Scale Economies in Rail Transit Systems

Ian Savage
Northwestern University, USA

ABSTRACT

The research uses Federal Transit Administration "Section 15" data to investigate the operating costs of 13 heavy-rail and 9 light-rail urban mass transit systems for the period 1985-91. A transcendental logarithmic technology is used to investigate various types of economies of scale. The principal findings are:

- Large economies of density. Adding additional trains, and passengers, to an existing network leads to a less than proportionate increase in short-run variable costs for nearly all systems. The exceptions are three larger systems that have track that is heavily utilized and offer a relatively flat level of service across the day and serve markets where passengers undertake short trips. When the cost of track maintenance and capital costs of way, structure and rolling stock are incorporated, economies of density become more pronounced for all systems.

- Constant returns to network size in short-run variable costs. A similar pattern of network economies persists when the cost of track maintenance and capital costs of way, structure and rolling stock are incorporated. The sole exceptions are newer heavy-rail systems catering to longer-distance commuter traffic. These systems have a high ratio of peak to off-peak service. Diseconomies of network size are found for these systems.

In making these calculations allowance was made for the effects of different levels of peaking of service, load factor and average journey length. Correction was also made for the use of light-rail technology. Newer systems which invested in high technology capital such as automatic train control and automatic ticketing systems have reduced their variable cost by 27% compared with comparable traditional systems.

There are four major public policy implications:

- Calculated economies of density can be used to estimate marginal cost. When comparison is made with marginal fare revenues, most systems are found to be pricing below marginal cost. Considerable welfare losses can therefore be expected.

- There has been considerable controversy about the accuracy of cost and revenue estimates used when seeking funding for extensions to existing systems and the building of entirely new systems. The equations estimated in this paper provide a possible method for the federal government to evaluate operating cost estimates.

- 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.

- The constant economies of network size for large systems suggests that there would not be cost disadvantages if the larger systems -- Boston, Washington, D.C., San Francisco, Chicago, and New York -- were divided into smaller operating units to permit privatization.

MOTIVATION

The cost function for rail mass transit systems has been investigated before (Pozdena & Merewitz, 1978; Viton, 1980, 1993). These authors were motivated by estimation of costs for new rail systems then being built, and especially for the BART system in San Francisco. The data for the estimations were for 1960-70. At that time there were very few transit systems. Eleven systems appeared in the pooled datasets used by these authors, of which two were Canadian systems.

The industry has changed considerably since that time. There has been a great expansion in construction of rail transit systems, aided by federal capital grants under the Urban Mass Transportation Act of 1964. Six cities have built new heavy-rail transit systems. Streetcar style light-rail systems have been constructed in nine cities, while five other cities have modernized historic streetcar systems with new cars, route extensions, and in certain cases subway operations in the center city. The considerable expansion of the industry warrants a revisiting of the estimation of cost functions.

There are also other troubling features to the previous literature. The first is a finding of diseconomies of density for certain systems which might seem a somewhat surprising result. The second has been a problem of defining "fixed factors" in the analyses which has meant that economies of system size have not been investigated. A third problem is that previous authors have assumed that output of the transit firm is determined exogenously. While it is certainly true that transit service provision is determined in a highly political environment, it would be too strong a statement to claim that management have no choice in the amount of service provided.

Aside from dealing with the technical problems discussed in the previous paragraph, there continues to be a public policy interest in cost estimation. There has been considerable controversy in recent times about the "optimistic" revenue and cost forecasting for new systems (Pickrell, 1989, 1992). The current analysis may be helpful in evaluating the costs of systems that are still at the proposal stage.

There is also interest in determining whether there are economies of scale in mass transit operation. While nobody would doubt that there are likely to be economies of density on individual routes, there is some doubt as to whether there are network size economies. Determination of economies of network size has important public policy implications. First, transit systems are currently proposed for smaller cities. The possible existence of network size economies effects the desirability of constructing such small systems. Second, there is a debate about privatization. All of the systems in the United States are publicly owned, generally by city or region-wide agencies. The absence of network size economies implies that, for example, a city with three distinct rail routes can have each route operated by a separate company without suffering any unit cost increases. Such a scenario allows the possibility for privatizing mass rail transit, using limited-time franchises, with the possibility for operating efficiencies commonly associated with the introduction of private enterprize.

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 technology (K). The last of these requires some explanation. Many recently-constructed transit systems have incurred high initial capital costs to install automatic ticket issuing and inspection systems, and sophisticated signalling systems to allow automatic train control. These investments have reduced the number of station staff required and led to the elimination of conductors.

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

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

Shephard's Lemma permits definition of a total operating cost function based on the assumption that firms minimize total cost (TC) for a given level of output:

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

where P_i are the factor prices. The derivation of the above function depends upon a number of assumptions. The first is that transit firms minimize cost. Much has been written about the objectives of transit managers and their political masters. Nash (1978) and others have suggested the transit firms maximize output, variously defined, subject to a subsidy budget constraint imposed by the political process. Even though the objectives of the firm may be different from the traditional economic model of profit maximizing, the firm should still be trying to minimize cost given output. For an output maximizing firm, minimizing cost ensures that more units of output can be produced within the budget constraint.

The second assumption is that both output and factor prices are exogenously determined. The exogeneity of factor prices is reasonably easy to argue. The cost of way and structure, rolling stock, and technology components are determined nationally. Electricity prices are clearly exogenous. Labor is hired in a competitive local marketplace. However, there are certainly people who would argue that management have acquiesced with transit unions to raise wages above those for comparable jobs.

The exogeneity of output requires that transit managers have no control over factors that influence demand: prices and levels of service. There is some justification for believing this is so. Urban transit is heavily subsidized. At best, farebox revenue covers half of operating costs, and often a much lower proportion. Subsidy levels are a major determinant of output, and fares and levels of services are decided in the political, rather than managerial, arena. However, this study will not assume exogeneity and will use a method of instrumental variables and three-stage least squares.

Capital versus Operating Costs

The theoretical model presented above has two types of inputs. Labor and electrical power are variable in the short run, and costs are incurred on a continuing

TABLE 1: CONSTRUCTION COST PER ROUTE MILE FOR DIFFERENT TYPES OF SYSTEMS

System Type		Cost per Route Mile	Examples
Heavy Rail	Subway	$210m	Los Angeles
Heavy Rail	Mixed aerial, subway, grade	$150m	Atlanta, Washington D.C.
Heavy Rail	Mixed aerial, existing railroad alignment	$ 55m	Chicago
Light Rail	Grade	$ 35m	Dallas, San Diego
Light Rail	Existing railroad alignment	$ 20m	St Louis

basis. Technology, way and structure, and cars are somewhat different in that there is a fixed capital cost of initially purchasing the items, and then some ongoing costs, such as routine car cleaning and maintenance, removal of snow from track, routine repainting of stations, and maintenance of way and structure.

For transit systems, the capital costs are substantial particularly if subway construction is necessary. 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). It would not be atypical to think that construction costs of light-rail systems are about $35 million per route mile, while heavy rail costs $100 million per route mile, and tunnelled track costs twice these amounts. Transit cars typically cost $1.5 million per heavy-rail car, and $2.5 million for a light-rail articulated unit. Therefore, if one excludes the unlikely possibility of steeply rising marginal operating costs, there must be economies of density in transit operation.

In the United States capital expenditures are heavily underwritten by the federal government. In contrast, the operating costs of employing labor, running the trains and upkeep of the existing way, structure, and rolling stock are primarily supported by farebox revenue and state and local funds.

Long-Run Versus Short-Run Cost Functions

The production and cost functions described in the previous sections assume that the quantities of all five factors of production are choice variables for the firm. It assumes that management can instantaneously decide on the optimal quantities of technology, track, cars, labor, and electrical power. In practice, history has already predetermined the quantities of some of these factors. There are long lead times to construct new track, or to receive permission to abandon existing trackage, and to deliver new rolling stock.

In the short run the capital expenditures have already been made and hence are a sunk cost. Sunk costs should not enter short-run decision making. We will assume that technology and way and structure are fixed in the short run. However, we should note that the considerable new-start construction and route expansions witnessed in the past 20 years suggests that way and structure is certainly variable in the medium term.

The short-run variable costs that a firm faces are the costs of labor and electrical power plus the maintenance of the rolling stock. Changes in demand for the transit system can result in cars being transferred from the active to the reserve fleets, with resultant changes in total maintenance expenses. The short-run production function is therefore:

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

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

$$SRTC = P_{tm}T^{*} + P_{cm}C + P_{l}L + P_{e}E \quad (3)$$

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 equation (3) 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 cots. Minimizing cost subject to a given level of output gives:

$$SRVC = f(Y, T^*, K^*, P_{cm}, P_l, P_c) \tag{4}$$

Fixed Factors in a Short-Run Cost Function

The fixed factors, technology and way and structure, are represented in equation (4) by their quantity rather than their price. If fixed and variable factors are substitutes, the estimated coefficient on the quantity of the fixed factor in a short-run variable cost function should be nonpositive. More of the fixed factor should lead to lower variable costs. The empirical work supports this assertion for the case of technology. Investment in technology does reduced short-run variable costs.

However, way and structure is a complementary factor of production rather than a substitute for 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, Christensen and Swanson (1981) found a similar result in their analysis of United States 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.

Study Objectives

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

$$ED = (\partial \ln SRVC / \partial \ln Y)^{-1} \tag{5}$$

where values of ED greater than unity, equal to unity, or less than unity indicate increasing returns to density, constant returns to density, and decreasing returns to density, respectively. Economies of network size in short-run variable costs can be investigated using a dataset that includes cross-sectional comparisons across firms and/or variations in network size over time for individual firms. Economies of network size (ES) are given by:

$$ES = ((\partial \ln SRVC / \partial \ln Y) + (\partial \ln SRVC / \partial \ln T))^{-1} \tag{6}$$

Again values of ES greater than unity, equal to unity, or less than unity indicate increasing returns to network size, constant returns to network size, and decreasing returns to network size, respectively. Basically, this calculation is investigating the proportionate effect on costs of operating, say, twice as many trains over a network that is twice as large.

Many cost function studies also investigate elasticities of substitution between inputs. This is not an objective of this work. The three 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. Typically, analysts have used mean values of variables as the point of estimation. This point of estimation is also used in this study. The output and factor prices are assumed to be separable. The general form of the estimated equation is:

$$\begin{aligned} lnSRVC = {} & \alpha_y lnY + \tfrac{1}{2}\alpha_{yy}(lnY)^2 + \alpha_t lnT + \tfrac{1}{2}\alpha_{tt}(lnT)^2 + \alpha_{yt}(lnY)(lnT) \\ & + \sum_i \beta_i lnH_i + \sum_i \gamma_{yi}(lnY)(lnH_i) + \sum_i \zeta_{ti}(lnT)(lnH_i) \\ & + \tfrac{1}{2}\sum_i \sum_j \eta_{ij}(lnH_i)(lnH_j) + \sum_i \theta_i D_i \\ & + \sum_i \lambda_i lnP_i + \tfrac{1}{2}\sum_i \sum_j \lambda_{ij}(lnP_i)(lnP_j) \end{aligned} \qquad (7)$$

where $\eta_{ij} = \eta_{ji}$, $\lambda_{ij} = \lambda_{ji}$, 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:

$$S_i = \frac{\delta lnSRVC}{\delta lnP_i} = \lambda_i + \sum_j \lambda_{ij} lnP_j \qquad (8)$$

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

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

It is assumed that (7) and (8) 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). For empirical estimation we only require i-1 share equations. For this analysis share equations were used for propulsion electricity and rolling stock maintenance.

Finally, the endogenous nature of output, Y, was attended to by instrumenting this variable and the use of three-stage least squares for the estimation of the system of equations.

PREVIOUS LITERATURE

The previous analysis of urban transit costs all used the same dataset. This contained eleven north american systems for the eleven years 1960-1970. Section 15 were not published at that time, therefore a research report by the Institute for Defense Analyses (1972) was used as a source. Pozdena and Merewitz (1978) estimated a short run cost function using a Cobb-Douglas technology but without factor share equations. Factor prices were formulated for electricity and labor. Track-miles were regarded as a fixed factor. However the annual expenditures on track maintenance could not be removed from the measure of operating costs, so a non-linear function was estimated with both a fixed component related to the amount of track and a Cobb-Douglas style function for the short run variable cost component. Output was measured in car miles but no hedonic output variables were used. Variable costs were found to display mild diseconomies of density. However, increasing network size reduced variable costs. There is a problem in that the estimated sum of factor shares was almost two. The authors also found diseconomies of scale in the maintenance of track. However, this may be explained by the fact that the larger systems were heavy rail and the smaller systems were light rail.

Viton (1980) used a sub-set of the same dataset to estimate a translog function complete with factor share equations. As in the earlier study, the measure of costs includes both short-run variable costs and the recurrent costs of maintaining the fixed infrastructure. At mean values there were considerable diseconomies of density. When estimates were made for individual systems, the large Philadelphia, Chicago and New York systems were found to have diseconomies of density, while the smaller systems all displayed economies of density. Increasing the size of the network was found to lead to reduced operating costs.

Viton reestimated the equation in 1993 using a translog form but assuming a different error structure which does not require the assumption that firms minimize cost. He estimates such a "frontier" cost function using the same variables as before, except that linear homogeneity was not enforced. However, he found that model predicted negative marginal costs for all observations. This problem was overcome by redefining the fixed factor from miles of track to the ratio of cars to miles of track. The fixed factor was found to be positively related to variable costs. Diseconomies of density were now found to exist for all systems.

Viton's findings on economies of density are at odds with evidence from mainline railroads. Using translog formulations, both the cross-sectional study by Caves, Christensen and Swanson (1981) and the time-series analysis by Braeutigam, Daughety and Turnquist (1984) find substantial economies of density.

DATA AND VARIABLES

Sample

Rail transit systems are basically of two types, commonly referred to as "heavy rail" and "light rail." The former are much closer to regular railroads and feature

TABLE 2: HEAVY RAIL SYSTEMS

City	Opened	Data For	Revenue Car Hours 1991 (million)	% Change over Period	Directional Route Miles 1991	% Change over Period	1991 Cost per Car Hour
New York Transit Authority	1904	85-91	16.22	+ 7	493	+ 3	$113
Chicago	1892	85-91	2.55	+34	191	0	$ 96
Washington, D.C.	1976	86-91	1.50	+29	156	+12	$133
San Francisco (BART)	1972	85-91	1.37	+22	142	0	$117
Boston	1901	86-91	1.14	+ 8	77	0	$144
Philadelphia (SEPTA)	1907	85-91	0.96	- 3	76	- 6	$113
New York (PATH)	1908	86-91	0.63	+13	29	+ 4	$188
Atlanta	1979	85-91	0.61	+42	67	+30	$ 75
Miami	1984	90-91	0.18	0	42	0	$181
Philadelphia (Lindenwold)	1969	85-91	0.15	+15	32	+ 3	$123
Baltimore	1983	90-91	0.15	-10	27	0	$140
New York (Staten Island)	1925	88-91	0.10	0	29	0	$141
Cleveland	1955	86-91	0.08	+ 5	38	0	$175

segregated right-of-way, subway construction, and/or heavy earthworks, and traditional operating practices and signalling systems. These systems were established in New York, Boston, Philadelphia, and Chicago around the turn of the century. Cleveland, plus Toronto and Montreal in Canada, constructed similar systems after the second world war. The systems that were constructed from the late 1960s featured a considerable changes in technology, with the introduction of automated train control and fare collection. Many systems were designed for suburban commuting rather than for distribution between the center city and the close-in neighborhoods. Stations became further apart, and hence average speed rose. The first of this new generation was the Lindenwold line into Philadelphia (1969), followed by San Francisco (1972), Washington, D.C. (1974), Atlanta (1979), Baltimore (1983), Miami (1984) and Los Angeles (1993). The latter system opened after the end of the period under analysis. As can be seen in table 2, the new systems have changed the nature of the heavy-rail dataset. Previously the New York City Transit Authority totally dominated any regression analysis. While it is still five times larger than its nearest rival, there is now a group of middle-sized systems.

Light rail evolved from the streetcar. Traditional streetcar operation with vintage cars still exists in New Orleans, Newark, and parts of Philadelphia. The cities of Cleveland, San Francisco, Boston, Pittsburgh, and Philadelphia have modernized historic streetcar systems with new cars, route extensions, and in certain cases subway operations in the center city. More recently new systems have opened in San Diego

TABLE 3: LIGHT RAIL SYSTEMS

City	Opened	Data For	Revenue Car Hours 1991 (million)	% Change over Period	Directional Route Miles 1991	% Change over Period	1991 Cost per Car Hour
Philadelphia (SEPTA)	1913	85-91	0.51	- 12	127	-29	$ 81
San Francisco (MUNI)	1912	86-91	0.39	- 3	50	+ 6	$137
San Diego	1981	85-91	0.23	+150	41	+28	$ 62
Pittsburgh	1890	88-91	0.15	- 4	65	+58	$110
New Orleans	1835	87-91	0.09	+ 18	17	+29	$ 51
Buffalo	1985	88-91	0.07	- 14	12	0	$114
Portland, Oregon	1986	88-91	0.07	- 5	30	0	$120
Cleveland	1920	86-91	0.05	+ 5	27	+ 3	$175
Newark	1935	85-91	0.05	+ 12	8	0	$ 80

(1981), Buffalo (1985), Portland, Oregon (1986), San Jose (1987), Sacramento (1987), Los Angeles (1991), Baltimore (1992), St Louis (1993), and Denver (1994).

The final three systems opened after the end of the time period analyzed. The San Jose, Sacramento and Los Angeles systems were dropped from the analysis because the data was found to be very misleading. These systems have constructed facilities and purchased rolling stock far in excess of current requirements because much larger systems are in the planning and/or construction process. The Boston system was also not used as the system was undergoing considerable changes during the 1985-91 period. Also not included in the analysis are two very small tourist oriented systems using vintage cars in Detroit and Seattle. The light-rail systems included in the analysis are shown in table 3. As can be seen some of the light-rail systems are larger than the very smallest heavy-rail systems.

The data for this analysis covered the years 1985 to 1991. The years in which individual systems are included are shown in tables 2 and 3. Data were not reported for some systems in 1985. Observations have been deleted for some systems in their early years of operation or during major system changes when reported data included unusual cost or output figures.

The resulting pooled dataset produced 124 observations on 22 systems. Tables 2 and 3 also show the size of the various systems in 1991 measured by the number of car hours operated in revenue service, and the number of directional route miles. In addition the percentage change in these two measures over the period for which data were used is shown. These four columns bode well for an estimation of density and scale economies as there are not only variations cross-sectionally in system size, but individual systems saw marked changes in output over the period. This is even true for traditional systems such as Boston and Chicago.

Data Source

In the United States 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 of a high quality and made available in an annual publication by the American Public Transportation Association. Canadian systems do report on a voluntary basis, but not all data items are reported, which meant that they are excluded from this analysis. In this analysis all prices have been inflated to 1991 dollars using the consumer price index.

Cost Variable

This analysis is concerned with estimating short-run variable costs. The data on costs are "total mode expense" less "nonvehicle maintenance." As a result, costs includes all operating expenses and maintenance of rolling stock. Maintenance of way and structure is excluded. Capital expenditures, such as new line extensions, new rolling stock or major station rehabilitations, and charges for such expenditures, are also excluded.

Output

The output measure is revenue car hours. 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 economies of density. This analysis takes the view that the operation of trains is the major determinant of expenses so that a supply side measure should be used. Car hours have been 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.

Some data cleaning was necessary. Some systems collect data on both car hours and car miles. Other systems evidently collect information on car miles and transform this into car hours by use of estimated average system speeds. Several systems which used the latter method changed the average speed that they used in calculating car hours during the time period under review. The data were corrected to remove such changes.

Fixed Factors

The measure of way and structure, and hence network size, is directional route miles. This is route length multiplied by two, which should approximate operational track miles for double-track systems.

The measure of technology is a discrete variable taking the value 1 for "high technology" 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 "high technology."

There are two other discrete variables used that relate to technology. The first is a dummy variable taking the value 1 if the system is a light-rail system. As described in an earlier section, these systems should have much lower costs than the heavy-rail systems with their mainline railroad characteristics. The final dummy variable is used to identify two small light-rail systems in Newark and New Orleans. As can be seen in the final column of table 3, they have unusually low costs. These two systems are the only ones to exclusively retain traditional streetcar operation with historic cars. These two systems have dummies variables identifying them as light-rail as well as "streetcar" operation.

Output Characteristics Variables

Four continuous variables are used to measure output characteristics. The first is a variable representing passenger usage. Clearly, there may be some costs associated with the number of passengers carried such as the number of ticket agents on duty. A load factor variable, calculated as passenger miles divided by revenue car miles, is used to capture this effect.

The second variable 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 third 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 costs as cars are used on average for fewer hours per day, and labor is less productive.

The final 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). Station staffing requirements will vary considerable depending on whether stations are at grade, elevated or underground.

Factor Prices

Section 15 data report the wages of train operators. Figures are available on the total train operators wages and salaries and the number of vehicle operator equivalents. These wages, when allowance is made for fringe-benefits which typically are equivalent to about half of direct wages, are about 20 percent of costs. In addition, 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.

Some discretion was necessary to ensure that data for the number of operators for individual systems were consistent from year-to-year. One system had a large change in the number of "operators" due to reporting categorization changes which led to the inclusion of ticket agents as "operators." A more common problem was reporting abnormal operator numbers in a couple of years, despite a constant level of service output and expenditure on operators wages. This is probably due to staff vacancies at the point that the data were collected. Data were corrected to remove this problem.

All of the systems use electric propulsion. Section 15 data report kilowatt hours (KwH) of propulsion electricity. Electric prices are available on a state by state basis (U.S. Department of Energy, annual). 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 1985-89 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¢ per KwH in California and the pacific northwest up to 15¢ per KwH in certain east coast states. Propulsion costs average about 10-15 percent of variable costs.

Section 15 data are available on total expenditures on car maintenance. A factor price is obtained by dividing by peak car requirement. Exceptions were made to this for San Diego and Pittsburgh where the calculations were per "active" car as there was an excessive stock of cars purchased awaiting service expansion. It may be argued this method of calculation produces a factor price that is not totally exogenous. For example, the transit company can choose the intensity with which it uses its rolling stock which will affect wear and tear. While this may be true, the calculated factor prices do seem to be consistent with intuitive observations about the age profile, design, complexity and technology of the rolling stock used by the various systems. Car maintenance costs average a quarter of total variable costs.

Instrumental Variables on Output

Two variables were used to instrument revenue car hours. 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. The second is a measure of automobile availability measured by annual vehicle miles travelled per head of population. Vehicle availability has continued to increase in recent years and abstracted ridership from transit. Both of these variables were obtained from Schrank, Turner and Lomax (1994). Figures were converted from kilometers to miles to be consistent with the rest of the dataset. Data were not given for Buffalo so the data for Cincinnati were substituted.

REGRESSION RESULTS

The results of the estimations are shown in table 4. 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 on cost functions. At 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 at mean values are measured by the inverse of the coefficient on car hours. 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 network leads to a less than proportionate increase in variable costs. This is not a particularly surprising result. Some costs, such as management, ticket agents, and certain aspects of car maintenance, may be invariant with marginal car hours. Economies of density will be discussed further in a later section.

Economies of Network Size

Economies of network size at mean values 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 network. The mean estimate of economies of network size is 0.988 with a standard error of 0.05. The estimate is insignificantly different from unity, which represents constant returns to network size.

Of course there are squared 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

TABLE 4: TRANSLOG REGRESSION ON LOGARITHM OF SHORT RUN VARIABLE COSTS

Explanatory Variables (logarithms except for dummy variables)	coefficient	t
Car hours	0.668	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
High Technology 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	116.6
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	

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 network size of 0.97, while the largest systems have economies of network size of 1.05. While, this implies an inverted U-shaped average variable cost curve, the extent of economies or diseconomies of network size would not be regarded as large and constant returns to network size cannot not be statistically rejected at any point. Further discussion is contained in a later section.

Technology Effects

Although there may be constant returns to network size, smaller systems do have lower average costs in absolute terms. This is because most of the smaller systems are of light-rail technology. Light-rail systems have 20% lower costs than comparable heavy-rail systems. The very basic streetcar systems 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.

Investments in modern automated ticketing and train control systems result in a 27% reduction in short-run variable costs compared with comparable older systems. Of course, without knowledge of the capital costs of these investments it is impossible to say whether these investments are justified.

System Characteristics

The effect of hedonic variables is often difficult to interpret in a translog function with many second order terms. At mean values increased load factor increases short-run variable costs, increased average journey length reduces cost, and a higher peak-base ratio increases cost. However, only the first of these is statistically significant. When allowance is made for second order terms these effects become stronger. A three-stage least squares estimation of a Cobb-Douglas function using the same variables produces t-statistics of 4.2, 8.2 and 13.5 respectively for these three hedonic variables.

The load-factor effect might be somewhat surprising given that it is well recognized that the marginal cost of an additional passenger are effectively zero. However, an increase in load factor, which is measured by passenger miles per car mile, represents a considerable increase in the absolute numbers of annual passengers. These passengers will have to be serviced by additional ticket agents and barrier staff at stations for heavy-rail systems. Light rail depends on on-car fare collection or ticket checking. Additional passengers may require additional conductors, roadside ticket machines, and ticket inspectors. At mean levels doubling load factor will increase short-run variable costs by 59%.

One would expect a negative coefficient for average journey length. If a system's total passengers miles consists of relatively few passengers travelling long distances, costs should be lower than a system serving a larger number of passengers travelling short distances. If nothing else, fewer ticket transactions would have to be made. In addition, longer average journey length is consistent with greater station spacings and higher operating speeds. At mean values, doubling average journey length, while holding total passenger miles constant, would reduce costs by 27%. Average journey length does vary considerably. For heavy rail, average journey length is under five miles for the traditional east-coast systems, but is almost twelve miles for the San Francisco BART system that serves longer distance commuters. For light rail, average journey length can be very short for systems that grew out of traditional streetcar operations. San Diego on the other hand has an average journey length of seven miles.

Systems with high peak-to-base operations are considerably more costly than systems with a more consistent level of operations across the day. Among the heavy-rail systems the peak-base ratio varies from 1.5 in Philadelphia, New York, and Atlanta up to 4 in Baltimore and 8 for the Lindenwold Line. Some light-rail systems, such as San Diego, do not offer enhanced peak service while Portland, Pittsburgh, and Cleveland have ratios in excess of 2.5. 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.

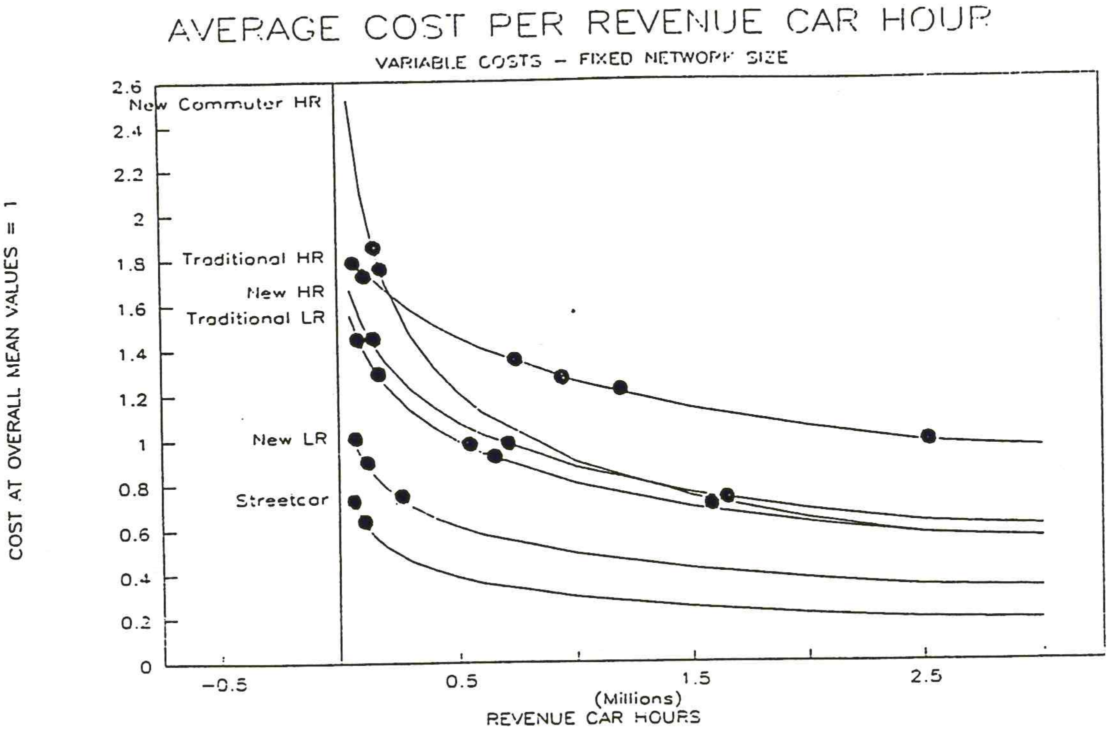
GENERIC ESTIMATES OF ECONOMIES OF DENSITY AND NETWORK SIZE

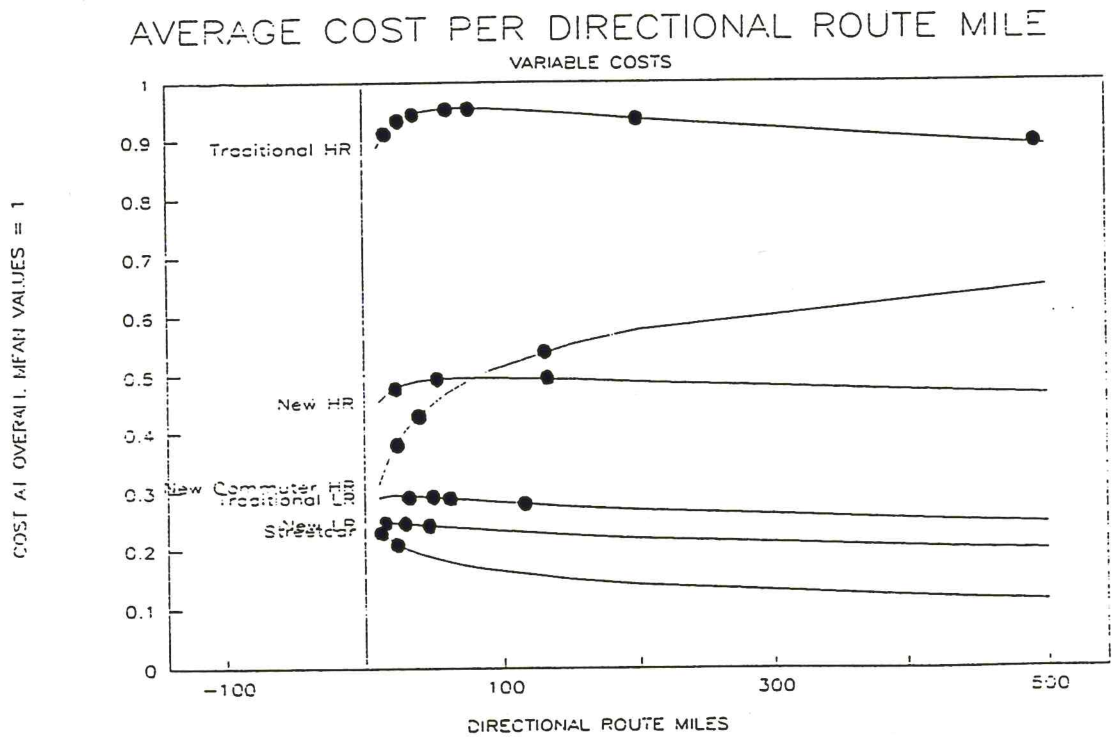
The preceding section discussed estimates of economies of density and network size at mean values. This gives a limited picture of the nature of the industry. However, more general inferences can be confusing because of the complex interaction of system characteristic and output variables in a translog function. Therefore, this section illustrates economies of density and network size by plotting of average variable cost curves for six generic system types. By selecting six generic types of systems, it is possible to graphically illustrate the effect of both system characteristic and output variables on costs.

TABLE 5: SIX GENERIC SYSTEM TYPES

	Streetcar	New Light Rail	Traditional Light Rail	New Commuter Heavy Rail	New Heavy Rail	Traditional Heavy Rail
Load Factor	19	25	26	22	22	21
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
High Technology	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
Directional Route Miles	12.6	27.8	66.9	71.9	83.2	133.0
Car Hours per Directional Route Mile	5.3	4.6	4.0	6.2	8.1	14.5
Systems	Newark New Orleans	Portland Buffalo San Diego	Cleveland Pittsburgh MUNI Philadelphia	Lindenwold Miami BART	Baltimore Atlanta Washington	Cleveland Staten Is. PATH Philadelphia Boston Chicago New York

The six types used are 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 twenty-two systems were divided into these six categories and average values calculated for each system type. These are shown in table 5.





FIGURES 1 AND 2: GENERIC AVERAGE VARIABLE COST CURVES

Two sets of average cost curves are plotted. The first is associated with economies of density. System characteristic, technology, factor price and network 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. The calculated average variable cost per car hour are shown in figure 1. When comparing the curves, the reader is reminded that the (fixed) network size is different for each generic system 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 solid circles represent the number of annual car hours of actual systems so as to give an indication of the ranges over which the average cost curves are applicable. The New York City Transit Authority is not indicated in this figure as its number of revenue car hours is very large at 16 million.

Figure 1 also illustrates the powerful effect of some of the system characteristic and technology variables. Light-rail systems are less expensive than heavy-rail systems, and streetcar systems are even more inexpensive. Investment in high technology also reduces costs.

The second set of cost curves, shown in figure 2, represent economies of network size. System characteristic, technology, and factor price variables are held at mean values for each of the six system types. The number of directional route miles and car hours were then varied. The ratio of car hours to directional route miles is held at the mean value for that type of system, which are shown in table 5. The plotted average cost per directional route mile curves are supportive of our earlier finding of constant returns to network size. While one can observe some evidence of an inverted U-shaped function, the cost functions are remarkably flat over substantial ranges. Again, the solid circles represent the directional route miles of actual systems so as to give an indication of the ranges over which the average cost curves are applicable.

The major exceptions are the newer heavy-rail systems built to serve longer-distance commuter traffic. Diseconomies of network 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 NETWORK SIZE

Point estimates can also be made for each individual system using data for 1991. These are shown in table 6 where the systems are organized in ascending order of annual car hours. It will be noted that economies of density apply to nearly all systems, except for three of the larger heavy-rail systems. Boston and the New York Transit Authority have very mild diseconomies of density while Philadelphia's heavy-

TABLE 6: POINT ESTIMATES FOR INDIVIDUAL SYSTEMS IN 1991

System	Light Rail	Economies of Density	Economies of Network Size
Newark	LR	1.63	0.99
Cleveland	LR	1.54	0.89
Portland	LR	1.35	0.94
Buffalo	LR	1.02	1.27
Cleveland		1.52	0.89
New Orleans	LR	1.06	1.32
New York (Staten Island)		1.79	0.87
Baltimore		1.87	0.89
Pittsburgh	LR	1.29	0.97
Philadelphia (Lindenwold)		2.04	0.78
Miami		1.05	0.99
San Diego	LR	1.13	1.04
San Francisco (MUNI)	LR	1.08	1.27
Philadelphia (SEPTA)	LR	1.12	1.31
Atlanta		1.22	1.08
New York (PATH)		1.52	0.99
Philadelphia (SEPTA)		0.71	1.49
Boston		0.99	1.28
San Francisco (BART)		1.61	0.86
Washington, D.C.		1.28	1.05
Chicago		1.92	0.94
New York Transit Authority		0.96	1.33

rail system is estimated to have considerable diseconomies of density. These three systems along with the PATH system and Chicago have the most dense service as measured by car hours divided by directional route miles. However, these latter two systems have a more "peaked" service which implies that trains are available to provide service out of the peak period at low marginal cost. The three systems with diseconomies of density systems also have comparatively short average journey lengths. Therefore, diseconomies of density in short run variable costs are likely to arise when track is heavily utilized, a relatively flat level of service is offered across the day, and the transit system is serving short distance trips such as occur in the center of a very large city. Of course it is unlikely that these three systems exhibit discconomies of density when allowance is made for the fixed costs of operation in terms of capital costs and the maintenance of track. We will return to this issue in the concluding section.

As with economies of density, point estimates can be made for economies of network size for each of the systems in 1991. These are shown in the final column of table 6. They are very supportive of the generic graphs shown in figure 2. Many systems have estimated economies of network size close to unity. One will note the estimated diseconomies of network size for the Lindenwold and BART systems, which are modern systems catering to longer-distance commuter traffic.

TRACK MAINTENANCE COSTS

The above cost estimation is for short-run variable cost, and excludes the ongoing costs of maintaining track, way and structure. For the present paper it is important to investigate whether there are economies of scale in track maintenance.

Regressions using Cobb-Douglas technology were conducted to observe whether total expenditures on nonvehicle maintenance increase more or less proportionately with the number of track miles. Separate regressions were made for heavy rail and light rail because the technology differs radically. In comparing different systems allowance was made for hedonic differences that may have important effects on track maintenance.

The first hedonic variable measured track usage. This is defined as annual revenue car miles divided by track miles. Other things remaining equal, more heavily used track should require more maintenance. For heavy-rail systems the average number of stations per track mile is used. More stations should, other things remaining equal, lead to higher total maintenance costs. This variable is not used for light rail where the definition of a station is far more ambiguous in the data. Some stops in street operation are nothing more than a pole and flag. Allowance was also made for the age of the system, measured by years from opening, as shown in tables 2 and 3. Many of the systems have opened relatively recently and should not require the major track and station rehabilitation that older systems need. Allowance was also made for factor price differences. The annual wage of train operators was used as a surrogate for the level of compensation paid to labor in individual cities. As discussed earlier, 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.

The most important hedonic 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 track miles that are elevated, and the proportion in tunnel. Miami has no tunnels and the Staten Island system is entirely at grade. For these two

TABLE 7: REGRESSION ON LOGARITHM OF NON-VEHICLE MAINTENANCE COSTS

Explanatory Variables (all in logarithms)	Heavy Rail		Light Rail	
	coeff.	t	coeff.	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	0.614	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	

systems, very small values were inserted to avoid recording a zero value. This was necessary as all variables were expressed in logarithms and normalized about the variable means.

Estimation results are shown in table 7. While the point estimates might suggest mild diseconomies of scale in track maintenance, constant returns to scale cannot be rejected for both heavy rail and light rail.

As to the other variables, whether the track is at grade, elevated or in tunnel has a powerful effect on costs. For light rail, at-grade track and stations are less expensive to maintain than tunnels. For heavy-rail systems tunnelled track is much more expensive than at-grade or elevated track. While there are costs to maintaining an elevated structure, there are savings in avoiding maintenance of extensive earthworks and right of way, and the elimination of grade crossings. As expected, the frequency of stations is significantly positively related to total track maintenance costs for heavy-rail systems.

Newer light-rail systems appear to have lower maintenance costs than older systems primarily because the stations and track have yet to reach mid-life or full-life refurbishment or replacement. The same effect is not so noticeable for heavy-rail where the newer systems have high technology signalling and ticketing systems that require expensive maintenance.

For the light-rail systems labor price is not significantly related to track maintenance, unlike heavy-rail systems. The amount of traffic is positively related to track maintenance cost, but the effect is only statistically significant for heavy-rail systems. For these systems doubling the density of service will increase track maintenance costs by 33%. Therefore correct calculation of economies of density should include an increase in track maintenance costs when service expands.

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 and the short-run average variable cost. Such a calculation does not include the effects of additional car hours on track wear-and-tear or the purchase of additional rolling stock. It is therefore a quite conservative estimate. 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 running of a marginal car hour. This calculation is also conservative in that Pratt, Pederson and Mather (1977) report a service frequency demand elasticity of around 0.65 for rail service. Therefore the marginal number of passenger miles generated by an additional car mile will 65% of the system average. However our estimate of economies of density was based on the assumption that average load factor is held constant.

The marginal cost can be compared with the marginal revenue collected per passenger mile. Marginal revenue is here equated with average revenue. Calculations on the revenue side are complicated because, unlike costs, data are reported for the whole transit agency and do not distinguished between revenue collected from bus and rail operations. Only five of the systems are primarily rail-only operators. However, providing a consistent pricing structure is used, our calculation based on passenger miles rather than passenger trips should provide a close estimate of marginal revenue on the rail system. It is possible that marginal revenue may exceed average revenue if new passengers pay the cash fare rather than purchase monthly tickets which often provide relatively low revenue per trip.

Table 8 shows the estimated marginal revenues and costs from expanding service on the current network using data from 1991. Bear in mind that the cost estimates are conservative. Even on these calculations few systems are pricing at marginal cost, and most systems are pricing at 20-60% below marginal cost. This is an extremely worrying conclusion. While there is an active debate about the value of transit subsidies (see Glaister, 1987), all economists would agree that pricing below marginal cost results in an inefficiently allocation of resources. The only possible justification could be that transit is priced below marginal cost to relieve road congestion caused by underpricing of automobile travel (Glaister and Lewis, 1977).

TABLE 8: COMPARISON OF MARGINAL COST AND MARGINAL REVENUE IN 1991

System	Light Rail	¢ per Passenger Mile		Ratio
		"Marginal" Revenue	Marginal Cost	
Newark	LR	0.18	0.22	0.83
Cleveland	LR	0.12	0.17	0.71
Portland	LR	0.12	0.16	0.77
Buffalo	LR	0.25	0.42	0.60
Cleveland		0.14	0.19	0.73
New Orleans	LR	0.17	0.24	0.71
New York (Staten Island)		0.13	0.19	0.67
Baltimore		0.19	0.18	1.07
Pittsburgh	LR	0.14	0.20	0.69
Philadelphia (Lindenwold)		0.15	0.09	1.66
Miami		0.15	0.27	0.56
San Diego	LR	0.11	0.10	1.09
San Francisco (MUNI)	LR	0.17	0.46	0.37
Philadelphia (SEPTA)	LR	0.21	0.36	0.58
Atlanta		0.10	0.11	0.92
New York (PATH)		0.21	0.29	0.74
Philadelphia (SEPTA)		0.21	0.41	0.52
Boston		0.13	0.32	0.41
San Francisco (BART)		0.11	0.11	1.00
Washington, D.C.		0.17	0.16	1.05
Chicago		0.17	0.15	1.13
New York Transit Authority		0.20	0.32	0.62

Economies of Density and Size in Total Costs

The calculations of economies of density earlier in the paper were focussed on short-run variable costs. These calculations ignored the costs of the fixed factor. Calculations were therefore made of economies of density in annual recurrent costs by including the annual costs of nonvehicle maintenance. Allowance was made for the

TABLE 9: ESTIMATION OF ECONOMIES INCLUDING TRACK AND CAPITAL COSTS IN 1991

System	Light Rail	Incorporating Track Maintenance	Incorporating Track Maintenance and Capital Costs	
		Density	Density	Network Size
Newark	LR	1.66	2.42	0.99
Cleveland	LR	1.64	1.86	0.95
Portland	LR	1.41	1.98	0.97
Buffalo	LR	1.18	1.67	1.08
Cleveland		1.80	3.22	0.96
New Orleans	LR	1.08	1.79	1.13
New York (Staten Island)		1.95	2.99	0.94
Baltimore		2.17	2.90	0.95
Pittsburgh	LR	1.50	2.03	0.98
Philadelphia (Lindenwold)		2.20	3.00	0.91
Miami		1.27	1.84	0.99
San Diego	LR	1.22	1.57	1.01
San Francisco (MUNI)	LR	1.14	1.35	1.13
Philadelphia (SEPTA)	LR	1.20	1.59	1.10
Atlanta		1.44	2.19	1.02
New York (PATH)		1.61	1.75	0.99
Philadelphia (SEPTA)		0.84	1.28	1.17
Boston		1.26	1.53	1.10
San Francisco (BART)		1.78	2.36	0.92
Washington, D.C.		1.50	2.04	1.01
Chicago		2.06	2.32	0.96
New York Transit Authority		1.14	1.32	1.15

findings in table 7 that increased track usage has a cost elasticity of 0.33 for heavy rail and 0.52 for light rail. The resulting calculation is shown in the third column of table 9. We find extensive economies of density for all systems except for the Philadelphia heavy-rail system.

Allowance can also be made for capital costs. Such calculations are of necessity much more speculative. 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. Our calculations of annual capital costs are therefore very rough and basic. 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. Way and structure is charged on an equal annual basis over 80 years. 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 charged equally over 25 years. Even though such replacement is currently funded by government grants, we are assuming that government should be inherently setting aside such amounts each year to ultimately replace the capital assets at the end of their useful lives.

The calculations in the fourth column of table 9 recalculate economies of density taking into account track maintenance and annualized capital costs. We have also assumed in this calculation that additional car hours will require a proportionate increase in fleet size. This column confirms the general assumption that rail transportation has considerable economies of density. This is true for all of the systems studied.

The final column of table 9 is a calculation of economies of network size when track maintenance and capital costs are taken into account. In making these estimates the point estimates of table 7 that there are diseconomies in track size of 0.936 for heavy rail and 0.951 for light rail are incorporated. A similar pattern of network size economies to that found in short-run variable costs is observed.

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 4 and 7 provide a possible method for the federal government to initially evaluate operating cost estimates provided by funding applicants. Readers are reminded that the explanatory variables in the regressions have been normalized about the mean value of that variable. These mean values are available on request from the author.

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 privatization of urban transit systems. The considerable economies of density make mass rail transit routes natural monopolies. Demsetz (1968) and Williamson (1976) discuss possible methods of introducing a competitive environment given that direct competition by rival companies over the same track is undesirable. They suggest the introduction of short term franchises, such as those that exist for cable television supply, the management of parking lots, and the provision of catering facilities at hospitals.

It is sometimes suggested that the larger systems -- Boston, Washington, D.C., San Francisco, Chicago, and New York -- could be divided into smaller operating units prior to privatization. Our results suggest that there would be negligible cost disadvantages of doing so. Indeed competitive pressures might lead to reduced factor prices and improved factor utilization.

ACKNOWLEDGEMENT

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. Presentation of the work in New Zealand has also been supported by the Federal Express Research Fund.

REFERENCES

American Public Transportation Association (annual). *Transit Operating and Financial Statistics*. Washington, D.C.: APTA.

Braeutigam, Ronald R., Andrew F. Daughety and Mark A. Turnquist (1984). "A firm specific analysis of economies of density in the U.S. railroad industry." *Journal of Industrial Economics*, Vol 33(1) pp 3-20.

Demsetz, Harold (1968). "Why regulate utilities?" *Journal of Law and Economics*, Vol 11 pp 55-65.

Caves, Douglas W., Laurits R. Christensen and Joseph A. Swanson (1981). "Productivity growth, scale economies, and capacity utilization in U.S. railroads, 1955-74." *American Economic Review*, Vol 71(5) pp 994-1002.

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

Glaister, Stephen (1987). *Transport Subsidy.* Newbury: Policy Journals.

Glaister, Stephen and David Lewis (1977). "An integrated fares policy transport in Greater London." *Journal of Public Economics*, Vol 9(3) pp 341-355.

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

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

Jane's (annual). *Jane's Urban Transport Systems.* Coulsdon: Jane's Publishing Company.

Nash, Christopher A. (1978). "Management objectives, fares and service levels in bus transport." *Journal of Transport Economics and Policy*, Vol 12(1) pp 70-85.

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

Pickrell, Don H. (1992). "A desire named streetcar: Fantasy and fact in rail transit planning." *American Planning Association Journal*, Vol 58(2), pp 158-176.

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

Pozdena, Randall J., and Leonard Merewitz (1978). "Estimating cost functions for rail rapid transit properties." *Transportation Research*, Vol 12(2) pp 73-78.

Schrank, David, Shawn M. Turner and Timothy J. Lomax (1994). *Trends in Urban Roadway Congestion - 1982 to 1991. Volume 2: Methodology and Urbanized Area Data.* Report 1131-6. College Station, TX.: Texas Transportation Institute.

Stern, James L, Richard U. Miller, Steven A. Rubenfeld, Craig A. Olson and Brian P. Heshizer (1977). *Labor Relations in the Transit Industry.* Report UMTA-WI-11-0004-77-2. Washington, D.C.: U.S. Government Printing Office.

United States Department of Energy, Energy Information Administration (annual to 1988). *Typical Electric Bills.* Washington, D.C.: U.S. Government Printing Office.

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

Viton, Philip (1980). "On the economics of rapid-transit operations." *Transportation Research A*, Vol 14A(4) pp 247-253.

Viton, Philip (1993). "Once again, the costs of urban rapid transit." *Transportation Research B*, Vol 27B(5) pp 401-412.

Williamson, Oliver E. (1976). "Franchise bidding for natural monopolies - In general and with respect to CATV." *Bell Journal of Economics*, Vol 7 pp 73-104.

Zellner, Arnold (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American Statistical Association*, Vol 57 pp 348-368.