptions that the tires will be ready for scrap or retreading at specific mileage increments and that only new replacement tires will be purchased. The transition matrix for conventional tires (P_C) is given below:

Phase	New	Retread 1st	Retread 2nd	Retread 3rd
New	0.375	0.625	0	0
Retread				
1st	0.65	0	0.35	0
2nd	0.90	0	0	0.10
3rd	1.00	0	0	0

The conventional tire transition matrix has fewer states because conventional tires do not have a fourth retread state. We have now defined all the values needed to perform the computations required to determine the number of tires in each state. Next, the number of time periods that are to be studied need to be estimated.

Fleet records indicated that trailers can be expected to be in service for from 8 to 9 years, at which time they, along with their tires, will be traded in for replacement trailers with new tires. During this period of service, a trailer can be expected to travel approximately 450 000 miles. This limits the number of future periods that need to be considered. For radials, only about four transitions (130 000 miles/period) are considered and for conventional tires only about seven transitions (80 000 miles/period). The number of miles in each period reflects the tread wear characteristics of the two types of tires.

The state values (Π_i) for both radial and conventional tires are given in Table 1. The data in Table 1 indicate the percentage of retreads that must be performed. The calculation of costs is accomplished by multiplying the data in Table 1 by the costs of new tires and recapping.

The costs for new tires and retreads for both types of tires for the number of mileage periods the trailers will be in service is given in Table 2. The number of new tires required for each period is equal to the percentage of new tires for that time period multiplied by 400 (the total number of tires in the fleet). The cost of new tires for each time period is simply the number of new tires needed times the cost of a new tire. Total recap costs are found by determining the number of recaps for each time period and multiplying that number by the cost of a recap.

Comparison of each type of tire period by period, as presented in Table 2, would yield inaccurate information because the periods are defined by different mileage intervals based on the tread wear of the tire. The periods shown in Table 2 were translated into miles and compared over the life of the trailer; the results are given in the table below:

Item	Cost ($) Conventional Tire	Cost ($) Radial Tire
Miles		
0	60 000	88 000
80 000	33 750	
130 000		31 900
160 000	41 100	
240 000	41 520	
260 000		47 960
320 000	39 840	
390 000		51 700
400 000	41 100	
450 000	25 425	22 033
Total cost	282 425	241 593
Savings		41 142

Table 1. Proportion of tires in fleet requiring retread or new replacements.

Time Period	New (%)	Proportion Needing Replacement (%) by No. of Retreads: 1st	2nd	3rd	4th	Total (1 + 2 + 3 + 4)
Radials						
1	100	–	–	–	–	0
2	15	85	–	–	–	85
3	36	13	51	–	–	64
4	45	31	8	16	–	55
5	39	38	18	3	2	61
6	39	33	23	5	–	61
Conventional						
1	100	–	–	–	–	0
2	37	62	–	–	–	62
3	55	23	22	–	–	45
4	56	34	8	2	–	44
5	52	35	12	1	–	48
6	55	33	12	–	–	45
7	54	34	12	–	–	46

Table 2. Tire costs per period.

Period[a]	Conventional Tire Cost ($) New	Retread	Total	Radial Cost ($) New	Retread	Total
1	60 000	0	60 000	88 000	0	88 000
2	22 500	11 250	33 750	13 200	18 700	31 900
3	33 000	8 100	41 100	31 680	16 280	47 960
4	33 600	7 920	41 520	39 600	12 100	51 700
5	31 200	8 640	39 840	34 320	13 420	47 740
6	33 000	8 100	41 100	34 320	13 420	47 740
7	32 400	8 280	40 680			

[a]A period is defined differently for conventional and radial tires; thus the cost streams are not directly comparable.

[Note: The costs at 450 000 miles were prorated to reflect the partial periods at the end of trailer use as follows: conventional = $40 680 x (50 000/80 000) and radials = $47 740 x (60 000/130 000).]

The data given above indicate that the purchase of radials rather than conventional tires results in a savings of $41 142 over the life of the trailers. By translating miles into years and assuming 50 000 miles/year, the present value of the savings can be determined (Table 3). Thus, the model has shown that radial tires possess an economic advantage over conventional tires when evaluated over the life of the trailer fleet.

CONCLUSIONS

It has been shown here that Markov analysis can be a useful technique for estimating the life-cycle costs of alternative tires to be purchased on new trailers. Such estimates are useful in measuring the life-cycle cost implications of alternative vehicle systems or subsystems. Although the methodology was demonstrated by using a truck trailer tire example, it could be used to evaluate tire alternatives for other types of vehicles, or to evaluate the life-cycle cost of other parts with probabilistic transitions from one state to the next.

As with many probabilistic models, the basic problem is to determine the probabilities to be used. Some fleet's maintenance records may reveal the probability of a part transcending from one state to the next, but many will not, particularly when a new brand or type of part or equipment is used. There are also some published experiments,

Table 3. Present value analysis of radial tire savings by mileage intervals.

Year	Miles (000s)	Conventional ($)	Radials ($)	Advantage of Radials ($)	Discount Factor (12%)	Discounted Advantage of Radials ($)	Cumulative Advantage of Radials ($)
1	0-50	60 000	88 000	-28 000	0	28 000	-28 000
2	50-100	33 750	-	33 750	0.893	30 139	2 139
3	100-150	-	31 900	-31 900	0.797	25 424	-23 285
4	150-200	41 100	-	41 100	0.712	29 263	5 938
5	200-250	41 520	-	41 520	0.636	26 407	32 385
6	250-300	-	47 960	-47 960	0.567	27 193	-5 192
7	300-350	39 840	-	39 840	0.507	20 199	25 391
8	350-400	41 100	51 700	-10 600	0.452	4 791	20 600
9	400-450	25 425	22 033	3 392	0.404	1 370	21 070

Note: The total costs for conventional and radial are $282 735 and $241 593, respectively.

but these are specific to those operations conducting the experiments.

Estimated probabilities can be derived and their sensitivity tested to aid in the decision-making process. For instance, in the case of the tire example, the data presented in the in-text table of retread possibilities and the transition matrices for radial and conventional tires reflect average values for all data elements.

The fact is that not all tires will wear down at precisely the same number of miles. Some will last 200 000 miles and others may last only a few thousand. Thus, there may be a chance, as shown in the transition matrices for radial and conventional tires, that a tire in the first retread state may actually remain there for more than a full period or for only a partial period. The matrix could be expanded to handle this by using smaller mileage increments and defining more states in the matrix. The desirability of adding this complexity to the model will depend on the sensitivity of the results to such factors.

Similarly, if used tires or retreads were to be purchased for replacements, there would be probabilities of moving to other retread or state-of-use stages. For example, as shown in the transition matrix for radial tires, the probability of moving from the 2nd retread to new would be divided between the 1st, 3rd, and 4th retreads, depending on the replacement purchase mix. Thus, it is possible to expand the level of detail in the analytical approach beyond that demonstrated in this paper.

We have attempted to show how the mathematical concept of Markov chains has been and can be applied to practical life-cycle costing problems to help decision makers arrive at an economically sound decision. There are, of course, other factors that influence this decision and must be brought into the analysis at some point. For example, steel-belted radials ride smoother and experience fewer punctures than conventional tires. These could be the overriding factor if damage and on-time deliveries are of paramount importance. There may be several other considerations if the methodology is to be used in evaluating life-cycle costs of other types of purchases. However, these factors are difficult to measure and difficult to include in any mathematical model.

ACKNOWLEDGMENT

We would like to thank Daniel C. Bolger, director of physical distribution, Anchor Hocking, Inc., for his support of the analysis and the data presented in this paper.

RERERENCE

1. F.S. Hillier and G.J. Lieberman. Introduction to Operations Research. Holden-Day, Inc., San Francisco, 1980, pp. 487-548.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Multiregional Input-Output Model: A Further Extension

CHONG K. LIEW AND CHUNG J. LIEW

A multimodal, multioutput, multiregional variable input-output (MMMVIO) model is introduced to evaluate the economic impact of an improvement to a transportation facility. The distinguishing feature of the MMMVIO model over conventional multiregional input-output models is its flexibility. The MMMVIO model is a price and cost-sensitive model, whereas conventional input-output models fail to share these properties. Regional coefficients, trade coefficients, modal split, and the composition of primary and secondary outputs become endogenous variables under the MMMVIO model. The conventional input-output models assume that regional coefficients and trade coefficients are fixed, regardless of changes in either input cost or output price, and each industry produces a single output. The modal split has never been considered explicitly within the context of the conventional input-output model. The MMMVIO model alleviates these unrealistic assumptions.

The conventional multiregional input-output models developed by Isard (1), Moses (2), Leontief and Strout (3), and Polenske (4) are able to describe the industrial transaction, trade flows, industrial output, income, and employment in regional and industrial details. However, the input-output model assumes that:

1. Each industry in each region produces a single output;
2. Regional input-output coefficients are fixed regardless of changes in output prices, input costs, tax structure, or transportation costs;
3. Neither input costs nor output prices nor transportation costs would affect the industry's decision on output, input mix, employment, income, transport modal choice, and trade structure (conventional input-output models fail to respond to cost and price changes);
4. Trade coefficients are independent of transportation costs and input costs; and
5. Transport modal choice has never been fully explained by conventional input-putput models.

To make the input-output model more flexible, we introduce a multimodal, multioutput, multiregional variable input-output (MMMVIO) model, which is not based on such unrealistic assumptions as imposed on conventional input-output models. The MMMVIO model assumes that each industry in each region may produce more than one output. The multioutput and multiinput relation is specified by the production frontiers. Under the MMMVIO model, the regional input-output coefficients become endogenous to the model. A change in output prices, tax structure, or transportation costs affects not only the input-output coefficients but also the trade coefficients. The MMMVIO model assumes that profit maximization guides every business decision on outputs, input mix, employment, income, modal choice of shipment, and trade flows.

The MMMVIO model is derived from the duality between production and price frontiers. The price frontiers are obtained from the dual relations. These price frontiers are expressed in terms of input elasticities, transportation costs, wage rates, service prices of capital, tax rates, and technical progress parameters. The equilibrium prices obtained from the price frontiers determine regional input-output coefficients, trade coefficients, and modal split of commodity shipment. The usual output, income, and employment of each industry in each region are determined by the balance equations.

A derivation of MMMVIO is given in the next section, and a brief description on the usefulness of the model is discussed afterwards.

MMMVIO MODEL

Consider an economy that has m regions and n industries. Each industry produces a primary output and several secondary outputs. Each commodity is shipped to each region by one of g shipping modes.

Industrial outputs in each region are produced by a linear logarithmic production frontier, i.e.,

$$\beta_{jj}^{r}\ln Y_{jj}^{r} + \sum_{i\neq j} \beta_{ij}^{r}\ln Y_{ij}^{r} - \alpha_{oj}^{r} - \sum_{k}\sum_{s}\sum_{i} \alpha_{ij}^{srk}\ln X_{ij}^{srk} - \gamma_{j}^{r}\ln L_{j}^{r} - \delta_{j}^{r}\ln K_{j}^{r} = 0 \qquad (1)$$

$\left(\text{Note, unless otherwise stated, } \sum_{k} = \sum_{k=1}^{g},\ \sum_{s} = \sum_{s=1}^{m},\ \sum_{i} = \sum_{i=1}^{n}, \text{ and } \sum_{i\neq j} = \sum_{\substack{i=1\\ i\neq j}}^{n}.\right)$

where

Y_{jj}^{r} = amount of primary output j produced by industry j in region r,

Y_{ij}^{r} = amount of secondary output i produced by industry j in region r ($i \neq j$),

X_{ij}^{srk} = amount of commodity i produced in region s and delivered to industry j in region r by shipping mode k,

L_{j}^{r} = labor employed by industry j located in region r, and

K_{j}^{r} = service of capital employed by industry j located in region r.

β_{ij}^{r}, α_{oj}^{r}, α_{ij}^{srk}, γ^{r}, and δ_{j}^{r} are parameters of the production frontier, and it is assumed to be a linear homogeneous function, i.e.,

$$\beta_{jj}^{r} + \sum_{i\neq j} \beta_{ij}^{r} - \sum_{k}\sum_{s}\sum_{i} \alpha_{ij}^{srk} - \gamma_{j}^{r} - \delta_{j}^{r} = 0 \qquad (j = 1, \ldots n; r = 1, \ldots m) \qquad (2)$$

The commodity i produced by all industries in region s is demanded by industries and final users of all regions, and the shipment of the commodity is made by transportation mode k:

$$\sum_{k}\sum_{r}\sum_{j} X_{ij}^{srk} + \sum_{k}\sum_{r} F_{i}^{srk} = \sum_{j} Y_{ij}^{s} \qquad (3)$$

The profit maximization with Equations 1-3 yields the following solutions. (Note: the full mathematical derivation is available from the authors.)

$$Y_{ij}^{r} = \beta_{ij}^{r}(1 - t_{j}^{r})P_{j}^{r}Y_{jj}^{r}/[\beta_{jj}^{r}(1 - t_{i}^{r})P_{i}^{r}] \qquad (\text{for } i \neq j) \qquad (4)$$

$$X_{ij}^{srk} = \alpha_{ij}^{srk}(1 - t_{j}^{r})P_{j}^{r}Y_{jj}^{r}/(\beta_{jj}^{r}C_{i}^{srk}P_{i}^{s}) \qquad (5)$$

$$L_{j}^{r} = \gamma_{j}^{r}(1 - t_{j}^{r})P_{j}^{r}Y_{jj}^{r}/(\beta_{jj}^{r}w_{j}^{r}) \qquad (6)$$

$$K_{j}^{r} = \delta_{j}^{r}(1 - t_{j}^{r})P_{j}^{r}Y_{jj}^{r}/(\beta_{jj}^{r}v_{j}^{r}) \qquad (7)$$

P_{j}^{r} is the equilibrium price of commodity j produced in region r, and C_{i}^{srk} is one plus the unit cost of delivering commodity i from region s to region r by shipping mode k. C_i^{srk} is called the transportation cost factor in this study.

From Equations 1, 2, 4-7, we obtain the price frontiers that can be conveniently presented as a matrix form:

$$(B - S)\ln p = \hat{\gamma}\ln w + \hat{\delta}\ln v - B\ln(1 - t) + W\ln c + A_o \qquad (8)$$

where

$$\underset{(nm,1)}{\ln p} = \begin{pmatrix} \ln p^1 \\ \vdots \\ \ln p^m \end{pmatrix} \quad \text{and} \quad \underset{(n,1)}{\ln p^r} = \begin{pmatrix} \ln p^r_1 \\ \vdots \\ \ln p^r_n \end{pmatrix}$$

$$\underset{(nm,nm)}{B} = \begin{bmatrix} \beta^1 & & 0 \\ & \beta^2 & \\ 0 & & \beta^m \end{bmatrix} \quad \text{and} \quad \underset{(n,n)}{\beta^r} = \begin{bmatrix} \beta^r_{11} & \cdots & \beta^r_{n1} \\ \vdots & & \vdots \\ \beta^r_{1n} & \cdots & \beta^r_{nn} \end{bmatrix}$$

$$S = A^1 + \ldots + A^g$$

$$\underset{(nm,nm)}{A^k} = \begin{bmatrix} \alpha^{11k} & \cdots & \alpha^{m1k} \\ \vdots & & \vdots \\ \alpha^{1mk} & \cdots & \alpha^{mmk} \end{bmatrix} \quad \text{and} \quad \underset{(n,n)}{\alpha^{srk}} = \begin{bmatrix} \alpha^{srk}_{11} & \cdots & \alpha^{srk}_{n1} \\ \vdots & & \vdots \\ \alpha^{srk}_{1n} & \cdots & \alpha^{srk}_{nn} \end{bmatrix}$$

$\underset{(mn,mn)}{\hat{\gamma}}$ = diagonal matrix of γ^r_j;

$\underset{(mn,mn)}{\hat{\delta}}$ = diagonal matrix of δ^r_j;

$\ln w$ = mn component vector of $\ln w^r_j$;

$\ln v$ = mn component vector of $\ln v^r_j$;

$\ln(1-t)$ = mn component vector of $\ln(1-t^r_j)$;

$\underset{(mn,nmm^g)}{W} = (W^1, W^2, \ldots W^g)$

$$\underset{(nm,nmm)}{W^k} = \begin{bmatrix} S^{1k} & 0 & \cdots & 0 \\ 0 & S^{2k} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & S^{mk} \end{bmatrix} \quad \text{and} \quad \underset{(n,nm)}{S^{rk}} = \begin{bmatrix} \alpha^{1rk}_{11} \cdots \alpha^{1rk}_{n1} & \cdots & \alpha^{mrk}_{11} \cdots \alpha^{mrk}_{n1} \\ \vdots & & \vdots \\ \alpha^{1rk}_{1n} \cdots \alpha^{1rk}_{nn} & \cdots & \alpha^{mrk}_{1n} \cdots \alpha^{mrk}_{nn} \end{bmatrix}$$

$(k = 1, \ldots g)$

$$\underset{(nmmg,1)}{\ln c} = \begin{pmatrix} \ln c^1 \\ \vdots \\ \ln c^g \end{pmatrix} \quad \text{and} \quad \underset{(nmm,1)}{\ln c^k} = \begin{pmatrix} \ln c^{11k}_1 \\ \vdots \\ \ln c^{11k}_n \\ \vdots \\ \ln c^{mmk}_1 \\ \vdots \\ \ln c^{mmk}_n \end{pmatrix}$$

A_o = nm component vector of A^r_{oj}.

(Note that the figures inside the parentheses indicate the size of the matrix.)

The price frontier (Equation 8) is expressed in terms of local wage rates (w^r_j), regional service price of capital (v^r_j), effective tax rates (t^r_j), transportation cost factor by each mode (C^{srk}_i), input-output elasticities (β^r_j, α^{srk}_{ij}, γ^r_j, δ^r_j), and technical progress parameters (α^r_{oj}).

The equilibrium prices (P^r_j) of the model are determined by the price frontier equation (Equation 8). The above equilibrium prices solve the output coefficients (d^s_{ij}) and the primary input coefficients (a^{srk}_{ij}); i.e.,

$$d^s_{ij} = Y^s_{ij}/Y^s_{jj} = \beta^s_{ij}(1 - t^s_j)P^s_j/[\beta^s_{jj}(1 - t^s_i)P^s_i] \qquad (9)$$

$$a^{srk}_{ij} = X^{srk}_{ij}/Y^r_{jj} = \alpha^{srk}_{ij}(1 - t^r_j)P^r_j/(\beta^r_{jj}C^{srk}_i \cdot P^s_i) \qquad (10)$$

Dividing Equation 3 by the primary output (Y^r_{jj}), the following balance equation is obtained, i.e.,

$$\sum_j d^s_{ij} \cdot Y^s_{jj} - \sum_r \sum_j A^{sr}_{ij} \cdot Y^r_{jj} = F^s_i \qquad (11)$$

where

$$A^{sr}_{ij} = \sum_k a^{srk}_{ij} \quad \text{and} \quad F^s_i = \sum_k \sum_r F^{srk}_i.$$

A matrix form of Equation 11 is as follows:

$$(D - A^*)Y = F \qquad (12)$$

where

$$\underset{(nm,nm)}{D} = \begin{pmatrix} D^1 & & 0 \\ & D^2 & \\ 0 & & D^m \end{pmatrix} \quad \text{and} \quad D^s = \begin{pmatrix} d^s_{11} & \cdots & d^s_{1n} \\ \vdots & & \vdots \\ d^s_{n1} & \cdots & d^s_{nn} \end{pmatrix}$$

$$\underset{(nm,nm)}{A^*} = \begin{bmatrix} A^{11} & \cdots & A^{1m} \\ \vdots & & \vdots \\ A^{m1} & \cdots & A^{mm} \end{bmatrix} \quad \text{and} \quad A^{sr} = \begin{bmatrix} A^{sr}_{11} & \cdots & A^{sr}_{1n} \\ \vdots & & \vdots \\ A^{sr}_{n1} & \cdots & A^{sr}_{nn} \end{bmatrix}$$

$$\underset{(nm,1)}{Y} = \begin{bmatrix} Y^1 \\ \vdots \\ Y^m \end{bmatrix} \quad \text{and} \quad Y^r = \begin{bmatrix} Y^r_{11} \\ \vdots \\ Y^r_{nn} \end{bmatrix} \qquad \underset{(nm,1)}{F} = \begin{bmatrix} F^1 \\ \vdots \\ F^m \end{bmatrix} \quad \text{and} \quad F^r = \begin{bmatrix} F^r_1 \\ \vdots \\ F^r_n \end{bmatrix}$$

The balance equation (Equation 12) determines the primary outputs of each industry in each region (Y).

Once the primary outputs (Y) and the equilibrium prices (P) are determined, the secondary products (Y^r_{ij} for $i \neq j$), intermediate purchases (X^{srk}_{ij}), labor demands (L^r_j), and capital demands (K^r_j) are determined by Equations 4-7.

The output produced by industry j in region r ($Y_{\cdot j}^r$) is computed as

$$Y^r_{\cdot j} = \sum_i Y^r_{ij} \qquad (13)$$

The regional input-output coefficients (a^{sr}_{ij}) are identified by the following equation:

$$a^{sr}_{ij} = \sum_k (X^{srk}_{ij})/Y^r_{\cdot j} \qquad (14)$$

The regional technical coefficient (a_{ij}^{r}) is the sum of the regional input-output coefficients over regions, i.e.,

$$a_{ij}^{r} = \sum_{s} a_{ij}^{sr} \tag{15}$$

The trade coefficients by each mode (t_{ij}^{srk}) are computed as

$$t_{ij}^{srk} = X_{ij}^{srk}/(\sum_{s} X_{ij}^{srk}) \tag{16}$$

Note that this definition of trade coefficients coincides with that of Moses (2), except that Moses did not break down the transport modal split, i.e.,

$$t_{ij}^{sr} = a_{ij}^{sr}/(\sum_{s} a_{ij}^{sr}) = X_{ij}^{sr}/(\sum_{s} X_{ij}^{sr}).$$

Following Moses (2), it is assumed that each industry in region r consumes some fraction of the import of commodity i from region s so that the trade coefficients of the transportation mode k (t_{ij}^{srk}) are the same regardless of the final users, i.e.,

$$t_{ij}^{srk} = t_{i}^{srk} \tag{17}$$

We impose this property by averaging the trade coefficients over industries, i.e.,

$$t_{i}^{srk} = (1/n)(\sum_{j} t_{ij}^{srk}) \tag{18}$$

An improvement of a transportation mode reduces the transportation cost factor (C_{i}^{srk}), which changes regional industrial outputs (Y_{ij}^{r}), trade coefficients (t_{i}^{srk}), regional coefficients (A_{ij}^{sr}), and modal choice ($t_{i}^{k} = \sum_{sr}\sum t_{i}^{srk}$).

POTENTIAL USEFULNESS OF MMMVIO MODEL

The MMMVIO model is capable of determining the feasibility of constructing new transportation systems such as highways, waterways, bridges, or railways. The model can be employed to evaluate the existing transportation system, measure the economic impact of an energy crisis, appraise the development impact of rail abandonment, and predict the economic conditions of a region that has a sustained shortage of essential resources.

The MMMVIO model is an extension of the multiregional variable input-output (MRVIO) model that has been in operation since 1979. The basic input data of the MMMVIO are the same as those of MRVIO. MRVIO was employed to evaluate an existing waterway (5,6), to appraise the feasibility of a new waterway (7), to measure the development impact of a water shortage (8), to evaluate the pollution impact of the relocation of an industry (9), and to assess the growth impact of an energy crisis (10). The sources of the data and the computer programs for the MRVIO model are described in the reports cited.

The MMMVIO model requires additional data besides those employed for MRVIO. The Make of Commodities by Industry (Survey of Current Business, April 1979) can be used to identify the primary and secondary products. The modal-split information may require a sample survey of commodity shipment. A rough estimation on the modal split can be made by using the 1972 Transportation Margin Tape (from the U.S. Department of Commerce), which identifies the transportation margin of goods delivered by each mode.

REFERENCES

1. W. Isard. Interregional and Regional Input-Output Analysis: A Model of a Space Economy. Review of Economics and Statistics, Vol. 33, Nov. 1951.
2. L. Moses. The Stability of Interregional Trading Patterns and Input-Output Analysis. American Economic Review, Vol. 45, No. 5, 1955.
3. W. Leontief and A. Strout. Multiregional Input-Output Analysis. In Structural Interdependence and Economic Development (T. Barna, ed.), St. Martin Press, Inc., New York, 1963.
4. K. Polenske. A Multiregional Input-Output Model for the United States. Harvard Economic Research Project, Economic Development Administration, Cambridge, MA, 1970.
5. C.K. Liew and C.J. Liew. Use of a Multiregional Variable Input-Output Model to Analyze Economic Impacts of Transportation Cost. TRB, Transportation Research Record 747, 1980, pp. 5-12.
6. C.K. Liew and C.J. Liew. The McClellan-Kerr Waterway and Regional Economic Development. Institute of Water Research, U.S. Army Corps of Engineers, Vicksburg, MS, 1981.
7. C.K. Liew and C.J. Liew. The Development Impact of the Coosa River Navigation Project on the Regional and National Economies. Institute of Water Research, U.S. Army Corps of Engineers, Vicksburg, MS, 1982.
8. C.K. Liew and C.J. Liew. The Development Impact of Water Shortage on Regional Economy. Proc., Business Economic Section, American Statistical Association Meeting, 1982.
9. C.J. Liew. Pollution-Related Variable Input-Output Model: The Tulsa SMSA as a Case Study. Journal of Urban Economics (in preparation).
10. C.J. Liew. The Impact of Higher Energy Prices on Growth and Inflation in an Industrializing Economy: The Korean Case. Journal of Policy Modeling, 1980.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Abridgment

Commuter Railroad Pricing in the New York Metropolitan Region

DAVID S. KESSLER AND WILLIAM SIMONSEN

A framework for examining commuter railroad pricing in the New York metropolitan region is presented. The Metropolitan Transportation Authority (MTA) operates two commuter railroads: the Long Island Railroad, which is the country's busiest, and the Metro-North, which consists of the Harlem, Hudson, New Haven, and the Hoboken-Port Jervis lines. It is shown that the distance component of the MTA commuter railroad pricing structure is fair: riders who travel longer distances pay a higher fare in relation to the benefits they receive from the incremental distance they travel; and it is efficient: the charges are related to the additional cost of carrying riders further distances. Peak-period pricing is another efficiency-based strategy that theoretically would move riders who have a choice to the off peak, thereby rationing expensive peak capacity to those who are most willing to pay for it. The current commuter railroad peak pricing policy has not charged the peak-period rider in accordance with efficient resource allocation. Restructuring of the relative prices of the different tickets along with offering a viable off-peak alternative for monthly commuters would go a long way toward pricing the peak riders in relation to the actual costs they impose while offering a workable off-peak alternative. The alternatives that are examined are those that are operationally feasible.

This paper deals with the commuter railroad fare pricing strategy at the Metropolitan Transportation Authority (MTA), which is headquartered in New York City. MTA is responsible for operating one of the largest transportation systems in the world, which encompasses subways, buses, commuter rail lines, tunnels, and bridges. A discussion is presented on pricing issues for two MTA commuter railroads: the Long Island Railroad (LIRR) and the Metro-North Commuter Railroad. These two railroads carried a combined average weekday ridership of 453 000 in September 1982.

PRICING MASS TRANSIT

Discussions of various types of fare structures often revolve around complications due to different pricing principles (e.g., economic efficiency versus social welfare), the market structure, and, finally, the role that subsidies play. Many other studies have detailed the efficiency and equity arguments of transit pricing, so we will only summarize them. Efficient pricing requires that riders pay in proportion to the costs they impose on the system. Theoretically, this would lead to true signals being sent to producers of transit services concerning how much the service is valued.

There are two different types of equity or fairness criteria that are generally considered: benefit equity, which requires that riders pay in relation to the benefits they receive, and ability-to-pay equity, which states that riders should pay according to what they can afford. Although ability-to-pay equity is clearly an important consideration and is always a priority when decisions are made, this paper only marginally deals with this issue. In accordance with established federal, state, and city legislation, MTA has provided discount fares for certain groups such as senior citizens and the handicapped. Studying the effect on various socioeconomic groups of the kinds of fare structure changes under consideration is a complex undertaking beyond the means of this paper. A separate study is being designed to better evaluate these issues.

The revenue implications of different options are clearly important considerations, especially during this time of decreasing federal assistance. In 1981, LIRR covered about 45 percent of its operating expenses through the fare box. The coverage ratio for the Metro-North Harlem-Hudson lines was about 37 percent in 1981, and it was about 56 percent for the New Haven line during the same period. The balance was provided through a variety of federal, state, local, and regional subsidies. Because the level of fares is an extremely sensitive political and economic issue, the utmost care is taken in evaluating the revenue implications of alternative fare structures.

Any modifications in the existing fare structure must be evaluated in terms of the facility and of their implementation. Changes that would make fares much more difficult to collect or place undue hardship on the administrative staff cannot seriously be considered. This includes measures that adversely affect ticket lines, on-board ticket collection, or revenue handling. Also, employees and riders should be able to easily understand the fare structure. These constraints limit the number of available alternatives and, therefore, this paper addresses only feasible alternatives that can be implemented in the short run.

MTA COMMUTER RAILROAD FARE HISTORY

Before 1980 there did not exist an independently determined rationale for pricing commuter railroad tickets. In general, the pricing relations that existed when MTA took control of these railroads, through ownership or contractual agreement, were the ones in effect until July 1980. Indeed, these were most likely inherited from the private managements of the Pennsylvania Railroad, the New York Central Railroad, and the New Haven Railroad. From a historical perspective, it appears that fares were correlated closely with distance--perhaps until around the time of the end of World War II--but that thereafter the flat rate increases in the one-way fares (a nickel, or later a dime, for each and every station on a line) distorted the relations. Discounts for commutation tickets were offered to the railroads' best customers, and deeper discounts were frequently offered to riders who traveled greater distances on the basis of a perception that there existed a rate above which the railroads would lose large numbers of riders and revenues. These notions were, at best, the tried and true rules-of-thumb of experienced railroad managers, although they were not necessarily based on economic theory. Table 1 gives a thumbnail sketch of the post-1970 fare structure changes on the LIRR.

When comparing the fares charged by distance, Commuter Rail Corporation (Conrail) fares had relatively higher monthly ticket prices than the LIRR but lower one-way ticket prices. Recent MTA policy has been to make the two MTA commuter railroad's fare structures more consistent with one another.

Thus, the pre-1980 MTA fare structures were characterized by (a) the one-way fare as the base for determining all fares; (b) a vague, informal relation of fares to distance traveled; (c) an irregular pattern of discounts for monthly commutation tickets (depending, in some cases, on local political arrangements made a long time ago); and (d) a relatively weak commitment to off-peak pricing as part of the overall fare structure.

MTA COMMUTER RAILROAD PRICING ISSUES

The following sections present some of the MTA commuter railroad pricing issues that merit review. (The arguments presented are our views, and may not necessarily reflect future MTA policy.)

Peak and Off-Peak Pricing

Public transportation in general, and commuter railroads in particular, are services characterized by considerable variability of demand based on both time of day and day of the week. Comparatively more people desire to travel during the peak periods than in the off peak, usually to commute to and from work. This group, which demands peak-period service, places the greatest burden on the system and thus imposes the greatest cost. Therefore, efficient or marginal-cost pricing requires that peak-period users pay for the additional capacity they require in order to allocate expensive peak space to those who value it the most.

Theoretically, higher peak-period charges have the desirable effect of moving some riders who have a choice to the lower-priced off peak, thereby rationing the peak capacity to those who are the most willing to pay for it. This would also make better use of excess capacity during the off peak and, in the long run, decrease operating costs in the peak to the degree they are variable. The magnitude of this shift depends, of course, on the price differential between the peak and off-peak fares. Clearly, a large differential would move more riders than would a small differential. The amount of the shift also depends on the sensitivity of peak riders to fare changes and how broadly the peak time period is defined.

The current MTA commuter rail peak pricing policy has not provided sufficient incentive to induce off-peak travel and has not priced services consistent with efficient resource allocation. Figure 1 portrays the extent of the peaking problems experienced by the railroads.

The main pricing inconsistency is that currently there exists no peak and off-peak fare alternative for monthly ticket holders who represent the vast majority of peak riders (approximately 90 percent of peak riders use some type of commutation ticket). Currently, there is a round-trip off-peak ticket designed to offer an off-peak alternative to the one-way peak ticket rider. The monthly commuter has no such off-peak alternative, since the current off-peak ticket costs more on a per ride basis than the per trip price of monthly commutation ticket holders (see Table 2). This situation grew out of the traditional view of railroad fares, which holds that the basic ticket is the one-way peak, with monthly, weekly, and off-peak ticket prices derived from the one-way ticket by using different formulas. The monthly price for a LIRR monthly commutation ticket, for instance, is discounted from the basic one-way fare times 42 rides/month.

When viewed in the traditional light, monthly tickets are not as economically efficient a manner of payment compared with one-way fares. Efficiency suffers, since monthly ticket holders tend to be peak riders, so discounts for monthly tickets lower the price for those who place the greatest burden on the system. In addition, efficiency is lessened to the extent that ridership is attracted to the peak travel times due to the discounts.

However, the monthly commutation ticket is likely to remain a fact of life, since returning to all one-way tickets would be operationally difficult. Current fare-collection methods on MTA commuter railroads are very labor intensive. There is no automatic fare collection in the offing; every ticket needs to be checked by a trainman. For instance, the LIRR has 140 stations on nine lines from which trips can originate. Under these circumstances, it is clear why a monthly flashticket makes operational sense. Currently, the railroads have no plans for installing a more capital-intensive fare-collection system. Thus, it is practical to assume that a monthly commutation ticket of some type will continue to be offered as long as fare collection remains labor intensive. Therefore, a more rational pricing policy would shift as many of these peak riders as possible, whether monthly or one-way, to periods of excess capacity and charge the ones who continue to ride in the peak a relatively higher price because of the cost they impose.

Two possible alternative methods would help achieve this goal and merit further detailed study of the revenue and operational implications. Both make the necessary assumption that offering a monthly commutation ticket is necessary for the smooth operation of the railroads.

The first option is to veiw monthly commutation ticket riders and one-way ticket riders as completely separate and distinct markets. Therefore, the first method would require that an off-peak commutation ticket be offered as an alternative for monthly commuters, as well as to continue to use an off-peak one-way equivalent. The second option calls for gradually lowering the relative price of the current off-peak ticket until it is below the per ride cost of a monthly commutation ticket. Thus, there is a single off-peak ticket that offers

Table 1. Recent fare structures, MTA commuter railroads (LIRR).

Date	Ticket Types Offered	Fare Changes	No. of Zones or Stations	Cost per Distance of Avg Trip[a] ($)	
				One-Way	Monthly
1/30/70	One-way, round trip, monthly, weekly, school (monthly), police and firemen, and ladies day	Flat fare increase of $0.20, $1.80, and $4.60 for one-way, weekly, and monthly, respectively	139 stations	1.85	47.10
1/29/72	One-way, round trip, monthly, weekly, school (monthly), and police and firemen	Up to a 16.67 percent increase	16 zones	2.00	54.85
9/1/75	One-way, one-way off peak, weekly, school (monthly), senior citizen, and Sunday round trip	23 percent across-the-board	16 zones	2.45	67.45
7/1/80	One-way, monthly, weekly, school, round-trip off peak	20 percent, monthly discount increased	11 zones	3.15	72.50
7/15/81	One-way, monthly, weekly, school, round-trip off peak	25 percent increase	10 zones	4.15	91.00
Proposed		Unknown	Unknown		

[a]This column uses the Bellmore run of 27.1 miles as an example.

Figure 1. Passenger arrivals and departures from Penn Station (LIRR).

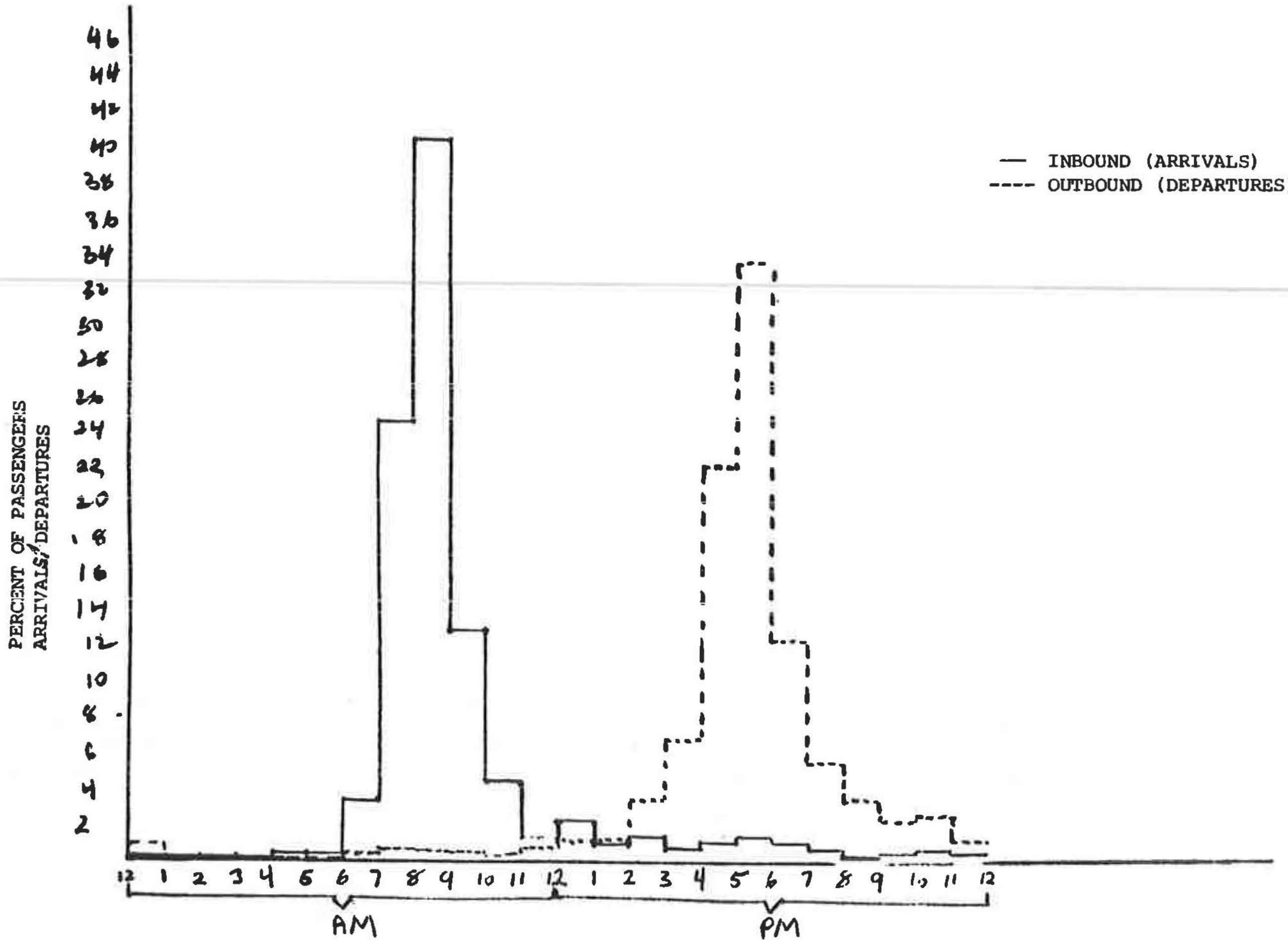


Table 2. Comparison of monthly per ride price and off peak per ride price.

Zone	Monthly Ticket Price ($)	Monthly Price per Ride (42 trips/ month) ($)	Round-Trip Off-Peak Price per Ride ($)
LIRR			
1	61.25	1.46	1.93
2, 3	69.50	1.65	2.30
4	81.25	1.93	2.78
5, 6, 7	91.00	2.17	3.13
8, 9	102.75	2.45	3.75
10	112.50	2.68	4.33
11	120.50	2.86	4.80
12	133.25	3.17	5.40
Metro-North Hudson Line			
A	47.50	1.13	1.48
B	49.00	1.17	1.50
C	51.50	1.23	1.58
D	54.25	1.29	1.78
E	59.50	1.42	2.00
F	62.50	1.49	2.10
G	67.75	1.61	2.23
H	72.50	1.73	2.53
I	78.25	1.86	2.85
J	81.25	1.93	3.00
K	82.75	1.96	3.15
L	87.50	2.08	3.45
M	91.00	2.17	3.60
N	103.50	2.46	3.98
O	110.75	2.64	4.63

a lower-priced alternative for both one-way peak riders and monthly commuters. This latter method would necessarily involve a substantial shrinking of the absolute price difference between the one-way peak ticket and the monthly ticket equivalent. This method has the advantage of offering one less ticket type than the first option. On the negative side, this option would consequently increase the number of tickets to be collected on the trains. However, collecting off-peak tickets would be done during times of less constrained capacity, and it would therefore have a smaller adverse impact on productivity. Bulk coupon booklet sales of off-peak tickets would also make sense under this option.

Both of these options may increase the number of step-ups necessary on peak trains for riders who hold off-peak tickets (riders who upgrade their tickets on the train by paying the difference between the two fares). This could be a potentially serious problem. However, both would offer an off-peak alternative for monthly ticket riders who currently have no such pricing alternative.

Other issues that merit further consideration include the following:

1. Replacing the off-peak round-trip ticket: Whether or not an off-peak monthly alternative is considered, the current off-peak one-way ticket has too many restrictions for it to be a viable alterna-

tive. For instance, the round-trip off-peak ticket needs to be used on the same day for both legs of the trip. If the return trip is made during a peak time or on the next day, a step-up fee is charged to make the total cost equivalent to two one-way peak tickets. Not only does this reduce the trainmen's productivity by forcing them to handle more fares on the trains, it generally fosters a good deal of ill will among passengers who simply do not understand the system. When the one-way off-peak ticket was offered on the Metro-North commuter railroad, 74 percent of the total one-way ticket riders bought the off-peak ticket. After the off-peak ticket became valid only for round trips, this percentage dropped to 28 percent. Returning to a one-way off-peak ticket would seem to be sensible.

2. Redefining outbound morning peak trains as off peak: The demand for seats on outbound trains during the morning peak is small compared with inbound peak demand. However, outbound service is limited in the extent that it can vary with demand. This is because trains need to be run outbound during the morning peak in order to make room for the inbound morning peak trains due to equipment storage constraints at the New York City terminals. This has led to a situation where there exists excessive capacity on these outbound morning peak trains, which can easily accommodate additional ridership. Lowering ticket prices on outbound trains would potentially attract modest increases in passengers who travel during the peak periods.

3. Monthly ticket price and one-way ticket price: The railroads are currently offering an average discount of 40-50 percent for monthly tickets when compared with using a series of one-way tickets for commutation. More analysis needs to be done to determine if this dramatic premium for using a one-way ticket during the peak period is consistent with what price breaks are necessary to discourage purchase of this type of ticket.

The thrust of these policies is to change the ticket prices of one-way peak riders and monthly commuters to reflect the true cost they place on the system while offering a viable off-peak alternative. These are much more rational policies that, if effective, would result in reducing the peak crush factors and make better use of off-peak capacity. In the longer run, they would lessen the need for future equipment purchases to meet the peak demand.

The standard objection to the policies outlined above is that they would produce revenue losses when compared with the current revenue yield. It is argued that offering cheaper tickets and inducing shifts to these cheaper tickets must necessarily lower total passenger revenue. This argument is short-sighted, since, as mentioned earlier, there could be longer-run cost reductions or revenue increases, depending on the latent peak demand. But, more importantly, instituting a peak pricing policy at the same time as a general fare increase would generate the needed revenue while maximizing the total system ridership, since off-peak ridership is more elastic than peak ridership. Stated another way, an across-the-board fare increase would move more riders off the system than would differential peak and off-peak increases. This involves raising peak charges high enough to offset the relatively cheaper off-peak price. This is as it should be under the efficiency criteria, since it rations the expensive peak capacity to those who are most willing to pay for it. Table 3 gives an example of the effects of various differential peak and off-peak fare increases on ridership and revenue, as compared with an across-the-board 25 percent fare increase.

Table 3. Ridership and revenue effects of differential peak and off-peak fare increases.

Fare Increase (%)		Revenue Increase	Ridership Change	Percentage of Ridership	
Peak	Off-Peak	(%)	(%)	Peak	Off-Peak
25	25	19.14	-4.69	75	25
26	23	19.44	-4.65	75	25
26.5	23	19.73	-4.71	75	25
26.5	22.5	19.66	-4.67	75	25
26	22.5	19.73	-4.61	75	25
26	23	19.38	-4.65	74	26
26.5	23	19.66	-4.71	74	26
26.5	22.5	19.59	-4.67	74	26
26	22.5	19.30	-4.61	74	26
26	23	19.31	-4.65	73	27
26.5	23	19.59	-4.71	73	27
26.5	22.5	19.51	-4.67	73	27
26	22.5	19.24	-4.61	73	27
26	23	19.11	-4.65	70	30
26.5	23	19.38	-4.71	70	30
26.5	22.5	19.30	-4.67	70	30
26	22.5	19.03	-4.61	70	30

Note: The following assumptions are used: Ep peak = −0.15, Ep off-peak = −0.30; and % Δ ridership = % Δ fare * Ep.

The peak and off-peak pricing strategy outlined in this section is not only more efficient but also may be more equitable. Benefit equity is served, since one-way riders in the peaks receive more frequent service than off-peak riders, and they are charged for it. However, benefit equity suffers to the extent that the peaks are more crowded and less comfortable. This may be somewhat eased by the fact that peak crowding may be reduced under a more rational peak pricing policy. On the other hand, there may be latent demand for peak service that would perpetuate the crowding, notwithstanding higher peak fares.

In summary, higher peak-period charges are more efficient than uniform fares, since they are based on cost and make peak space available to those who are most willing to pay for it. The current commuter railroad pricing strategy does not operationalize these concepts particularly well. Modifying the ticket structure in a more rational way would help move toward this end. Peak pricing is also somewhat consistent with the doctrine of benefit equity.

Distance-Based Component

Distance-based fares are the next important component of the MTA commuter rail pricing structure to be considered. Distance fares, which relate the price of a trip to the distance traveled, are more efficient than uniform fares, since they address the increased cost of carrying passengers longer distances.

Both MTA commuter railroad divisions are well suited for distance fares due to the radial commuting patterns and the clearly defined Manhattan central business district (CBD) where most riders terminate. The railroads have a zone fare structure, where the price of a ticket increases with the distance from Manhattan. The smaller and more numerous the zones, the greater is the opportunity to charge each rider the cost he or she actually imposes on the system. However, small zone sizes must be traded off against whatever productivity and operational gains are associated with larger zones, such as handling fewer different ticket types less frequently.

Figure 2. Distance and monthly ticket prices.

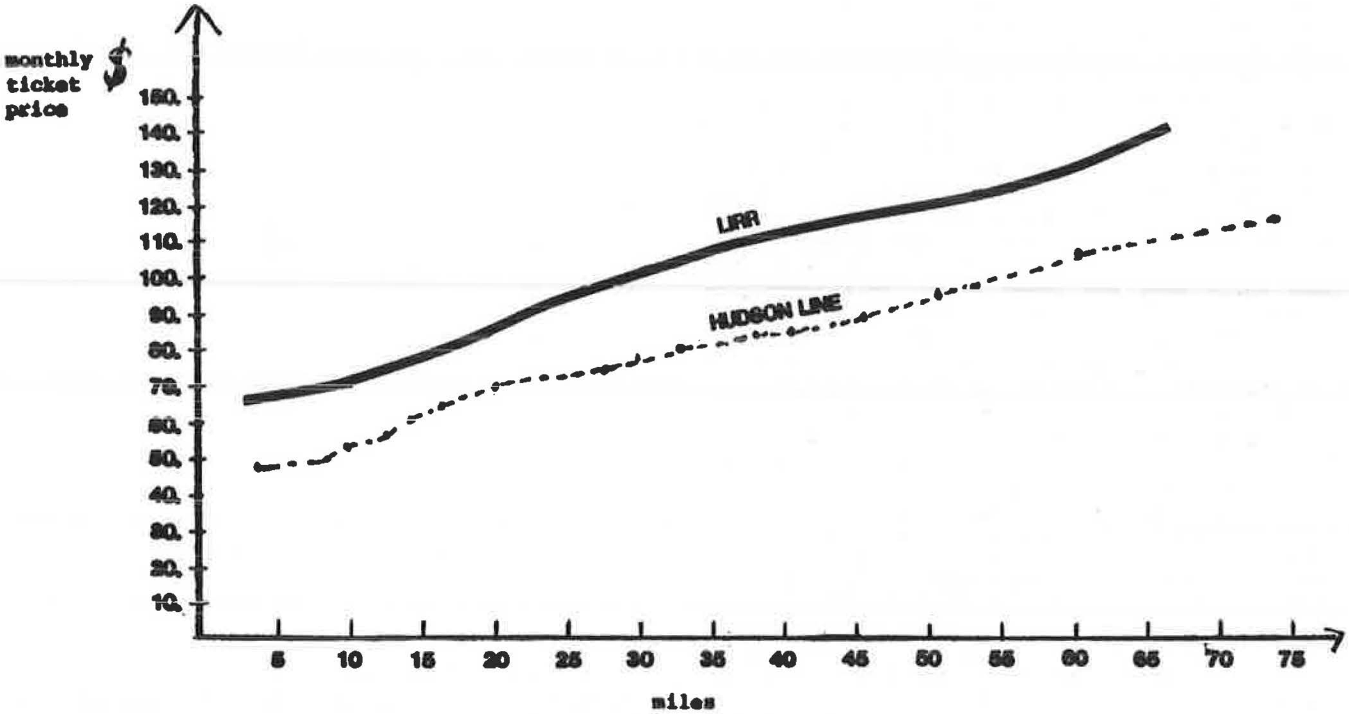
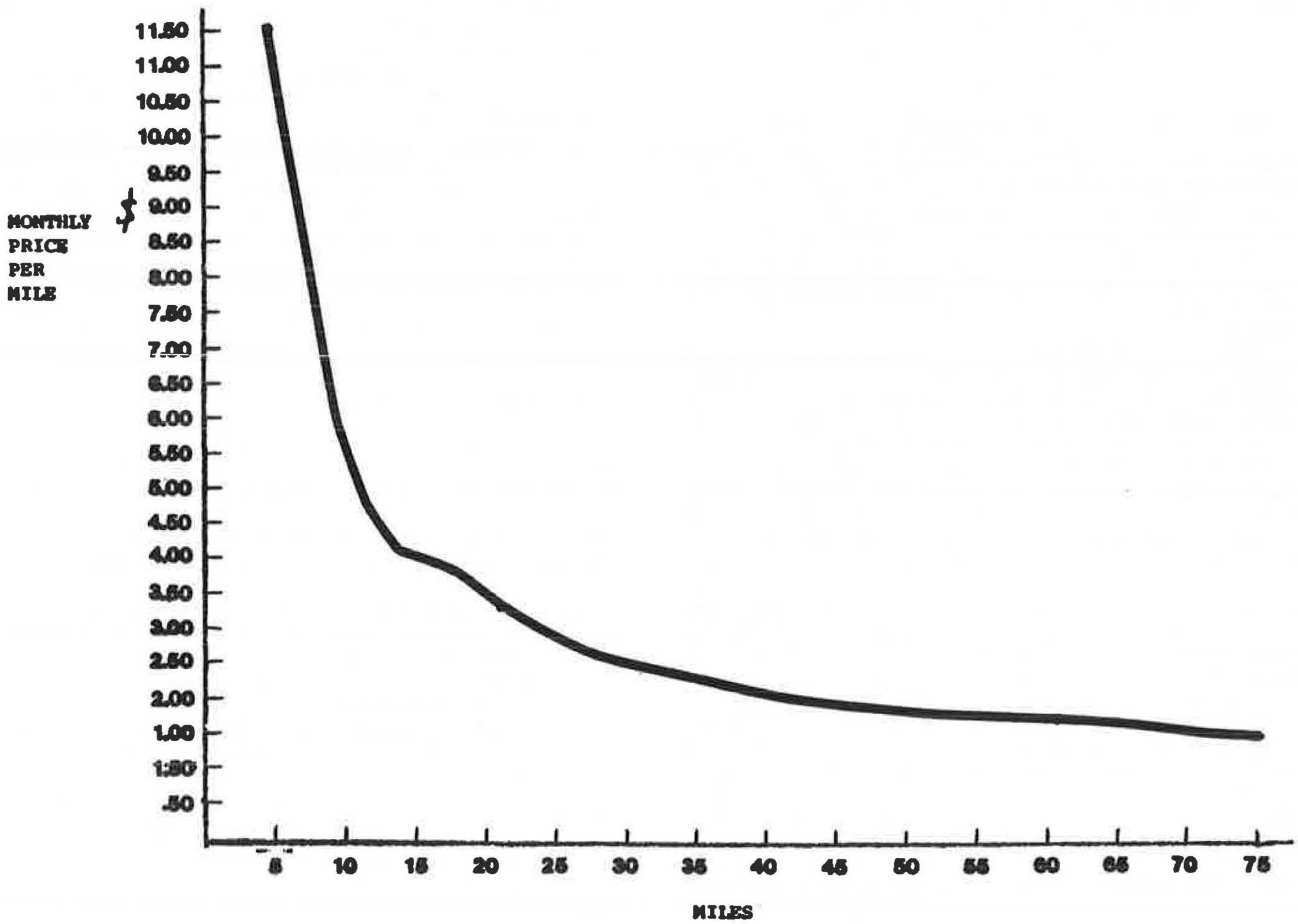


Figure 3. Monthly price per mile, Metro-North Hudson line.



Empirical Evidence on Determinants of Mass Bicycle Commuting in the United States: A Cross-Community Analysis

MICHAEL D. EVERETT AND JOHN SPENCER

A nationwide study of determinants of mass bicycle commuting (10 percent or more of trips) is discussed. Numerous studies in specific cities and states have isolated important determinants of mass bicycle commuting, such as separation from high-speed, high-volume motor vehicle traffic and relative costs (including time). However, considerable political controversy exists over the proper policies for stimulating mass bicycle commuting, and no study systematically quantifies where mass cycling takes place in the United States or the correlates of mass cycling. Therefore, the data in this paper attempt to fill that research gap and reduce the area of policy controversy by reporting all the available examples of mass bicycle commuting in the United States. The data find almost 200 examples of mass cycling for educational institutions, but fewer than 10 examples of mass cycling to work and shopping destinations. Separation from high-speed, high-volume traffic correlates with mass cycling, although examples of mass cycling on wide moderate-speed, moderate-volume arteries exist. The relative cost of cycling, which includes time costs, correlates less strongly. However, correlation does not prove causation. The overwhelming majority of fatal accidents reported occurred on arteries and not on separate bicycle facilities or residential-type roads. Nevertheless, cycling appears to remain more hazardous than driving over a given route.

Short-distance bicycle commuting provides an example of appropriate technology for a substainable economy and society in a world of increasing scarcity, congestion, and pollution. Bicycles theoretically can provide rapid, flexible, low-cost, pollution-free transportation with consistent exercise for short trips in congested urban areas. Several groups have an interest in stimulating bicycle commuting: the U.S. government to save petroleum and reduce air pollution, the bicycle manufacturer's association to stimulate new bike sales, and local transportation planners and bike clubs.

All of these groups need solid information on the determinants of mass bicycle commuting (10 percent or more of trips), but unfortunately vigorous controversy has led to considerable misinformation and confusion. For example, the U.S. Department of Transportation (DOT) published a national comprehensive bicycle transportation program (1) that emphasized promotion and education and deemphasized separate bicycle facilities to shift 15 to 30 percent of a target group of short-distance urban drivers to safe bicycle commuting.

In a review of such an ambitious bicycle-marketing program, little or no support for its assumptions in the replicable, empirical, bicycle modal-choice or marketing literature could be found (2). The literature concentrated on several determinants of safe mass bicycle commuting. First, numerous surveys were found (including sophisticated logit models) that indicated that the overwhelming majority of actual and potential commuter cyclists wanted separation from high-speed, high-volume traffic (2, p. 38). Second, relative costs, which include time costs, played a major role in modal choice. Finally, evidence that separation would reduce the risk of fatal bicycle accidents, but not eliminate it, was presented (2, p. 38).

In the review of the DOT study, it was observed that known examples of mass cycling in the United States and Europe tended to support this literature. The cities of Davis and Santa Barbara, California; Madison, Wisconsin; and Amsterdam and Utrecht, Netherlands, incorporate substantial separation from high-speed, high-volume traffic along with short trips in areas where bicycles often provide faster and more flexible transportation than other modes. However, no statistical analysis of where mass bicycling takes place exists to support or refute these observations.

Therefore, data on the percentage of cycling were collected, and determinants of cycling (separation, distance, and relative time) from nearly 300 college communities in the United States were hypothesized in order to provide a quantifiable description of those areas in which mass bicycle commuting takes place. After the methodology for collecting the data is described, the data are analyzed in light of the available literature, and finally a nontechnical discussion of the results and their implication for planning are presented.

METHODOLOGY FOR DATA COLLECTION

The study sent a mail-back questionnaire to key respondents in all major college communities in the United States as part of several senior-level marketing research classes. The students made valuable contributions by reviewing, checking, and criticizing results. They also funded the survey with $10/student, or about $600 overall. East Tennessee State University provided the computer facilities and released time for this research.

A one-page (front and back) questionnaire was developed to collect data on the percentage of cycling and on the key variables that the literature suggested would affect the level of cycling and safety. The questionnaire was field tested on approximately 20 institutions where the levels for most of the variables were known. The respondents--typically university police, junior high school principals, and city traffic engineers--made estimates of the level of the variables that roughly coincided with the known knowledge of the institutions.

The sampling strategy focused on college and university communities. First a census was taken of all junior colleges, colleges, and universities [higher education (HE)], excluding technical, seminary, and other such specialized schools, that had a student enrollment (including part time) of 9000 or more based on the College Blue Book (3). Then about 95 percent of the junior high schools (JHSs) were sampled in the same communities as were the HEs and that had populations of approximately 300 000 or less based on Patterson's American Education (4). To double-check these responses and collect data on bicycle commuting to work and shopping, questionnaires were sent to traffic engineers (TEs) in all of the latter communities.

This sampling strategy was based on the assumption that most examples of mass cycling take place in smaller cities and college or university communities. Censuses and other published studies showed relatively little cycling in large cities (1 to 2 percent of the traffic flows on routes used by cyclists). Also, rather than lightly sampling a

The current one-way charge for each zone is based on a terminal charge plus a mileage charge. Specifically, the one-way fares that went into effect in July 1981 are based on the following formula: $2.25 plus $0.075 per each mile from New York to the center of each zone. The terminal charge theoretically represents a fixed cost applicable to every zone. The mileage charge represents the variable cost of moving trains and people over different distances. Figure 2 shows how monthly fares on the LIRR and on Metro-North's Hudson line increase with distance.

The $0.075 standard mileage charge is based on an average cost, and not a marginal cost in the pure sense. The marginal cost and the average cost are equivalent only to the extent that variable costs are uniform across all distances. For instance, if it costs more to move people in the city zones due, perhaps, to higher power costs, this would not be reflected in the price. Instituting a true marginal-cost distance pricing strategy would further complicate an already complicated pricing structure with apparently only small efficiency gains.

An interesting footnote is that the use of fixed and variable charges tends to cause closer zones to have an overall higher per mile charge than more distant zones, since the fixed cost is a larger proportion of the total (see Figure 3). This is consistent with another MTA policy, which is that the commuter railroads should not be price-competitive with the New York City Transit Authority for intra-New York City trips.

Distance fares are also consistent with benefit equity, since riders who travel longer distances and receive additional benefits when compared with riders who travel shorter distances pay an incremental charge related to the additional benefits they receive. Thus, in this sense, distance fares on the commuter railroads are both efficient and equitable.

Weekly Tickets

In addition to monthly commutation tickets, both MTA commuter railroads currently offer weekly commutation tickets priced at 31 percent of the monthly ticket fare. There are two traditional arguments in favor of offering weekly tickets. First, it is thought that weekly tickets provide an alternative for commuters who cannot afford the capital outlay at the beginning of the month necessary for the purchase of a monthly ticket. In a sense, this provides a public service for these riders. Second, weekly tickets are an alternative for commuters who do not expect to ride the required number of times to make a monthly ticket economical due to vacations, illness, etc. Tradition and the convenience factors mentioned above appear to be the main reasons for continuing to offer this type of ticket.

SUMMARY

Currently, both commuter railroads charge fares that are based on distance traveled and have a peak and off-peak pricing strategy for one-way riders. However, there is no peak pricing strategy for monthly commuters who represent the vast majority of riders. The distance component is fair, to the extent that riders pay in relation to the benefits they receive, and it is efficient, since the charges are related to cost. The current peak pricing policy could be improved by offering an off-peak alternative for monthly commuters and replacing the round-trip off-peak ticket with a more flexible one-way off-peak ticket. This strategy prices all peak tickets to better reflect the actual burden the riders impose while offering a viable off-peak alternative.

Further work in this general area, which is beyond the scope of this study but merits future attention, includes analyzing the burden of alternative fare structures on various socioeconomic groups and geographic locations, examining the benefits and costs of the different taxes collected to subsidize operations, and generating more reliable fixed and variable cost estimates for pricing purposes.

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

number of strata, it was decided to collect a large number of observations from one stratum to help smooth out respondents' estimates and approximations and to pick up as many examples of mass cycling as possible.

The strategy yielded a reasonable response rate and a large data set of more than 600 observations that had minimal nonresponse problems. The initial questionnaires went out by the end of September 1981; therefore, good weather for cycling was at least a recent memory in the northern states. By mid-November 1981 there was a 27 percent response rate for the JHSs and a 35 percent rate for the HEs and TEs combined. Follow-up questionnaires at that time netted an overall response rate of 51 and 58 percent, respectively. Given that TEs and HEs filled out the same questionnaire, the overall coverage of HE reached about 75 percent. The relations between the percentage cycling and key variables, such as type of access and percentage living near campus, remained similar for the original and follow-up groups, which indicates no serious biases from the nonrespondents in exploring the key determinants of mass cycling. However, budget constraints precluded a telephone sample of nonrespondents. Their lack of cooperation suggests they contained a disproportionately large number of schools with little or no cycling.

The students edited, coded, and checked the data, which resulted in a large, final, and usable data set with minimal opportunity for investigator bias. The students also checked and double-checked the questionnaires for errors. The researchers and students resolved serious conflicts between TE and HE responses on the percentage of cycling and other key variables by follow-up telephone calls. Finally, computer printouts were examined for outlayers and coding errors. This yielded a usable data set of 216 for HE, 308 for JHS, and 91 for TE, most of whom reported on several cycling areas in their community.

FINDINGS

Where Does Mass Bicycle Commuting Take Place in the United States?

Through more than a decade of observations in many parts of the United States, mass bicycle commuting (10 percent of trips or more) has been observed only around large institutions of higher education and public schools. Bicycles constituted less than 1 percent of vehicles in traffic counts on journey-to-work bike routes in large cities such as Washington, D.C., and Chicago.

Published censuses and studies corroborate these observations. A 1977 census report on travel to work (5) found only 11 cities with around 1 percent of the workers reporting the bicycle as their principal mode of transportation to work. Only Madison (4.5 percent) and Sacramento (3.2 percent) reported substantially higher levels. Moreover, a worker included persons 14 years and older who had a part-time job (5, p. 20). Other studies of specific cities and routes give similar results, which showed bicycle traffic as about 1 percent (6,7) of vehicles on heavily traveled bicycle routes for central business district (CBD) commuters.

The cross-community data also support these observations. TEs report that most of the high-percentage cycling is to schools and only 9 observations of mass cycling is to work or shopping destinations, even when the criterion was dropped to 5 percent of vehicles or more (Table 1). Only a maximum of 16 work and shopping examples of mass cycling in the 277 college communities of less than 300 000 population can be inferred from these data (Table 1). The overwhelming majority of mass cycling involved students.

Responses from the schools themselves corroborate this finding. The table below gives 63 examples of mass cycling to HE and 116 to JHS, with a maximum inferred level of 84 and 228 in the 277 college communities (note that high numbers are inferred maximums, and the percent cycling figures are in response to the following question: Approximately what percent of the total student body regularly uses a bicycle to commute to classes at this school during good weather?).

	No. of Examples Reported			
	HE		JHS	
Percent Cycling	Low	High	Low	High
0-4	103	137	132	259
5-9	42	56	57	112
10-19	31	41	60	118
>20	32	43	56	110
Total mass cycling (10 and over)	63	84	116	228

Surprisingly, these examples of mass cycling were spread more or less evenly across the country, with most states having at least one example of each (HE and JHS); most have two examples, and only six or seven have more than two.

Levels of cycling high enough to theoretically bestow net social benefits on society have not yet appeared for the work or shopping commuter in large cities and few examples exist in small cities. A nucleus of visible adult bicycle commuters exists in

Table 1. TE reports on bicycle commuting to work, shopping, and schools in college communities.

	No. of Examples Reported[b]							
	HE		Public Schools		Work Place		Shopping	
Percent Cycling[a]	Low-Side Estimate	High-Side Estimate	Low-Side Estimate	High-Side Estimate	Low-Side Estimate	High-Side Estimate	Low-Side Estimate	High-Side Estimate
1-4	28	48	11	19	25	43	10	17
5-9	18	31	6	10	5	9	3	5
10-19	9	15	5	9	1	2	0	0
≥20	5	9	6	10	0	0	0	0

Notes: Low-side estimates represent actual reported number. High-side estimates indicate inferred maximum number of observations, which were inferred from the respondents to the overall target population, assuming nonrespondents had similar levels of cycling. They probably had lower levels, so high-side estimates represent inflated estimates.
In response to the question, "If no areas of substantial commuter cycling exist in your community, please just write in community name and return questionnaire," the total blank questionnaires were 24 and 41 for the low-side and high-side estimates, respectively.

[a]Responses to the following question, "During peak cycling hours bikes represent about what percent of total vehicles on, or alongside, the road or roads providing access to these commuter bicycle destinations?"
[b]Responses to the following question, "Please list the major commuter bicycle destinations in your community such as schools, work places, or shopping areas."

some large U.S. cities, whereas Europe and Asia have numerous examples of mass commuter cycling to work and shopping.

Determinants of Cycling: Separation from High-Speed, High-Volume Traffic

Numerous national, state, and local surveys show that the overwhelming majority of actual and potential commuter cyclists want separation from high-speed, high-volume traffic (HSHVT) and consider such separation a precondition for bicycle commuting (2,8). Sophisticated logit analyses, which are used in marketing research to go beyond what consumers say they want to actual prediction of their behavior, find that separate bikeways would substantially increase the propensity to cycle (9). Also, several studies of existing bikeway systems provide direct observation of separation diverting and increasing commuter cycling (10-12). Observations in many states and countries failed to provide examples of mass commuter cycling mixing with HSHVT.

Thus, it was hypothesized that separation from HSHVT constitutes an important determinant of mass commuter cycling. Note that separation does not necessarily mean grade or physical separation with a bikeway, raised beam, or even a striped-off lane. Low-speed, low-volume roads can provide the separation, particularly in combination with barrier-breaking bicycle facilities along major arteries.

The data generally support this hypothesis and help refine it. First, in Table 2, the data in column 1 reveal that the higher the traffic speeds and volumes, the fewer the examples of mass cycling. The table records 125 examples of mass bicycle commuting in the United States taking place on separate bike paths or lanes or low-speed, low-volume, nonarterial residential streets (answers a and b in Table 2). Only 48 examples of mass cycling exist where access involves higher-speed, higher-volume residential through streets or wide (including shoulder) moderate-speed, moderate-volume arterials (answers c and d). No HEs and only six JHSs reported mass cycling along narrow high-speed arteries without shoulders or heavily traveled multilane arterials (answer e). On follow-up calls to JHSs, moreover, it was found that students were cycling out of residential areas and crossing busy arteries to reach school rather than cycling along HSHVT arteries.

Furthermore, on average, schools with bikeways have a much higher percentage of students cycling than do schools that rely only on the road system for access. In Table 2, the data in column 2 reveal that, for all road categories except residential, bikeways along the road are associated with more than double the percentage of students cycling. Cycling averages 16 percent with bikeways and only 7 percent without. Without bikeways along high-speed, narrow, or congested multilane arteries, the percentage of students cycling drops to an average of 3 percent. With bikeways, the percentage stays up at mass cycling levels.

Finally, the data further suggest that mass cycling will mix with motor vehicle traffic up to and including wide (including shoulder) moderate-speed, moderate-volume arterials (Table 2, answer d). Although most examples of mass cycling have access on separate facilities, the other examples were spread out evenly over residential to wide, moderate-speed, moderate-volume arteries. The data do not separate the moderate arteries from through residential streets and do not provide a more detailed description of road width and surface or traffic volumes. However, Table 3 provides a list of seven universities that have high levels of mass cycling mixing with moderate to heavy, but slow-moving, motor vehicle traffic. European cities such as Amsterdam or Ursula, Sweden, also provide examples. These communities constitute laboratories for making precise field measurements of road types and traffic volumes to assess the outer limits of mass bicycling mixing with motor vehicle traffic.

The observed relation between separation from HSHVT and level of cycling does not prove that separation causes cycling to increase. First, causation could run either way. The emergence of mass cycling on a road could help motivate officials to construct a separate bicycle facility. Second, the data include a number of schools with separate facilities that provide access, but with less than 10 percent of the student body cycling (i.e., less than mass cycling). Third, Table 3 provides some examples of high levels of cycling mixing with moderate- to high-volume traffic. Finally, Pearson's squared and Kendall's tau rank-order correlation between type of access and percentage cycling explain only about 20 percent of the variation in cycling. Thus, other variables must play an important role in stimulating mass cycling.

Economic Determinants of Mass Cycling

Most studies of passenger transportation modal choice find that relative costs, including the time costs of the modes, play a major role (13). Computer simulations of cycling versus driving find that, although bicycles cost far less than automobiles to own and operate, the generally slower overall travel time of bicycles (cruising speeds as well as preparation time) cannot compete with vehicle savings (14). In these studies, only college students who live within 2 to 3 miles of a campus with limited convenient parking find cycling substantially less expensive than driving (14, p. 597). These findings remain consistent with the observation that few white-collar commuters cycle the relatively long distances from suburban areas to CBDs, but many examples exist of mass bicycle commuting of college students in small university cities. Thus, it is hypothesized that the costs (including time costs) of cycling constitute a major determinant of mass cycling.

The cross-community data, however, give only partial support to this hypothesis. Table 4 shows that the mean percentage cycling to classes during good weather usually associates strongly with relevant proxy variables for low costs of cycling relative to other modes. About twice as many students cycle to classes where cycling appears to provide a quicker, cheaper, or more convenient mode of transportation.

These variables, however, explain only a minor part of the variation of cycling between schools. The percentage of the student body living on or within 2 to 3 miles of campus explains 19 percent of the variation in the HE cycling, and speed of cycling relative to driving explains 7.7 percent (Table 4). The other variables explained only a small percentage of the variations and generally did not reach statistically significant levels. Taken together in stepwise linear regression models, these proxy economic variables explain about 25 percent of the variation in cycling among HEs ($R^2 = 0.266$) but only 6 percent of the variation among JHSs ($R^2 = 0.065$). Again "explain" or "correlate" do not necessarily mean cause. For example, people who do not own cars and who wish to cycle may choose to live close to campus.

These results suggest that either our proxy variables do not capture the real relative costs of cycling or that other variables explain the major part of the variation in the percentage of students cycling to class.

Other Possible Determinants of Mass Bicycle Commuting

The DOT study (1) assumed that bicycle promotion and education, along with safe bicycle parking and minor modifications in the road (wider lanes and safer drainage grates), constitute the major determinants of bicycle commuting. Other observers and practitioners have hypothesized that variables such as culture and weather play a major role. The cross-community data, however, fail to support these hypotheses.

The data suggest that promotion and education do play a role in existing mass bicycle transportation systems, but not the major role. The data in Table 5 reveal that miles of bike paths or lanes, number of bike racks, and dollars spent on bicycle promotion and education all correlate positively with the percentage of students cycling to classes. Promotion does explain more than 13 percent of the variation in the percentage of students cycling at different universities. But other variables explain

Table 2. Number of schools with mass bicycle commuting and average percentage of students cycling to class for all schools by type of bicycle access.

Type of Access[a]		Column 1: No. of Schools with Mass Cycling		Column 2: Avg Percentage of Students Cycling to All Schools[b]	
Answer	Description	HE	JHS	With Bikeways	Without Bikeways
a	Bikeway system with paths or lanes	40	54	-	-
b	Low-speed, low-volume, nonarterial residential streets	9	22	14	9.5
c	Combination of b and d	5	3	24	8.5
d	Higher-speed, higher-volume residential through streets or wide (including shoulder) moderate-speed, moderate-volume arterials	9	31	18	8.5
e	Narrow high-speed arterials without shoulders or heavily traveled, multilane arterials (or any combination that includes such arterials)	0	6	10	3
	Total	63	116	16	7

[a]Description of access type in questionnaire, which asked which description best fits the respondent's situation.
[b]Analysis of variance showed all the difference between percentage cycling with and without bikeways are statistically significant beyond the 0.01 level.

Table 3. Universities with mass cycling by moderate arterial access and no reported bike systems.

HE	Percentage Cycling	Type of Access
University of Wisconsin at Eau Claire	50	Moderate to high-speed, high-volume traffic
University of Kansas, Lawrence	45	Very congested; low speed (less than 20 mph)
University of Southern California, Los Angeles	35	Residential (apartments) to moderate arterials
Indiana University, Bloomington	30	Moderate- to high-volume arteries; lanes on campus
University of Kentucky, Lexington	15-50	Moderate speed, narrow, little room to cycle, but high level of protection on campus
Bowling Green State University, Bowling Green, Ohio	20-25	Residential to moderate arteries
Auburn University, Auburn, Alabama	15	Moderate

Table 4. Percentage of students cycling and percentage variation in cycling explained by proxy variables for relative cost of cycling, including time costs.

	Mean Percentage Cycling to		Percentage of Variation in Cycling that Proxy Variable Explains[b]	
Proxy Variables for Relative Cost[a] (including time cost)	HE	JHS	HE	JHS
"Approximately what percent of the student body lives on campus, or within 2-3 miles of this school?"			19	4.8
<50 percent	7	8		
≥50 percent	20	15		
"Does this school attempt to discourage automobile commuting?"			1[c]	NA
No	7.6	NA		
Yes	13.1	NA		
"What does a yearly student parking permit cost?"			0.06[c]	NA
0-$25	7.5	NA		
$26-$100	14.0	NA		
"Generally, for a student living within 2-3 miles of school, does bicycling to classes take less time than driving, parking, and/or walking?"			7.7	0.0[b]
No	6.6	13.8		
Yes	13.0	13.8		
"Does an adequate bus system provide access to classes?"			0.0[b]	0.5[b]
No	8.1[b]	9.9		
Yes	9.4[b]	17.0		

[a]From questions used in survey.
[b]Pearson's R^2 was used for numerically scaled data, such as percent and dollar cost, and Pearson's R and Kendall's tau for the discrete yes-no data. Although Kendall's tau represents the statistically correct procedure for discrete data, it does not give the percentage of variation explained and it closely approximated Pearson's R. Also, the results were checked by using two statistical packages [Statistical Package for the Social Sciences (SPSS) and Statistical Analysis System (SAS)] and little difference was noted; therefore, an average was presented.
[c]Indicates differences or percentage variation explained that are not statistically significant at the 0.05 level. Rest are significant well beyond the 0.01 level based on analyses of variance.

Table 5. Other determinants of mass cycling: percentage of total variation in cycling between schools explained (R^2).

Determinants of Mass Cycling[a]	Percentage of Variation (in cycling) Explained for HE	JHS
"Miles of paths and stripped-off lanes on school's campus (or for JHS feeding into schools)"	28	6.7
"This school has bicycle racks for approximately ________ bikes"	28	26.5
"This school and/or community spends approximately ________ dollars per year on the following programs[b] (e.g., maps and education) to promote cycling: ________."	13.5	2.2[c]

[a]Questions from mail-back survey.
[b]Answers that obviously included construction of bicycle facilities were removed or reduced to $10 000/year, which probably also includes some construction. Twenty-two HEs reported $2500 or more spent at the school or community level on programs that typically included maps, registration, bicycle clubs, bike week, and bike rodeos. Three had bicycle patrols, five reported education, and several mixed in bike racks, signs, routes, and planning in a way that could not be separated out.
[c]Not statistically significant at the 0.05 level; rest significant at well beyond the 0.01 level.

Table 6. Number of bicycle-related fatalities by location.

	No. of Fatalities for			
	HE		JHS	
Fatality Location[a]	All	With Mass Cycling[b]	All	With Mass Cycling[b]
On campus bike paths or lanes	2	2	5	0
On bike paths or lanes that provide access to the school	3	2	4	1
On campus streets	11	1	5	0
On streets that provide access to the school	44	15	28	10
On the general road system in the community	137	36	68	12

[a]Response to the following question, "How many bicycle-related fatalities can you recall in the last 5 years or so?"
[b]Includes separate bike system.

more. Miles of bike paths and lanes and number of bike racks explain 28 percent of the variation (Table 5), and the percentage of students who live near the campus explains 20 percent (Table 4). The direction of causation, moreover, remains unclear because increased cycling may motivate officials to spend more on safety programs, bike paths, and bike racks. Also, most institutions with mass cycling do not report any money spent on promotion and education.

Second, the data provide no evidence that culture constitutes a major determinant on bicycle commuting. Mass cycling is spread evenly over most of the country, with two or three examples from most states. Those states that did report substantially more examples of HE and JHS mass cycling represent diverse geographical regions: California, 50; Illinois, 15; Florida, 11; Wisconsin, 10; and Oregon, 7. These five states may share a similar high income, high education, and modern culture, but then why is mass cycling in the industrialized northeast, such as New York State, not found? Further, the existence of mass cycling among the university students in a community does not necessarily create a social climate for JHS students to cycle. Only about 20 percent of the JHSs have mass cycling in those communities where HEs have mass cycling, even when controlling for safe access.

Finally, the data provide no evidence that weather explains the difference in mass cycling. The questionnaire asked about the percentage of students who cycled during good weather. Some of the highest levels of cycling exist in northern schools, such as the University of Wisconsin at Madison and at Eau Claire. Moreover, some of the states that reported the most examples of mass bicycle commuting are located in the northern parts of the country that have severe winters: Wisconsin and Illinois. At the other extreme, Florida represents a climate that is too hot for commuter cycling.

Mass Cycling and Safety

The fear that mass cycling will lead to higher traffic fatality rates has focused attention on mass cycling and safety. Estimates from England put the per mile risk of a fatal accident on a bike at 10 times greater than in a car [see Everett (15) for citations to the safety literature]. The Dutch estimated a 3.5 times higher risk for cyclists (15). Many planners assume that separation, particularly with paths and lanes, will reduce this risk. However, some cyclists and planners who oppose bikeways theorize that bike paths and lanes only protect against the overtaking accident but expose cyclists to awkward positions at intersections, where most accidents occur (16).

No support for this latter position in the replicable, empirical bicycle literature could be found. There was a survey of bike club members who reported more accidents on bikeways than in the road (17). This and other studies indicate that bikeway accidents can cause serious injury (18). However, Wheatley and Cross, in their rigorous and well-funded nationwide study of bicycle fatalities (19), found that the largest group of fatal accidents (more than 37 percent of the total) entailed motor vehicles overtaking bicyclists. By definition, a separate bikeway should substantially reduce that type of fatal accident. Reports on studies in Europe (20) indicate that separate facilities reduced most types of intersection fatalities and overall fatalities.

The cross-community data also fail to support the notion that separate bicycle facilities increase the overall risk of fatal accidents. The data in Table 6 reveal that key informants recalled only 14 bicycle fatalities on separate paths or lanes for all 524 reporting schools. The informants reported that the overwhelming majority of the 307 fatalities occurred on the general road system. This, of course, may have resulted from more cycling taking place in the roads than on paths. Therefore, the number of fatalities for roads and facilities only in those schools with mass cycling and separate bike systems was calculated. It was assumed that most of the commuting to these schools takes place on the bikeways. Here the overwhelming majority of fatalities still occurred on the roads (Table 6). Moreover, the fatalities on the general road system apparently occurred on arteries or collector streets. None was reported on noncollector residential streets.

But the data, which are based on key informants' memory or record checks, remain crude. A number of respondents failed to specify the type of street where the accident occurred. For example, informants reported 17 fatalities on campus streets, but some of these included high-speed, high-volume arteries through the campus. Thus, much work remains before understanding the determinants of safe mass bicycle commuting.

DISCUSSION AND CONCLUSIONS

The data on bicycle commuting around schools across the United States tend to support the researchers' observations and hypotheses and the replicable, empirical literature. Few or no examples of mass bicycle commuting to work or shopping anywhere in the United States were found. The overwhelming majority of schools with mass bicycle commuting (10 percent or more of the students cycling to class regularly during good weather) have bicycle access separated from HSHVT. Note that separation does not necessarily mean a separate bicycle facility. Although most schools with mass cycling did have separate facilities, many relied on low-speed, low-volume, residential-type roads and 20 or so may have relied on moderate-speed, moderate-volume arteries. The bicycle also tended to provide the quickest and least expensive mode for students at schools that had mass cycling. The overwhelming majority of reported fatalities apparently took place on the arterial road system rather than on bikeways or residential streets, even when attempting to control for miles cycled.

No reasonable evidence was found to support the DOT study that hypothesized that promotion and education with minor road modifications would shift 15 to 30 percent of short-distance urban automobile drivers to bicycles for journey-to-work and shopping trips. First, only a few examples were found of mass cycling mixing with the kind of high-speed or high-volume traffic many drivers must use to reach urban work and shopping centers. Second, although dollars spent on promotion did correlate with mass cycling in this study, only a few schools reported such expenditures, and causation could run either way. Thus, no evidence currently exists that promotion or education played a major role in stimulating existing mass cycling.

This, of course, does not mean that aggressive, well-funded promotion and education along with minor road modifications could not generate mass cycling in urban areas. Theoretically, they could play an important role by making potential cyclists aware of favorable conditions, although education and promotion that point out the probabilistic hazards of cycling might substantially discourage the mode. Currently available data suggest that bicycle commuting, even with extensive education, traffic law enforcement, and separate bicycle facilities, remains much riskier than driving per mile. For example, the Dutch, who have instituted all of these bicycle program inputs, estimated the risk of a fatal accident on a bicycle at 3.5 times greater than in a car per mile traveled (15).

The preponderance of evidence suggests that bicycle planners who want to generate mass cycling generally will have to find ways of separating cyclists from HSHVT. In addition to the data, study after study (2), including sophisticated logit models (9), find that the overwhelming majority of actual and potential cyclists want separation and that separation can increase cycling substantially in certain situations. Observations of conditions under which mass cycling to work and shopping takes place in European cities suggest the same. Again, separation may involve use of existing low-speed, low-volume roads; widening of lanes and roads; or building separate bicycle facilities--a combination of all of these approaches would likely be involved.

This study does not present a tight predictive model for precise planning guidelines. First, the social costs and benefits of mass cycling are not addressed. Only the determinants of mass cycling are considered [for literature on the cost and benefits, see Everett (2)]. Second, although a correlation between the percentage cycling and inputs (such as separation, relative costs of mode, bike racks, and promotion) was found, correlation does not mean causation. Moreover, enough of the variation in cycling can be explained to build a model that would predict the impact of a change in one policy variable on percentage cycling. In some places, separation might have a strong impact; in another, changes in the relative cost of cycling might provide the greatest increase of cycling; and in yet another, promotion might be the most effective. Finally, the road, traffic, relative cost, and other conditions under which mass cycling takes place could not be precisely measured and defined.

Nevertheless, the study does isolate a number of communities for developing more precise measurements and guidelines on mass cycling. For example, on-site studies that measure the exact road types and traffic volumes in those communities where mass cycling mixes with moderate-speed, moderate-volume to congested low-speed arteries could indicate the limits of such mixing. This would provide much sounder guidelines for when to use minor road modifications or build separate bikeways than the current speculations.

Also, detailed on-site studies in these communities on safety and other determinants of cycling (such as time costs, promotion, and education) probably could yield valuable insights. The methodology description and questions in the tables provide a basis for replicating and extending the current study. (The actual questionnaires and data sets may be obtained from the authors at cost.) Although these detailed studies would require on-site data collection and cost more than the mail-back survey, they should cost less than recent government reports [such as the DOT reports (1,16)].

Discussion

Steven Faust*

Everett and Spencer state that they are attempting to identify the determinants of mass bicycle commuting in the United States. In their paper they

1. Introduce and define mass bicycle commuting,
2. Define and evaluate substantially separated bicycle facilities,
3. Determine the volume of cycling at a number of HEs and JHSs,
4. Determine modal choice and accident rates based on their data, and
5. Compare this work with the findings of the 1980 DOT study (1).

DOT REPORT

To begin with the last point, the authors have misstated both the intent and findings of the DOT bicycle energy conservation report. The mandate of this report was to develop an implementable program to conserve energy by reducing the share of trips taken by automobile in favor of the bicycle. DOT's findings support expenditures for both fixed-facility improvements as well as for education and promotion as part of a comprehensive regional transportation program. The DOT report is faulted for failing to address issues that were in fact covered, or issues, such as major capital investments, that were beyond the original mandate.

*Urban Mass Transportation Administration, Region 2, 26 Federal Plaza, Room 14-130, New York, NY 10007

MASS BICYCLE COMMUTING

The authors introduce a new concept to transportation planning: mass bicycle commuting. This term is defined as 10 percent or more of trips, and again later as 5 percent of vehicles, with no further explanation as to why these arbitrary figures are useful or meaningful. However, the authors also imply that this mass level of cycling is the trigger point for bestowing net social benefits on society. This, of course, presupposes that a valid cost/benefit analysis could be performed for the entire transportation system, including the bicycle mode.

Further, disaggregate data on the volume of bicycle use for all purposes are both limited and unreliable. Traffic counts omit bicycle traffic unless well-trained personnel directly observe the roadways. This is confirmed by work in such disparate environments as Boston, New York City, and Eugene, Oregon. One must note that heavily supported public mass transit ridership in cities of 300 000 population rarely reaches the 10 percent level, even for rush-hour work trips.

The authors further confuse their definition by using the term 10 percent of all traffic, without controlling for long-distance through traffic. More than 20 percent of all motor traffic in lower Manhattan's CBD is bridge traffic that connects Long Island with New Jersey. The DOT energy report focused on affecting only a portion of locally oriented traffic. No source for the authors' statement that the DOT report claims a 15 to 30 percent shift from driving to bicycling could be found.

The authors have correctly identified a need for better bicycle volume and origin and destination data. Unfortunately, the introduction of a new term--mass bicycle commuting--does not add to that data or to the understanding of events.

SEPARATED FACILITIES

The major premise of the paper revolves about the value of substantially separated bicycle facilities as the key determinant for the increase in bicycle use, including grade separation, physical separation with a bikeway, raised beam, a striped-off lane, and even low-speed, low-volume roads. These all met the authors' criteria for substantial separation. This list is so all-encompassing as to be practically meaningless for effective cross-community evaluation.

The use of a totally ambiguous definition of separate facilities results in a flaw that invalidates the analysis of reported data. Without a consistent and clear definition of right-of-way conditions, there can be no comparison of the various data collected or of the reports in the literature. Without uniform criteria, one traffic engineer's designated wide curb or bicycle lane is another's high-speed, high-volume roadway that is unfit for nonmotorized traffic. Even if the authors' generalization "that high-speed, high-vehicle traffic constitutes a serious barrier to mass cycling" were to be accepted, one cannot identify that condition or its absence from this study's criteria. The authors themselves confuse the use of separate bikeways along existing major arteries with special barrier-breaking facilities that provide totally new direct access where none existed before.

The paper cites the four Williamette River bridges in Eugene, Oregon (12) for increasing commuter cycling. Three of these bridges create entirely new gateways that cross a barrier that was otherwise at least 2 miles apart by any other route. Combined with the bridges is a riverfront path system, which is also a barrier edge route. These are site-specific, capital-intensive projects that have as much regional recreation benefits as transportation benefits. The Williamette River Greenway is far more an example of Olmstead's original linear park-parkway concept coordinated with short segments of barrier-breaking right-of-way.

Eugene is also an example of where citizen interest in cycling created a community organization that worked for more than 10 years to see these improvements put into place. Clearly, the cycling attitudes came before the cycling infrastructure.

Current bicycle design practice has attempted to move beyond simplistic rigid definitions of three classes of bikeways. The 1981 AASHTO bicycle design guidelines present a more functionally oriented approach to providing both dedicated and shared rights-of-way for bicycle travel.

ACCIDENT ANALYSIS

Bicycle accident analysis is seriously complicated by the authors' ambiguous definition of a bicycle facility. With limited exceptions, designated urban area bicycle routes either share streets with motor vehicles or with on-grade cross streets at frequent intervals. Due to limitations in police and motor vehicle department data-collection methods, virtually all accidents are reported as located on the motor vehicle roadway. Furthermore, police traffic data rarely include accident or fatality information for nonmotorized vehicle incidents. The result is that all formal accident reports will systematically underreport bicycle path involvement in bicycles-to-automobiles, as well as bike-to-bike, bike-to-pedestrian, bike-to-animal, or solo bicycle incidents.

Furthermore, the authors rely on the highly subjective memory of their respondents to document accidents. Nowhere was there discussion of whether a given accident occurred to a nonstudent such as a child, or whether the bike trip was in any way related to work or school commuting. At no point does the paper present reliable data for the volume of cycling compared to accidents at given points necessary to develop an accurate accident rate.

The authors cite European data and an Institute of Transportation and Traffic Engineering report, both a decade out of date, as solid and replicable bicycle literature. Neither European cycling nor motoring conditions are reliably transferable to U.S. urban areas.

DATA COLLECTION AND ANALYSIS

Data collection and analysis will get limited review here because, first, it is complex and detailed, and second, because both the bikeway and accident conditions are flawed; therefore, most of the conclusions are invalid.

Setting up the questionnaire to be answered by a single person opens the results to substantial uncontrollable variation. The questions themselves appear highly subjective because they focus on the respondents' opinions and memory of events.

In brief, the use of a two-page questionnaire to document detailed variables of conditions, as well as the student bodies' sociodemographic background, would appear to require some simplistic questions.

PARKING AND NON-RIGHT-OF-WAY INFRASTRUCTURE

The authors generalize from the literature that traffic conditions are a serious barrier to mass cycling. Two studies in the New York area find that safe bicycle parking is the limiting factor by more than half of the respondents, whereas traffic and roadway conditions are far less serious. In two different situations--a midtown Manhattan commuter bicycle study and a study at New Jersey commuter rail stations--commuters required safe and secure parking for any commuter cyclist. Note that secure parking was considered (i.e., lockers, not racks) unless full-time security was provided.

The authors repeatedly ignored all nonroadway facilities required to support cycle commuting. This is the same as encouraging automobile commuter park-and-ride programs by building the feeder highways and leaving out the parking lots. Commuters must expect their vehicle to be intact at the end of the day. The issue of bicycle access to commuter bus and rail park-and-ride stops was never raised in this paper. There are already substantial examples in Connecticut; New Jersey; Washington, D.C.; and the San Francisco Bay area of a shift to cycle access to transit when secure parking is provided.

CONCLUSIONS

The authors have repeatedly stated that separated bicycle facilities are the key determinant to generate a condition called mass cycling. Unfortunately, their research was not supported by real-world facts. The study has no reliable control for local bicycle volumes, a reporting bias toward roadways, a simplistic evaluation of campus transportation alternatives, and a preconceived hypothesis that a moderate-cost engineering, education, enforcement, and encouragement (4E) program would be counterproductive.

Yet the authors conclude that they could only find a correlation, but not a causal direction, between a number of relevant variables and percentage cycling. Moreover, their findings "cannot explain enough of the variation in cycling to build a model that would predict the impact of a change in one policy variable on the percentage cycling." This does not appear to support their sustained attack on the DOT report and its authors.

Although Everett and Spencer have found the DOT proposals unsatisfactory, what alternative program have they put forth? Do they propose a massive investment in a network of separate bicycle facilities, or do they propose that all encouragement of cycling be deferred until such a comprehensive system is in place? Their study fails to show how such a program can be financed, built, or maintained under current economic realities when the U.S. urban infrastructure has fallen into a state of total disrepair.

As noted before, the DOT mandate (1) was to develop an implementable and cost-effective program. To this end, Everett and Spencer's paper does not refute the DOT proposals, provide a viable alternative, or appreciably add to the body of bicycle planning knowledge.

Authors' Closure

Faust's comments excellently illustrate the vigorous and often emotional controversy that surrounds the role of separate bikeways in bicycle transportation systems. Commentators from the TRB Committee on Bicycling and Bicycle Facilities also made similar sweeping rejections of the study. Indeed, a sensitive nerve has been touched.

Planners who attempt to formulate rational, utility-optimizing transportation systems need to understand this controversy to avoid biases and distortions in the bicycle literature. It is believed that one basis of the controversy stems from special-interest group conflict. Historically, a relatively small group of cyclists often associated with bike clubs in the United States and England have vigorously opposed separate bikeways (2, p. 39). These cyclists fear law or custom would force them to use bikeways, which they consider generally slower and inferior to roadways. Their more extreme positions argue that bikeways would discourage cycling and make it more dangerous.

On the other hand, survey after survey shows that the overwhelming majority of actual and potential commuter cyclists want separation from HSHVT (a list and summary of surveys are available from the authors). This appears to imply that simply building bikeways would generate substantial safe bicycle travel. Although replicable, empirical studies suggest that a number of inputs, ranging from facilities to education, could play a role in generating increased cycling, the controversy continues to intensify.

It was hoped that the cross-community analysis of where mass cycling occurs would help end the more extreme arguments in the controversy and focus research and analysis on narrower issues such as the limits of mass bicycle and motor vehicle mixing. The vigorous opposition of the discussant, however, forces us to reconsider our work. Does it simply represent an attempt to rationalize our previously held working hypotheses? Or does it represent a reasonably sound attempt to observe systematically where mass cycling takes place and the correlates of that mass cycling?

After double-checking the data again, it was still found that mass cycling generally takes place where low-speed, low-volume residential streets or bikeways separate cyclists from high-speed, high-volume motor vehicle traffic. Continued data analysis and follow-up interviews have reduced the number of reported observations of mass cycling mixing with moderate traffic to HSHVT. Thus, the data strongly support our hypothesis, and our critics should replicate these studies if they do not have confidence in the data. However, the following analysis of each major criticism shows that a proper interpretation of the tables and text should remove most of their objections to the data.

INTERPRETATION OF DOT REPORT

Faust believes that the DOT report (1) "supports expenditures for fixed facilities." However, we strongly disagree and believe that the DOT and workshop reports (1,16) basically represent an attempt to propagate the positions of the antibikeway movement. For example, the DOT report (1, p. 33) repeats, with no support, the old argument that bikeways only help novice and recreational cyclists and do not protect cyclists at the intersections, where most accidents occur. The DOT report also recommends that the government publish a guide for state and local facilities that "would highlight the desirability of making minor modifications to the existing street system as a top priority with the construction of special bicycle facilities viewed only as a last resort" (1, p. 99). Finally, the DOT report based its conclusions on serious misinterpretations of two contracted studies [see Everett (2, p. 38) for support to this statement].

MASS CYCLING

The discussant criticizes the use of the mass cycling concept. A proper interpretation of the tables should overcome or reduce this objection. Mass cycling is defined as 10 percent or more of trips for schools, and the in-text table on bicycle commuting to HEs and JHSs refines that definition to 10 percent or more of students cycling to class during good weather. For work and shopping trips, 5 percent of vehicles along the road is used to adjust for the longer distance and through-the-city commuter (Table 1).

We do not accept the implication that one cannot generate and use bicycle volume data. First, several studies in the United States (6,7) and abroad (10) have reported such data; censuses (5) have collected the data; and we have personally made bicycle counts. Second, although the exact threshold to mass cycling cannot be agreed on (i.e., cycling that bestows net social benefits), most researchers can agree that massive cycling in college communities like Davis and Madison have substantially different impacts than the trickle of cyclists along roads in Chicago or Washington. Benefit-cost studies indicate that mass cycling generates large net social benefits, whereas a small group of cyclists may impose net social costs. Although the 10-percent-of-trips threshold remains somewhat arbitrary, changing the definitions to 5 or 15 percent of trips makes little difference in the statistical results and conclusions.

ACCURACY OF REPORTS ON SEPARATE FACILITIES

Faust's major criticism of the study involves the possible inability of respondents to distinguish consistently between the various types of bicycle access listed in the questionnaire. Although some inconsistency in categorizing bicycle access undoubtedly occurred, it could not invalidate the entire study. The range of possible accesses are quite wide--from separate paths and lanes to narrow high-speed arteries (see Table 2 for categories of access). The questionnaire explained formal bikeway systems as having separate paths or striped-off lanes. On field testing, the questionnaire respondents correctly categorized bicycle access. A large number of observations were made to help smooth out possible errors. Obviously deviant cases were double-checked with follow-up telephone calls, and the data results generally coincided with the replicable, empirical literature.

From a planning standpoint, formal bikeway systems (paths and lanes) and low-volume, nonarterial residential streets clearly characterize most mass cycling systems, whereas high-speed, high-volume arteries carry virtually no mass cycling. The middle category--moderate-speed, moderate-volume arteries--however, does create a problem. This was pointed out, 7 locations for on-site study (see Table 4) were isolated, and at least 20 others can also be shown. Analyzing this subset involves feasible, cost-effective, on-site research. Detailed questionnaires that ask for voluntary measurements undoubtedly would have suffered from low response rates.

ACCIDENT DATA

The discussant finds the accident data weak. We agree and pointed out the weaknesses. However, it was believed that the data would provide some valid insights to planners who attempt to assess controversies over bicycle safety. First, the data coincide with the informal field interviews where we could probe for bikeway relatedness. Second, the respondents overwhelmingly report fatalities that occur in the roadway, so that even considerable failure to report bikeway relatedness could still lead to the same basic conclusions. Third, no other cross-community data on bikeway versus road fatalities exist. Finally, the data coincide with other replicable, empirical studies.

The well-funded and rigorous Cross study (19), for example, found that the overtaking accident constituted the major category of bicycle fatalities (more than 37 percent). Bikeways should substantially reduce this type of fatal accident. The European studies, which find bikeways reducing intersection fatalities, remains less verifiable. Thus, only the available studies were stated in the paper, and the data failed to support the notion that bikeways increase the risk of fatal accidents. It is believed that the government reports should have looked more objectively at all the data and drawn similarly circumscribed conclusions.

BICYCLE PARKING

The discussant states that we "repeatedly ignored all nonroadway facilities required to support cycle commuting." But Table 5 clearly includes bicycle racks and promotional and educational programs and the text discusses these in the section on Other Possible Determinants to Mass Bicycle Commuting. Bike racks did correlate well with the percentage of students cycling to class, but causation obviously could run both ways. It was accepted as a reasonable proposition that, in some areas, safe bicycle parking would constitute a major determinant of mass cycling.

Bicycling interacting with mass transit and park-and-ride was not explicitly mentioned because of space and data limitations. However, it is believed that bicycles theoretically could play an important role in such systems if perceived safe bicycle access and secure parking exist. In essence, such systems could provide the short distances in congested areas where bicycles provide faster and more flexible transportation.

CONCLUSIONS

Basically, Faust takes us to task for emphasizing separate bicycle facilities as the key determinant of mass cycling and for rejecting DOT's moderate-cost 4E program.

Again, a proper interpretation of the study should reduce this criticism. Substantial evidence was found to indicate that separation from HSHVT with residential roads and bikeways correlates strongly with mass cycling. However, a number of communities were isolated where mass cycling appears to mix with moderate-speed, moderate-volume traffic and at times heavy-volume traffic. Relative cost, including time, was considered as important as separation, and considerable space was devoted to analyze cost. Evidence does suggest that education and promotion may play a role, particularly in safety, but no evidence that they play a major role could be found.

Proposing a comprehensive bicycle program is outside the scope of this paper. We believe the available evidence does not allow us to predict the impact on any set of variables with any degree of confidence. Given this uncertainty, we believe prudent bicycle planning would involve a reasonable balance of all inputs, including separation and education. Nevertheless, planners can no doubt generalize mass cycling from campuses to urban commuting without radically changing the relative costs of cycling and perceived safety by separating cyclists from HSHVT and probably from most moderate-speed, moderate-volume traffic. But there is the fear that, even with extensive education and traffic law enforcement, commuters who shift to bicycles will face substantially higher risks. To further test these hypotheses, a more detailed analysis of the limits to mass cycling and motor vehicle mixing in the communities that have been isolated is recommended.

REFERENCES

1. K. Moran; Mountain Bicyclists' Association, Inc. Bicycle Transportation for Energy Conservation. U.S. Department of Transportation, April 1980.
2. M.D. Everett. Marketing Bicycle Transportation: A Critique of the National Comprehensive Bicycle Transportation Program. TRB, Transportation Research Record 851, 1982, pp. 37-40.
3. The College Blue Book: Tabular Data. MacMillan, New York, 1981.
4. Patterson's American Education. Education Directories, Inc., Mount Prospect, IL, 1980.
5. Selected Characteristics of Travel to Work in 21 Metropolitan Areas, 1977. Bureau of the Census, U.S. Department of Commerce, Current Population Reports, Series P-23, No. 105, Jan. 1981.
6. De Leuw, Cather and Company. Bikeway Design Evaluation. District of Columbia Department of Transportation, Washington, DC, 1979.
7. C.A. Buckley. Bicycle Traffic Volumes. TRB, Transportation Research Record 847, 1982, pp. 93-102.
8. F.O. Robinson, J.L. Edwards, and C.E. Ohrn. Strategies for Increasing Levels of Walking and Bicycling for Utilitarian Purposes. TRB, Transportation Research Record 743, 1980, pp. 38-48.
9. W. Hyman and others. Development of Wisconsin Urban Work Trip Models for Forecasting Modal Choice. Wisconsin Department of Transportation, Madison, 1982.
10. Cycling as a Mode of Transportation. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1980.
11. D.F. Lott, T.J. Tardiff, and D.Y. Lott. Evaluation by Experienced Riders of a New Bicycle Lane in an Established Bikeway System. TRB, Transportation Research Record 683, 1978, pp. 40-46.
12. S.G. Lipton. Evaluation of the Eugene, Oregon, Greenway Bicycle Bridge. TRB, Transportation Research Record 739, 1979, pp. 29-37.
13. Transportation Choices for Urban Passengers: Measures and Models. Organization for Economic Cooperation and Development, Paris, France, 1980.
14. M. Everett. Commuter Demand for Bicycle Transportation. Traffic Quarterly, Vol. 28, No. 4, Oct. 1974, pp. 585-602.
15. M. Everett. Bicycle Safety and Ethics. Presented to TRB Committee on Bicycling and Bicycle Facilities, Summer Meeting, San Diego, 1979.
16. V.S. Darago. Regional Workshops on Bicycle Safety, Final Report. NHTSA, U.S. Department of Transportation, Rept. DOT HS-803 658, Sept. 1978.
17. J. Kaplan. Characteristics of the Regular Adult Bicycle User. FHWA, Rept. HNG-25, July 2, 1976.
18. P.L. Wheatley and K.D. Cross. Causal Factors of Non-Motor-Vehicle-Related Bicycle Accidents. TRB, Transportation Research Record 743, 1980, pp. 20-30.
19. K. Cross; Anacapa Sciences, Inc. Bicycle Safety Education: Facts and Issues. American Automobile Association Foundation for Traffic Safety, Falls Church, VA, 1978.
20. Bikeway Planning Criteria and Guidelines. Institute of Transportation and Traffic Engineering, Univ. of California, Los Angeles, 1972.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Statistical Cost Analysis of the Regulated Household-Goods Trucking Industry

WILKIE W. CHAFFIN AND WAYNE K. TALLEY

An investigation of whether the household-goods (HG) trucking industry, which is regulated by the Interstate Commerce Commission, will become concentrated (i.e., fewer HG truck carriers controlling a larger percentage of the industry's market) during the current deregulatory environment is presented. The likelihood of concentration is investigated by alternatively investigating the existence of economies of scale in the industry. It is concluded that the HG trucking industry exhibits economies of scale and therefore will likely become concentrated during the current deregulatory environment.

In July 1980 President Carter signed into law the Motor Carrier Act of 1980. This Act provided for deregulation of the Interstate Commerce Commission (ICC) regulated trucking industry. For example, the Act increased opportunities for new carriers to enter the trucking industry, established a zone of rate freedom, and expanded the number of commodities to be exempt from ICC regulation. One type of ICC truck carrier that was excluded from the Act was the household-goods (HG) carrier. Given the unique nature of HG carriers, regulatory reform for these carriers was considered by Congress apart from the Motor Carrier Act of 1980. In fall 1980 the Household Goods Transportation Act of 1980 was passed by Congress. This Act reduced unnecessary government regulation of HG truck carriers and furnished additional pricing options for the carriers and their customers.

An investigation of whether the deregulated HG trucking industry will become concentrated (i.e., fewer HG truck carriers controlling a larger percentage of the industry's market) during the current deregulatory environment is presented. The likelihood of concentration occurring in a deregulated industry has traditionally been investigated by alternatively investigating whether there exist economies of scale in the industry; this is the approach adopted in this paper. Economies of scale refer to a less-than-proportional increase in cost when all inputs are increased equiproportionally. The likelihood of concentration occurring in the HG trucking industry is an important issue because the occurrence of concentration will be contrary to an objective of regulatory reform, i.e., to promote a competitive HG trucking industry.

Although numerous studies have investigated the existence of economies of scale for general-freight trucking firms or a combination of general-freight and HG trucking firms, no study (to our knowledge) has investigated separately the existence of economies of scale for HG carriers. The general conclusion of previous studies has been mixed. By using statistical cost analyses, Nelson (1) and Roberts (2), in separate studies, conclude that economies of scale do not exist in the U.S. trucking industry. By using a statistical production-function approach, Ladenson and Stoga (3) conclude that economies of scale do exist. This conclusion is also supported by Dicer (4), Johnson (5), and Rakowski (6). However, Spady and Friedlaender (7) conclude that economies of scale disappear when shipment characteristics such as lengths of haul and types of loads are taken into account.

Although HG carriers share many characteristics with general-freight carriers, the peculiar nature of the demand facing HG carriers has made their operations distinct from those of general-freight carriers. The origins and destinations for HG shippers are geographically dispersed. With shipper demands being nonrepetitive in nature, HG carriers are also prevented from providing scheduled service over regular routes. By comparison, general-freight carriers transport freight between a limited number of origin and destination points on a regular basis.

Because of the irregular, nonrepetitive nature of demand for HG carriers, the probability of an empty backhaul is great. As a result, nationwide and large regional HG carriers have established solicitation agents in local communities to serve geographically scattered shippers in order to minimize empty backhauls. Also, the carrier's fleet of vans is used to provide irregular route, nonscheduled moving service throughout a territory without respect to a home base of operations. The routes taken by moving vans are determined by a central dispatcher who attempts to match shipments booked by local agents with the available capacity of vans. Alternatively, local carriers who have no representation at potential destination points are thus limited to shorter-haul outbound shipments that can be handled profitably under backhaul conditions.

Due to the distinctive nature of HG carriers, it is therefore reasonable to investigate separately the existence of economies of scale for HG carriers and that of general-freight carriers. An investigation of the existence of economies of scale for HG carriers by means of a statistical cost analysis is conducted. In addition, cost elasticity estimates for various characteristics (such as weight and length of haul) of HG shipments are obtained. Furthermore, the results are analyzed and compared with that of previous research.

This investigation is conducted as follows. First, the specification of the cost function to be estimated is developed. Then the statistical cost results, along with a comparison of previous research, are presented and analyzed. Finally, conclusions are presented.

SPECIFICATION OF COST FUNCTION

In return for its operating certificate, an HG carrier is obligated to carry forthcoming traffic at established ICC rates (8). With HG carriers being under legal and economic pressure to abide by this obligation, the level of output produced by an HG carrier, at least in principle, is taken out of the control of the firm and placed in the hands of its customers. Thus, profit-maximizing HG carriers seek a combination of inputs that minimize the cost of transporting an exogenously determined volume of freight.

Assume that the cost (C) to be incurred for inputs (X_1, X_2, ..., X_n) by a given HG carrier may be expressed as

$$C = P_1X_1 + P_2X_2 + \ldots + P_nX_n \quad (1)$$

where P_i is the price of the ith input ($i = 1, 2, \ldots, n$). Further assume that the above inputs can be combined efficiently to transport Q volume of freight, or

$$Q = f(X_1, X_2, \ldots, X_n) \quad (2)$$

Thus, from the above discussion, a profit-maximizing HG carrier will seek those amounts of inputs that will minimize cost in Equation 1 that are subject to an exogenously determined volume of freight Q. In solving such a problem, a cost function that represents the minimum cost to be incurred in transporting Q volume of freight can be derived; i.e.,

$$C = C(P_1, P_2, \ldots, P_n, Q) \quad (3)$$

In attempting to estimate the parameters of Equation 3, a problem arises, as it does in all empirical studies in transportation: how to measure output. The measurement used most often for freight output is the ton-mile. This measurement, however, has been criticized, because it considers a shipment of 1 ton transported 1000 miles as being equivalent to a shipment of 1000 tons transported 1 mile. These shipments are not equivalent because "a firm with heavy loads and long hauls is able to produce a ton-mile more cheaply than its light-load, short-haul counterpart" (9, p. 58).

Warner (10, p. 15) states: "It is clear that if all shipments were alike, there would be no difficulty in the choice of an output variable. The variable, number of shipments, would itself be a natural measure of output. A firm whose shipments were twice those of another would clearly have twice the output of the other." However, shipments differ due to weight, length of trip, time in transit, pickup time, delivery time, and so on. If Q in Equation 3 were defined as shipments, and if these shipments differ according to the above characteristics, then the cost equation (Equation 3) for an HG carrier may be rewritten as

$$C = g(P_1, P_2, \ldots, P_n, Q, S_1, S_2, \ldots, S_m) \quad (4)$$

where S_k is the kth characteristic of a given shipment ($k = 1, 2, \ldots, m$).

If shipments are used as a measure of output, then ideally all distinguishing characteristics of nonhomogenous shipments (or the S_k's) should be considered in the estimation of a HG carrier's costs. Although such data are not ordinarily available, some aggregate measures are available that partly reflect the composition of shipments transported by HG carriers. Following Warner (10), we shall consider the following aggregate characteristics: average weight per shipment and average length of haul per ton. Assuming further that HG carriers pay the same prices for given inputs, the general stochastic version of Equation 4 that will be estimated by using HG carrier data thus becomes

$$C_j = h(Q_j, W_j, H_j, \epsilon_j) \quad (5)$$

where

C_j = cost incurred by the jth HG carrier in transporting Q_j shipments,
Q_j = number of shipments transported by the jth HG carrier,
W_j = average weight (total tons/total number of shipments) per shipment transported by the jth HG carrier,
H_j = average length of haul (total ton-miles/total tons) per ton transported by the jth HG carrier, and
ϵ_j = stochastic error term for the jth HG carrier.

EMPIRICAL RESULTS

In order to investigate the existence of economies of scale for HG carriers, Equation 5 was estimated by assuming linear and logarithmic functions. Because the statistical fit for the logarithmic cost function was superior to that of the linear function, only the results of the logarithmic estimation will be reported here. Although it would be desirable to estimate a translog cost function so as to take advantage of all the relevant information it offers, the available data base does not permit this degree of cost disaggregation. The translog cost function would require a better data base, one that had expenditure information on each factor input in the production process. Estimation of a translog cost function, for example, appears in Spady and Friedlaender (7).

The data used in the estimations were based on a 1975 cross-sectional sample of 32 HG carriers and were taken from the Trinc's Bluebook (11). The average number of shipments per carrier (in the sample) was 24 000 shipments with an average weight of slightly more than 4 tons/shipment.

The logarithmic formulation of Equation 5 that was estimated is

$$C_j = Q_j^{\beta_1} W_j^{\beta_2} H_j^{\beta_3} D_j^{\beta_4} e^{\epsilon_j} \quad (6)$$

where D_j is a dummy variable and e is the base of natural logarithms.

In an analysis of a linear cost function, a constant term (α) would be included, because the presence of economies of scale could be inferred by an estimate of α that is significantly greater than zero. However, in the logarithmic function analysis, the presence of economies of scale is inferred by the estimates of the β_1 coefficients being significantly less than 1. Thus, the inclusion of a constant term is not warranted in terms of detecting economies of scale. The α value, if included, would reflect the influence of all omitted factors on cost during the period of study. It is believed that all costs are variable in the true model. If there are variable costs that have not been included in this model, then the effect of these costs would still be reflected by the dummy variable coefficient (β_4). In addition, an estimated constant term [as concluded by Warner (10)] would be biased upward. Because any information reflected by the constant term would be of secondary interest and would be suspect because of estimation bias, no constant term is included in Equation 6.

The dummy variable is included in Equation 6 as a proxy for those characteristics of shipments not otherwise considered. It is assumed that the characteristics for class 1 HG carriers are distinguishable from those of class 2 carriers. Hence, we assign a 1 to the dummy variable of a class 1 carrier and a 0 for a class 2 carrier.

The parameters β_1, β_2, and β_3 in Equation 6 can be interpreted as cost elasticity coefficients; i.e., they represent the percentage change in cost with respect to a percentage change in the corresponding explanatory variable. Parameter β_1 is of particular interest to this study, because if its value is less than 1 (but positive), it can be concluded that economies of scale exist among HG carriers. This conclusion follows because a given percentage change in shipments will result in a lesser percentage change in costs if β_1 is positive as well as less than 1.

In Table 1, estimates of the parameters of Equation 6 are presented. Estimates were found by using total cost as the dependent variable as well as various components of total cost. By using various cost components, Equation 6 was estimated to investigate the impact of the explanatory variables on these costs.

HG carriers' total costs are broken down into administrative salaries and wages, general operating costs, depreciation and amortization, insurance, communication, and purchased labor and transportation costs. Purchased labor and transportation include the cost of leased vehicles and the cost of temporary help at the destination for unloading and at the warehouse for periods of abnormal demand.

Heteroscedasticity is frequently present in cross-sectional studies of this type. By using each of the cost components, Equation 6 was tested for heteroscedasticity with respect to each of the explanatory variables. Based on the Goldfeld-Quandt test (12, pp. 104-106), the equation for administrative costs and the equation for operating costs were both found to be heteroscedastic with respect to average length of haul. No other equation was found to be heteroscedastic with respect to any explanatory variable.

The administrative costs and operating costs equations were reestimated by using transformed data in order to correct for the heteroscedasticity; i.e., data were obtained by dividing each firm's observations by its average length of haul. Based on the Goldfeld-Quandt test, the corrected equations were found to be free of any significant heteroscedasticity. The results of the regression analysis on these two corrected equations, as well as the equations for the other cost components, are given in Table 1. In this table the t-statistics test for nonzero coefficients for the explanatory variables, and b_1, b_2, b_3, and b_4 represent estimates of the parameters β_1, β_2, β_3, and β_4, respectively.

For the total-cost equation, the presence of economies of scale is suggested because the coefficient of the shipment variable Q is less than 1 and almost identical to the 0.947 value obtained by Warner (10, p. 21) by using general-freight carrier data. The estimated standard error for b_1 was

Table 1. Regression results when estimating Equation 6.

		b_1 (Q)		b_2 (W)		b_3 (H)		b_4 (D)	
Cost Component	R^2	Estimated Regression Coefficient	t-Statistic	Estimated Regression Coefficient	t-Statistic	Estimated Regression Coefficient	t-Statistic	Estimated Regression Coefficient	t-Statistic
Administrative	0.4795	0.6781	3.755[a]	-1.073	2.411[a]	0.9764	10.84[a]	-0.329	0.544
Purchased labor and transportation	0.6183	0.8042	2.412[a]	0.5255	0.6395	0.7141	4.295[a]	0.597 1	0.5464
General operating	0.7020	0.8993	6.144[a]	-0.8327	2.302[a]	0.7947	10.887[a]	-0.890 2	1.856
Salaries and wages	0.3910	0.8490	5.659[a]	-0.9444	2.554[a]	1.062	14.206[a]	-1.535	3.122[a]
Depreciation	0.5671	0.734	4.838[a]	-0.216	0.5780	0.568	7.496[a]	-0.922 2	1.855
Insurance	0.8816	0.9364	8.245[a]	-1.214	4.339[a]	0.7186	12.687[a]	-0.005 95	0.0158
Communication	0.8050	0.9267	6.174[a]	-0.782	2.299[a]	0.5647	8.202[a]	-0.225 9	0.4995
Total cost	0.7168	0.9464	6.189[a]	-0.9104	2.416[a]	1.2456	16.334[a]	-0.920 8	1.838

[a]Significant at the 0.05 level.

Table 2. Regression results when average weight variable is omitted.

		b_1 (Q)		b_3 (H)		b_4 (D)	
Cost Component	R^2	Estimated Regression Coefficient	t-Statistic	Estimated Regression Coefficient	t-Statistic	Estimated Regression Coefficient	t-Statistic
Administrative	0.3715	0.6912	3.547[a]	0.7849	17.114[a]	-0.6942	1.125
Purchased labor and transportation	0.6124	0.7977	2.418[a]	0.8079	10.406[a]	0.7795	0.7466
General operating	0.6453	0.9095	5.799[a]	0.6461	17.503[a]	-1.1792	2.376[a]
Salaries and wages	0.2419	0.8606	5.2599[a]	0.8940	23.216[a]	-1.862	3.5965[a]
Depreciation	0.5619	0.7367	4.194[a]	0.5286	14.982[a]	-0.9972	2.101[a]
Insurance	0.8020	0.9513	6.595[a]	0.5019	14.782[a]	-0.4277	0.9362
Communication	0.7682	0.9364	6.335[a]	0.4251	12.220[a]	-0.4974	1.063
Total cost	0.6577	0.9575	5.800[a]	1.083	27.874[a]	-1.237	2.367[a]

[a]Significant at the 0.05 level.

0.1529, which yields a t-statistic of -0.3506 for testing the hypothesis H_0: $\beta_1 > 1$ versus H_1: $\beta_1 < 1$. This t-value corresponds to a level of significance of approximately 0.365. Although not statistically significant at the more commonly chosen values for level of significance, this t-value does indicate some statistical evidence of economies of scale.

In addition, note that the estimated value for β_1 is greater for the total-cost equation than for any of the cost-component equations. This may indicate some aggregation bias, which suggests that the true value of β_1 is actually somewhat less than 0.947. Furthermore, the conclusion of economies of scale for HG carriers is also supported by the fact that the cost elasticities (i.e., the estimates of β_1) are less than 1 in each of the cost-component equations.

The estimated coefficients for average weight, with the exception of the purchased labor and transportation equation, were found to be negative. Although weight should not have a large effect on costs, an increase in weight should not cause a decline in costs. The problem may well be one of multicollinearity. Average weight was defined as total tons per number of shipments, which is obviously related to the shipments variable. Because length of haul is the total ton-miles per total tons, weight and distance may also be collinear.

In order to determine if multicollinearity is the source of the problem, another regression set was estimated without the average weight variable in order to examine the effect on the standard errors of the coefficients. The estimated standard error of the average length of haul variable declined substantially, thus indicating that a relation between average length of haul and average shipment weight may have existed. The t-statistics for length of haul also greatly increased, and the R^2 values declined only slightly (see Table 2).

With the weight variable being deleted, the b_1 value for the total-cost equation in Table 2 still indicates the presence of economies of scale for HG carriers (because it is less than 1). Furthermore, the cost elasticities with respect to the shipment variable are less than 1 in each of the cost estimations. None of these individual cost elasticities is significantly less than 1 in a statistical sense. However, the fact that all seven cost-component coefficients and the total-cost coefficient are less than 1 does provide substantial evidence that economies of scale for HG carriers do exist. Thus, if shipments increase by a certain percentage, we would expect the cost to be incurred by HG carriers to increase by a smaller percentage.

Because the coefficients on the dummy variables are negative for every cost estimation except for purchased labor and transportation costs, it is concluded that, for these cost estimations, class 1 HG carriers are expected to experience lower costs than class 2 carriers (other things remaining the same). With the dummy coefficient being positive for purchased labor and transportation costs, it further appears that class 1 HG carriers are expected to experience higher costs for this category than class 2 carriers.

The major difference between our cost analysis, which used HG carriers, and that of Warner's, which used general-freight trucking firms, is in the estimated value of the coefficient on length of haul. Warner (10, p. 21) obtains a value of 0.321 for this coefficient in his total-cost equation as opposed to our value of 1.083. Thus, Warner (10) concluded that if length of haul increased for general-freight carriers, total cost would increase by a smaller percentage.

Based solely on the size of our estimate (1.083), we conclude that cost will increase at a faster rate than length of haul. In fact, the null hypothesis that the length-of-haul coefficient for total cost <1 can be rejected at the 0.05 level by using the HG data. However, the length-of-haul coefficient is less than 1 for each of the cost-component equations and considerably less than 1 for most of these cost-component equations. This indicates that the length-of-haul parameter (β_3) for total cost is overestimated because of aggregation bias. Thus, conclusions about economies of scale for length of haul must be based on cost-component coefficients.

Based on these coefficients, it can be concluded that economies of scale do exist for length of haul. Still, with the length-of-haul coefficients for the cost-component equations being substantially greater than Warner's (10) coefficient of 0.321, a proper conclusion would be that a percentage increase in length of haul for HG carriers would be expected to result in a greater percentage increase in costs for these carriers than for general-freight carriers.

CONCLUSIONS

The purpose of this paper has been to investigate the existence of economies of scale for HG carriers by means of a statistical cost analysis. The general conclusion is that economies of scale do exist for HG carriers. Also, the extent of economies is almost identical to that found by Warner (10) for general-freight carriers. Hence the irregular, nonrepetitive nature of demand for HG carriers does not appear to be a hindrance to economies of scale for these carriers. Our analysis also suggests that HG carriers receive substantially less economies from length of haul than that found by Warner (10) for general-freight carriers.

From our analysis of various cost categories, it is further concluded that class 1 HG carriers are expected to experience lower costs for these categories than class 2 carriers (other things remaining the same). One exception was purchased labor and transportation costs. This conclusion is reasonable because larger carriers are more likely to purchase labor and transportation services than smaller carriers.

Because our analysis supports the existence of economies of scale in the HG trucking industry, we can further infer that concentration (i.e., fewer HG truck carriers controlling a larger percentage of the industry's market) will likely occur during the current deregulatory environment. Existing HG carriers will be seeking to increase their market share and size in order to take advantage of the lower unit costs attributed to the existence of economies of scale.

ACKNOWLEDGMENT

We wish to thank James D. Voorhies for help rendered in the statistical analysis.

REFERENCES

1. R.A. Nelson. The Economic Structure of the Highway Carrier Industry in New England. Presented at New England Governors' Conference on Public Transportation, Boston, 1957.
2. M.J. Roberts. Some Aspects of Motor Carrier Costs: Firm Size, Efficiency, and Financial Health. Land Economics, Aug. 1956, pp. 228-238.
3. M.L. Ladenson and A.J. Stoga. Returns to Scale in the U.S. Trucking Industry. Southern Economic Journal, Jan. 1974, pp. 390-396.
4. G.N. Dicer. Economies of Scale and Motor Carrier Optimum Size. Quarterly Review of Economics and Business, Spring 1971, pp. 31-37.
5. J.C. Johnson. Trucking Mergers: A Regulatory Viewpoint. Heath, Lexington, MA, 1973.
6. J.P. Rakowski. Economies of Scale in U.S. Trucking. Journal of Economics and Business, Spring/Summer 1978, pp. 166-176.
7. R.A. Spady and A.F. Friedlaender. Hedonic Cost Functions for the Regulated Trucking Industry. Bell Journal of Economics, Spring 1978, pp. 159-179.
8. D.A. Breen. The Monopoly Value of Household-Goods Carrier Operating Certificates. Journal of Law and Economics, April 1977, pp. 153-185.
9. R. Koenker. Optimal Scale and the Size Distribution of American Trucking Firms. Journal of Transport Economics and Policy, Jan. 1977, pp. 54-67.
10. S.L. Warner. Cost Models, Measurement Errors, and Economies of Scale in Trucking. In The Cost of Trucking: Econometric Analysis (M.L. Burstein and others, eds.), William C. Brown Co., Dubuque, IA, 1965, pp. 1-46.
11. Trinc's Bluebook of the Trucking Industry. Trinc Transportation Consultants, Washington, DC, 1977.
12. R.S. Pindyck and D.L. Rubinfeld. Econometric Models and Economic Forecasts. McGraw-Hill, New York, 1976.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.

Quantitative Methods for Evaluation and Selection of TSM Project Alternatives

DAVID REINKE AND DAVID CURRY

The evaluation of transportation system management (TSM) projects should ideally include a ranking of their relative desirability. Project ranking requires a consistent method of summarizing the evaluation of each project. Three methods of presenting the results of a TSM project evaluation are compared. These methods are quantitative techniques that were specified for evaluation and selection of TSM project alternatives in a 1982 study for the California Department of Transportation. The following findings are discussed. First, simple displays of project outcomes are useful adjuncts to cost-benefit information but are by themselves insufficient for aiding project decision making. Second, cost-benefit data clearly facilitate economic assessment of project alternatives. Third, cost-effectiveness information is highly prone to arbitrary assumptions and misinterpretations, especially when more than one effectiveness criterion is used, unless (a) the criteria can be expressed in a formula that relates noncostable outcomes to project cost and (b) no cost-benefit relations can validly be defined. These results are applicable to other states and can be used to evaluate construction and TSM projects.

Three ways to present the results of an evaluation of transportation system management (TSM) project alternatives are compared. At the simplest level, referred to here as an outcome display, TSM project results can be organized and listed. Two other ways--cost-benefit and cost-effectiveness analysis--can be used to aggregate and summarize information from the evaluation so that it is easier to interpret. Examples from actual TSM evaluations illustrate the three approaches and provide guidelines for each approach. A combination of outcome display and cost-benefit information is recommended in most cases.

The research for this paper was developed for a particular study (1), but its results can be applied to states other than California and to construction projects and TSM projects.

OUTCOME DISPLAY

A simple display of project outcomes by evaluation criteria can be a convenient way to summarize and compare projects. Table 1 (1, p. E-8) is an abridged version of a display of project outcomes from a prototype TSM study of a section of an urban arterial (2).

Although the table rates project outcomes only as positive, negative, or no effect, numerical results or rating scales could be displayed in place of the +, -, and o signs. A simple rating scale is often useful because the results can then be added--assuming that care is taken to avoid double counting and nonlinear rating scales. An example of a numerical performance scale is 0 = unacceptable (a fatal flaw), 1 = poor, 2 = good, and 3 = excellent.

We recommend a scale of no more than five points in order to keep the rating simple. Considerable creativity is possible in the choice of adjectives or numbers represented by a numerical scale, and the adjective or number can differ by outcome. For example, air quality effects could be rated by the scale given above, while noise level ratings could be expressed in dbA, and equity of financing by a scale for which 0 = very discriminatory, 1 = discriminatory, 2 = somewhat nondiscriminatory, and 3 = nondiscriminatory.

The advantage of outcome displays is that they allow easy comparison between projects according to any set of evaluation criteria. The format shown also provides ready reference back to the original problem statement because outcomes related to specific project objectives are themselves specified as criteria. The disadvantage of such a table is that there is no single figure of economic merit; therefore, choices among alternatives may have to be made on highly subjective grounds. For example, the alternative in Table 1 that is marked not recommended has more o and - ratings than the recommended alternatives. But it does not require much imagination to visualize a group of project alternatives among which the choice is not obvious.

Two issues that the outcome display helps to illuminate are the choice of evaluation criteria and consideration of the effects of trade-offs between different objectives. The evaluation criteria should be based on the transportation system objectives, and their number should be kept small (1). They should address all important objectives of the project in question but be omitted for minor objectives or for outcomes that are not significant.

Trade-offs among project features could be analyzed by varying the scale, location, timing, or focus of a project and noting the incremental effects on cost and other outcomes in other columns of the same table or in a separate table. Consideration of trade-offs is one way to generate additional project alternatives, which is not often done in evaluations of TSM projects. Generally, the alternatives can most readily be considered in the order of increasing cost, with each increment of cost (compared with other acceptable alternatives) considered separately.

The outcome display should be used as a first step in any evaluation because it is easy to generate, it may serve the purposes of the decision in question, and it provides an intuitively useful summary. Whether to proceed with the greater quantification requirements of cost-benefit or cost-effectiveness analyses will depend on the value of the information they add. The original outcome display

Table 1. Example of outcome display.

	Candidate TSM Project		
Evaluation Criteria	Signal Interconnect	Eliminate 10 Curb Cuts	Expand Park-and-Ride Lots
Corridor mobility			
Transit use	+	o	+
Commercial vehicle trips	+	-	o
Peak-period trips	+	+	+
Travel-time delay	+	+	+
Safety: accident rate	+	+	
Social and environmental			
Air quality	+	o	+
Energy use	+	o	
Transit rider comfort and convenience	o	o	+
Existing land use: local access to local commercial and industrial center	+	-	
Cost ($)	150 000	66 000	100 000
Result	Recommend	Not recommend	Recommend

Note: + = positive effect, - = negative effect, and o = no effect.

should also be used to complement a cost-benefit or cost-effectiveness summary.

COST-BENEFIT ANALYSIS

Cost-benefit analysis is a method of aggregating outcomes that can be assigned a monetary value into a single measure. A frequently used criterion that summarizes the results of an economic evaluation is the benefit/cost ratio, which is computed as follows:

1. Add up all project or program costs,
2. Assign dollar values to outcomes when possible (e.g., value of time saved, value per accident reduced) and compute a total dollar figure to represent the value of the benefits, and
3. Find the ratio of benefits to costs.

Benefit/cost ratios of 1.0 or greater are judged to be favorable. Equivalent criteria are the cost per dollar of benefits, for which amounts under $1 are judged to be favorable, or the internal rate of return, for which rates above the minimum attractive rate of return are favorable. With any of these criteria, important results that cannot readily be valued in dollars can still be considered in the form of the outcome display just described.

The authoritative guide to highway cost-benefit analysis is the 1977 AASHTO report (3). Cost-benefit analysis has also been applied to TSM projects according to the guidelines in that report. Two examples are shown in Tables 2 (4, p. 2-15) and 3 (4, p. 2-19), which deal, respectively, with parking management and flextime promotion programs of Seattle Commuter Pool, a regional ridesharing agency (4). The tables are self-explanatory, moving in sequence from outcomes to benefits to costs to the calculation of benefit/cost ratios.

The source report also evaluates Commuter Pool's vanpool and ride-matching programs in the same manner, obtaining benefit/cost ratios of 11 to 21 for vanpools and 53 for the ride-match services. With ratios of 11 to 14 for parking management (in Table 2), these indicate impressive economic justifications for ridesharing programs. The ratio of 101 for flextime in Table 3 is unusually high due to inclusion of productivity benefits (line d). For the Seattle evaluation, the economic merit of these programs was the principal evaluation criterion of interest, so no additional information was presented except for the efficiency measure in line h of Table 2 and the footnote regarding outside use of the flextime manual in Table 3.

Users of cost-benefit analyses should, however, be aware of several points. Whenever a cost-benefit analysis is used to evaluate projects whose outcomes are considered over more than 5 years, future costs and benefits should be discounted in order to compute their equivalent present or annual value. This is especially important when the projects being compared have different patterns of costs and benefits over time. The interest rate for discounting should generally be 4 percent [the approximate long-range cost of capital, assuming the use of constant dollars (no inflation)]. If future costs and benefits are inflated, the discount rate and the inflation rate should be combined. For example, if an inflation rate of 10 percent is used, the combined rate will be (4 percent x 10 percent) + 10 percent, or 10.4 percent.

If a project entails any significant risks or uncertainty, there are three simple ways to allow for it:

1. Add 1 to 2 percent to the discount rate,
2. Increase the minimum acceptable benefit/cost ratio to between 1.1 and 1.2, or
3. Estimate the range of possible outcomes rather than the most likely single numbers.

Table 2. Parking management evaluation.

Evaluation Criteria	Description	Value
Outcomes	a. New downtown parking carpool registrations	292
	b. New park-and-pool carpools: 1500 spaces maintained x 35 percent occupancy rate	525
	c. New high-occupancy vehicle (HOV) priority parking spaces facilitated at employment sites (estimate)	300
Benefits	d. User benefits per new carpool ($)	4830
	e. Land use benefits per new carpool = 0.94 space saved per pool x $1.80/day x 250 working days/year x 12.66 (present worth factor for 18 years at 4 percent) ($)	5355
	f. Total benefits = (a + b + c) x 20 percent influenced to pool x (d + e) ($)	2 275 300
Cost	g. 1980 cost of parking management element ($)	161 000
Efficiency measure	h. Program cost per new HOV space = g/(a + b + c) ($)	155
1980 benefit/cost ratio	i. Benefit/cost ratio = f/g	14
Typical benefit/cost ratio	j. Benefit/cost ratio with b reduced to 167 (b ÷ 2.7 years) to reflect replacement carpools only	11

Table 3. Flextime promotion.

Evaluation Criteria	Description	Value
Outcomes	a. Commuter Pool survey results: 3374 employees in Seattle area firms assisted to convert to flextime in 1980 x 0.5 to discount for other influences on cooperating employers	1687[a]
	b. Estimated persons induced to rideshare by flextime introduction = a x 0.096	162
Benefits	c. Average daily time saved per flextimer = 2.3 min/trip (one-half of Boston experience) x 2 trips/day x $0.05/min value of time ($)	0.23
	d. Daily value of increased productivity per worker ($)	0.50
	e. One-time employer implementation cost per worker ($)	100
	f. Total benefits = a(c + d) x 250 working days/year x 15.62 [present worth factor for 25 years at 4 percent (total, $4 809 000)] + b x $2100 benefits per carpooler (total, $34 000) – a x f (total, $168 700) ($)	4 674 300
Costs	g. 1980 cost of flextime promotion ($)	46 500
Benefit/cost ratio	h. Benefit/cost ratio = f/g	101

[a]In addition, the Commuter Pool flextime manual was sold to other companies outside of the Seattle area that have adopted flextime, including Crocker Bank in San Francisco with 17 000 employees.

More sophisticated ways of dealing with risk entail assigning probabilities to different outcomes, but this is unlikely to be necessary in TSM studies.

The value of time will be an important issue in the economic evaluation of many TSM projects. First, there is no definite standard for the value of time to be used. Various studies of traveler behavior show that travelers tend to value in-vehicle time (e.g., driving time and on-board transit time) between 20 and 130 percent of their wage rate, and out-of-vehicle time (e.g., waiting time for transit) by a factor between 2 and 3 times higher than in-vehicle time. A reasonable standard would be to use half the average wage rate for in-vehicle time and the full wage rate for out-of-vehicle time. A related problem is that the relative value of time for travel under different conditions has not been clearly identified. For example, there is probably a higher value placed on driving than on riding in a carpool or vanpool, and a higher value on standing in a transit vehicle than riding in a comfortable seat where reading is possible; but no one knows by how much.

Another issue in valuing time savings is that research has clearly shown that the perceived value of travel-time savings varies with the purpose of the trip and with the amount of time saved per trip (3). Savings under 5 min/trip have low values and only savings of 15 min or more are fully valued at the rates cited above. Many transportation providers ignore this finding or argue that the data for applying it are not always available. We recommend either a precise or an approximate method of valuing time savings, depending on the rigor required in the study. The precise method is to ignore time differences per trip of 5 min or less, use straight-line interpolation for savings between 5 and 15 min, and use the full values for savings of 15 min or more per trip. The approximate method is to ignore savings under 10 min/trip and use the full value for savings of 10 min or more, which will avoid the need to value time in all but the most dramatic types of improvements. Whatever the standard used, it should be applied uniformly across the region; this is another coordination task for the regional transportation planning agency.

Benefit/cost ratios can be misleading if there is no standard way to categorize costs and benefits. For example, one of the outcomes of a ridesharing program will be that some transit users will become carpoolers. Depending on the amount of transit fares lost as a result, the benefit/cost ratio could be different if this value is treated as a benefit to users rather than as a cost to the transit agency. The treatment should depend on whose point of view is being considered. If it is the traveler's point of view, which is usual, the savings in fares are clearly a benefit and offset any similar costs for the ridesharing journey. A definite standard for classifying such outcomes should be used for all analyses in the region.

Like all aggregate measures, the computation of a benefit/cost ratio results in some loss of information. There may be other problems with using this measure, particularly how to value various outcomes. Cost-benefit analysis is, however, a useful technique for quickly summarizing large amounts of information, especially when there are many different types of outcomes to consider in the evaluation. Moreover, use of this method does not preclude the consideration of other outcomes that cannot be valued in dollars or are not quantifiable; in fact, it can help bring these to the forefront because a large number of other outcomes will have been aggregated. Therefore, this method should be used only under the following conditions:

1. Several outcomes must be considered, and cost-benefit analysis can usefully summarize some of them; or there is interest in the economic merit of the project or in the relative economic merits of alternative projects; and
2. Standard procedures are followed to resolve issues about valuation of outcomes, interest rates, and classification of outcomes.

Cost-benefit analysis does not relieve the planning agency of its responsibilities to note all significant project outcomes--quantifiable or not--and to identify and analyze significant trade-offs. The use of a simple outcome display, as discussed in the previous section, can therefore be a useful supplement.

COST-EFFECTIVENESS ANALYSIS

Cost-effectiveness analysis entails the calculation of one or more indices for a project, each of which is the ratio of project costs to some outcome measure. If there is a single predominant goal for the project, such as reducing delay or increasing capacity, total project costs can be assigned to a single associated cost-effectiveness index such as cost per passenger-minute saved or cost per added vehicle per hour of capacity.

The table below (5, p. II-22) gives an example of a single cost-effectiveness index--cost per vehicle mile of travel (VMT) reduced--for the Golden Gate Bridge, Highway, and Transportation District (GGBHTD) vanpool project (5); the table also gives an alternative index--program cost per dollar of user benefit--which is simply an inverse benefit/cost ratio:

Evaluation Criteria	Value
Eligible users	45 000
Program characteristic (annual)	
One-way trips served	312 500
Program cost ($)	264 300
VMT reduced	6 800 000
User benefits ($)	1 079 800
Performance measure ($)	
Program cost per VMT reduced	0.039
Program cost per dollar of user benefit	0.24

[Note that costs are expressed in 1980 dollars, and all costs and benefits (including VMT reductions) are present values obtained by using a 10 percent discount rate over a 5-year program period.]

It is immediately apparent from the latter index (program cost per dollar of user benefit) that this is an attractive project economically because only $0.24 in program costs produced $1 in user benefits. By comparison, the $0.039 cost per VMT reduced is less clear and requires more information before it can be understood, in particular:

1. What is a reduction of one VMT worth?
2. Is $0.039 an attractive cost per VMT in comparison with its value?
3. Is VMT reduction the only goal of the GGBHTD vanpool program? If there are other goals, such as reducing air pollution or energy consumption, should not part of the program cost be allocated to the other goals?

A usual practice is to allocate program costs among different goals in calculating multiple cost-effectiveness measures in order to avoid double counting. But such allocations are arbitrary because there is no intuitive or commonly accepted way to arrive at the correct allocation. Moreover, the resulting cost-effectiveness measures are usually difficult to interpret and may produce conflicting results unless a fortunate choice of cost allocations has been made.

Table 4 (6, p. 120) gives an example of such a cost allocation for an evaluation of four alternatives for mixed-mode operations on the San Bernardino Freeway Busway. Option A is the addition of two unrestricted freeway lanes only, and option B is the busway as actually constructed. Option C is a lower-cost busway with less-cost-effective features omitted, and option D is the same as option C with reversible, contiguous lanes (which are similar to the Shirley Highway Busway approaching Washington, D.C.). The allocation is made by assigning a relative importance to each cost, and then allocating the costs of each option among the results according to these weights.

The cost-effectiveness indices for the first two goals in Table 4, measured respectively by person-trips per assigned dollar and assigned dollars per person-hour saved, are shown in Figures 1 (6, p. 121) and 2 (6, p. 123). Figure 1 shows that option D is superior to the other options in person-trips per assigned dollar (note that lined blocks are based on the peak hour and the total is based on the peak 4 hr). Figure 2 shows that options C and D have a lower assigned cost per person-hour saved--on the order of $4.20 compared with $5 for stage 2 of option B.

But the analysis begs the question: What is a reasonable cost per person-hour saved? If a reasonable cost is $4, then all options are too expensive; or if a reasonable cost if $6, then all options are acceptable by this criterion. If only 10 percent rather than 20 percent of total costs were assigned to improved level of service, the assigned costs per person-hour saved would be only half of the numbers shown in Figure 2.

This example shows the hazards of cost-effectiveness analysis where there are two or more goals. In contrast, the cost-benefit analysis adds up the dollar value of travel-time savings, reduced travel costs, improved safety, energy saving, and, if possible, air pollutant emissions. This would combine the value of the outcomes for five of the seven goals given in Table 4. If benefits exceed costs based on these outcomes, added capacity and provision for future contingencies can simply be regarded as nonpriced fringe benefits. If total benefits still do not exceed total costs, then only one question remains to be answered: Is the value of any added capacity or added provisions for future contingencies offered by an option large enough that benefits would exceed costs? This may not be a simple question, but dealing with it is easier than dealing with seven independent goals and corresponding criteria in a cost-effectiveness framework.

There is one valid way of including multiple measures of effectiveness in a cost-effectiveness framework that avoids the procedure of allocating project costs among different goals. This is the practice of expressing the criterion in a formula that contains two or more terms, where each term identifies an outcome not readily valued in dollars. For example, the following cost-effectiveness index is used by the California Department of Transportation (Caltrans) for ranking roadside noise barriers:

$$\text{Noise attenuation index} = [R \times (E - 70\ \text{dbA})^2 \times N]/C \qquad (1)$$

where

R = noise reduction achievable by sound barrier (dbA),

Table 4. Relative cost of options assigned to each goal for San Bernardino Freeway busway.

Goal	Relative Importance (%)	Equivalent Annual Cost by Option ($000s) A	B	C	D
Added capacity	20	310	1528	1245	1232
Improved level of service	20	310	1528	1245	1232
Reduced cost of travel	20	310	1528	1245	1232
Improved safety	15	232	1146	934	924
Reduced environmental impacts					
Air pollutants	10	155	764	622	616
Energy savings	10	155	764	622	616
Future contingencies	5	77	382	311	308
Total		1549	7640	6224	6160

Figure 1. Capacity cost-effectiveness.

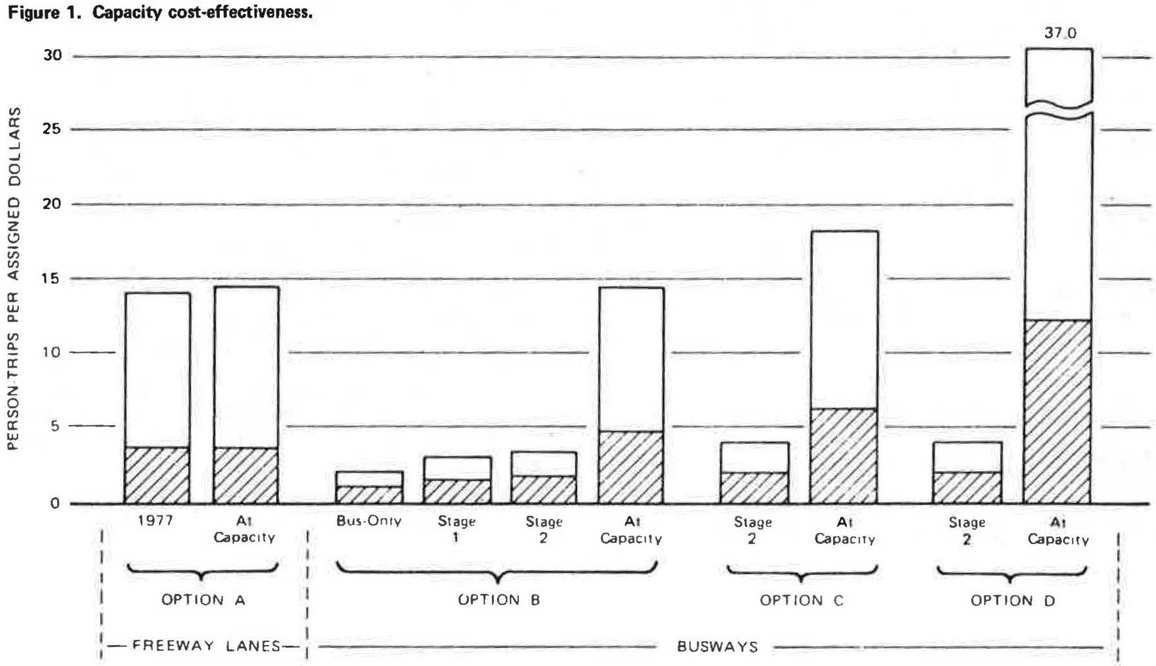
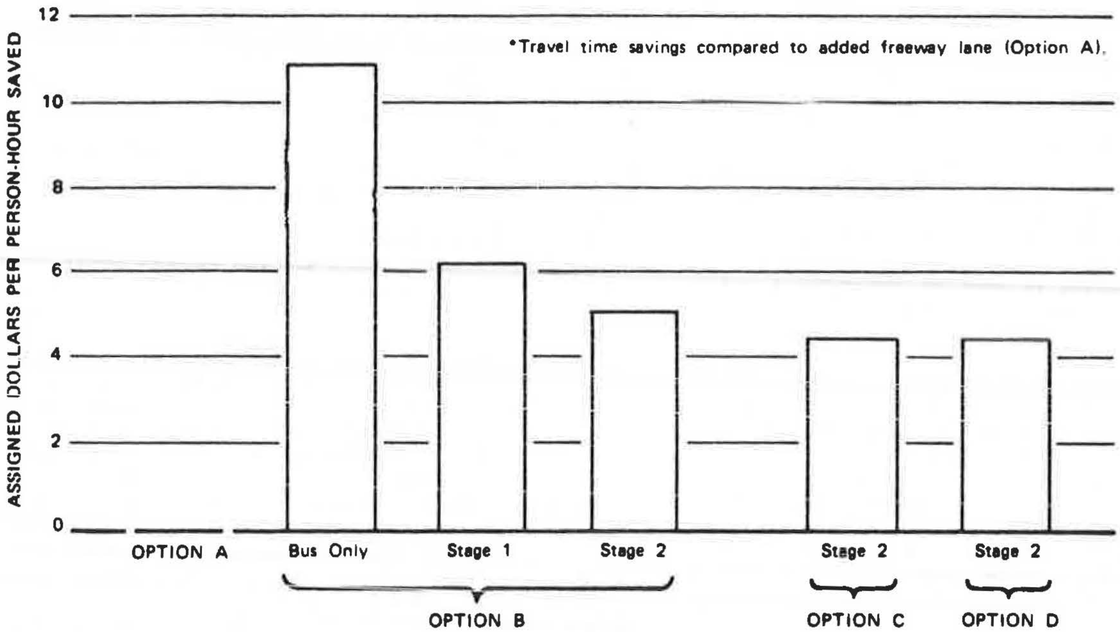


Figure 2. Travel-time cost-effectiveness.



E = existing noise level (dbA) at the first row of houses from the highway,
N = number of dwelling units benefited by noise barrier, and
C = project cost ($000s).

There are no known examples of this approach to TSM projects, and the approach can be recommended only when a cost-benefit analysis is not feasible.

In summary, cost-effectiveness has the appeal that it can be simpler than cost-benefit analysis when only a single effectiveness measure is used because benefits do not have to be valued in dollars. But a cost-effectiveness analysis has several serious disadvantages:

1. When there is more than one important result, project costs must be allocated among the different results in some arbitrary way (unless the formula approach just illustrated for a noise attentuation index is used).
2. Cost-effectiveness criteria do not permit selecting or ranking of project alternatives with multiple outcomes unless, by chance, one project alternative is clearly superior for all outcomes.
3. Cost-effectiveness criteria do not show whether or not a project is economically attractive unless thresholds of desirability (e.g., $5/person-hr saved) are set for all criteria. But doing that would enable direct computation of the benefits and a much simpler cost-benefit display of results.

CONCLUSIONS

Ranking of TSM projects requires a consistent method for summarizing the results of the evaluation of each project alternative. We have discussed three methods for summarizing the evaluation results: outcome display, cost-benefit analysis, and cost-effectiveness analysis. A simple display of outcomes is a useful first step in summarizing the evaluation and is also a useful supplement to any further analysis. We prefer cost-benefit analysis as a consistent way to combine project outcomes that can be valued in dollars; however, use of this method does not relieve the planner of the responsibility for considering other important outcomes that cannot be conveniently included in the cost-benefit analysis. We recommend cost-effectiveness analysis only for evaluating TSM project alternatives that have a single important outcome that cannot be readily valued in dollars.

REFERENCES

1. D. Curry, B. Crowell, D. Reinke, and G. Shearin. TSM Planning Guidelines for California. California Department of Transportation, Sacramento, Sept. 1982.
2. G. Spanovich. Transportation System Management Prototype Planning Study: Portland. Office of Planning Assistance, UMTA, Nov. 1979.
3. D. Anderson, D. Curry, and R. Pozdena. A Manual on User Benefit Analysis of Highway and Bus Transit Improvements. AASHTO, Washington, DC, 1977.
4. D. Curry, G. Shearin, and N. Klein. Evaluation of Seattle/King County Commuter Pool Program. Crain and Associates, Los Altos, CA, June 1981.
5. R.F. Kirby and G.K. Miller. A Casebook of Short-Range Actions to Improve Public Transportation. Urban Institute, Washington, DC, Rept. 3072-02, Nov. 1981.
6. L.J. Glazer. San Bernardino Freeway Express Busway: Evaluation of Mixed-Mode Operations. Southern California Association of Governments, Los Angeles, July 1978.

Publication of this paper sponsored by Committee on Application of Economic Analysis to Transportation Problems.