Pergamon

Transpn Res.-A, Vol. 31, No. 6, pp. 459–473, 1997

0965-8564/97 $17.00 + 0.00

PII: S0965-8564(97)00003-7

SCALE ECONOMIES IN UNITED STATES RAIL TRANSIT SYSTEMS

IAN SAVAGE

Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, Illinois 60208, U.S.A.

(Received 31 October 1995; in revised form 23 November 1996)

Abstract—A three-staged least squares translog cost function is estimated for 13 heavy-rail and nine light-rail United States urban mass transit systems for the period 1985–1991. Firm output is taken to be endogenous. Large economies of density are found in operating costs. These economies become even more substantial when the cost of way and structure maintenance and capital costs are incorporated. However, there are constant returns to system size in short-run variable costs. © 1997 Elsevier Science Ltd

INTRODUCTION

This paper estimates operating cost functions for urban mass transit rail systems in the United States. A pooled dataset is used representing 13 heavy-rail (subway) systems and nine light-rail (tramway) systems for the period 1985–1991. Estimates are made of the existence of economies of density and system size.

ECONOMIC AND ECONOMETRIC THEORY

Long-run cost function

Consider a production function where provision of units of transit service, Y, are produced by five factors of production: way and structure (T), rolling stock (C), train operations labor (L), propulsion electricity (E) and automation (A). Many recently-constructed systems have used automated ticket issuing and inspection systems, and automatic train control to reduce the number of station staff and eliminate conductors.

$$Y = f(T, C, L, E, A) \tag{1}$$

We will assume that $f(\bullet)$ is continuously differentiable and strictly quasi-concave in inputs.

Duality theory permits definition of a total cost (TC) function:

$$TC = c(Y, H, P_t, P_c, P_l, P_e, P_a) \tag{2}$$

where P_i are the factor prices and H is the hedonic characteristics of output commonly used when a cost function is estimated using data from several firms. While Nash (1978) and others have suggested that objectives of transit firms may be different from the traditional profit maximization, duality theory still holds as it merely requires that the firm minimizes cost given output. For example, an output maximizing firm will wish to minimize cost to ensure that more units of output can be produced within a budget constraint.

Duality theory assumes that factor prices are exogenously determined. The price of way and structure, rolling stock, electricity and automation equipment are determined nationally or regionally and are clearly exogenous. Labor may be an exception in that it is commonly believed that management have acquiesced with transit unions to raise wages above those for comparable jobs. This paper will assume that the price of labor is exogenous.

This paper breaks with the prior literature in assuming that output is endogenous. While decisions concerning fares and levels of services are often decided in the political arena, it is unreasonable to assume that there is no managerial discretion. The assumption of endogenous output does not affect duality theory, but does require the use of instrumental variables in the econometric estimation of the cost function.

Capital versus operating costs

Inputs are of two types. Labor and electrical power are variable in the short run and costs are incurred on a continuing basis. Automation, way and structure, and cars require initial capital purchasing of a fixed quantity of the input, and then ongoing routine maintenance.

The capital costs are substantial. Table 1 presents some typical construction cost per route mile for new start or extension projects that have recently been completed or are in the final planning phase (Federal Transit Administration 1992; *Railway Age*, February 1994). Construction costs per route mile of at-grade track are about \$35 million for light-rail and \$100 million for heavy-rail. If tunnelling is required, these costs could double. Transit cars typically cost \$1.5 million per heavy-rail car and \$2.5 million for a light-rail articulated unit.

Long-run versus short-run cost functions

In the short run, the firm is not free to choose the quantities of all factors of production. The degree of automation, the quantity of way and structure, and the number of cars have been pre-determined. However, we will assume that changes in output can result in cars being transferred from the active to the reserve fleets, with consequent changes in maintenance expenses. The short-run production function is therefore:

$$Y = f(T^*, A^*, C, L, E) \tag{3}$$

where the superscript * indicates that the quantity of the input is fixed. Assuming that automation is regarded as a purely capital item, total short-run costs (SRTC) are represented by:

$$SRTC = P_{tm}T^* + P_{cm}C + P_1L + P_eE \tag{4}$$

where P_{tm} is the factor price of way and structure maintenance and P_{cm} is the factor price of car maintenance. The final three terms of eqn (4) will be defined as short-run variable cost (SRVC). The data used in this analysis separates out expenditures on way and structure maintenance from other short-run variable costs. Minimizing cost subject to a given level of output gives:

$$SRVC = f(Y, H, T^*, A^*, P_{cm}, P_1, P_e) \tag{5}$$

Fixed factors in a short-run cost function

If fixed and variable factors are substitutes, the coefficient on the quantities of way and structure and automation in an econometric estimation of eqn (5) should be non-positive. More of the fixed factor should lower variable costs. The empirical work supports this assertion for the case of automation.

However, way and structure is a complementary factor of production with the variable factors. Expanding the quantity of way and structure by extending the length of the system increases rather than reduces variable costs. A larger system, even when holding output constant, will require more stations and require more ticket agents, station and cleaning staff, dispatchers and signalling staff. There may be more power losses from the electrical supply system and the stations will have to be lit. Caves *et al.* (1981) found a similar result in their analysis of U.S. class I railroads. Their result was based on more ideal data on the value of the stock of way and structure rather than the purely physical quantities available for this study.

Table 1. Construction cost per route mile for different types of systems

System type	Cost per route mile	Examples
Heavy-rail (subway)	\$210 million	Los Angeles
Heavy-rail (mixed aerial, subway, grade)	\$150 million	Atlanta, Washington, D.C.
Heavy-rail (mixed aerial, existing railroad alignment)	\$55 million	Chicago
Light-rail (at grade)	\$35 million	Dallas, San Diego
Light-rail (existing railroad alignment)	\$20 million	St Louis

Definition of output

There has been a discussion in the literature as to whether passenger miles 'demand related output' or car hours 'technical output' should be used to measure output. Clearly the ultimate 'output' of a transit system is the number of passenger miles. However, a cost function is derived by duality theory from a production function, and production functions are an expression of an engineering relationship between physical inputs and physical outputs. Therefore this analysis takes the view that the operation of trains, which we measure as car hours in revenue service, is the appropriate primary measure of output. However, there are expenses that are driven by the number of passengers carried. A load factor variable, calculated as passenger miles divided by revenue car miles, is used as a second output measure. Clearly these two measures are related as a greater number of car hours will induce higher demand for the system.

Study objectives

The objective of the research is to investigate the economies of both density and system size. Economies of density (ED) are found by varying the amount of output over a fixed system:

$$ED = (\partial \ln SRCV/\partial \ln Y)^{-1} \quad (6)$$

where values of ED greater than unity, equal to unity, or less than unity indicate increasing, constant and decreasing returns to density, respectively. While our primary measure of output is car hours, the correct definition of economies of density will also incorporate the effect car hours has on load factors and hence on cost (Jara-Díaz and Cortés, 1996). Because demand is typically believed to be inelastic with respect to service frequency, an increase in car hours will depress average load factor (Pratt *et al.*, 1977).

Economies of system size (ES) are given by:

$$ES = ((\partial \ln SRVC/\partial \ln Y) + (\partial \ln SRVC/\partial \ln T))^{-1} \quad (7)$$

Again values of ES greater than unity, equal to unity or less than unity indicate increasing, constant and decreasing returns to system size, respectively. Basically, this calculation is investigating the proportionate effect on costs of operating, say, twice as many trains over a system that is twice as large. The load-factor effect is not relevant here because service frequencies, and hence the attractiveness of the system, do not change.

This work does not investigate elasticities of substitution between inputs. The three variable factors of production are complementary in nature, and measurement problems in defining factor prices would make any calculated elasticities very approximate.

Functional form

In common with most analysts of transportation costs, the flexible transcendental logarithmic, translog, function has been used. This functional form provides a second-order numerical approximation to almost any underlying cost function at a given point on that cost function. Like most studies, mean variable values are used as the point of approximation. Output and factor prices are assumed to be separable. This restriction carries with it the assumption that the factor shares are invariant with firm size. That is to say, we are requiring that, for example, the proportion of total expenses represented by labor are the same, *ceteris paribus*, for both large and small transit systems. This is a major restriction, but is justified in the present work for reasons of conserving degrees of freedom, and of ensuring concavity on factor prices for nearly all of our observations.

The general form of the estimated equation is:

$$\begin{aligned}\ln SRVC = &\sum_i a_i \ln Y_i + \beta_i \ln T + \sum_i \gamma_i \ln H_i + \sum_i \theta_i D_i + \frac{1}{2}\sum_i\sum_j \alpha_{ij}(\ln Y_i)(\ln Y_j)\\ &+ \frac{1}{2}\beta_{tt}(\ln T)^2 + \frac{1}{2}\sum_i\sum_j \gamma_{ij}(\ln H_i)(\ln H_j) + \sum_i \phi_{it}(\ln Y_i)(\ln T)\\ &+ \sum_i\sum_j \phi_{ij}(\ln Y_i)(\ln H_j) + \sum_j \phi_{tj}(\ln T)(\ln H_j) + \sum_i \lambda_i \ln P_i + \frac{1}{2}\sum_i\sum_j \lambda_{ij}(\ln P_i)(\ln P_j)\end{aligned} \quad (8)$$

where $\alpha_{ij}=\alpha_{ji}$, $\gamma_{ij}=\gamma_{ji}$, $\lambda_{ij}=\lambda_{ji}$, H_i are continuous output characteristics and D_i are discrete output characteristics. All variables, except for the discrete variables, have been expressed as a ratio to their means prior to taking of logarithms. Use of Shephard's Lemma gives the following share equations:

$$S_i = \frac{\delta \ln SRVC}{\delta \ln P_i} = \lambda_i + \sum_j \lambda_{ij} \ln P_j \tag{9}$$

To ensure that the cost function is homogenous of degree one in factor prices the following restrictions were imposed:

$$\sum_i \lambda_i = 1, \sum_j \lambda_{ij} = 0 \tag{10}$$

It is assumed that eqns (8) and (9) have classical additive disturbances, and that they can be estimated as a multivariate equation system. The system of equations can be estimated using a technique proposed by Zellner (1962). Only i–1 share equations are required for the estimation. Share equations were used for propulsion electricity and rolling stock maintenance. The endogenous nature of both output variables was attended to by instrumenting and the use of three-stage least squares for the estimation of the system of equations.

DATA AND VARIABLES

Data source and sample selection

In the U.S. all transit operators are required to file a standard annual operating and financial data report to the Federal Transit Administration (formerly the Urban Mass Transportation Administration). These 'section 15' data are available in an annual publication by the American Public Transportation Association.

A pooled dataset has been constructed for 22 systems over the period 1985–1991. All dollar values have been inflated to 1991 dollars using the consumer price index. Because the 1985 data were incomplete for some systems, and because other systems commenced operation after 1985, there are 124 observations. Descriptive statistics on the systems are shown in Table 2 indicating whether they are of heavy- or light-rail construction, the system size (measured in directional route miles) and the number of car-hours operated. The sample should provide fertile data for estimation of economies of density and system size because in addition to variations cross-sectionally in system size, individual systems have seen changes in output over the 1985–1991 period.

Cost variable

This analysis is concerned with estimating short-run variable costs. The data used are 'total mode expense' less 'non-vehicle maintenance.' Maintenance of way and structure is therefore excluded. The data do not include any capital expenditures or charges.

Output

The primary output measure is revenue car hours. Car hours are used in preference to car miles because many cost items, particularly labor, are incurred on an hourly basis. Only car hours incurred in revenue service are counted because there are anomalous data on total car hours for some systems regarding the operating of maintenance trains which may operate for hours without moving very far. Data for individual systems were checked for consistency over time by comparing car hours and car miles. In a couple of cases, anomalies were found for individual years and some adjustment of data was necessary. The second output measure represents passenger usage. A load factor variable is used, calculated as passenger miles divided by revenue car miles.

Fixed factors

The measure of way and structure, and hence system size, is directional route miles. This is calculated as miles of road multiplied by two if traffic moves in both directions. This measures approximate mainline track miles. This measure does not double count when two transit lines share the same track.

The measure of automation is a discrete variable taking the value 1 for 'highly automated' systems. For heavy-rail this is defined as systems with automatic ticketing systems, automatic train control and one-person-operation of trains. This applies to the Washington D.C, San Francisco BART, Atlanta, Miami, Philadelphia (Lindenwold) and Baltimore systems. For light-rail systems the three new systems (San Diego, Buffalo and Portland) are also defined as 'highly automated.'

Engineering characteristics

Three variables represent the predetermined engineering nature of the system. Transit systems are basically of two types: heavy-rail and light-rail. Heavy-rail systems have a great affinity to regular railroads and feature segregated right-of-way, subway construction and/or heavy earthworks, and traditional operating practices and signalling systems. These systems should be more costly to operate than the light-rail systems that have evolved from the streetcar. A dummy variable taking the value 1 for light-rail is used.

Most light-rail systems have either been constructed recently or are refurbishments of older streetcar systems. However, traditional streetcar operation with vintage cars still exists in New Orleans and Newark. As can be seen in the final column of Table 2 they have unusually low costs. For these two systems a dummy variable, called 'streetcar' is used in addition to the light-rail dummy variable.

A third, continuous, variable is the proportion of track miles that are at grade rather than elevated or in tunnel. Information on individual systems was obtained from Jane's (annual) and UITP (1985).

Output characteristics variables

Two continuous variables are used to represent output characteristics. The first is average journey length, calculated as passenger miles divided by passenger journeys. Systems serving long-distance commuting markets may well have different cost characteristics from systems providing short-trip inner-city markets. While average journey length has been used in this analysis, an average speed variable, revenue car miles divided by revenue car hours, would serve equally well in representing this effect.

The second variable is used to investigate the effect of excess 'peaking' on costs. This is measured by the ratio of morning peak car requirement to midday car requirement. We will refer to this as the peak-to-base ratio. Highly peaked systems should incur higher unit costs as cars are used on average for fewer hours per day and labor is less productive.

Table 2. System descriptions—1991

System (LR = light-rail)	Data for	Revenue car hours (million)	Directional route miles	Variable cost per car hour
Newark (LR)	1985–1991	0.05	8	$80
Cleveland (LR)	1986–1991	0.05	27	$175
Portland, Oregon (LR)	1988–1991	0.07	30	$120
Buffalo (LR)	1988–1991	0.07	12	$114
Cleveland	1986–1991	0.08	38	$175
New Orleans (LR)	1987–1991	0.09	17	$51
New York (Staten Island)	1988–1991	0.10	29	$141
Baltimore	1990–1991	0.15	27	$140
Pittsburgh (LR)	1988–1991	0.15	65	$110
Philadelphia (Lindenwold)	1985–1991	0.15	32	$123
Miami	1990–1991	0.18	42	$181
San Diego (LR)	1985–1991	0.23	41	$62
San Francisco (MUNI) (LR)	1986–1991	0.39	50	$137
Philadelphia (SEPTA) (LR)	1985–1991	0.51	127	$81
Atlanta	1985–1991	0.61	67	$75
New York (PATH)	1986–1991	0.63	29	$188
Philadelphia (SEPTA)	1985–1991	0.96	76	$113
Boston	1986–1991	1.14	77	$144
San Francisco (BART)	1985–1991	1.37	142	$117
Washington, D.C.	1986–1991	1.50	156	$133
Chicago	1985–1991	2.55	191	$96
New York Transit Authority	1985–1991	16.22	493	$113

Factor prices

The wages of train operators can be calculated from Section 15 data which report the total train operators wage bill and the number of vehicle operator equivalents. Operator wages, when allowance is made for fringe benefits, represent about 20% of variable costs. Operator wages can be used as a surrogate for the level of wages of other labor such as conductors and station staff. Stern *et al.* (1977) report that in union negotiations the agreed operators' wage is used as the benchmark for all other wages.

All of the systems use electric propulsion. Section 15 data only reports the number of kilowatt hours (KwH) of propulsion electricity. Price per KwH, on a state by state basis, are available from the U.S. Department of Energy. Prior to 1991 price data are divided into domestic and commercial customers. From 1990 a third category was introduced for sales to "public authorities, railways, railroads and interdepartmental sales." For 1990 and 1991, a calculation was made of the ratio of the prices in this category to commercial prices for each state. The ratio was then applied to the commercial prices for 1985–1989 to obtain a complete series of state level prices relevant for sales to transit systems. There is considerable variation in prices from state to state, from 5 cents per KwH in California and the pacific northwest up to 15 cents per KwH in certain east coast states. Propulsion costs average about 10–15% of variable costs.

A factor price of car maintenance was obtained by dividing total expenditures on car maintenance by the peak car requirements. Objections can be raised that the resultant 'price' is not totally exogenous. However, the calculated prices do seem to be consistent with intuitive observations about the age profile, design and complexity of the rolling stock used by the various systems. Car maintenance represents about a quarter of total variable costs.

Instrumental variables on output

Two variables were used to instrument the output variables. Both were exogenous variables that influence demand for transit service. The first is the density of population per square mile of the urban area. Highly suburbanized cities, such as Atlanta, have low average density and residences and workplaces are located in places difficult to serve by transit. This contrasts with more traditional cities such as New York, Chicago, Philadelphia and San Francisco.

The second variable also represents urban form by measuring annual vehicle miles travelled per head of population. While one can argue that this variable is endogenous in that the quality of transit service affects vehicle ownership and use decisions, investigation of the data indicate that they accord more with the urban form of the various cities.

Both variables were obtained from Schrank *et al.* (1994). Figures were converted from kilometers to miles to be consistent with the rest of the dataset. Data were not given for Buffalo, so data for another city were substituted. The nearest city, both geographically and structurally, for which data were given was Cincinnati.

REGRESSION RESULTS

The results of the estimations are shown in Table 3. Results for the factor share equations are not shown as the estimated coefficients for these equations are repeated in the main equation. The estimated equation was consistent with economic theory. For all observation points predicted marginal costs were positive. In addition, at 118 out of the 124 observations the estimated cost function was concave in factor prices. The exceptions were 1990 for the San Francisco BART system and for five of the six years for the San Francisco MUNI light-rail system. The cause of the trouble appears to be in the electricity factor prices for these systems. The following paragraphs interpret the regression results.

Economies of density

Economies of density will depend both on the effects on costs of changes in car hours and any resulting impact on load factors. The first component can be measured by the inverse of the coefficient on car hours. At the sample mean, a value of 1.50 is calculated which is statistically greater than unity and implies economies of density. Increasing the number of car hours operated over a fixed system leads to a less than proportionate increase in variable costs.

Table 3. Translog regression on logarithm of short run variable costs

Explanatory variables (logarithms except for dummy variables)	Coefficient	t
Car hours	0.688	6.14
Directional route miles	0.380	5.13
Load factor	0.592	2.75
Average journey length	−0.266	1.25
Peak-base ratio	0.209	0.91
Proportion at grade	−0.337	1.95
Highly automated dummy variable	−0.272	5.01
Light-rail dummy variable	−0.199	3.72
Streetcar dummy variable	−0.278	3.50
Car hours^2	−0.076	0.52
Directional route miles^2	−0.159	0.62
Load factor^2	−1.052	1.82
Journey length^2	−0.485	2.49
Peak-base ratio^2	0.061	0.21
At grade^2	−0.129	1.69
Car hours × directional route miles	0.099	0.52
Car hours × load factor	0.421	2.30
Car hours × journey length	−0.163	0.79
Car hours × peak-base ratio	−0.248	1.28
Car hours × at grade	−0.143	0.71
Directional route miles × load factor	−0.583	2.14
Directional route miles × journey length	0.410	1.59
Directional route miles × peak-base ratio	0.397	1.45
Directional route miles × at grade	0.200	0.63
Load factor × journey length	0.047	0.17
Load factor × peak-base ratio	0.800	2.34
Load factor × at grade	0.167	1.39
Journey length × peak-base ratio	−0.368	1.87
Journey length × at grade	−0.340	1.49
Peak-base ratio × at grade	0.068	0.38
Labor factor price	0.629	116.60
Electricity factor price	0.115	36.24
Car maintenance factor price	0.256	61.30
Labor factor price^2	0.108	6.47
Electricity factor price^2	0.059	9.13
Car maintenance factor price^2	0.091	9.17
Labor price × electricity price	−0.038	4.37
Labor price × car maintenance price	−0.070	6.06
Electricity price × car maintenance price	−0.021	3.67
Number of observations	124	
Adjusted R^2—main equation	0.99	
Adjusted R^2—electricity share equation	0.41	
Adjusted R^2—car maintenance share equation	0.28	

If the service frequency elasticity of demand is inelastic; the magnitude of the density economies will increase when the effects of increased car hours on load factor are included. The estimated cost function indicates that load factor is positively related to cost (doubling load factor is predicted to increase costs by 59%). Increased car hours will reduce average load factors and reduce variable costs. The 0.65 frequency elasticity calculated by Pratt *et al.* (1977) implies that a 10% increase in car hours will reduce average load factor by 3.2%. Economies of density at mean values will be:

$$1/(0.668 + (-0.32 * 0.592)) = 2.08 \tag{11}$$

Calculations in the previous paragraph allow us to ask a related question concerning economies of density when the number of passengers changes but output of car hours remains constant. Fares changes or changes in land use that result in increased passenger numbers will lead to a reduction in average variable cost per passenger mile. Economies of density are calculated by the inverse of the coefficient on the load factor variable. At the sample mean, the value is 1.70. Conversely, loss of ridership due to exogenous factors such as suburbanization will lead to increased average variable costs for the remaining passengers.

Economies of system size

Economies of system size are found by looking at the inverse of the sum of the coefficients on the car hours and directional route miles variables. The intuition is that the size of the fixed factor is expanded with a similar level of service offered on the new trackage as on the existing system.

The calculation is complicated because there are square and cross-terms in output and the fixed factor. If these are varied, with car hours changing at a 35% faster rate than route miles to account for the higher density of service found in large systems, but with hedonic characteristics held at mean values, the smallest small systems are estimated to have diseconomies of system size of 0.97, while the largest systems have economies of system size of 1.05. While this implies an inverted U-shaped average variable cost curve, the extent of economies or diseconomies of system size would not be regarded as large and constant returns to system size cannot be statistically rejected at any point.

Automation and engineering effects

Investments in modern automated ticketing and train control systems result in a 27% reduction in short-run variable costs compared with comparable older systems.

Although there may be constant returns to system size, smaller systems do have lower average costs in absolute terms. This is because most of the smaller systems are light-rail. Light-rail systems have 20% lower costs than comparable heavy-rail systems. The very basic streetcar systems, which are represented by both the light-rail and streetcar dummy variables, have costs 42% below a heavy-rail system. This is before allowing for the fact that most light-rail systems are at grade level while heavy-rail is often in tunnel or elevated. Incorporating the at-grade effect, at mean levels, an at-grade light-rail system will have short-run variable costs that are 57% less than a tunnelled or elevated heavy-rail system.

Output characteristics

The effect of hedonic variables is often difficult to interpret in a translog function with many second order terms. At mean values, average journey length and peak-base ratio are insignificantly related to cost. When allowance is made for second order terms, strong relationships are found. As an illustration, a three-stage least squares estimation of a Cobb–Douglas function produces t-statistics of 8.2 and 13.5, respectively, for these hedonic variables.

At mean values, doubling average journey length while holding total passenger miles constant reduces costs by 27%. Systems with longer average journey length have greater station spacings and higher operating speeds. Average journey length does vary considerable. Heavy-rail systems which provide circulation in the downtowns of traditional east-coast cities have an average journey length of under five miles. In contrast, the commuter-orientated BART system in San Francisco has an average journey length of 12 miles.

Systems with high peak-to-base operations are considerably more costly than systems with a more consistent level of operations across the day. Some systems do not offer any enhanced peak service, but normally the peak-base ratio varies from 1.5 up to 4 or more for commuter-oriented systems. At mean values, doubling peak-base ratio while holding car hours constant increases short-run variable costs by 21%. This cost disadvantage would be compounded when allowance is made for the capital costs of the additional cars that are used in the peak only.

GENERIC ESTIMATES OF ECONOMIES OF DENSITY AND SYSTEM SIZE

The above calculations of economies of density and system size at mean values only provide a limited picture. More general inferences can be made by plotting average variable cost curves for six generic system types: streetcar systems, new light-rail systems, traditional light-rail systems, new commuter heavy-rail systems, new heavy-rail systems and traditional heavy-rail systems. The 22 systems were divided into these six categories and average values for the explanatory variables are calculated. These are shown in Table 4.

Figure 1 deals with economies of density. System characteristics and system size (directional route miles) variables are held at mean values for each of the six system types. The number of car hours was then varied and the load factor adjusts based on a service frequency elasticity of 0.65. Average variable cost curves for the six generic system types are plotted over the range of output

Table 4. Six generic system types

	Streetcar	New light-rail	Traditional light-rail	New commuter heavy-rail	New heavy-rail	Traditional heavy-rail
Mean values of characteristics that will be varied						
Car hours	65 650	123 700	272 900	564 600	752 400	3 098 000
Directional route miles	12.6	27.8	66.9	71.9	83.2	133.0
Load factor	19	25	26	22	22	21
System characteristics that will be held fixed						
Journey length	2.5	5.0	4.3	9.5	5.0	5.3
Peak-base ratio	1.7	1.9	2.6	4.5	2.7	2.2
At-grade	0.81	0.74	0.89	0.21	0.48	0.40
Highly automated	0	1	0	1	1	0
Light-rail	1	1	1	0	0	0
Streetcar	1	0	0	0	0	0
Labor price	28 400	41 000	29 600	35 100	35 200	33 600
Electric price	0.114	0.060	0.083	0.077	0.077	0.097
Car maintenance	54 900	95 000	97 600	87 000	81 000	85 200
Systems	Newark New Orleans	Portland Buffalo San Diego	Cleveland Pittsburgh MUNI Philadelphia	Lindenwold Miami BART	Baltimore Atlanta Washington	Cleveland Staten Island PATH Philadelphia Boston Chicago New York

applicable to actual systems of that type. The downward sloping average cost curves indicate that economies of density can be expected at all output levels and for all types of systems. The figure also illustrates the powerful effect of some of the system characteristic variables. Light-rail systems are less expensive than heavy-rail systems and streetcar systems are even more inexpensive. Investment in automation also reduces costs.

Figure 2 represents economies of system size. System characteristics are held at mean values while directional route miles and car hours were varied. The ratio of car hours to directional route miles is held constant at its mean value. The plotted average cost per directional route mile curves support the earlier finding of constant returns to system size. While there is some evidence of an inverted U-shape, the cost functions are remarkably flat over substantial ranges.

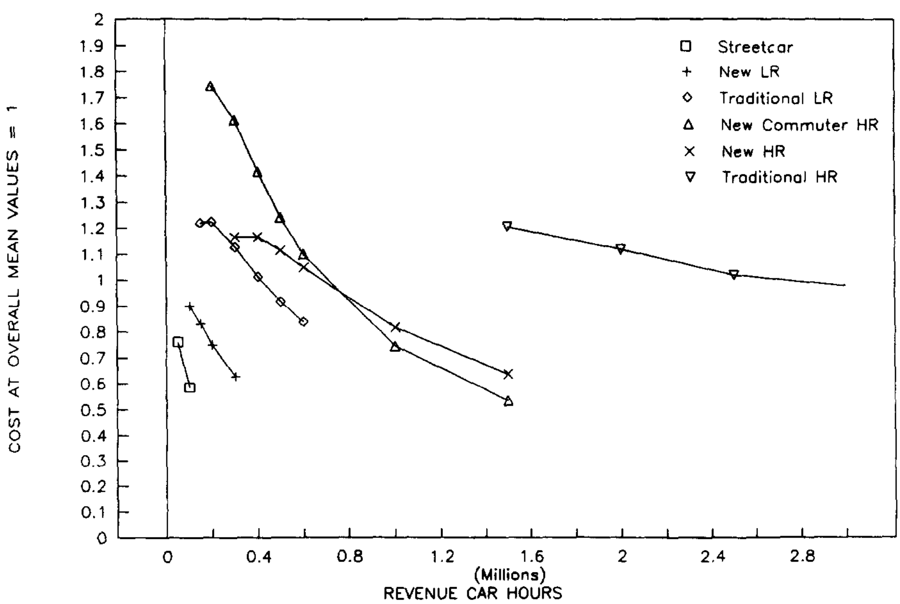

Fig. 1. Generic average variable cost per car hour for a fixed system size.

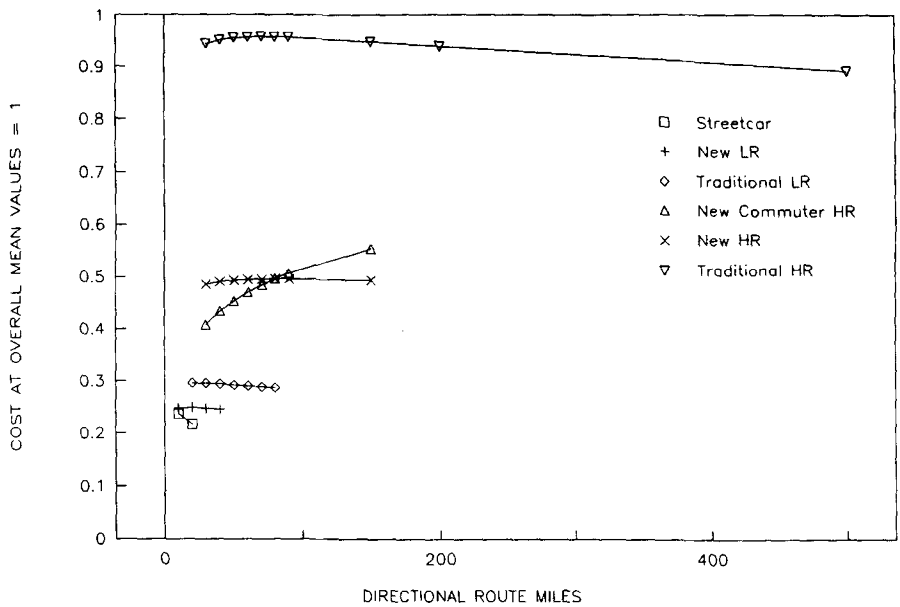

Fig. 2. Generic average variable cost per directional route mile.

The major exceptions are the newer heavy-rail systems built to serve longer-distance commuter traffic. Diseconomies of system size are found for these systems. This result is driven by the cross-terms in the translog equation between directional route miles and average journey length, and directional route miles and peak-to-base ratio. These types of systems are characterized by long average trip length and very peaked operation. As these systems get larger, they appear to become less efficient in dealing with the large number of cars and staff that are required for comparatively short periods of time each day.

LOCAL ESTIMATES OF ECONOMIES OF DENSITY AND SYSTEM SIZE

Table 5 shows point estimates of economies for each of the systems. It shows economies of density, measured both with and without the load factor effect, and economies of system size. Remember that the economies of density including the load factor effect are based on an assumed service elasticity of demand.

Constant returns or economies of density apply to nearly all systems, with the marked exception of Philadelphia's heavy-rail system. The systems which show least evidence of economies of density in short-run variable costs have the following characteristics in common: track that is heavily utilized (car hours divided by directional rout miles) and a relatively flat level of service across the day (low peak to base ratio). The point estimates for economies of system size are very supportive of the generic graphs shown in Fig. 2. Many systems have estimated economies of system size close to unity, except those systems catering to longer-distance commuter traffic.

ECONOMIES OF DENSITY AND SYSTEM SIZE IN TOTAL COSTS

Calculations thus far deal with short-run variable costs and ignore capital costs and the annual maintenance costs for the fixed factor. Incorporation of the latter costs permits estimation of the full magnitude of economies of density.

Way and structure maintenance costs

Investigation of way and structure maintenance is necessary to find out: (1) whether there are economies of scale in way and structure maintenance, which affects economies of system size and (2) whether changing the density of service changes way and structure maintenance costs which

Table 5. Point estimates for individual systems in 1991

System (LR = light-rail)	Economies of destiny		Economies of system size
	Car hours	Car hours and load factor	
Newark (LR)	1.63	3.13	0.99
Cleveland (LR)	1.54	1.72	0.89
Portland (LR)	1.35	1.34	0.94
Buffalo (LR)	1.02	1.09	1.27
Cleveland	1.52	1.60	0.89
New Orleans (LR)	1.06	1.04	1.32
New York (Staten Island)	1.79	2.30	0.87
Baltimore	1.87	4.70	0.89
Pittsburgh (LR)	1.29	1.21	0.97
Philadelphia (Lindenwold)	2.04	7.81	0.78
Miami	1.05	1.05	0.99
San Diego (LR)	1.13	0.97	1.04
San Francisco (MUNI) (LR)	1.08	1.05	1.27
Philadelphia (SEPTA) (LR)	1.12	1.06	1.31
Atlanta	1.22	1.23	1.08
New York (PATH)	1.52	2.59	0.99
Philadelphia (SEPTA)	0.71	0.63	1.49
Boston	0.99	1.08	1.28
San Francisco (BART)	1.61	1.91	0.86
Washington, D.C.	1.28	1.39	1.05
Chicago	1.92	3.86	0.94
New York Transit Authority	0.96	0.98	1.33

affects economies of density. Cobb–Douglas cost functions were estimated because the engineering characteristics of heavy-rail and light-rail systems vary so markedly as to require separate regressions with consequent reduction in possible degrees of freedom.

The output measure was the number of track miles with a hedonic output variable representing track usage. This is measured as annual revenue car miles divided by track miles. The engineering characteristics of systems were represented by a number of variables. The first is the average number of stations per track mile used. This variable only appears in the heavy-rail regression as the definition of a station is far more ambiguous for light-rail where some stops in street operation are nothing more than a pole and flag. A variable is used to represent the age of the system, measured by years from opening. Many of the systems have opened relatively recently and should not require the major track and station rehabilitation that older systems need.

The final engineering variables relate to the proportion of the system that is at grade, elevated or in tunnel. Information on individual systems was obtained from Jane's (annual) and UITP (1985). Light-rail systems are predominantly at grade with limited center-city tunnels on some systems. For light-rail a variable indicating the proportion of track miles at grade is used. For the heavy-rail systems two variables are used: the proportion of tack miles that are elevated and the proportion in tunnel.

A factor price variable for labor was included. The annual wage of train operators was used because the operator's wage is often used as a benchmark in union wage bargaining. Other maintenance costs, such as materials, tend to have prices that are consistent nationwide.

Estimation results are shown in Table 6. While the point estimates might suggest mild diseconomies of scale in way and structure maintenance, constant returns to scale cannot be rejected for both heavy-rail and light-rail. The amount of traffic is positively related to way and structure maintenance cost, but the effect is only statistically significant for heavy-rail systems. For these systems doubling the density of service will increase way and structure maintenance costs by 33%.

Other variables take expected signs. At-grade track is less expensive to maintain than tunnels but more expensive than elevated structures. The latter do not require maintenance of extensive earthworks and right of way, and permit the elimination of grade crossings. The frequency of stations is significantly positively related to total way and structure maintenance costs for heavy-rail systems. Newer light-rail systems appear to have lower maintenance costs than older systems.

Table 6. Regression on logarithm of non-vehicle maintenance costs

	Heavy-rail		Light-rail	
Explanatory variables (all in logarithms)	Coefficient	t	Coefficient	t
Constant	−0.361	7.72	0.157	1.13
Track miles (coefficient is compared with 1 for t-test)	1.068	0.95	1.052	0.25
Proportion at grade	—	—	−0.986	3.17
Proportion elevated	0.006	0.19	—	—
Proportion in tunnel	0.088	2.83	—	—
Car miles per track mile	0.334	2.60	0.521	1.17
Stations per track mile	0.641	4.29	—	—
Years since opening	0.033	0.61	0.205	1.80
Labor factor price	1.149	4.01	−0.572	0.95
Number of observations	74		50	
Adjusted R squared	0.95		0.77	

Annualized capital costs

Calculations of annual capital costs are very rough and basic. Nearly all capital expenditures of transit systems are supported by federal and local grants. Transit systems therefore do not show allowance for capital replacement in their annual accounts in the same way that a commercial corporation does. Based on Table 1 we will assume that heavy-rail construction is $200 million per route mile for tunnel track and $100 million a mile for elevated or grade track; the equivalent figures for light-rail are $100 million and $35 million, respectively. We will assume that provision for replacement of way and structure is made on an equal annual basis over 80 yr. Rolling stock costs $1.5 million for a heavy-rail car, $2.5 million for a light-rail unit and $1 million for a streetcar. Rolling stock is taken to have a 25 year life before it requires replacement or major refurbishment.

Estimated economies of density and size in total cost

The second column of Table 7 shows local estimates of economies of density when the costs of way and structure maintenance are included. Allowance was made for the findings in Table 6 that increased track usage has a cost elasticity of 0.33 for heavy-rail and 0.52 for light-rail. In the third column annualized capital costs are also included, where additional car hours are assumed to

Table 7. Estimation of economies including way and structure (W&S) maintenance and capital costs in 1991

	Incorporating W&S maintenance	Incorporating W&S maintenance and capital costs	
System (LR = light-rail)	Density	Density	System size
Newark (LR)	2.97	3.56	0.99
Cleveland (LR)	1.78	1.96	0.95
Portland (LR)	1.39	1.96	0.97
Buffalo (LR)	1.24	1.74	1.08
Cleveland	1.87	3.31	0.96
New Orleans (LR)	1.06	1.76	1.13
New York (Staten Island)	2.44	3.46	0.94
Baltimore	3.94	4.25	0.95
Pittsburgh (LR)	1.43	1.95	0.98
Philadelphia (Lindenwold)	5.66	4.73	0.91
Miami	1.27	1.84	0.99
San Diego (LR)	1.06	1.43	1.01
San Francisco (MUNI) (LR)	1.12	1.33	1.13
Philadelphia (SEPTA) (LR)	1.14	1.54	1.10
Atlanta	1.45	2.21	1.02
New York (PATH)	2.61	2.53	0.99
Philadelphia (SEPTA)	0.75	1.16	1.17
Boston	1.35	1.63	1.10
San Francisco (BART)	2.08	2.65	0.92
Washington, D.C.	1.61	2.16	1.01
Chicago	3.65	3.27	0.96
New York Transit Authority	1.16	1.34	1.15

require a proportionate increase in fleet size. Considerable economies of density are found, even for the Philadelphia heavy-rail system. To illustrate graphically the magnitudes of the economies of density, Fig. 3 shows plots of average total cost curves for the six generic system types used earlier in the paper.

The final column of Table 7 is a calculation of economies of system size when way and structure maintenance and capital costs are taken into account. In making these estimates the point estimates of Table 6, that there are maintenance diseconomies in track miles of 0.936 for heavy-rail and 0.951 for light-rail, are incorporated. A similar pattern of system size economies to that found in short-run variable costs is observed.

COMPARISON WITH PREVIOUS LITERATURE

The three previous analyses of North American urban transit costs used an Institute for Defense Analyses (1972) dataset on 11 North American systems for the 11 years 1960–1970. Pozdena and Merewitz (1978) and Viton (1980, 1993) found diseconomies of density in short-run variable costs for all of the large systems and in the latter paper for all of the systems. This is at odds not only with this paper but also with evidence from mainline railroads. Using translog formulations, both the cross-sectional study by Caves *et al.* (1981) and the time-series analysis by Braeutigam *et al.* (1984) find substantial economies of density.

It is my opinion that the earlier results are explained by the dataset used. The dataset is very unbalanced in that it contains some very large heavy-rail systems and some very small streetcar-type systems, with no intermediate observations. The past 25 years have seen construction of a number of middle-sized systems which help to balance the dataset. In Viton's 1993 paper a cost function with track miles as the fixed factor predicted negative marginal costs for all observations.

PUBLIC POLICY IMPLICATIONS

Pricing implications

The marginal cost of an additional car hour can be calculated for each system using the product of the inverse of the point estimate of economies of density (incorporating the load factor effect) and the short-run average variable cost. Such a calculation does not include the effects of additional car hours on way and structure wear-and-tear or the purchase of additional rolling stock. Dividing by the average passenger miles per car hour for the system permits the calculation of the marginal cost per passenger mile generated by the marginal car hour.

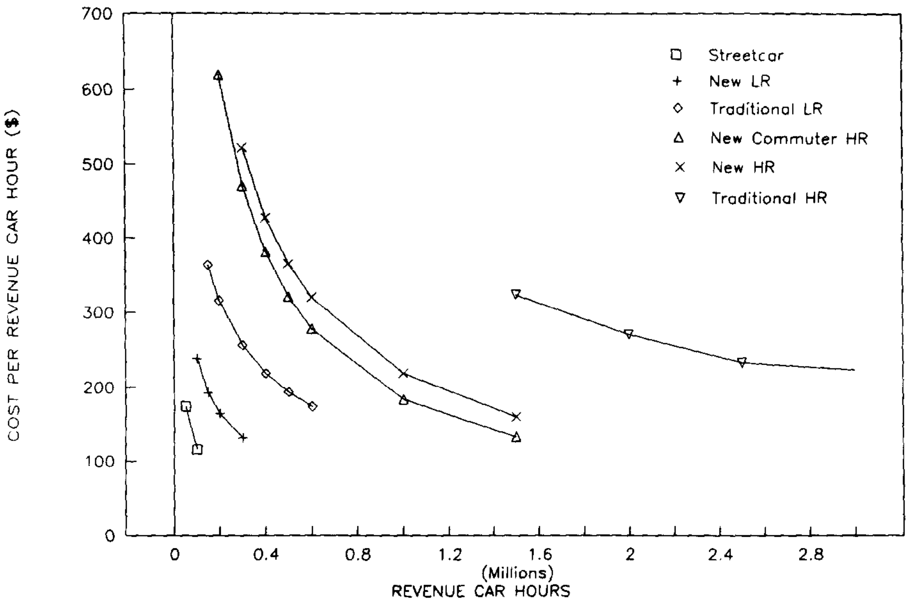

Fig. 3. Generic average total cost per car hour for a fixed system size.

The marginal cost can be compared with the fare per passenger mile. Calculations on the revenue side are complicated because, unlike costs, data are reported for the whole transit agency and do not distinguish between revenue collected from bus and rail operations. Only five of the systems are primarily rail-only operators. However, providing a consistent structure is used, our calculation based on passenger miles rather than passenger trips should provide a close estimate of price per passenger mile on the rail system.

Figure 4 plots price against marginal cost for the 22 systems in 1991. Systems with low marginal costs are able to price close to, or in excess of, marginal cost. However, more expensive systems are unable to pass on their high costs in prices. This suggests that political pressures may be capping fares at about 20 cents a mile.

Implications for transit construction

There are currently numerous proposals for construction of extensions to existing heavy-rail and light-rail systems, and the building of entirely new light-rail systems. There has been considerable controversy over the accuracy of cost and revenue estimates used when seeking funding for these projects (Pickrell, 1989, 1992). The equations estimated in Tables 3 and 6 provide a possible method for the federal government to initially evaluate operating cost estimates provided by funding applicants. Currently some smaller communities are proposing limited light-rail schemes. These very small schemes should be able to operate with similar average costs to those systems found in larger cities.

Economies of size and the privatization debate

The findings of this research provide input to the continuing debate about the competitive contracting or privatization of urban transit systems. The considerable economies of density make mass rail transit routes natural monopolies. Short-term franchises are a possible method of introducing a competitive environment given that direct competition by rival companies over the same track is undesirable.

It is sometimes suggested that the larger systems could be divided into smaller operating units prior to privatization. While our analysis does not directly address the issue of economies of scope between lines, the constant returns to system size suggest that there would be negligible cost disadvantage to breaking up firms into smaller component parts.

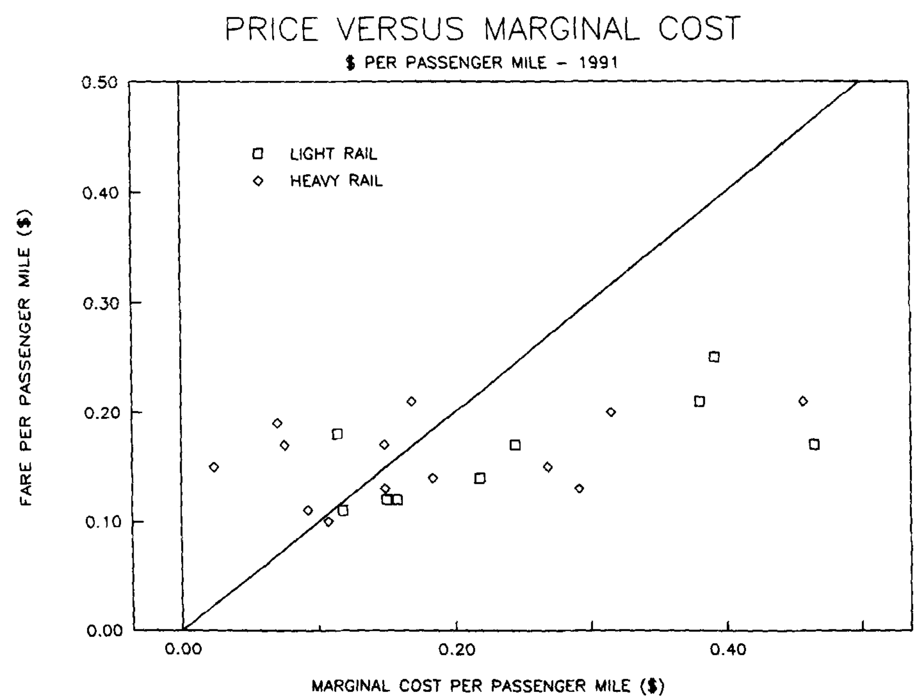

Fig. 4. Fare and marginal cost per passenger mile.

Acknowledgements—This work was financially supported by the U.S. Department of Transportation through a University Transportation Centers grant to the Great Lakes Center for Truck and Transit Research, University of Michigan, Ann Arbor. The excellent comments of several referees greatly improved this paper.

REFERENCES

American Public Transportation Association. *Transit Operating and Financial Statistics.* APTA, Washington, DC, (Annual).

Braeutigam, R. R., Daughety, A. F. and Turnquist, M. A. (1984) A firm specific analysis of economies of density in the U.S. railroad industry. *Journal of Industrial Economics* **33**, 3–20.

Caves, D. W., Christensen, L. R. and Swanson, J. A. (1981) Productivity growth, scale economies, and capacity utilization in U.S. railroads, 1955–1974. *American Economic Review* **71**, 994–1002.

Federal Transit Administration (1992) *Report on Funding Levels and Allocations of Funds.* Report FTA-TBP-10-92-2. U.S. Government Printing Office, Washington, DC.

Institute for Defense Analyses (1972) *Economic Characteristics of the Urban Public Transportation Industry.* U.S. Department of Transportation, Washington, DC.

International Union of Public Transport (UITP) (1985) *UITP Handbook of Public Transport.* UITP, Brussels, Belgium.

Jane's. *Jane's Urban Transport Systems.* Jane's Publishing Company, Coulsdon, (Annual), U.K.

Jara-Díaz, S. R. and Cortés, C. E. (1996) On the calculation of scale economies from transport cost functions. *Journal of Transport Economics and Policy* **30**, 157–170.

Nash, C. A. (1978) Management objectives, fares and service levels in bus transport. *Journal of Transport Economics and Policy* **12**, 70–85.

Pickrell, D. H. (1989) *Urban Rail Transit Projects: Forecast versus Actual Ridership and Costs.* Office of Grants Management, Urban Mass Transportation Administration. U.S. Government Printing Office, Washington, DC.

Pickrell, D. H. (1992) A desire named streetcar: Fantasy and fact in rail transit planning. *American Planning Association Journal* **58**, 158–176.

Pratt, R. H., Pederson, N. J. and Mather, J. J. (1977) *Traveller Responses to Transportation System Changes: A Handbook for Transportation Planners.* Urban Mass Transportation Administration. U.S. Government Printing Office, Washington, DC.

Pozdena, R. J. and Merewitz, L. (1978) Estimating cost functions for rail rapid transit properties. *Transportation Research* **12**, 73–78.

Schrank, D., Turner, S. M. and Lomax, T. J. (1994) *Trends in Urban Roadway Congestion — 1982 to 1991. Volume 2: Methodology and Urbanized Area Data.* Report 1131-1136. Texas Transportation Institute, College Station, Texas.

Stern, J. L., Miller, R. U., Rubenfeld, S. A., Olson, C. A. and Heshizer, B. P. (1977) *Labor Relations in the Transit Industry.* Report UMTA-WI-11-0004-77-2. U.S. Government Printing Office, Washington, DC.

United States Department of Energy, Energy Information Administration (Annual to 1988). *Typical Electric Bills.* U.S. Government Printing Office, Washington, DC.

United States Department of Energy, Energy Information Administration (Annual from 1989). *Electric Sales and Revenue.* U.S. Government Printing Office, Washington, DC.

Viton, P. (1980) On the economics of rapid-transit operations. *Transportation Research A* **14A**, 247–253.

Viton, P. (1993) Once again, the costs of urban rapid transit. *Transportation Research B* **27B**, 401–412.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* **5**, 348–368.