6 Patronage Analysis

In chapter 5, patronage was treated as a parameter in the cost effectiveness analysis. Such a procedure is useful for two reasons: (1) it separates the problem of analysis and discussion of cost effectiveness from the complex and controversial problem of determination of the patronage by treating patronage as a parameter; (2) it gives the systems analyst a good feeling for the range of patronage needed to recommend proceeding with detailed planning and design of a proposed system, for the variation of cost effectiveness with patronage, and for the accuracy with which patronage must be determined in specific cases; and (3) it enables the system analyst to explain the cost effectiveness behavior of the system to cognizant decision-making bodies with patronage viewed as a policy variable, which indeed it is in many cases.

In the state of transit development at the time of this writing, many engineers choose simply to ignore the problem of determining the patronage, and usually implicitly, to have faith that their system design will attract sufficient patronage to make it worthwhile. In the author's opinion, this attitude is at the root of most of the intense controversy over various transit options.

Patronage analysis is behavioral analysis, and is outside the range of knowledge and experience of most engineers. But the transit systems engineer simply must understand something of the technique of patronage estimation, and in planning the development of new systems he must understand the various behavioral factors that will influence people either to ride or not to ride his system.

The details of patronage analysis are very complex and are best left to specialists; however, the systems analyst must be able at least to make rough estimates to satisfy himself that the detailed calculations are reasonable. A good overview of the techniques of patronage analysis is given by Hutchinson[1]. References [1 through 8] will give the interested reader a good grasp of the problem of patronage analysis or demand estimation. In the design of new conventional systems for which operating experience can be used to calibrate the patronage models, the theory has been found to yield good results; however, in the planning and design of new transit systems thought to be able to increase patronage markedly, the extrapolation of existing models is risky and imprecise at best. Nonetheless, if progress is to be made toward solution of pressing transport problems, the problem must be treated in a variety of ways. Construction and operation of new systems in urban areas is of course the only acceptable final proof, but it is probable that much useful information can be obtained by carefully designed behavioral experiments and observations of human behavior in analogous situations, by analysis of all the steps the patron must take in making a trip on the new system, and by use of opinion surveys. An annotated bibliography of the literature is given in reference[9].

This chapter is intended as a first exposure to the problems of patronage analysis for new systems for which no operational experience is available. The material presented will assist the systems analyst to make rough preliminary estimates; however, again he should be cautioned to consult experts when making detailed estimates upon the basis of which decisions to invest funds are to be made.

6.1 Relationship between Yearly, Daily, and Peak-Hour Patronage

Three patronage parameters appear in chapter 5: the number of trips per year, per work day, and per peak hour. The first is the parameter of significance in determination of cost effectiveness and the third is needed for the estimation of capacity requirements. The second is a convenient intermediary value. In chapter 5 the following assumptions were made:

$$\text{Trips per year} = 300 \times \text{Trips per work day}$$

$$\text{Trips per work day} = 10 \times \text{Trips per peak hour.}$$

Since there are about 254 week days per year not counting holidays, the factor of 300 tacitly assumes that the traffic on an average one of the 111 weekend days or holidays is 46/111 or 41 percent of the traffic on a typical week day. The number 300 is close to that assumed by many consultants (some use 299 which looks more precise, but probably is not), but to determine it precisely would require far more extensive traffic surveys than usually can be afforded. In planning new systems in certain institutions, the factor of 300 may not be appropriate. For example, hospitals experience a more uniform traffic flow, and universities usually operate on fewer than 254 regular school days per year.

The ratio of daily to peak hour travel can be determined from graphs of traffic volume as a function of time of day. Such graphs are given by Meyer, Kain, and Wohl[10] for city-wide travel. The data presented shows that the factor of ten assumed in chapter 5 is high for auto drivers, but low for conventional transit. One may assume that the ratio for a new automated system may lie in between; however, that depends on the use of the system. If it is a line-haul system, primarily used to take people between home and work, the ratio may be closer to seven or eight. But if the system is a collector-distributor used more uniformly throughout the day, the ratio of daily to peak-hour travel may be higher than ten.

As indicated in chapter 5, the ratio of daily to peak-hour travel is used to estimate capacity requirements. Therefore, the value used is to some extent a matter of policy because it determines how much effort is to be expended to stagger the use of the system in rush periods.

6.2 Mobility

Equation (5.3.12) gives the trip density as a product of three terms: the modal split to transit, discussed in section 6.6; the daytime density of people in the service area of the transit system; and the factor τ_d called the mobility. The mobility is the number of trips per person per day, or some multiple of it. Zahavi, in reference[11], defines mobility as the number of trips per day per 100 residents. In table 6-1, his data is retabulated per resident. The table shows that τ_d varies from 1.65 in the high density area around New York City to 3.18 in Oklahoma City. The population weighted average value for three intermediate cities, Baltimore, Cincinnati, and Washington, is 1.99, and for the remaining smaller cities is 2.45. Thus, mobility is correlated with city size and density. Zahavi also shows that mobility increases with average trip speed in such a way that the most nearly constant parameter is the daily travel time budget. In other words, if the average speed of travel reduces, the average person takes fewer trips, that is, the mobility declines;

Table 6-1 Mobility in Various Cities
(Mobility = τ_d = trips per resident per day)

Tri-State	1.65	Springfield	2.17	Pulaski	3.09
Baltimore	1.72	Salt Lake City	2.48	South Bend	3.04
Cincinnati	2.17	Orlando	2.58	Columbia	2.79
Kansas City	2.00	St. Petersburg	2.17	Monroe	2.99
Indianapolis	2.14	Peoria	3.03	Fort Smith	2.26
S. E. Virginia	2.25	Baton Rouge	2.51	Rapid City	2.49
Oklahoma City	3.18	Knoxville	2.49	Washington	2.07

Source: reference [11].

or if the average speed of travel increases the mobility increases. The conclusion is that the estimate of patronage on a new automated system should not be based on the same center city mobility that exists prior to its installation but on a mobility adjusted according to the average speed provided by the new system compared with the average speed of travel prior to its construction. Thus, the assumption of a mobility of three trips per person per day, used in figure 5-7, is felt to be justified as a basis for preliminary estimates.

6.3 Required Precision of Patronage Estimates

Before spending a great deal of time estimating a difficult variable, it is important to estimate how accurately the variable must be known. Such an estimate can be obtained for the patronage variable by examining of cost per trip versus trip density, such as shown in figures 5-2, 5-7, and 5-8; or curves of present value versus trip density such as shown in figure 5-9. It is interesting to note from figures 5-7, 5-8, and 5-2 that, in the region of trip densities in which the system is cost effective in comparison with the bus or auto system, the cost per trip varies slowly with trip density; whereas in the low trip density region where the system is not comparatively cost effective, the cost per trip is very sensitive to changes in the estimate of trip density. For the automated network system assumed in figures 5-7, 5-8, 5-9, the transition occurs at about forty trips per day per hectare, and for the bus system of figure 5-2, the transition occurs at about twenty trips per day per hectare.

The meaning of the transition at forty trips in regard to the conditions in which it can occur is elucidated in figure 6-1, in which equation 5.3.12 is

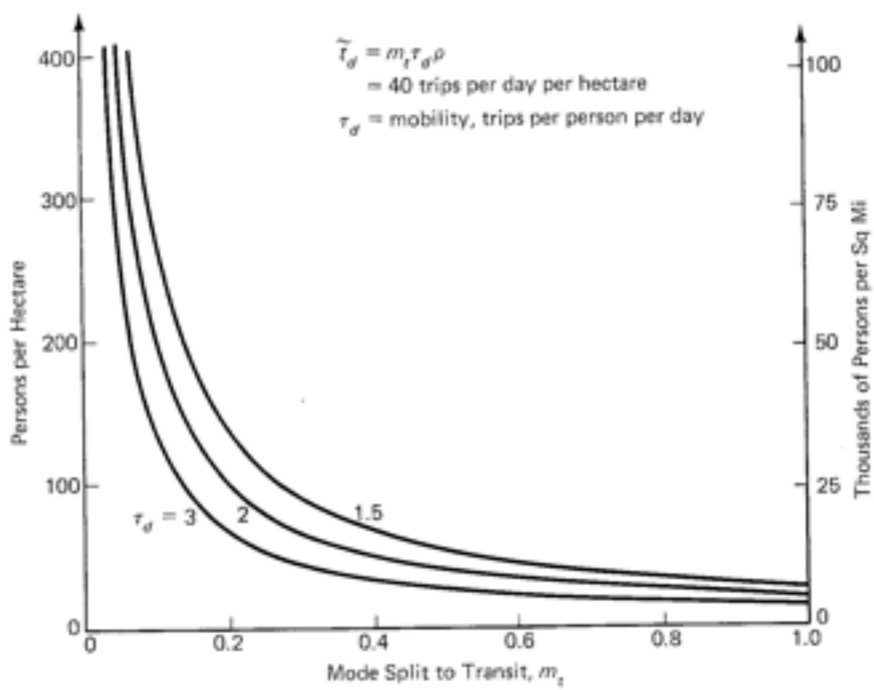

Figure 6-1. The Population Density Required to Achieve $\tilde{t}_d = 40$ Trips per Day per Hectare

plotted for $\bar{t}_d = 40$ trips per day per hectare, for three values of the mobility factor τ_d. Based on the discussion of section 6.2, it is reasonable to assume that the mobility lies within the range of the curves plotted. Then, the curves give the daytime population density in the transit service area required as a function of mode split to achieve a trip density of forty trips per day per hectare. It is seen that the required population density rises very rapidly after the mode split falls below about 20 percent.

The analyst can examine data on daytime population density independently to determine if, with reasonable mode splits, the system can be close to the range of cost effectiveness. If the judgment can be made that the mode split cannot be high enough for cost effectiveness, then the project can be abandoned without going through the expensive procedure of accurate mode split estimation. If the system is in the cost effective range, then fortunately the computation of patronage need not be precise for the economic analysis. It must of course be accurate enough to determine if capacity limitations will be approached.

In some cases, previous experience will indicate that with free competition from the automobile, the system cannot be cost effective. A policy decision can then be made in regard to the imposition of auto disincentives such a high parking fees to increase the transit mode split. However, as Zahavi points out, one must not assume that raising the mode split will automatically raise the transit trip density; because, if the transit system is too slow, auto disincentives may reduce mobility by a greater factor than the mode split to transit is raised. On the other hand, if the new automated mode has speed and service characteristics superior to the auto in the downtown situation, auto disincentives may not be necessary to attract an adequate mode split.

6.4 Trip Generation

The first step in patronage analysis is to estimate the total number of trips that could be served by the proposed transit system. In chapter 5, total trip density was defined as the product of the mobility and the density of people who live, work, shop, and seek recreation in the area served by the automated system. If the automated system is to serve a major activity center, this total person density can be many times the resident population density. The total travel is then the product of person density, mobility with the new system in place[a] and a suitably defined transit service area. The transit-service area includes at least the area within walking distance of stations, typically considered to be a quarter of a mile (0.4 km). This is

[a]Here the mobility is the ratio of the total number of trips within the transit service area to the daytime population of that area, and may bear no relationship to the values in table 6-1.

because data on bus travel indicates that only a very small fraction of bus trips either originate or teriminate more than this distance from bus stops. At an average walk speed of 2 mi/hr (3.2 km/h) a quarter of a mile is a 7.5 ~~fifteen~~-minute walk[12].

If the transit service area is taken to be a larger area, then it is a multimodal area and can be treated as such. If the patronage is based on the total daytime person density in the transit service area, then account is already taken of the fact that the people upon which the patronage estimate is to be based have somehow arrived within the transit service area. If they take the transit system under consideration, they are making a multimodal trip, at least for their trips into and out of the transit service area. This kind of trip is discussed in the following section. If the transit system is anticipated to use a feeder mode regularly and if it is region wide, then a secondary service area around each station should be defined consisting of the area beyond walking distance but within a distance from which trips can reasonably be expected to be drawn.

The outer boundary of the secondary service area is of course not sharply defined, but an indication can be obtained from data on specific operating systems. In the case of BART, data taken in 1975 showed that of all the people that ride the system, 26 percent arrive at the stations by foot, and 16 percent by bus. All but 2 percent of the remainder arrive by auto. On the other hand, of all of the people that leave BART stations, 68 percent walk to their destinations and 26 percent leave by bus. Thus, the effective service area of the destination station is substantially smaller than that of the origin station. In reference[13], Figure 1 (reproduced in this book as figure 6-2), access mode split curves are shown which are calibrated based on BART data. They show that very little bus patronage comes from beyond about 5 km (3 mi) of a station, and that the bus mode split is maximum at about 1.6 km (1 mi) from the stations. It can be argued that these data may not be representative because the schedules of bus service to BART stations are in need of improvement. Substantial improvements would, however, be expensive and, in the face of the need to improve BART patronage, have not been implemented. The curves of Figure 3, reference[13], show how the transit, auto driver and auto passenger mode split varies with the size of the access and egress transit service areas. If, for example, the mode split is computed for all trips in which both the origin and the destination are within 1 km (0.6 mi) of a transit station, only 9 percent of the trips can be expected to use transit with the assumptions made. Thus, Figure 3 shows how the transit mode split decreases as the size of the service areas increase.

The purpose of the preceding discussion was to indicate that the concept of a transit service area is useful in roughly estimating the potential transit patronage if account is taken of the fact that the mode split decreases if the service area around each station increases. Figure 4 of reference[13]

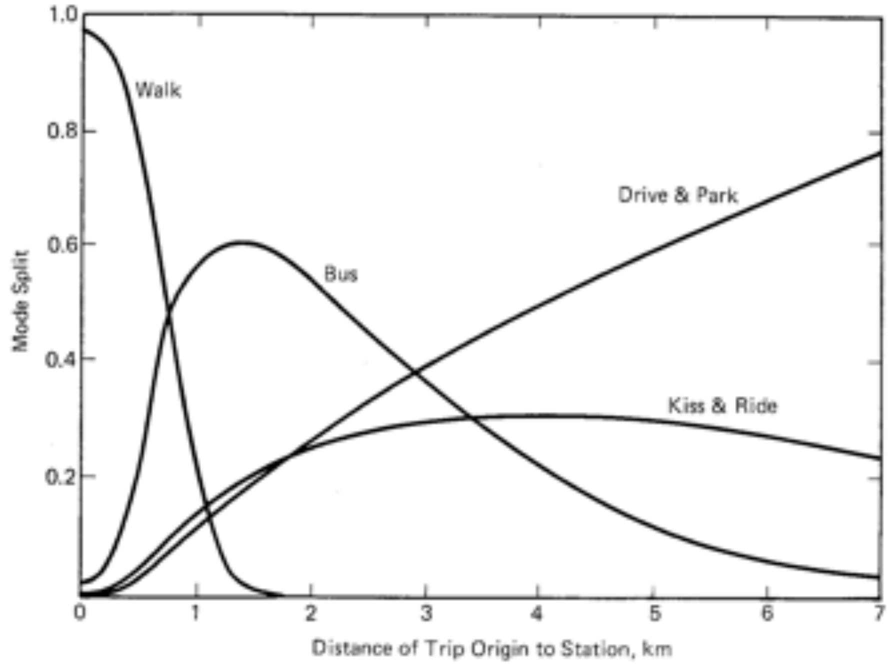

Figure 6-2. Access-Mode-Split Functions

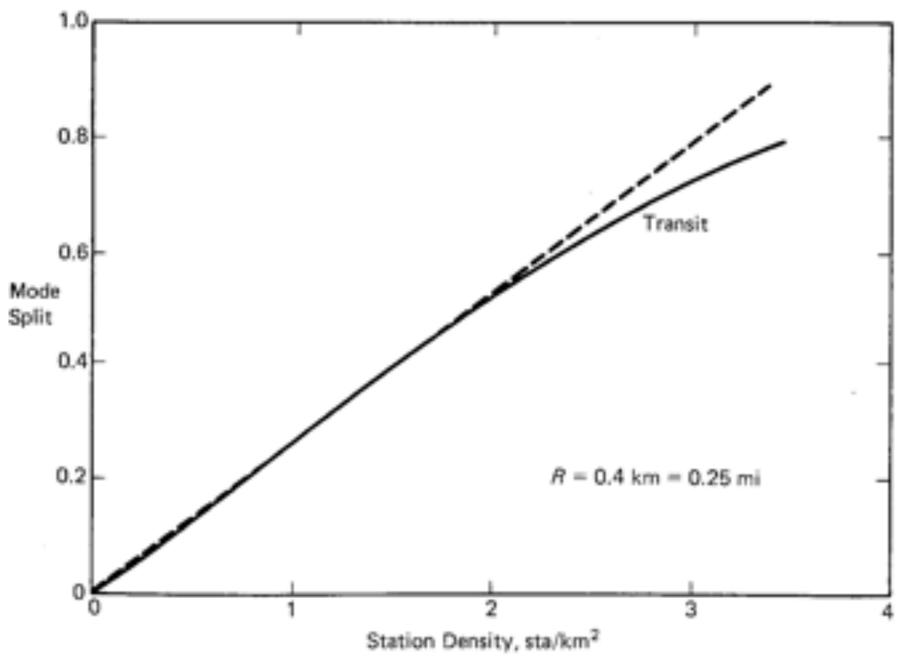

Figure 6-3. Transit Mode Split as a Function of Station Density

gives further insight into this phenomenon (reproduced here as figure 6-3). Here the transit mode split is plotted as a function of the inverse of the service area around each station, and it is shown that for service areas larger than that corresponding to the walking service area, the mode split is inversely proportional to the service area around each station. Thus, if A_s is considered to be the area around each station from which trips are drawn, and n_s is the total number of stations, the total transit service area is $A_{ts} = A_s n_s$. The mode split is approximately of the form $m_t = k/A_s$, in which k is independent of A_s. Then the total number of trips per day attracted to the transit system is

$$t_d = m_t \tau_d \rho A_s n_s = k \tau_d \rho n_s$$

Thus, the total patronage depends approximately on the number of stations and the mean population density at the stations, and not on the size chosen for the service area.

It is seen that the discussion of trip generation could not be divorced from a discussion of mode split, and that the problem of selecting the area from which transit trips are generated is secondary to the problem of determining the mode split for one particular service area.

6.5 Trip Distribution

The trip distribution is a matrix of origins and destinations of all trips, usually tabulated in terms of traffic assignment zones. In a macrosense, the trip distribution is needed in the analysis of a limited area automated transit system because the mode split may be different for trips with only one end in the transit service area, than for those with both ends so located. Thus, in general, the total number of trips on the transit system can be written

$$\text{Total trips} = m_{ii} t_{ii} + m_{io} t_{io}$$

in which m_{ii} is the mode split for the t_{ii} internal trips, and m_{io} is the mode split for the t_{io} trips which have one end outside the service area. Unless there is an auto disincentive within A_{ts}, it is likely that the inconvenience of transferring between modes will cause m_{io} to be considerably less than m_{ii}.

In a microsense, the trip distribution is the distribution to and from specific stations in the transit system. This distribution is needed to determine if capacity limitations are a problem at any of the station locations, or on any of the links. If this is the case, capacity may be increased by rerouting, or stations may have to be enlarged, or more stations and links may have to be added.

The full theory of trip distribution analysis is complex and extensive, and in any detailed analysis experts should be consulted. A good review of the methodology is provided by Hutchinson[1].

6.6 Mode Split Analysis—A Probability Argument

Anderson[13] has introduced the argument that the mode split to a transit system is proportional to the product of two probabilities: (1) the probability that the origin of the trip is within a reasonable distance of a station of the transit system; and (2) the probability that the destination is within a reasonable distance of a transit station. "Reasonable distance" does not necessarily mean walking distance, but it is the distance relative to the total length of the trip that will cause the traveler to feel it is worthwhile in comparison to an auto trip to go to a station by some means (auto, bus, walking), wait for and ride the major transit system, and then go to the destination by foot or by feeder bus. A reasonable distance will generally be small with respect to the length of the trip and less than the station spacing, otherwise the traveler will take a more direct route. Thus, if the average trip length is say, ten miles, it seems reasonable that the "reasonable distance" will not be more than one or two miles.

Each of the above probabilities is the total number of trip ends within reasonable distance of a station divided by the total number of trip ends in the urbanized area. If the trip end density is uniform, then each of the above probabilities is simply the station density, and it follows that the mode split is proportional to station density squared. With home-based trips, if the residential density is substantially uniform and the non-home ends of the trips are all concentrated at stations through judicious selection of station locations and concentrated development of work-shop-recreation locations, then the non-home trip end probability is unity and the modal split is proportional to the first power of station density. In the extreme of concentrated development of housing and other structures, both probabilities are unity and the mode split is maximum. For an automated system in a central business district in which the concentration of activity is relatively uniform, it would be expected that the mode split to the system will be proportional to the station density squared. For an entire urban area, a reasonable first approximation is that mode split is proportional to the first power of station density. Equation (4.5.5) shows that station density is proportional to L^{-2}, where L is the line spacing. With this assumption, equation (5.7.3) shows that for network systems, the cost per trip is a linearly increasing function of line spacing.

The above argument is useful to give a feeling for the gross behavior of mode split of network systems with line spacing, but it seems legitimate to apply it for gross estimates only for cases in which the behavioral attributes of the two modes compared are nearly the same. If significant differences exist in average speed, waiting time, availability, cost, physical or psychological comfort, or the like, the probability argument cannot be expected to be useful. These additional attributes form the subject of section 6.8.

6.7 Mode Split Analysis—The Logit Model

Experience has shown that the logit model for determination of choice between two or more alternatives is, at least in the transit mode choice situation, more satisfactory than other models[1]. The logit model is therefore gaining popularity in practical applications. It has the advantage that it is not ad hoc, but can be derived from fundamental considerations. It can be calibrated based on experience in such a way that it has been used successfully for predictions of mode split in some cases. The reader will appreciate the model, its strengths and limitations, and the discussion of section 6.8 on factors that influence patronage much better after having studied the following derivation of the logit model.

Let m be the mode split to the subject transit mode, that is, the fraction of the total number of trips taken by the subject mode. The mode split m is a function of various attributes $x_1, x_2, \ldots, x_q$, which are perceived by the individual traveler to a greater or lesser extent in all modes. For some modes, a specific attribute may be insignificant. For convenience and consistency, let each of the x_i be chosen in such a way that m is a monotone, continuous, decreasing function of x_i for all i. By definition, the function $m(x_1, x_2, \ldots, x_q)$ is bounded between zero and one, and by choice of the meaning of each attribute,

$$\frac{\partial m}{\partial x_i} < 0 \text{ for all } i$$

Then, the following postulates lead to the logit model:

1. The attributes can be treated as independent variables, that is, it is possible to vary only one of them while holding all others constant.
2. The mode split $m(x_1, x_2, \ldots, x_q)$ does not reach 0 or 1 for any finite value of any of the x_i, that is, m approaches 0 only as x_i approaches $+\infty$, and m approaches 1 only as x_i approaches $-\infty$, for all i.

Because of postulate 1, it is possible to consider the function $m(x)$, where x is any of the q attributes. Consider $m(x)$ in the neighborhood of a value of $x = x_1$ for which $m(x_1)$ is much less than 1. Then, because of postulate 2, an increase in x to $x_1 + dx$ causes dm to decrease but never to become negative. Therefore dm must be of the form

$$dm = -\alpha_1 f(m) dx \tag{a}$$

in which α_1 is a positive constant; $f(m)$ is a monotone, increasing, continuous function for which $f(m) > 0$ for $m > 0$; and $f(0) = 0$. Then, $f(m)$ can be expanded into a series of the form[b]

$$f(m) = m^{\nu}(1 + c_1 m + c_2 m^2 + \ldots)$$

$$\approx m^{\nu} \tag{b}$$

for very small values of m. If $\nu \neq 1$, equation (a) can be integrated into the following form for $0 < m_1 < m << 1$.

$$\alpha_1(x - x_1) = -\int_{m_1}^{m} m^{-\nu} dm = \frac{-1}{1-\nu}\,(m^{1-\nu} - m_1^{1-\nu})$$

from which

$$m = [m_1^{1-\nu} - (1-\nu)\alpha_1(x - x_1)]^{1/1-\nu}$$

There is always a finite value of $x = x_2(\nu)$ for which the term in brackets vanishes. Thus, if $\nu < 1$, $m(x_2)$ vanishes; and if $\nu > 1$, $m(x_2) = \infty$. Neither of these forms is admissible according to postulate 2. Therefore $\nu = 1$ and equation (a) becomes

$$dm = -\alpha_1 m\, dx \qquad m << 1 \tag{c}$$

which satisfies postulate 2 for all dx for which $\alpha_1\, dx < 1$.

It is useful to note that equation (c) makes sense from the behavioral viewpoint: If t_t is the number of transit trips, and t_T is the total number of trips,

$$m = t_t/t_T$$

---

[b]By permitting the constant ν to take any value, $f(m)$ can approach zero with any slope from zero to infinity. The coefficient of the first term in the power series expansion can be taken equal to unity because $f(m)$ is multiplied by an arbitrary constant in equation (a).

For $t_t << t_T$, t_T is affected very little by changes in t_t. Therefore

$$dm = \frac{dt_t}{t_T} - \frac{t_t\,dt_T}{t_T^2} \approx \frac{dt_t}{t_T}$$

Therefore, equation (c) becomes

$$\frac{dt_t}{t_t} = -\alpha_1\,dx$$

This equation states that, if t_t is much less than t_T, a given change in attribute x causes a certain percentage change in t_t regardless of the size of t_t, that is, the portion of people who change their travel modes as a result of the change dx is proportional to the number of people who use the transit mode before the change dx. This is exactly what is to be expected if people make their decisions based on self interest and independent of one another.

For values of x for which $1 - m(x)$ is much less than 1, exactly the same line of reasoning that led to equation (c) can be applied, and the result is

$$d(1 - m) = +\alpha_2(1 - m)dx \qquad 1 - m << 1 \qquad \alpha_2 > 0$$

or

$$dm = -\alpha_2(1 - m)dx \tag{d}$$

But, since in equation (d) $m \approx 1$, it may be approximated by

$$dm = -\alpha_2 m(1 - m)dx \tag{e}$$

Similarly, since $1 - m \approx 1$ in equation (c), it may be approximated by

$$dm = -\alpha_1 m(1 - m)dx \tag{f}$$

Equations (e) and (f) satisfy the postulates both near $m = 0$ and near $m = 1$, and lead to the same curve only if $\alpha_1 = \alpha_2 = \alpha$. Thus the differential equation of the modal choice curve is

$$\frac{dm}{m(1 - m)} = \frac{dm}{m} + \frac{dm}{1 - m} = -\alpha\,dx \tag{g}$$

which integrates to

$$ln\left(\frac{1-m}{m}\right) = \alpha x + \gamma \tag{h}$$

in which γ is a constant of integration. Solving for m, equation (h) becomes

$$m = \frac{1}{1 + e^{\alpha x+\gamma}} \tag{j}$$

If the modal choice is between two modes, x may be considered to be the difference in the attributes of the two modes, and one may in general write

$$\alpha x + \gamma = \alpha_1 x_1 + \gamma_1 - \alpha_2 x_2 - \gamma_2 = -U_1 + U_2$$

in which

$$U_i = -\alpha_i x_i - \gamma_i \tag{k}$$

Thus, if the two modes are equal in all respects, $m = 0.5$. Using equation (k), equation (j) can be written

$$m_i = \frac{e^{U_i}}{\sum_{k=1}^{2} e^{U_k}} \qquad i = 1, 2 \tag{6.7.1}$$

in which m_i is the mode split to the ith mode, and it is clear that $m_1 + m_2 = 1$.

The form of equation (6.7.1) is readily extended to n modes. Thus

$$m_i = \frac{e^{U_i}}{\sum_{k=1}^{n} e^{U_k}} \tag{6.7.2}$$

The quantity U_i is called the utility of the ith mode, because m_i increases as U_i increases.

The analysis thus far has only considered one attribute. If there are q attributes, equation (h) can be extended to the form

$$\ln\left(\frac{1-m}{m}\right) = \sum_{j=1}^{q}(\alpha_j x_j + \gamma_j)$$

Hence the utility function for the ith mode can be written in the form

$$U_i = -\sum_{j=1}^{q}(\alpha_{ij}x_{ij} + \gamma_i) \qquad (6.7.3)$$

in which x_{ij} is the value of the jth attribute for the ith mode, α_{ij} is a weighting factor for the jth attribute when applied to the ith mode, and γ_i is a bias factor for the ith mode.

The above derived solution for $m(x)$ is not unique because, without violating the two postulates, equation (j) can be modified into the form

$$m(x) = \frac{F(x)}{1 + e^{\alpha x+\gamma}} + G(x)$$

in which, on substituting this equation into equation (g), as x approaches $\pm\infty$, $G(x)$ and $G'(x)$ vanish; as x approaches $+\infty$, $F(x)$ and $F'(x)$ remain finite; and as x approaches $-\infty$, $F(x)$ approaches 1 and $F'(x)$ approaches 0. Also, these functions are constrained by the fact that $0 < m(x) < 1$ for all x. Generalization of the logit model by introduction of the functions $F(x)$ and $G(x)$ permits the analyst to avoid the following two unrealistic properties of the logit model ~~in cases in which~~ it is applied to a situation in which there are more than two modes. The first is seen from equation (6.7.2) by dividing m_i by m_j. Thus

$$\frac{m_i}{m_j} = e^{U_i - U_j}$$

Since, U_i is a function of the attributes of the ith mode only, the ratio of the mode splits to two modes depends only on the properties of those modes and not on the properties of any other modes present. If a third mode is added, the model says that it attracts patronage from the other two modes in a strict proportion independent of the properties of the third mode. The first of these properties is called the "irrevelevant alternatives property," and the second the "new mode problem." The theory of avoidance of these problems is developed by McLynn[14], but as yet insufficient work has been done on calibration to determine the degree of improvement possible.

Thus far, nothing has been said about the nature of the attributes x_{ij}, which appears in the logit model. In most applications of the model, the only attributes taken are time and cost, and in others the attribute of auto ownership is included. Auto ownership is, however, not a continuous variable. It is better to account for it by doing separate mode splits in two groups of people: those with access to automobiles, and those without. The composite mode split is then

$$M = m_a f_a + m_{na} f_{na} \tag{6.7.4}$$

in which f_a is the fraction of people with access to autos, $f_{na} = 1 - f_a$, and m_a, m_{na} are the corresponding mode splits. Experience with alternative transit modes gives one a great deal of unease in relying on a simple time-cost model, as it would seem that many other behavioral and attitudinal variables may play a significant role in determining the mode split. Recker and Golob (references[7] and [8]) give recognition to this difficulty and derive a logit model in which the attributes are descriptive ratings chosen to represent latent perception factors. These authors' analysis indicates closer predictions to observed behavior than those obtained using only the attributes of time and costs. While the increased mathematical difficulty of the model will reduce the access of transportation planners to it, it is a welcome step into a direction of greater reality.

The author developed a ten-dimensional logit model[13], based on only the time and cost attributes. The model includes both access and egress modes, and features a rapid means for rough calibration of the logit coefficients; however, insufficient data was available at the time to confirm its overall performance. Moreover, the magnitude of the errors that may result from the irrelevant alternatives problem are not known, and the neglect of behavioral attributes felt to be important may limit its usefulness to that of a mathematical structure for conveniently handling access, line-haul, and egress modes for a model of the type developed by Golob and Recker.

6.8 Factors That Influence Patronage

The personal decision as to whether or not a trip will be taken and, if so, by what mode depends on the characteristics of both the individual and of the transportation mode. When the individual has access to an automobile and the only alternative is a transit mode that is much slower, the decision is easy—the transit mode isn't given a second thought. In most U.S. cities the mode split to transit is in the range of from 3 to 5 percent; the vast majority of people give no thought to the possibility of using a bus or even to car pooling regularly, and changes in mode split are due to attracting or discouraging the marginal transit user—the person who doesn't have such easy access to an automobile, who changes residences or jobs to a location in which the use of transit is particularly convenient, or for whom auto travel is unusually unpleasant. As indicated in section 6.3, this low mode split range is the circumstance in which the cost per trip is particularly sensitive to policies or service features that encourage or discourage patronage. The cost effectiveness analysis of chapter 5 shows that automated guideway transit systems are worth considering only if the potential exists for increasing the mode split several fold. To accomplish such an increase requires careful consideration of all factors that influence patronage: the characteristics of the transit system, the characteristics of competing modes, and the attitudinal and behavioral characteristics of the potential patron. Increases in transit patronage can be produced by either of two ways: by making the transit system more attractive, or by making the alternatives less attractive. A third method is also tried: by marketing techniques to make the transit mode seem more attractive or acceptable compared with the alternatives without actually changing the physical characteristics of either. The third method was effective for a short while during the energy crisis of 1973, but soon the mode splits returned to their precrisis values. The promise of automated transit systems is that it may be possible to make them significantly more attractive than present transit systems, therefore the following discussion relates to methods of increasing patronage on transit by improving the characteristics of the transit system.

Availability of Information

If route and schedule information can be easily found so that potential patrons can feel at ease about getting to their destination on time, and equally important, if they are satisfied that they can get back again without being stranded, they may take transit much more often. If it is too difficult or too time consuming to find reliable information, transit will not be considered even though the alternative may be considerably more expensive. In many European cities, complete guides to the transit systems are easily available at newstands and elsewhere. They are easy to use even if the language is not understood, and often the complete schedule is posted at each bus or trolley stop. It is foolish to spend a great deal of money on transit improvements if an ample advertising budget is not to be provided.

Character of the Information

If the information received does not indicate that I can get where I need to go when I need to go, that I can return when necessary, and that the trip is sufficiently convenient and comfortable, I will choose an alternative which more closely fulfills my needs, or, as Zahavi reference [11]) points out, I may forego the trip. In other words, the transit network must be sufficiently comprehensive to meet a wide range of travel demands; the schedule of service must be frequent enough for a sufficiently large fraction of the day; and the service must be comparatively convenient and comfortable. A system that can take the patron directly to work but makes it impossible to make necessary side trips at lunch or on the way home will lose substantial patronage as a result. Some will counter that the appropriate recourse is to relocate all necessary services in concentrated clusters so that the side trips can be taken by walking. This solution has considerable merit in principle, yet the range of destinations that can be reached by walking is limited in most cities, and it is difficult in a free society to contemplate restructuring the city to a significant extent just to accommodate the needs of a transit system.

Some transit systems provide good rush-hour service, but, because of the high operating cost of keeping vehicles moving empty or almost empty in nonrush periods, the schedule frequency is reduced at those times from perhaps a vehicle every five minutes to one every half hour or one an hour. A person who works on a fixed schedule may accommodate to such an arrangement, but in circumstances in which the inconvenience or cost of driving to work is not too great, the flexibility of leaving work when desired is a strong deterrent to use of transit. Patronage analysis should take into account schedule variations, but all too often do not.

Perhaps the major inconvenience factor associated with conventional transit is the transfer. Consider the fraction of possible destinations in a city that can be reached without a transfer: To simplify the problem, consider a square city of area A and side $A^{1/2}$, and assume that a transfer is not needed if the origin or destination (trip end) is within a distance w of a transit line. Then, from a given point in the city, a trip can be made without transfer if the destination lies within either of two mutually perpendicular strips each of width $2w$ and length $A^{1/2}$, that is, the fraction of the area of the city that can be reached without transfer is $4w/A^{1/2}$. If the city streets are predominately curved, the length of each strip is longer than $A^{1/2}$, therefore the area that can be reached without transfer is larger. If the transit system is a network of lines with stops every two blocks or every quarter mile, if as indicated in section 6.4 the maximum walking distance is one quarter mile, and if walking distance is measured along city streets parallel and perpen-

Table 6-2 Fraction of Area Reachable without Transfer

$A^{1/2}$ (mi)	Percent of A Reachable without Transfer
5	15%
10	7.5
20	3.75
30	2.5

dicular to the transit lines, then the boundary of the strip within walking distance of bus stops is a sawtoothed line and the average value of w is (3/4)(0.25 mi) = 3/16 mi. Thus, the fraction of the area of the city that can be reached without transfer is

$$4w/A^{1/2} = 0.75/A^{1/2}$$

where $A^{1/2}$ is in miles. Some values are given in table 6-2.

For most major metropolitan areas, the value of $A^{1/2}$ is of the order of 15 to 30 miles. It is thus interesting to note that the fraction of the area reachable without transfers is in the same range the bus mode splits in most of those areas[15]. Navin[2] reports data synthesized from travel data in many cities which indicates that the transit travel behavior can be accounted for if it is assumed that people weigh the time required for a transfer in their choice of travel mode such that one minute of transfer time is equivalent to 6.6 to 10 minutes of riding time. Weighted thus, most mode split studies will indicate that only a small fraction of transit riders will regularly transfer. Consequently, the elimination of transfers in a network of guideways through the introduction of automatic switching capability removes a major deterrent to significant increases in mode split.

Station Accessibility

In the previous paragraph, it was assumed that the stations or stops of a network transit system were accessible by walking. With some transit systems, such as conventional line-haul systems, the stations must be widely spread to attain a sufficiently high average speed, and the cost per mile is too high to permit the system to be constructed in any but the highest density corridors. Thus, as discussed in section 6.4, to attract sufficient patronage the line-haul system must work in conjunction with a system of feeder buses or a background network of bus lines, ample parking facilities must be created at each station, and the process of transfer to and from the line-haul system must be as simple as possible. In the analysis of cost effectiveness, all of these types of facilities must be fully taken into account throughout the analysis and not as an afterthought.

Negotiation of the Station

Transit patronage is influenced by three basic station-related factors: the simplicity of paying the fare and finding the right vehicle, the feeling of personal security, and the waiting time. All of these factors require careful consideration from the viewpoints of the physiology and psychology of the patrons, and are dealt with in references [16]-[20]. In reference [18], the authors deal with the design of stations from the operations point of view, and call attention to a wide range of complexity in various automated systems which were under consideration in Denver. Their findings indicated that the scheduled, group-riding systems which use off-line stations and which were popular in conception at the time of the study provide considerable difficulty for the patron because the berth at which the vehicle must stop cannot generally be predicted before it must switch off the main line and into the station. Thus the patron cannot know in advance where to stand to wait for his vehicle. On the other hand, if the stations are on-line and the vehicles stop at every station, or if they are off-line and the service is on demand with each vehicle travelling directly to the destination, the patron may wait anywhere on the platform and board any vehicle. In these cases, the system operational design provides the patron maximum convenience in finding the right vehicle and removes the anxiety that he may not be headed towards the desired destination.

Fear of assault is often given as a reason for not riding a transit system. Thus, stations which have secluded corners and in which the potential patron may have to wait long periods in the off-peak periods provide an environment in which criminals may lurk awaiting their prey. On the other hand, stations with small, well-lighted areas easily monitored by television provide fewer opportunities for assault, particularly if the service is on demand so that in the off-peak hours particularly one boards a vehicle and leaves the station immediately. In this situation, loiterers are clearly identified.

In an automated transit system, wait time is of concern therefore for two reasons: the increased anxiety as a result of fear of assault, and the uncertainty in the total trip time. According to Navin[2], one minute of waiting time is perceived to be equivalent to 4.2 to 6.3 minutes of riding time.

Because the above factors are difficult to quantify in a patronage analysis, they are often ignored, but only at the peril of those responsible for the system design. It is necessary that all aspects of the design of transit stations and their operation be studied carefully by competent human factors specialists and social psychologists and that their recommendations be carried out. The analysis of cost effectiveness indicates that station costs are a relatively minor part of the overall cost of the system[13], therefore attention to human factors in station design and operation appear to offer the potential of significant dividends.

The Vehicle

The portion of the total trip spent in the vehicle can increase or decrease patronage depending on the physical and psychological comfort associated with it. Physical discomfort is due mainly to the spectrum of acceleration and jerk of the vehicle. In section 3.6 it was shown that the maximum lateral jerk due to a perturbation in the guideway is proportional to the cube of the speed. Therefore, for the sake of ride comfort the maximum speed should be held as low as possible; but to minimize travel time, the average speed must be as high as possible. These two contradictory requirements can best be resolved, as indicated in section 2.4, by minimizing the number of intermediate stops and by maximizing the normal rate of acceleration and deceleration.

The psychological comfort of the vehicle depends on the number and arrangement of seats and the degree to which a patron must be confronted with strangers. Fried and deFazio[19] shed some light on this question by observations on riders of the New York subway, and found that people do indeed arrange themselves in such a way as to minimize the possibility of eye contact with strangers. The whole question of personal space or psychic space in various cultures is discussed by Hall[20], and from the examples given in his book it is clear that the designers of transit vehicles need the assistance of the social psychologist before proceeding too far into expensive system development programs.

The two distinctly different psychological environments proposed in transit system design are the group-riding systems and the personal systems in which one cab is occupied only by people travelling together by choice. The moderately small group-riding systems, in which the cab holds six to twelve people, provides a transit environment similar to that of a dial-a-ride bus or an airport limousine, but without a driver. In semi-off-peak periods, the probability of being required to occupy such a cab with only one stranger is high and creates considerable anxiety in the minds of many people. The possibility of an uncomfortable encounter may be a strong deterrent to use of the system, and should be thoroughly investigated before proceeding too far into the development of such a system. On the other hand, the system in which each cab is used only by a group travelling together nonstop to the destination provides maximum personal security and freedom from anxiety, but is thought by some to foster loneliness in society. It is, however, the environment of the private automobile, but without the need or opportunity to drive. The possibility should not be overlooked that such an environment will provide a quantum jump in patronage and hence a significant increase in cost effectiveness. This is a particularly promising possibility for increased cost effectiveness since the small private party vehicle permits minimum wait time and minimum riding time because of the elimination of intermediate stops.

Egress

The final leg of the trip takes the patron from the destination station of the primary transit system to the final destination of the trip. Unless the patron is aware in advance of an acceptable way to make this trip, the total trip will not be taken by transit. Since a car will be available for this trip only rarely, the patron will have to rely on walking, a feeder bus, or a taxi. As indicated in section 6.4, in the BART system, 68 percent of the people who ride the system leave the stations by walking. Thus the effective transit service area for destination stations in the BART system cannot be much larger than the area accessible by walking.

6.9 Summary

This chapter is an introduction to the subject of patronage analysis and is intended to give the reader an intuitive feel for the subject. Detailed patronage analysis required for policy decisions about the deployment of particular transit systems is beyond the scope of this book. The first topic of discussion is the numerical values assumed in other chapters for the ratios of yearly to daily patronage and daily to peak-hour patronage, and it is emphasized that verification of figures used for the ratio of yearly to daily patronage requires much more detailed trip-making surveys than usually, if ever, undertaken. This is an irreducible error in most patronage analyses. Next, the concept of mobility is introduced and data is given on its values. It is emphasized that, in the design of new systems of high service level, the mobility is likely to be higher than existed before the new system is introduced.

The required precision of patronage estimates is the next topic of discussion. It is shown that, with guideway systems, the required precision is high in the range of trip density in which the system is uneconomical, and low when the system is economical. Next the problem of trip generation is discussed, and it is shown that the choice of boundary of the transit service area is unimportant in comparison to obtaining a good analysis of mode split for the area chosen, and therefore that the area within walking distance of stations is a good choice. The mode split is considered in two ways: First, by means of a basic probability argument; and, second, by deriving the logit mode split model from two postulates. The probability argument shows that if the trip distribution is close to uniformity, the mode split is proportional to the station density squared; and, in a real situation, that the mode split should be at least proportional to the first power of the station density, thus leading to the conclusion that the cost per trip in a network system increases at least linearly with the line spacing for line spacings greater than that which places all trip ends within walking distance of a station. Next, it is shown that the logit mode split model can be derived from two postulates: (1) That the mode split is a function of independent attributes, and (2) that the mode split, bounded between zero and one, does not vanish for any finite value of an attribute. It is also shown that the resulting formulation leads to the conclusion that, near zero and one, mode split decisions are made independently, consistent with the behavioral assumption that people act in their own interest. Finally, a series of factors that influence patronage are discussed, and, in particular, it is shown that there is a strong correlation between the percentage of the area of a city that can be reached by fixed route, fixed schedule bus and the actual bus mode splits achieved.

References

1. B.G. Hutchinson, *Principles of Urban Transport Systems Planning*, Scripta Book Company, McGraw-Hill Book Company, New York, 1974.

2. F.P.D. Navin, "Time Costs in Personal Rapid Transit," *Personal Rapid Transit II*, University of Minnesota, Minneapolis, Minn., 1974, p. 589-602.

3. "Behavioral Demand Modeling and Valuation of Travel Time," Special Report 149, Transportation Research Board, National Research Council, Washington, D.C., 1974.

4. "Travel Behavior," Report Number 446, Transportation Research Board, National Research Council, Washington, D.C., 1973.

5. "Transportation Demand and Analysis Techniques," Report Number 392, Transportation Research Board, National Research Council, Washington, D.C., 1972.

6. "Choice of Travel Mode and Considerations in Travel Forecasting," Report Number 369, Transportation Research Board, National Research Council, Washington, D.C., 1971.

7. T.F. Golob and W.W. Recker, "Attitude-Behavior Models for Public Systems Planning and Design," General Motors Research Publication GMR-1906, Warren, Michigan, June 19, 1975.

8. W.W. Recker and T.R. Golob, "An Attitudinal Mode Choice Model," *Transportation Research*, Vol. 10, No. 5, Oct. 1976, pp. 299-310.

9. Margaret E. Shepard, "Annotated Bibliography of the Application of Behavior and Attitude Research to Transportation System Planning and Design," General Motors Research Publication GMR-2089, Warren, Michigan February 1976.

10. J.R. Meyer, J.F. Kain, and M. Wohl, *The Urban Transportation Problem*, Harvard University Press, Cambridge, Mass., 1966, chapter 8.

11. Y. Zahavi, "Traveltime Budgets and Mobility in Urban Areas," Federal Highway Administration Report FHWA PL 8183, Washington, D.C., May 1974.

12. J.J. Fruin, *Pedestrian Planning and Design,* Metropolitan Association of Urban Designers and Environmental Planners, New York, 1971.

13. J.E. Anderson, "The Development of a Model for Analysis of the Cost-Effectiveness of Alternative Transit Systems," *Personal Rapid Transit III,* Department of Audio Visual Library Services, University of Minnesota, Minneapolis, Minn., 1976.

14. J.M. McLynn, "Disaggregate Modal Choice Models of Fully Competitive Type," Contract DOT-UT-20021, Office of Research and Development, Urban Mass Transportation Administration, Washington, D.C., December 1973.

15. *Urban Transportation Factbook,* American Institute of Planners and Motor Vehicle Manufacturers Association of the U.S., Detroit, Mich. 1974.

16. L.L. Hoag and S.K. Adams, "Human Factors in Urban Transportation Systems," *Human Factors,* 1975, 17(2), 119-131.

17. C. Tehan and M. Wachs, "The Role of Psychological Needs in Mass Transit," *High Speed Ground Transportation Journal,* Vol. 9, No. 2, 1975.

18. R.E. Johnson, H.T. Walter, and W.A. Wild, "Analysis and Simulation of Automated Vehicle Stations," *Personal Rapid Transit III,* Department of Audio Visual Extension, University of Minnesota, Minneapolis, Minn., 1976.

19. M.L. Fried and V.J. DeFazio, "Territoriality and Boundary Conflicts in the New York Subway," *Psychiatry,* Vol. 37, February 1974.

20. Edward T. Hall, *The Hidden Dimension,* Anchor Books, Garden City, New York 1969.