

URBAN BUS AND RAIL TRANSIT  
SUPPLY FUNCTIONS:  
MODELS AND EMPIRICAL ESTIMATES

by  
Edward K. Morlok  
and  
Oliver Scheuftan

Civil Engineering Department  
University of Pennsylvania  
Philadelphia, Pennsylvania

July 1975

This research was supported by the Program of University Research of the U.S. Department of Transportation, through contract no. DOT-OS-40092, and the Illinois Department of Transportation. This support is gratefully acknowledged, but implies no endorsement of the findings by the sponsors.

## ABSTRACT

### URBAN BUS AND RAIL TRANSIT SUPPLY FUNCTIONS: MODELS AND EMPIRICAL ESTIMATES

By Edward K. Morlok and Oliver Scheuftan

In order to treat public transportation more adequately in the network flow equilibrium models commonly used in transportation planning, it is necessary to model the supply of such services in a manner consistent with the supply-demand paradigm. In this paper, previously reported models for the estimation of the departure frequency of peak period bus service and the introduction of express service are extended to the non-peak periods, and similar models for the estimation of rail service frequency, passenger capacity, and service pattern (e.g., all-stop, skip-stop and express) are developed. All of these models are tested with data for the Chicago transit routes.

These models provide the basis for predicting major levels of service characteristic of transit routes in network equilibrium models. Specifically, departure frequency can be used to estimate waiting time, operating plan can be used to estimate travel time, and seated capacity relative to traffic can be used to estimate fraction seated. The various uses of these models in planning and policy evaluation are noted, along with modifications to existing network equilibrium models, which might be required in order to more accurately model public transportation.

## Introduction

The purpose of this paper is to present supply functions for the provision of public transportation services in metropolitan areas. These models will permit the estimation of certain characteristics of the quality of service and the capacity that would be provided on each route as a function of the volume of traffic and other characteristics of each such route. These models have been developed in order to improve the treatment of public transportation modes within the network flow equilibrium models, which are almost universally used in urban transportation planning. While supply functions for road traffic have been very well developed for use in such models, and are in common use, there are no adequate methods for estimating the level of service or other supply characteristics of public transport. The focus of this paper is upon supply functions for conventional, fixed route, bus and rail transit lines. Since the theory underlying these supply functions is presented in detail in a previous paper by Morlok [1], the primary focus of this paper will be upon the empirical estimation and testing of such functions, with only a brief discussion of the underlying theory.

### Supply Functions in Urban Transportation Planning

There are many different forms that the supply functions for public transportation may take. At one extreme, it may be desired to model the decisions of transit management with respect to all characteristics of transit service in the metropolitan area which are potentially under management's control, including such major decisions as where to locate new routes and which mode or technology to use on these routes. This would not seem appropriate for urban transportation planning because many of the decisions regarding transit service represent options which are to be identified and evaluated within the planning process. In particular, the location of new routes and the technology to be used on them would generally seem to fall within the purview of planning decisions. On the other hand, it seems undesirable and impractical to treat all aspects of transit supply, including the detailed schedules, as options which are to be explicitly considered and evaluated within the planning process. One reason is that the number of options becomes extremely large as soon as detailed schedule and operating policy variations are to be considered, and hence for practical purposes these cannot be adequately evaluated in the context of large-scale, expensive, network equilibrium models designed for long-range planning. Another reason is that transit management usually has a great deal of discretion regarding many operating variables such as the frequency with which buses or trains will be operated on a particular route, the number of cars per train and the size of buses operated (although there are limited options in this instance), and often whether the operating pattern will be all local stop, zone, express, or skip-stop service.

Thus, for purposes of planning it may be desirable to attempt to predict what service transit management would provide on various routes which might be operated, in much the same manner that we attempt to predict the mode choice behavior of travelers. For these reasons, probably the most usable form of public transport supply functions in urban transportation planning would be ones which assume that the route and mode (or

technology) used on that route are given, but that the decisions regarding the manner in which to operate the route, the frequency of trips, the capacity provided, and the operating pattern, be determined through a model which represents the behavior of transit management. (A discussion of other possible types of supply functions is contained in reference [1].)

## Theory of Transit Route Supply Functions

Although the underlying theory for fixed route bus and rail transit supply functions has been presented earlier [1], it is appropriate that this be briefly reviewed here before presentation of the empirical models and other results.

### Bus Transit

In the provision of bus transit service, once the route and its location are specified, the decisions open to transit management regarding the service to be provided include: 1) the schedule or frequency with which buses will be operated, 2) the capacity, as influenced by the seating and standee capacity of buses assigned to each trip, and 3) the extent to which the service will be operated as all-stop or local service, a skip-stop service (uncommon for buses), or a zone express service. The latter is a service in which the route has been divided into two or more sections or zones with buses making all stops in one zone, and then running non-stop to another zone, usually in the central business district or other major activity center.

Based upon interviews of transit management in selected cities, a model regarding their decisions with respect to the operating choices was hypothesized, as described in detail in reference (1). This model for peak period, peak flow direction bus frequency is:

$$f = \max\left(F, \frac{p}{Q}\right) \quad (1)$$

where:

f = frequency of bus departures in one direction, buses per unit time

F = minimum acceptable frequency, buses per unit time

p = passenger flow past peak load point on route, passengers per unit time

Q = capacity of bus, passengers per bus

Thus a minimum departure frequency criterion and the passenger volume determine the frequency of service, such that a frequency is required by capacity considerations.

This model was tested with the data from the Chicago Transit Authority for the 133 bus routes it operated in 1972. The data consisted of observations of the number of bus trips and passengers on board past the peak load points on each of its 133 routes during each of the two weekday peak periods, thus yielding 266 data points. The frequency of bus trips per two hours was regressed on the number of passengers observed past the peak load point during the same two peak hours, yielding the linear relationship given below:

$$f = 9.29 + 0.0136p \quad (2)$$

$\bar{F} = 22.04$  buses/period

$\bar{p} = 940.8$  passengers/period

$R^2 = 0.822$

Period = 2 hours peak

SE = 4.15 buses/period

N = 266 observations

As can be noted, the coefficient of determination was extremely high and the standard error low – 18.8% of the mean. As can be seen from the plot of the data contained in Figure 1, there was no evidence of adherence to a minimum frequency between zero volume and a passenger volume corresponding to filling that frequency of buses to capacity. Rather there seems to be a constant increase in bus trips operated with increasing volume, along with a minimum threshold number of buses operated (of the order of nine per two hours).

Unfortunately, no data were available at the time of this prior study on the services provided and the volume of traffic during periods other than the two weekday peak periods, so no estimates of non-peak periods bus transit supply functions could be made. Nor could there be any independent test of the peak period bus frequency supply relationship.

Zonal express services are often operated on high volume bus routes typical following the pattern shown in Figure 2. It was found that the decision to operate in such a manner can be easily explained by reference to the effects of operating such service on travel time and operating costs. On routes which have a high volume of passengers traveling to or from a particular point, most typically a central business district, the advantage of operating a zonal express service is that many travelers will reach their destination more rapidly than with an all-stop service. Even if the service frequency is reduced in each zone, in order that the buses will continue to be filled to capacity, the reduced running time of express services can still more than offset the increase in waiting time. The only travelers who would not benefit from the change are 1) those who are adjacent to the CBD for whom there is a reduced frequency of service but no reduction in travel time and 2) those who travel between zones such that they must transfer between buses. From the standpoint of the operator the cost reduction resulting from a zone express service can be substantial, because higher average bus speeds with zone operation will generally reduce driver requirements and vehicle requirements as well as reducing vehicle-miles of operation. Knowing the speed which buses can attain in all-stop and various zonal express configurations on any given route, it is possible to determine the volume at which it is advantageous from any one of these viewpoints to institute the zonal express service. The specific equations for this determination are contained in [1].

On the basis of approximate data for such express and all-stop travel times from Chicago, it was found that the transit authority there did operate zonal express service on those CBD-oriented routes where such service would yield benefits to both (most) travelers and to the operator. All eighteen Chicago bus routes, which comprise the seven zonal express services (some have two zones, others three), operated there had volumes greater than the criterion value, which varies by route around a volume of about 650 passengers per hour. On twelve other routes the volume exceeded the criterion value, but these were either very short routes or ones which have dispersed origin-destination patterns.

Thus, prior empirical examination and testing of the theory regarding bus service supply on a given route seems to indicate that the theory is valid at least during peak periods. However, at the time of that research, testing of the theory and the empirical equations against an independent data set was not possible, nor was it possible to test the theory in non-peak period services.

## Extensions of Theory of Transit Route Supply Functions

The concern of this paper is primarily with the extension of the theory of bus transit route supply to non-peak periods, along with additional testing of the theory regarding peak services, and the development of the theory and its testing for rail transit supply. Each of these will be discussed in turn.

### Non-Peak Period Bus Service

Transit management has a clear motivation to minimize the cost of providing its services, subject to providing a satisfactory level of service, this pressure being particularly acute because operating deficits are increasingly common in the industry. Since the number of drivers and the number of buses required to operate any service are largely if not entirely determined by the peak period requirements, such a cost orientation leads naturally to operating the frequency required by either service or capacity considerations, whichever is binding. In the case of non-peak period services, however, such a cost orientation would not generally lead to such a strong adherence to operating the minimum frequency of trips. The reason is simply that once a bus has been obtained for use during the peak period it is, of course, available for operation during non-peak periods at no additional vehicle ownership cost. Similarly, for drivers, most transit labor agreements provide for the paying of the driver for at least eight hours once he has been called for duty, thus in effect providing many free hours of driver labor once the driver has been called for operation of the vehicle during a peak period. For these reasons transit management is more likely to operate more frequent service than would be required purely on capacity grounds during non-peak periods than during peak periods, since the marginal costs of providing additional trips are much less in the former. If management were to attempt to maximize profit or to maximize revenue (the latter being suggested as an operational criterion by at least one manager, see [3]) the operation of many more bus trips than would be required to simply accommodate the traffic may be warranted due to the elasticity of demand with respect to service frequency or waiting time. Thus, one might expect the off-peak period frequency of transit services to be determined by passenger volume in a manner somewhat analogous to that observed for the peak period, but with a tendency to provide a greater frequency of service for the same volume during the off-peak period than during the peak period. Thus a plausible non-peak period supply function would be:

$$f_t = A_t + B_t + p_t \quad (3)$$

where:

$f_t$  = frequency of bus operation, each direction, in period t buses per period  
 $A_t, B_t$  = empirically determined parameters for period t  
 $p_t$  = passengers past peak load point on route in period t, passengers per period

The coefficients  $A_t$  and  $B_t$  would vary by period of the day, reflecting different degrees of availability of “free” vehicles and labor and perhaps different frequency elasticities of demand as perceived by management.

### Rail Service Supply Functions

The provision of rail service on a given route is even more indeterminate on purely *a priori* grounds than bus service. The reason is that the capacity of the vehicle can be easily varied in the case of rail, by changing the number of cars per train, and thus even passenger volume considerations would not lead to a unique specification of the frequency of train trips required. In addition, of course, there are other operating options just as on bus routes. Although most rail lines are only two tracks, limiting and often precluding the operation of zonal express type services, trains can be operated making all-stops or skipping stops, the most common form of skip-stopping being one in which trains stop at alternate stations.

In the case of bus transit we have seen that management seems to balance considerations of level of service, which would lead to a high frequency regardless of traffic volume, and cost, which would dictate operating a frequency sufficient only to accommodate the passenger volume. A similar behavior is likely to underlie rail transit decisions. Thus, plausible hypotheses regarding transit management decisions on train operation are 1) that the number of cars operated past the peak load point on a rail route would be an increasing function of passenger volume, tending toward capacity equal to volume, and 2) the number of trains would also be an increasing function of passenger volume with the operation of single car (or two cars if all equipment were in permanently coupled pairs) trains at low volumes and longer trains at higher volumes. Mathematically these relationships would be:

$$r_t = D_t + E_t p_t \quad (4)$$

$$c_t = F_t + G_t p_t \quad (5)$$

where:

$r_t$  = train frequency past peak load point in period  $t$ , trains per period  
 $c_t$  = car frequency past peak load point in period  $t$ , cars per period  
 $p_t$  = passenger flow past peak load point in period  $t$ , passengers per period  
 $D_t, E_t, F_t, G_t$  = empirically determined constants

The operation of zonal express trains would generally be an option only where two tracks were available in one direction, unless the frequency of service is extremely low. Where the option does present itself, transit management probably would decide to operate zonal express and local trains in much the same manner that it seems to decide this in the case of bus services. Once the volume exceeds a certain amount, the benefits of zonal express service in reducing travel time for those destined to the CBD (or other focus of such travel) and the reduced operating and equipment costs would offset the

possible negative effects on short distance riders, resulting in the decision to institute such service. This suggests a volume criterion similar to that found for buses.

A more common option in the case of rail service is to operate a skip-stop service, in which half the trains stop at one set of stations (often called “A” stops) and the remainder of the trains would stop at the remainder (“B” stops), with some stations being stops for all trains – usually the higher volume stations. Although there are intricacies in attempting to introduce skip-stop service, such as insuring that the selection of stations yields approximately the same number of passengers on “A” and “B” trains and insuring that the running times of the two trains would be the same (to avoid bunching) basically the service is instituted once the travel time savings resulting from such service more than offsets the increased waiting time at the “A” and “B” stops. In addition, there would be some gain in the form of reduced vehicle requirements and labor costs since the average speed of trains would increase slightly. Thus, there perhaps is a volume criterion for introducing skip-stop service. The existence of such a criterion could be checked in a manner similar to the check on zonal express bus operations described above.

### The Data

The new data available for this study are comprised of two sets. The first is data collected by the Chicago Transit Authority (CTA) [4], which includes counts of total passengers, the number of buses or rail cars, flowing past the peak load point in the peak direction on all bus and rail routes during two peak hours in the morning (7 to 9 a.m.) and in the evening (4 to 6 p.m.). This data set [5] also provides the information on number of cars per train, yielding with [4] information on trains past the peak load points on rail routes during each peak period. These data are probably subject to considerable error in the passenger counts because they are made manually, counters observing the passage of each train or bus and estimating by visual inspection the number of passengers on board. However, these counts are those used by the carrier to adjust the schedules and services operated on each route, and hence, the management must feel that they are adequate for schedule development purposes.

The second set of data is from the Chicago Area Transportation Study (CATS) [6], and these data include the number of passengers on board transit vehicles leaving many of the stations or stops along almost all of the transit routes in Chicago. In general on rail route all stops are included, and on bus routes stops approximately every mile are included. These data are also obtained by manual counts, and hence, are subject to limitations with respect to accuracy of the count as well as representiveness of an average traffic level on a weekday. The advantage of the CATS data over the CTA data is that most routes were observed for twenty-four hour periods, the data being tabulated for five periods: 6 to 9 a.m., 9 a.m. to 4 p.m., 4 p.m. to 7 p.m., 7 p.m. to midnight and midnight to 6 a.m. In addition to the passenger counts, the number of buses or rail cars passing each of the counting stations in each period was tabulated. In some instances the sample was 100%, while in others a fraction of the vehicle movements past the observation points were observed, the data being expanded to represent total flow during the period in question.

## Bus Transit

### Peak Period Functions

The earlier study by Morlok (1) contained an estimation of peak period bus supply functions for the Chicago Transit Authority routes. The relationship for frequency obtained is given above (equation (2)). An important question is the extent to which this bus frequency supply function remains stable over time, that is, the extent to which transit management would continue to behave in a manner consistent with this function in the future. There is a two-year gap between the set of data on which the supply function was estimated and the most recently available data, and hence, this provides one test of the relationship.

In 1972 supply function was tested in the following manner. Given the actual observed passenger counts on each route in 1974, the 1972 function was used to estimate the frequency of buses that would be provided. The estimated frequency was then compared with the actual 1974 frequency and the standard error of the differences for all routes calculated. The standard error is 4.43 buses per period, which is 19.4% of the mean 1974 frequency of 22.87 buses per two hours. This is a good fit, the standard error of the 1972 function with respect to the 1972 data on which it was based being 18.8% of the 1972 mean. The 1974 data have been plotted in Figure 3, along with the 1972 function. As can be seen, the distribution of errors around that estimating relationship seems to reveal no major change in the function. An indication of this can also be obtained from the linear regression equation based on the 1974 data, which is:

$$\begin{aligned} f_1 &= 8.19 + 0.0130p_1 & (6) \\ \bar{f}_1 &= 20.71 \text{ buses per period} & \bar{p}_1 = 961.4 \text{ passengers per period} \\ R^2 &= 0.870 & \text{Period} = \text{two hour peaks} \\ SE &= 3.41 \text{ buses per period} \end{aligned}$$

The differences between the 1972 and 1974 equations are rather minimal, the threshold frequency dropping by 1.10 buses per two hours and the slope changing by less than four percent.

While this test seems to indicate the relationship probably has remained stable for the two-hour peak period, it certainly is not an ideal test for such a supply function. The ideal test would cover a much longer period but the data for other years is not available and hence a test of a longer period is impossible at the present time.

Another possible test would be to compare the changes of passenger traffic from 1972 to 1974 on each route with the change in frequency over this period. Such changes would reveal the extent to which transit management actually adjusts bus frequency according to the coefficient of the slope of the supply function. However, these data do not lend themselves to such attempts because there actually have been many passenger counts made between 1972 and 1974 (but not published or tabulated in an accessible manner), and many adjustments in the schedule of each route made on the basis of these. The supply function tested thus represents one which has incorporated in it the likely deviation between the frequency on a route at any point in time and the frequency the management would like to operate if it had known the passenger traffic at that time.

These deviations due to variations in passenger traffic for which the appropriate modification to bus frequency has not yet been made. Such deviations would of course be expected to exist at any point in time, since passenger traffic cannot be continuously monitored and bus frequency continuously adjusted to suit that traffic.

### Non-Peak Period Relationships

The new CATS data permitted the estimation of base period (9 a.m. to 4 p.m.) and evening period (7 p.m. to midnight) frequency functions for bus service. As discussed above, the most likely form of these relationships was hypothesized to be of the form of equation (3). Data on only twenty-five routes (fifty data points, one in each direction) were available for the base period, and eleven routes (twenty-two data points) for the evening period. Regression of directional frequencies against directional passenger volume past the peak load point yielded the following relationships:

Base (period two) seven hours:

$$f_2 = 30.5 + 0.016p_2 \quad (7)$$

$$\bar{f}_2 = 48.5 \text{ buses per period} \quad \bar{p}_2 = 1126 \text{ passengers per period}$$

$$R^2 = 0.63$$

$$SE = 8.2 \text{ buses per period} = 0.162 \bar{f}_2$$

Evening (period three) five hours:

$$f_3 = 21.7 + 0.0126p_3 \quad (8)$$

$$\bar{f}_3 = 26.8 \text{ buses per period} \quad \bar{p}_3 = 401 \text{ passengers per period}$$

$$R^2 = 0.52$$

$$SE = 3.77 \text{ buses per period} = 0.141 \bar{f}_3$$

The statistical fit of these relationships is reasonably good, although clearly far inferior to that obtained during the peak period. Perhaps one reason for this is simply that the low marginal cost of operating non-peak period bus trips leads to acceptance of more “sloppiness” in those operations with much less concern for adjusting frequency exactly to that which the traffic seems to justify.

Various other forms of the non-peak period supply functions were hypothesized. Attention was directed toward identifying other variables which might determine the number of bus trips operated during the non-peak period. In principle, the peak period frequency might affect the number of off-peak trips, since if particular buses were associated with a single route, then the number of available drivers and buses would be determined by peak frequency. But this is not the case, because in fact buses can be easily transferred from one route to the other. Also, the extent to which a route was patronized by travelers who travel short distances might affect the frequency, with higher frequencies being provided on routes with short distance travelers for whom the waiting time would be more important than for long distance travelers. Unfortunately, line length and line round trip time had to be used as surrogates for average trip length since no actual trip length data were available. Multiple regressions with frequency as the dependent variable and peak period frequency, line length in miles and line length in

round trip time (hours) as independent variables were performed, but no significant additional explanation of variations in frequency was obtained. Thus, the equations presented above seem to be the best which can be developed on these data for estimating non-peak period frequencies.

The owl service (midnight to 6 a.m.) seems to bear no identifiable relationship with passenger volume, the reason being simply that service on those routes which are operated in the midnight to 6 a.m. period is provided at essentially the same headway throughout the system. In no case is the volume of passengers such that it would tax the route capacity at a frequency of two buses per hour.

## Rail Transit Functions

### Peak Period Relationships

For estimating peak period supply functions, the CTA data were used rather than the CATS data. The primary reason for this choice is that two hours seems to be a more reasonable definition of the peak period than the three hours of the CATS data. Also, two of the CTA rail rapid transit routes operate in one manner during the peak period of slightly less than two hours duration, the manner of operation changing substantially outside of these periods, posing some question as to the definition of operating characteristics during a three hour peak period.

Peak period train frequency and number of cars operated past the peak load point are hypothesized to vary with volume in the manner given by equations (4) and (5). Care must be taken in defining the passenger flow on which train and car frequency would be determined in the case of rail lines because many route which have different designations in public time tables and maps are in fact paired together and operated as a single train route. For example, the Dan Ryan and the Lake Street routes in Chicago are really operated as a single route for purposes of train movements. A train begins the outer end of one line, proceeds through the central business district, and then outbound along the other to the end of that line.

For purposes of determining train requirements over any line the following procedure was used. The maximum passenger volume which that train would have to accommodate along its route – which may encompass two lines in the public designation of routes – presumably is the volume on the basis on which the train frequency and car flow are determined. Thus, the number of trains operated say, northbound on the Dan Ryan and then westbound on the Lake Street line would be determined by the passenger flow past the peak load point on the train's route, that point being on either the Dan Ryan line or the Lake Street line. Much more capacity might be provided past the peak load point on one of the two lines that is necessary, that capacity being determined by requirements on the other line or leg of the train's journey.

Also, it was observed from initial inspection of the data that the number of trains operated in one direction on a line was almost identical to that operated in the other. This is of course necessary on lines which have one terminal in the CBD, with no train storage provisions there, trains having to be shuttled largely empty in the reverse peak direction. However, in the case of paired lines, with both terminals containing storage facilities,

such extra train movement is not mandatory. But since these are made, the reverse flow peak movements are determined by peak train flow requirements. Thus only peak direction train, car and passenger flows are contained in the regressions. The reverse peak flows would be simply those required to get equipment back to its starting terminal.

Following these procedures, regressions were performed between the trains (and cars) operated during the two peak hours and the passenger flow past the peak load point on the train's route in the peak direction. These regressions yielded the following relationships:

$$\begin{aligned} \text{Peak (period one) two hours:} \\ r_1 = 19.88 + 0.00058p_1 \end{aligned} \tag{9}$$

$$\bar{r}_1 = 27.5 \text{ trains per period} \qquad \bar{p}_1 = 13108 \text{ passengers per period}$$

$$R^2 = 0.86$$

$$SE = 2.00 \text{ trains per period}$$

$$c_1 = 20.54 + 0.0114p_1 \tag{10}$$

$$\bar{c}_1 = 172 \text{ cars per period}$$

$$R^2 = 0.97$$

$$SE = 16.6 \text{ cars per period}$$

As the statistics on the goodness of fit indicate, both equations approximate the data very well. The data and the equations are plotted in figures (4) and (5).

The number of cars per train can be derived by dividing equation (9) by equation (10). The resulting equation is:

$$\text{Cars per train}_1 = \frac{20.54 + 0.0114p_1}{19.88 + 0.00058p_1} \tag{11}$$

indicating the expected increase in train length, from one car per train at very low volumes to, in the case of CTA operations, eight cars per train at the highest volume. Of course, in any case, train length would be limited by station platform length.

#### Non-Peak Period Rail Service Functions

Since passenger flow past the peak load seems to explain virtually all the variations in frequency of train operation and number of cars operated on all CTA rail rapid transit routes, it was felt appropriate to begin with a similar functional form for the provision of off-peak train service. It is necessary to use the CATS data for this, because no CTA data on off-peak passenger flows are available. The CATS data contain information on the number of cars operated past each observation point along the route, but no survey information on the number of train operated. However, the CTA employees' timetables indicated that two-car trains were operated during most non-peak period times on all transit routes, with the exception of the Evanston shuttle and Skokie lines which operate one-car trains in such periods, and the North-South line which operates four-car trains during the midday period. Thus, train length is known, but there

is insufficient variation in train length to derive any generally useful relationship for this as a function of passenger volume.

The regression of cars operated against passengers past the peak load point during the seven hour base period in each direction on each of the seven routes for which data were available yielded the following equation:

$$\begin{aligned}
 &\text{Base (period 2) 7 hours:} \\
 &C_2 = 63.02 + 0.0242p_2 \qquad \qquad \qquad (12) \\
 &\bar{c}_2 = 231.0 \text{ cars per period} \qquad \qquad \bar{p}_2 = 6945.8 \text{ passengers per period} \\
 &R^2 = 0.925 \\
 &SE = 29.7 \text{ cars per period}
 \end{aligned}$$

The low value of the standard error, 12.9% of the mean, and the high coefficient of determination indicates a good fit to the data. As was indicated above, all trains included in this regression are two car trains except those on two routes, the Evanston route which has the lowest volume with one car trains, and the North-South “el”, which has the highest volume and mostly four car trains. The data for these routes are circled on the figure.

A similar regression was performed for the evening period (7 p.m. to midnight) and this yielded the following relationship:

$$\begin{aligned}
 &\text{Evening (period 3) 5 hours:} \\
 &c_3 = 42.67 + 0.0217p_3 \qquad \qquad \qquad (13) \\
 &\bar{c}_3 = 82.6 \text{ cars per period} \qquad \qquad \bar{p}_3 = 1836.1 \text{ passengers per period} \\
 &R^2 = 0.738 \\
 &SE = 58.0 \text{ cars per period}
 \end{aligned}$$

The explanation of the variation in cars operated is not nearly as high for this equation as in the case of those developed for other periods. One of the reasons for this is probably the wide variation in actual flows during these five hours, flows probably being their maximum at the beginning of the period and also peaking at some later time in the evening after the termination of evening entertainment in the central business district and other regional centers. Also, one data point seemed to add greatly to the variance, this data point corresponding to one leg of a paired set of routes on which layout and manner of operation of routes seems to require the same number of trains and cars on both legs regardless of traffic conditions. However, it was not felt appropriate to reject this data point from the regression.

The regression for the owl period (midnight to 6 a.m.), yielded the following relationship:

$$\begin{aligned}
 &\text{Owl (period four) six hours:} \\
 &c_4 = 22.2 + 0.0317p_4 \qquad \qquad \qquad (14) \\
 &\bar{c}_4 = 47.4 \text{ cars per period} \qquad \qquad \bar{p}_4 = 791.0 \text{ passengers per period} \\
 &R^2 = 0.784 \\
 &SE = 12.3 \text{ cars per period}
 \end{aligned}$$

Considering the likely wide variation in actual volume between hours during this period, and the likelihood that minimum frequency considerations require trains to be operated even with extremely low volumes, the strength of the relationship with volume is rather surprising. During this period of time, all trains are two cars except for those on the low volume Evanston line, which has one-car trains.

### Operating Plans

As was mentioned above, there are three main options for the operation of rail rapid transit lines: all-stop service, skip-stop service, and zonal express service. The Chicago Transit Authority operates some skip-stop service on all its lines but two. In all cases the stations are divided into three groups – those at which “A” trains stop, those at which “B” trains stop and those at which stops are made by both groups of trains. On the Skokie branch, which only has two stations – one at each end of the line – all trains of course make all stops, and on the Evanston shuttle line, all trains stop at all stations, although some of these trains operate as expresses to the central business district during the morning and evening peak periods and to a very limited extent during the midday base period. Other lines are operated in an all-stop manner during certain periods of the day, and in the A-B skip-stop manner during other periods. As discussed previously, it is hypothesized that, in the balancing of the gains to certain travelers and reductions in cost to the operator resulting from operation of skip-stop service with the losses to certain other travelers from such operations, transit management is likely to introduce skip-stop service during the higher volume periods. Such a hypothesis can be examined with the data available from CATS by examining the volume of traffic on those routes and during those periods in which each of the two services is offered.

The first difficulty, which arises in examining such a hypothesis, is that of specifying precisely which volume is relevant as the criterion. In examining the CTA’s service pattern, one observation is that whenever skip-stop services are operated in one direction, skip-stop services are also operated in the other direction. This would suggest that the criterion might be the passenger volume in the peak flow direction at least during periods when traffic is unbalanced by direction as during the weekday rush hours. The second observation is that on paired routes, skip-stop service is sometimes offered in one of the routes but not on the other. Thus each route which is distinctly designated in public information might be the passenger volume in the peak flow direction at least during periods when traffic is unbalanced by direction as during the weekday rush hours. The second observation is that on paired routes, skip-stop service is sometimes offered on one of the routes but not on the other. Thus each route which is distinctly designated in public information might be considered individually for purposes of determining whether or not skip-stop service is to be provided even though the train’s route might encompass two such public schedule routes.

The CATS data on passenger traffic on non-peak period flows are available for each of seven routes, by direction, for five separate periods during each weekday. The CTA data on peak flows cover the same seven routes with separate data for the peak direction during each of the two daily peak periods. These yield a total of seventy data points.

During the peak periods in the peak flow direction, two of the routes are operated in the all-stop manner and four routes are operated in the skip-stop manner. The maximum volume of traffic in the peak load direction during these periods for the local routes is 2,021 passengers per hour while the lowest volume on those routes operated as skip-stop is 2,378 passengers per hour. This would indicate that the transit management decides to operate skip-stop service when the volume reaches approximately 2,100 to 2,200 passengers per hour in the peak direction during peak periods. The highest volume on any route is 10,309 passengers per hour. The one express service, that between Evanston and the central business district, carries approximately 1,600 passengers per hour in each direction during the peak periods. If this service were not operated as an express, and all passengers carried on the skip-stop trains of the North-South "el", the volume on that route would reach approximately 10,000 passengers per hour. Thus, there might also be a volume criterion for the provision of express service, although such service would be difficult to operate elsewhere on the CTA system simply because all routes except this northern route are limited to two tracks.

During the non-peak periods local services are characterized by volumes no greater than 931 passengers per hour, while all skip-stop services have volumes greater than 1,129 passengers per hour. This suggests that during non-peak periods, when trains are shortened and any given passenger volume results in a greater frequency than would be the case during peak periods, the criterion for provision of skip-stop service probably is in the vicinity of 1,000 to 1,100 passengers per hour.

Thus the management seems to be providing local and skip-stop services in a manner which is consistent with the hypothesis of a criterion passenger volume which is felt to justify the provision of higher quality services.

## Discussion

### Use in Planning

The use of supply functions such as those developed above seems to offer substantial advantages in urban transportation planning network models. The first of these is that will tend to overcome the necessity of making rather simplistic assumptions regarding the kind of service to be operated under various conditions, as is now the case. In most planning efforts, it seems that bus frequency is determined by assuming that each bus operated during the peak period would be filled to its normal capacity past the peak load point. The empirical evidence presented above clearly indicates that transit management in one large city does not do this, rather filling buses to an approximate load of about forty-five passengers per bus. Furthermore, management seems to behave in such a manner as to provide a minimum level of service on each route, increasing frequency and hence the capacity provided as traffic increases.

In the case of rail transit the necessity of making assumptions in the absence of supply functions is even more drastic because there, train length must be assumed before the train capacity can be adjusted to traffic volume. As the above relationships indicate, at least on the Chicago transit lines, the management seems to balance service considerations with the economies resulting from operation of longer trains, the actual

operating policy being to operate one or two car trains at low traffic volumes, increasing train length to achieve economy as traffic increases, with a slight increase in train frequency with increasing volume.

During non-peak periods, traffic on routes is generally so light that it would result in excessively long headways if each vehicle were to be filled to capacity. Buses and trains are consistently operated at less than full seated capacity in a manner closely related to volume, however. Thus the use of supply functions, as those outlined above, would seem to overcome the many problems inherent in having to make assumptions regarding management decision-making behavior.

A possible objection to the use of these supply functions is that in the context of transportation planning it is often desirable to explore the effects of different levels of service in transit patronage and other impacts, including levels of service which are different from those which transit management would ordinarily provide. This is certainly a very important class of options to be explored, and the construction of the supply functions is not meant to suggest that it should not be undertaken. However, it may be useful to have the supply function information available as part of such an exploration of alternative level of service policies for transit routes. In the context of exploring such alternative service policies, one policy to be explored should be that of management's current policy. In this manner the current policy of transit management could be evaluated, to determine if any significant change in that policy is desirable. If it is desirable, then the supply functions will provide a basis for determining the extent to which a higher level of service, such as more buses or trains per hour, would in fact be required. Knowing what the required shift in the supply function is, then various mechanisms such as subsidies or changes in regulations, could be explored. A closely related benefit of using the supply functions is that it does provide a way of predicting what quality of service transit management will provide on any new routes. One of the problems which seems to be characteristic of planning for major new transit routes is that it is difficult to predict what quality of service will be provided when those routes are actually opened for use. This provides a way of predicting that service, in a manner compatible with network equilibrium models. If that service will not lead to the derivation of maximum benefits from the project, then plans can be made to induce changes in the service provided.

These supply functions provide not only a basis for estimating the level of service but also for estimating the costs associated with providing the service. As has been revealed, the service frequency and capacity provided may in fact be in excess of that required on the basis of operating all vehicle trips at maximum rated capacity during peak periods. The operation of excess capacity and more frequent trips clearly results in the need for more vehicles and operators. In general this will increase both the capital cost and the operating cost associated with that route. These functions will provide a basis, therefore, for a more accurate estimation of the future costs for any given route.

### Network Equilibrium Modeling

The preceding analysis has revealed that the supply of public transit service on fixed routes is such that the user's travel time is generally a decreasing function of passenger volume. Theory suggests that waiting time will decrease once the minimum

frequency of vehicle trips are filled, while the empirical evidence suggests waiting time is a decreasing function of volume over the entire range. The only exception is due to higher vehicle speeds while shorter distance travelers may experience some increase in travel time due to reduced frequency. Nevertheless, the overall trend is definitely toward reduced travel time with increasing volume. This would be further reinforced if fixed plant changes were considered, such as special lanes for buses, priority signals for buses, busways, and transit – all changes which are more likely as traffic increases.

Virtually all existing network equilibrium models are solved using algorithms or heuristic rules which are derived using the assumption of increasing user costs as a function of volume. This assumption is of course quite correct in the case of automobiles on roads, but is inappropriate in the case of public transit. For the most part, transit time will decrease with volume, given that management has sufficient time to adjust service. Since most carriers revise schedules at least twice a year, the time lag is quite short in comparison to the duration of most planning horizons.

This is not to imply that transit time cannot be increasing with volume. Obviously in the very short run – a period in which management can not respond to traffic changes – travel time will increase if the traffic increase is sufficient to lengthen dwell times for loading and unloading and perhaps affect running speeds. In a system with little slack in the schedules at turn around points, such lengthening of the running time of one trip will affect later trips, resulting in longer headways, more passengers boarding at each stop, resulting in further delays, etc. Of course, management will attempt to minimize such deviation of service from that planned.

Of perhaps more significance for bus services is the congestion on the streets used. The running times of transit vehicles will be affected by the prevailing traffic flow conditions, and hence these must be considered in estimating the running times. From the standpoint of transit supply, other traffic can be considered an independent input. Transit vehicle flow is usually so small compared to total vehicle flow that the volume increases probably will not significantly affect overall flow and hence bus speeds. However, if the bus volume becomes very large, running time will be affected. It may increase, as would perhaps be expected on intuitive grounds. On the other hand, the heavy volume of buses may in effect preempt a lane, and perhaps move more rapidly than if the bus volume were less. The question of bus running time and the factors which influence it requires much research.

Although consideration of these factors will further complicate supply functions in public transit, it is clear that service will for the most part improve with increasing volumes. This is provided that appropriate schedule adjustments can be made, as seems reasonable in the time frames for most transportation planning efforts. In this context, the current procedures and underlying theory for finding network equilibrium will have to be examined for applicability to modes in which service can improve with volume. Although it is impossible to prejudge which methods will be applicable to public transit and multi-modal networks, it is possible to conjecture some of the problems that may be encountered. One is likelihood of multiple equilibria, which raises interesting mathematical as well as social policy questions about the selection of a “best” equilibrium and how to achieve it. It is likely that realistic modeling of the demand for transport will require consideration of a vector of level of service measures, including time, price, comfort and perhaps others. This will require the seeking of an equilibrium in

many dimensions, a problem undoubtedly much more complex than that for a single dimension, such as the travel time in most models developed for the road context. Thus there is a wealth of interesting and challenging problems posed for network equilibrium modeling by the treatment of public transportation.

### Conclusions

The supply of public transportation or mass transit service in an urban area seems to be amenable to quantitative treatment in a manner broadly compatible with current urban transportation network models. However, the supply functions are quite different from those used for automobile transport. Transit management determines the supply of transit service, subject to factors beyond its control such as road traffic conditions and regulatory constraints.

This research has just touched the surface of potential research into this aspect of urban transportation modeling. Much more needs to be done in the testing and refinement of models for services in which the route structure is given the major focus of the work reported. Data for other cities and operating agencies need to be explored, as do questions of estimating bus journey times in maxed traffic. Also, questions of route structure and possibly choice of technology should be addressed.

### Acknowledgements

The authors wish to acknowledge the assistance of Ronald Eash of the Chicago Area Transportation Study, Thomas Luglio of the Southeastern Pennsylvania Transportation Authority, and Mehmet O. Akyilmaz and John A. Warner of the University of Pennsylvania.

## References

E.K. Morlok. Supply functions for Public Transportation: Initial Concepts and Models. In Proceedings of the First International Conference on Transportation Network Equilibrium Methods (M. Florian, ed.), Springer-Verlag, Berlin and New York, 1975.

A.K. Sen et. al. On the Statistical Estimation of Transit Supply Functions. Research Report, College of Urban Sciences, University of Illinois, Chicago, 1975.

D.A. Quarmby. Implementing Research in a Transit Operating Agency. In Proceeding of the First International Conference on Transportation Research. The Transportation Research Forum, Chicago, and The College of Europe, Bruges, Belgium, 1973, pp. 875-881.

Chicago Transit Authority. CTA Bus and Rapid Transit Operating Facts, Winter 1973-74 Fact Sheet no. 0P-y 74308, July, 1974.

Chicago Transit Authority. Rapid Transit System Operating Facts, Sheet no. RP-x 73198, October, 1973.

Chicago Area Transportation Study, Computer Tabulations of Transit Route Service and Ridership. Chicago, 1974.

## List of Figures

- Fig. 1            1972 Chicago Peak Period Bus Route Frequency and Traffic
- Fig. 2            Types of Bus Service Operation
- Fig. 3            1974 Chicago Peak Period Bus Route Frequencies and Passenger Traffic
- Fig. 4            1974 Chicago Peak Period Rail Route Train Frequencies and Passenger Traffic
- Fig. 5            1974 Chicago Peak Period Rail Route Car Frequencies and Passenger Traffic

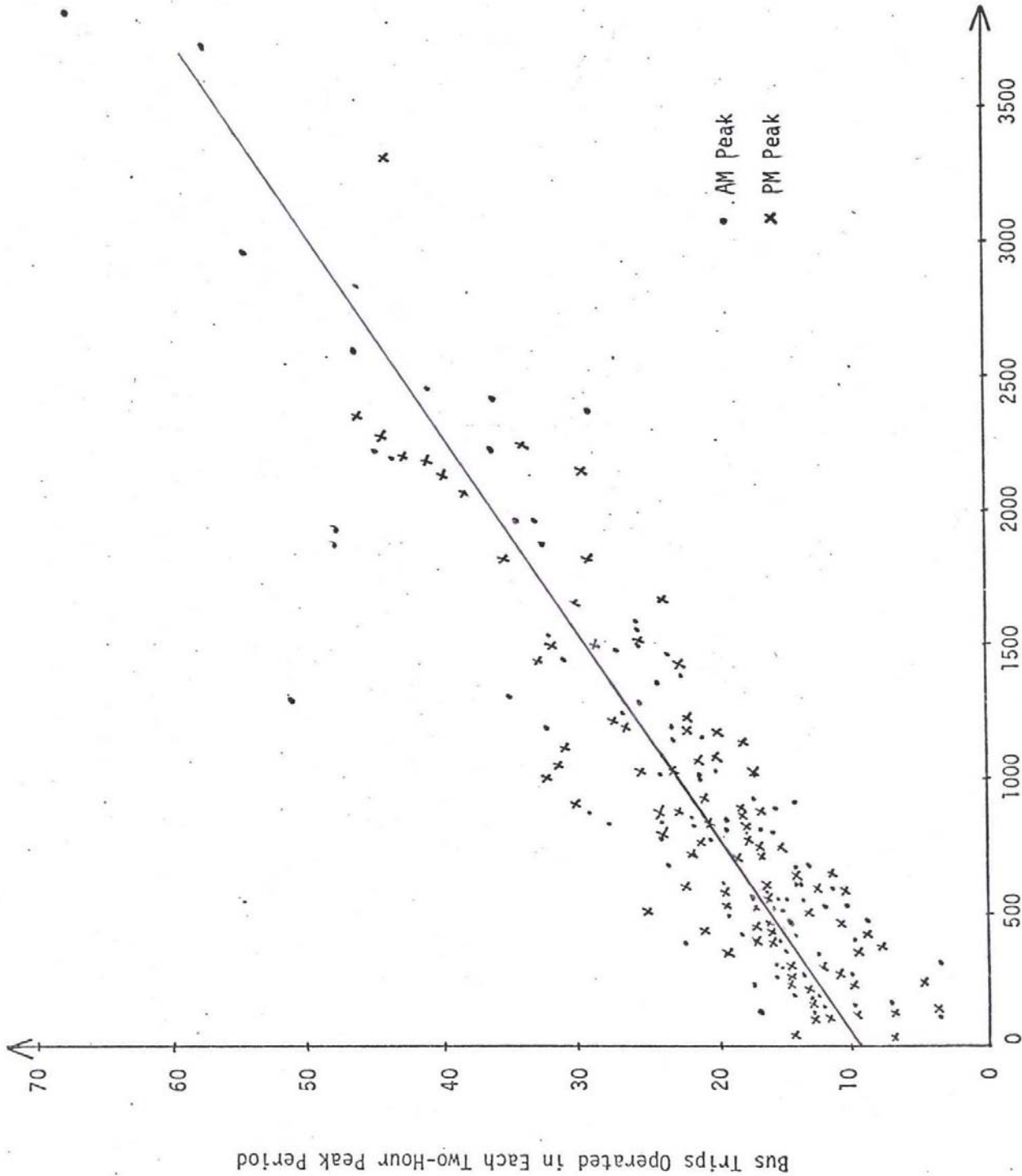
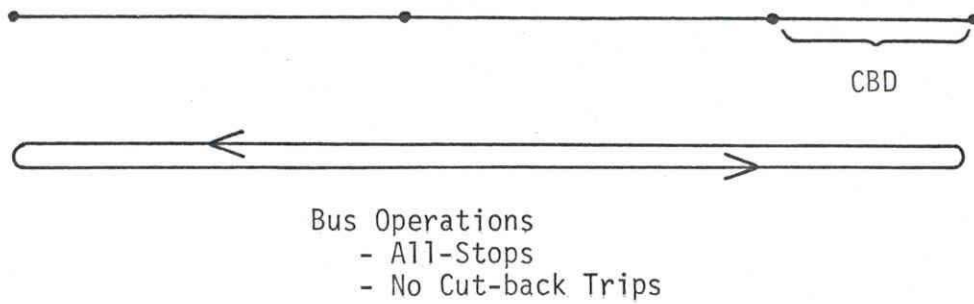
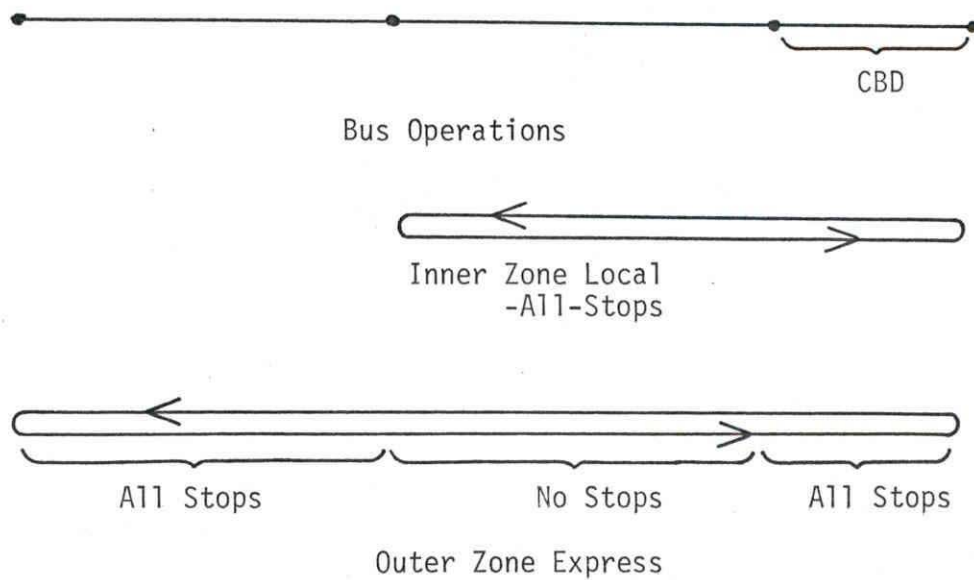


Figure 1. 1972 Chicago Peak Period Bus Frequency and Traffic Data and Regression Lines.



2a. All-Stop Local Service



2b. Zone Type Service

Figure 2. Types of Bus Service Operation.

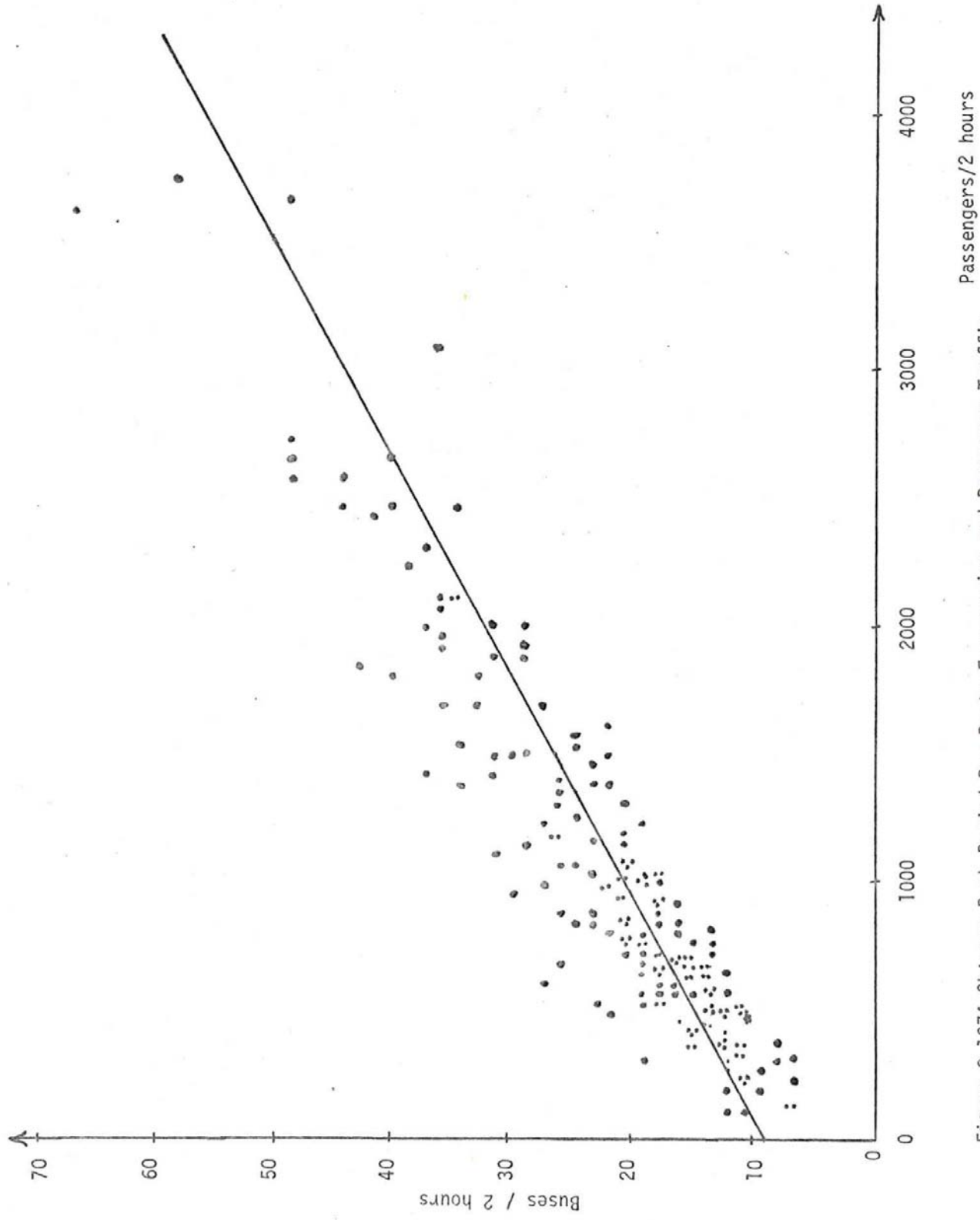


Figure 3.1974 Chicago Peak Period Bus Route Frequencies and Passenger Traffic

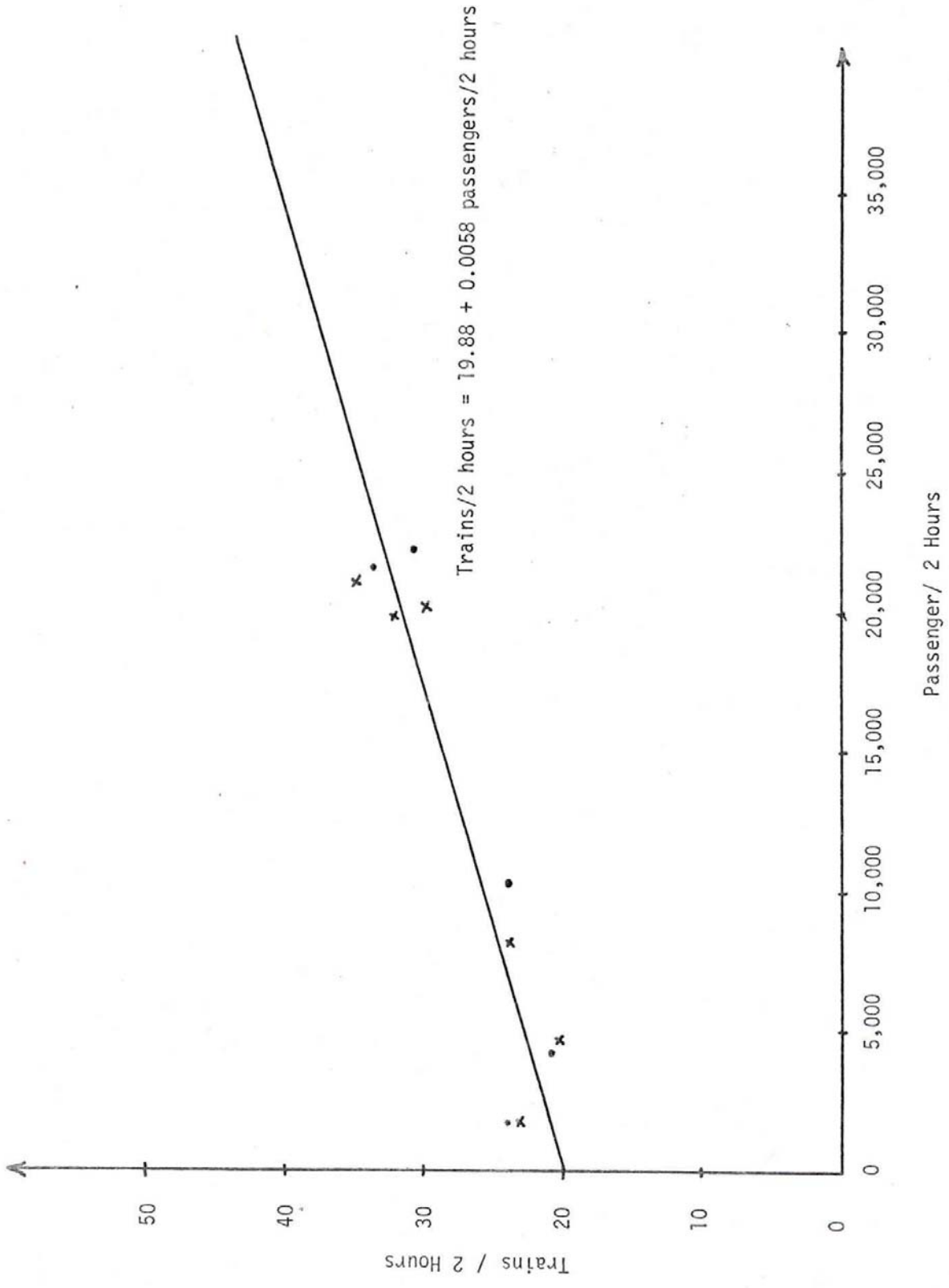


Figure 4. 1974 Chicago Peak Period Rail Route Train Frequencies and Passenger Traffic

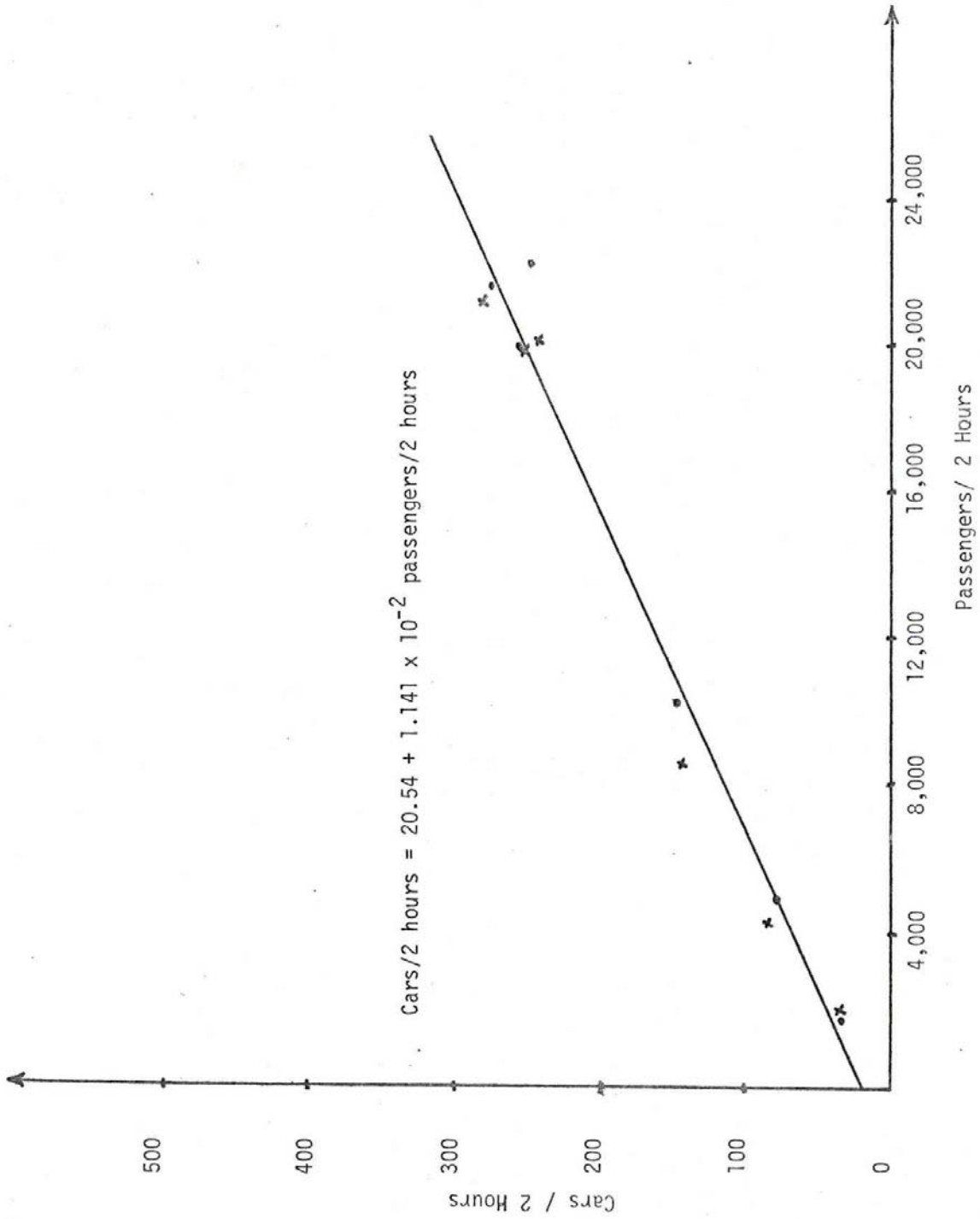


Figure 5. 1974 Chicago Peak Period Rail Route Car Frequencies and Passenger Traffic