COMPARISON OF COST MODELS FOR URBAN TRANSIT

James H. Miller and John C. Rea, Pennsylvania State University

Urban transportation planning often requires a comparative evaluation of alternate modes of transportation. One of the key criteria for comparing systems is the cost of operation. Some of the most common cost descriptors used to compare different modes are capital costs, total annual operating costs, total costs including depreciation and debt service, and cost expressed in terms of units of output such as route-miles, ton-miles, passenger-miles, or vehicle-miles. To evaluate proposed urban transport system alternatives, it is normally necessary to completely describe a hypothetical system and then determine each cost element item by item. Although this approach is useful and in many cases necessary in making final system selection, it is unnecessarily burdensome when attempting to screen a large number of alternatives. A more useful transportation planning tool would be a relatively simple model that would predict costs based on a limited set of input data. A review of the literature revealed several efforts at modeling urban transportation mode costs. It is the purpose of this paper to review these models and to evaluate their suitability for use in urban transportation planning. It should be noted that this study results from work done to establish the viability limits for urban transit modes. It does not purport to be an exhaustive study of operating cost models.

•SIX models will be considered in the following sections. These models differ in two important ways: the type of independent variables considered and the method used to determine the coefficients assigned to the dependent variables. Three types of independent variables are used. They are system output measures, e.g., vehicle-miles, vehicle-hours, and passenger-miles; system characteristics, e.g., number of vehicles, length of station platform in feet, and length of right-of-way in miles; and system environment factors, e.g., age of city and density of land use. Most models use combinations of the types of variables. Parameters for the dependent variables are either estimated by regression analysis or determined on a unit-cost basis. This latter term will be defined in the next section. Table 1 gives the six models by type.

As given in Table 1, the models have been developed to determine operating costs for bus systems. Operating costs are defined as the variable costs of operation excluding allowances for depreciation, interest payments, and taxes. Exclusion of fixed costs from bus cost analysis is not a serious omission. For modes without the high capital costs of separate right-of-way, operating costs constitute 90 percent or more of the total cost. Variable costs for exclusive right-of-way systems such as the proposed Pittsburgh Skybus account for less than half the total cost; therefore, operating cost comparisons are less conclusive than for bus or similar modes that are less capital-intensive.

BUS COST MODELS

Four-Variable Unit-Cost Models

Operating cost models for bus operations were evaluated by Alan M. Voorhees and Associates (1). These models were designed to reflect certain operating costs such as

Sponsored by Committee on Transportation Systems Design.

equipment maintenance and garage expense, transportation expenses, traffic and advertising expenses, insurance and safety costs, and administrative and other general expenses. They do not reflect moneys needed for depreciation, interest payments, operating taxes, special fare-collection equipment, or income taxes. Two models evaluated are identical in general form, the only difference being in the method of estimating parameters for the independent variables. Both cost models have the following general form: Cost = A x vehicle-hours + B x vehicle-miles + C x peak vehicles + D x revenue passengers, where A = cost related to an hour of vehicle operation, B = cost related to a mile of vehicle operation, C = cost related to a peak vehicle, and D = cost associated with a revenue passenger other than transportation. Parameters A through D were estimated in two ways, first by the unit-cost method and then by multiple regression analysis.

To determine the values of parameters A through D, we use the unit-cost system to first divide the various accounting cost items in a manner given in Table 2 and then calculate the cost of each item on the basis of the output most directly associated with it. For example, transportation expenses of supervision, operator wage, and other transportation expenses, as well as welfare expenses for the employees, are most closely related to vehicle-hours; whereas insurance and safety costs, for example, are mostly a function of either the number of passengers or the total amount of revenue received. Using March 1968 data from D. C. Transit System, Inc., the cost for fuel was $0.0312 per mile. (Records indicate that a total of 2,786,050 vehicle-miles were traveled, and $86,786 was spent for fuel.) Assuming that during 1969 D. C. Transit System, Inc., operated 30 million vehicle-miles, the cost for fuel would be $0.0312 x 30,000,000, or $936,000. A unit-cost model for that situation would be as follows: Cost = $0.0312 x vehicle-miles (2).

The unit-cost bus models (1) for the four transit operating companies in the Washington, D. C., area are as follows:

1. D. C. Transit System, Inc.—

   Annual cost ($) = (6.431 VH + 0.2187 VM + 1,802 PV + 0.01067 RP) Y

2. Washington-Virginia and Maryland Coach Company, Inc.—

   Annual cost ($) = (5.825 VH + 0.1364 VM + 850 PV + 0.01872 RP) Y

3. Alexandria, Barcroft and Washington Transit Company—

   Annual cost ($) = (4.244 VH + 0.1824 VM + 1,402 PV + 0.01442 RP) Y

4. WMA Transit Company—

   Annual cost ($) = (2.985 VH + 0.1264 VM + 3,114 PV + 0.03281 RP) Y

In these models VH = annual vehicle-hours, VM = annual vehicle-miles, PV = number of scheduled peak-hour vehicles, RP = annual revenue passengers, and Y = contingency at 2½ percent. It can be seen that the parameters vary widely. In some cases more than 100 percent variation is seen among the various transit operations. Based on this example and other experiences of this type of model, it must be recognized that generalization to other transit operations would not be warranted because of the wide variation in the sizes of the companies, age and condition of equipment, and labor conditions faced by each company.

The advantage of the four-variable model is the relatively small amount of information required to estimate parameters for the equation. The only information that might not be directly available from operating records would be the number of vehicle-hours operated. However, this could be obtained by analysis of schedules. After comparison of the results of the unit-cost model with the regression model, the Voorhees consultants determined that the unit-cost model more accurately predicted costs for its study.

The unit-cost model can be used to predict future costs for a particular system; however, any future estimates of cost should be stated in terms of the base-year dollars. If actual dollar costs are to be estimated, the various parameters should be inflated by a suitable index of the change in cost of that item.

Four-Variable Regression Model

The second type of parameter estimation for the WMATA study (2) was based on a multiple linear-regression analysis of the system characteristics and annual operating costs. Data for the period 1962-1970 were used to estimate parameters for the equation. Costs were stated in terms of 1970 dollars (Table 3); seven different variations of the basic equation were tried. However, as given in Table 3, at least four of the regression models did not do an acceptable job of forecasting even though they had a high coefficient of determination. This is an indication of the sensitivity of the models to changes in relation among operating characteristics. The consultant concludes as follows (2, p. 4):

> The models were developed using data for a period of operation when passengers were declining [and] efficiency was constant. The application of the equations to date where there is an increase in both passengers and system speed results in cost estimates which are questionable.

The advantage of the regression equation model is that the parameters are estimated based on several observations rather than on one and, therefore, should be a more reliable indication of the relation among the variables. However, as just explained, regression equation coefficients are based on one set of operating characteristics, which may change in the future. It is especially true when the model is being used to aid in planning new transit systems.

Daily Cost Four-Variable Model

Ferreri developed a similar cost model for use in a Metropolitan Dade County Transit Authority study. The cost model used in this study is in the following form: $C = A_1M + A_2H + A_3R + A_4V$, where A_1 through A_4 = parameters determined by the unit-cost method, C = average daily cost of route operation, M = average daily vehicle-miles of service on route, H = average daily vehicle-hours of service on route, R = average daily passenger revenue on route, and V = peak vehicles needed on route (3). As can be seen, the model differs from the Voorhees model in two ways:

1. The costs and other characteristics are determined on a daily basis instead of an annual basis, and
2. Total passenger revenue for a day is used instead of number of passengers.

For the purpose of estimating parameters, accounting cost items were again assigned to the various output measures or system characteristics. To determine the validity of this approach, Ferreri performed separate regression analysis comparing the transportation expenses and measure of output for 11 transit properties.

As a more comprehensive check on these assignments, cost data from a sample of 66 transit properties were analyzed by the authors using regression analysis. Figures 1, 2, and 3 show the scatter diagrams and the related R^2 values. The cost items shown in these figures correspond to accounts prescribed by the ICC Uniform System of Accounts for Class I Common and Contract Motor Carriers of Passengers.

As can be seen, assignment of operating costs in a manner similar to the system used by Voorhees (Table 2) is certainly justified. In addition to the four-variable formulas, a two-variable formula was also tested by Ferreri. This formula only included vehicle-hours and vehicle-miles as the independent variables. The premise behind this comparative investigation is that, for planning purposes, the simpler the formula the easier the application if a sufficient degree of accuracy can be maintained. He also made a route-by-route estimate of costs using the two- and four-variable versions of the formula for the Miami Transit Authority. When compared to actual

Table 1. General typology of cost models.

Model	Type of Independent Variable	Method for Parameter Determination	Mode Described	Type of Cost Estimated
WMATA four-variable (2)	Output and system characteristic	Regression	Bus	Total annual operating cost
WMATA four-variable	Output and system characteristic	Unit-cost	Bus	Total annual operating cost
Slowness function (4)	Output	Regression	Bus	Cost per mile
D.R. Miller (6)	Output and system characteristic, environment	Regression	Bus	Cost per mile
M.G. Ferreri (3)	Output and system characteristic	Regression	Bus	Average daily cost per route
DOT-IDA (9)	Output and system characteristic	Regression	Bus	Total cost

Table 2. Allocation of account items to operating cost model.

Item	Vehicle-Hours	Vehicle-Miles	Peak Hour Vehicle	Revenue Passengers
Equipment maintenance and garage expenses				
Supervision		X		
Maintain service equipment			X	
Maintain buildings and grounds			X	
Maintain revenue equipment		X		
Tires and tubes		X		
Others		X		
Transportation expenses				
Supervision	X			
Operators' wages	X			
Fuel and oil		X		
Station expenses			X	
Others	X			
Traffic and advertising		X		
Insurance and safety				X
Administration and general				
Officers' salaries			X	
Employees' wages			X	
Legal expenses			X	
Welfare expenses	X			
Others			X	

Table 3. Regression models of bus costs.

Model	R^2	Percentage of Original Estimated Costs	
		1975	1990
1 Y = 28.3462 VH + 0.5433 VM + 976.582 PV - 0.1841 RP - 52,235,544	0.997	68.8	44.2
2 Y = 33.2481 VH - 0.1971 RP - 47,658,800	Not given	71.0	39.0
3 Y = 1.0717 VM + 10,669 PV - 10,437,042	0.691	91.8	99.0
4 Y = 1.0104 VM + 2,729,693	0.649	94.5	104.6
5 Y = 25.1411 VH - 0.0532 VM + 2,385 PV + 821,010 (VH/RP) - 68,389,056	0.995	75.7	57.1
6 Y = 0.6028 VM + 274,656 (VM/RP) + 8,911,720	0.738	91.9	99.2
7 Y = 25.4827 VH + 793,072 (VM/RP) - 68,004,768	0.993	76.2	58.3

operating results, it was concluded that the four-variable model was substantially more accurate and warranted the gathering of additional data. The only information that would not be readily available when proposing a new transit system would be the number of vehicles needed for peak-hour operations on a route. However, by making several assumptions about schedules and spare bus requirements, it is possible to estimate peak vehicle needs. The consultant concluded that, for long-range estimates, the two-variable model might be adequate. However, for short-range planning and fiscal planning, the additional effort required to use the four-variable model would be warranted (3, p. 9).

Slowness Function

A slightly different approach to modeling operating costs was taken by Miller and Holden (4). They have formulated what they call a slowness function, based on the premise that vehicle-miles operated and number of hours operated are the two most important determinants of operating costs. They have combined these two variables into one that they call slowness stated in minutes per mile (4). Starting with Ferreri's model, which has the form $C = A_1M = A_2H + A_3R + A_4V$, they divide this expression through by M and get $C/M = A_1 + (A_2H/M) + (A_3R/M) - (A_4V/M)$, where H/M is slowness in terms of hours per mile. If $B_1 = 60A_2$, the second term becomes B_1S_3, where S_3 = minutes per mile or slowness.

The last term in V is also a function of slowness because $V = S_3L/H$, where L is the total round-trip mileage of the route under consideration, and H is a peak-hour headway in minutes. The term R represents the cost of injuries and damage because insurance is often based on revenue. For purposes of this paper, it is assumed, however, that large transit properties may operate in a self-insured manner; therefore, this variable is dropped. The form of the final cost model then is $C/M = B_1 + B_2S_3$.

To estimate the parameters B_1 and B_2, Holden and Miller used data mainly from the New York City Transit Authority (NYCTA) and the San Francisco Municipal Railway. Some of the data from the Southern California Rapid Transit District and Chicago Transit Authority were also used. Table 4 gives parameter values determined in subsequent applications of this slowness function as reported elsewhere (5).

The advantage of this model is its simplicity. With only one independent variable, and depending on only two measures of output, it is the simplest of all the models evaluated. Holden and Miller have also included a factor to reflect variations in labor costs. However, even with this, the model does not adequately consider differences in operating systems and, therefore, is only valid for the particular transit system whose data were used for calibration. As will be seen in a later section that compares these various models, the slowness function cannot be used to generalize operating costs for systems other than those with identical characteristics to the one used for estimation of parameters.

Urban Environment Cost Model

Figure 4 (8) shows that, for bus operations in 1970, operating costs per mile varied greatly, from \$0.29 per mile to \$1.97 per mile. Obviously, managerial efficiency alone does not account for this wide range in operating cost. An analysis of variance study indicates that per-mile operating costs are related to fleet size (F = 54.3, significant at 0.001 level), possibly suggesting diseconomies of scale. However, because large fleets usually operate in large congested cities, there can be little doubt that the cost is more directly related to environmental operating context than to fleet size, which explains the variation in operating costs.

Miller (6) has attempted to identify factors that explain this wide variation in cost. Variables that help explain variations in cost include population density of the urban area, the age of fleet, the age of city, scheduled speed, and labor costs in a particular area. Inclusion of these variables improved the explanatory power of the regression model. Labor rate, scheduled speed, and city age were all significant. Miller's conclusion was that the other variables, such as fleet age and population density, influenced operating costs; however, because of inaccurate or inappropriate measurement of the variables, they were not statistically significant in the regression model.

Figure 1. Relation between administrative expenses and peak number of buses.

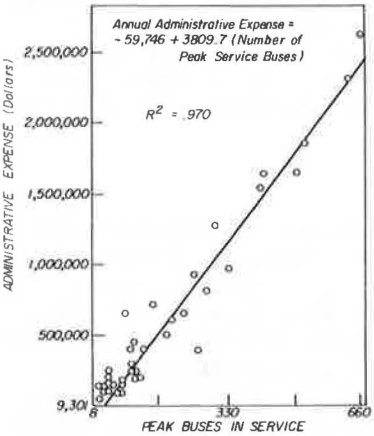

Figure 2. Relation between annual equipment and garage expenses and bus-miles.

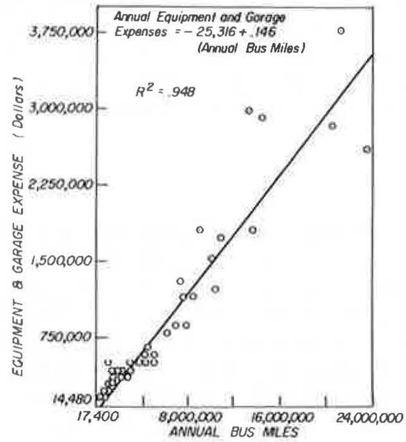

Figure 3. Relation between annual transportation expense and total bus-hours.

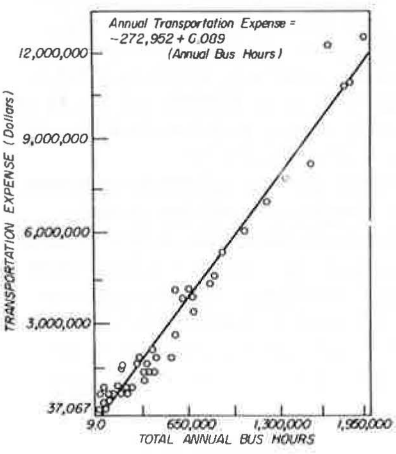

Figure 4. Frequency distribution of operating costs (8).

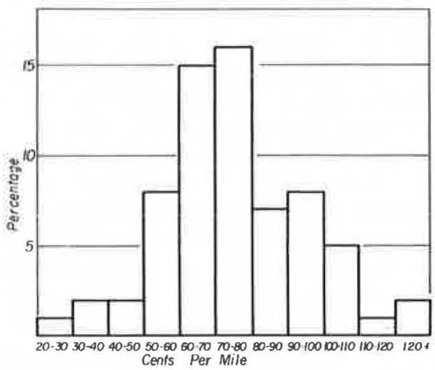

Table 4. Summary of reported data on slowness function (5).

Property	Date	Slowness (S_3)	Constant (B_1)
NYCTA 1969	Oct. 1969	19.38	5.28
Rapid busway	Nov. 1969	15.65	34.56
Bus 1967-1968	Nov. 1970	9.21	22.3
Bus 1970	Mar. 1971	18.15	30.83
San Francisco streetcars	Aug. 1971	11.9	85.10
San Francisco trolley bus	Aug. 1971	7.89	89.30
San Francisco diesel bus	Aug. 1971	11.61	56.50
Bus 1970	Aug. 1971	9.66	87.79

Another effort (9) to model bus costs is of interest for several reasons. First, it models total cost including depreciation and debt service. Second, it includes a variable that reflects prevailing wage rates. In addition, a dummy variable reflecting form of ownership (publicly owned or otherwise) is included because this influences capital costs. The model is of the following form: $\ell n C = a_1 + a_2 \ell n B + a_3 \ell n W + a_4 \ell n VEL + a_4 A + a_5 S + a_6 (PUB/a_7 s)$, where C = total cost, B = bus-miles, W = hourly wage rate of operating personnel, VEL = bus-miles per bus-hour attained by the firm, A = average age of fleet, S = average seats per bus, PUB = one for publicly owned firm and zero otherwise, and s = proportion of fleet purchased with capital grant.

These variables explained virtually all of the variation ($R^2 = 0.990$). Note that the VEL variable is the reciprocal of Holden's slowness variable. This variable was found to have a strong effect on costs, but a decrease in cost is slightly less than proportional to the increase in VEL. The dummy variable for ownership was significant. In fact, public agencies enjoyed total costs that were 10 percent lower than those of private firms (9). Fleet age, as in Miller's model, was only weakly significant. Possibly the same measurement problems he noted are the cause.

COMPARISON OF THE MODELS

Three distinct approaches to modeling costs have been presented (unit cost, four-variable regression, and slowness regression). To evaluate the relative merits of each, we assembled several sets of data and applied them to each model. Because it is not possible to check directly the predictive ability of the regression models, the criteria used to evaluate the model were (a) the degree to which the model explained variation in cost (R^2) and (b) the data requirements for estimating parameters and using the model.

Data from the D. C. Transit System, Inc., for the years 1962-1970 were used to compare the four-variable regression cost model with the slowness model. The resulting equations are as follows:

1. Four-variable model—$C = 28.095\ VH + 0.5488\ VM + 1.438\ DV - 183\ RP$ and $R^2 = 0.9966$, where VH = annual vehicle-hours in thousands, VM = annual vehicle-miles in thousands, PV = peak vehicles required, RP = annual revenue passengers in thousands, and C = annual operating cost.
2. Slowness function—$C/M = 0.9975 + 0.15\ S_3$ and $R^2 = 0.997$ (b for S_3 not significantly different from 0.0), where C/M = operating cost per mile, and S_3 = slowness expressed in minutes per mile.

Although the four-variable regression model explains variations in costs, when applied to future operating characteristics of D. C. Transit System, Inc., it underestimated costs by as much as 68.8 percent. It was thus rejected for the Washington study.

The slowness function cannot be calibrated from these data because the value of the coefficient assigned to S_3 is not significant. This is undoubtedly due to the nature of the data. Ten observations over time for the same company were used. If we assume that operating practices, technology, and operating environment remain constant, any variations in cost would be random or at least not explainable by a variable measuring speed of operation. These results are in contrast to Holden and Miller's analysis of the New York Transit operation. Using line-by-line data for the NYCTA (5), the following equation was determined: $C/M = 8.585 + 19.097\ S_3$ and $R^2 = 0.9888$. Apparently, because of the lack of change in transit systems over time, the slowness function can only be calibrated from data on individual routes of a single transit property at one point in time. The time series data were adjusted to a constant dollar; therefore, this should not be a source of variation.

On the basis of explanatory ability, the four-variable model is clearly superior for time-series data. For cross-sectional line-by-line data, the slowness function explains substantially the same amount of variation as the four-variable model. Neither model is especially suitable for explaining variations in costs when data from several properties are combined for regression analysis. Application of the four-variable model to the American Transit Association's sample of 69 firms yielded an R^2 of 0.52. The slowness function provided less explanatory capability.

The second criterion of data for evaluating the models is less important than the first. Size of data requirements is a factor only if a simpler model does an equally adequate job of prediction. In this case, the slowness function requires fewer pieces of input data but often does a poorer job of predicting. Using the Miami Transit Authority data presented by Ferreri, an R^2 of 0.45 was obtained with the slowness function, whereas using a two-variable cost function of the form C = a VM + b VH led to explanation of 99.5 percent of the variation in total cost. For the NYCTA data, the slowness function explained 99 percent of the variation. The four-variable cost function explained virtually 100 percent of the variation. For the Washington, D.C., data, the slowness function could not be used, whereas the four-variable function explained 99.66 percent of the variation. Annual mileage alone explains 65 percent of the variation.

The conclusion reached from this comparison is that the four-variable regression model is equal to, and usually superior to, the slowness function. Additional data requirements for the four-variable model are easily satisfied; therefore, there is no reason not to use the four-variable version.

To check the accuracy of the unit-cost model, data from the Pittsburgh Skybus and D.C. Transit System were used. The resulting equation for 1970 D.C. Transit System, Inc., data is C = 7.885 VH + 0.254 VM + 2,124 PV + 0.0121 RP. When the 1970 model was used to predict 1990 costs, the figure computed came within 5 percent of an estimate made in a detailed engineering cost study.

To further check the suitability of the unit-cost model for planning purposes, data from the Pittsburgh Skybus project were evaluated. Based on 1970 estimates, the following unit-cost model was determined (7): C = 0.496 VH + 0.148 VM + 0.008 RP + 10,321 PV. Applied in 1980 engineering estimates, this model predicted costs to within 3 percent of actual estimates of operating costs. The conclusion reached is that the unit-cost method of determining parameters appears to be an accurate method when used to predict future costs for the same system.

SUMMARY

The purpose of this comparative review of cost models for urban transit has been to identify a model, or models, that can estimate costs for alternative transit systems with sufficient accuracy for use in transportation planning. Five of the models estimate operating costs and are best suited for bus transit systems or other systems where operating costs constitute a substantial part of the total cost of operation. Capital costs, although difficult to model, are a relatively insignificant part of the total cost of operation for buses and are included in the sixth model. The simplest model in terms of data requirements is the slowness model; however, it is the least desirable for general application. The single independent variable only describes one of the cost determinants. It can only be used for cost estimation for additions to the property for which it was calibrated due to its stringent "ceteris paribus" requirements. The four-variable unit-cost model is much better suited to generalization even though it too does not encompass all factors influencing costs. The data requirements are not limiting. Recent experience of the authors using this model indicates that it is well suited for short-range planning purposes. As suggested by the Washington, D.C., study (1) described in this paper, the four-variable regression model is not as useful.

Efforts to formulate a general cost model, such as that by Miller, have been only partially successful. Further research is necessary to refine measurements of environmental variables such as urban density. Also, as Miller points out, better measure of fleet age and route structure might lead to a model better able to reflect cost variation from property to property.

REFERENCES

1. Traffic, Revenue, and Operating Costs. Alan M. Voorhees and Associates and W. C. Gilman and Co., 1969.

2. Working Paper No. 10: Bus Operating Cost Models. Alan M. Voorhees and Associates, McLean, Virginia, unpublished.
3. Ferreri, M. G. Development of a Transit Cost Allocation Formula. Highway Research Record 285, 1969, pp. 1-9.
4. Miller, D. R., and Holden, W. H. T. An Algorithm for Determining Transit Operating Costs. Transit Research Foundation of Los Angeles, unpublished.
5. City and Suburban Travel. Issue 105, Oct. 1969; Issue 106, Nov. 1969; and Issue 122, March 1971.
6. Miller, D. R. Cost Functions in Urban Bus Transportation. Northwestern Univ., PhD dissertation, 1967.
7. Transit Expressway Report. MPC Corporation, 1967.
8. 1970 Transit Operating Report. American Transit Association, Washington, D.C., 1971.
9. Wells, J. D., et al. Economic Characteristics of the Urban Public Transportation Industry. U.S. Govt. Print. Office, Washington, D.C., Feb 1972, Ch. 4, pp. 75-79.