CHAPTER 4

# A Theory of Population Travel Demand Behavior

## 4.1. Introduction

In this chapter we establish the link between the theory of individual behavior discussed in the previous chapter and aggregate demand data obtained from sampling an urban population. In addition to providing a general theoretical framework for this analysis, we develop concrete models with a sufficiently simple structure and small number of unknown parameters to permit the application of practical statistical estimation methods. Discussion of specific statistical procedures is given in the following chapter. However, we note that the choice of the demand model sets rather narrow limits on the range of feasible estimation techniques, and that there is some trade-off between the theoretical plausibility of the demand model and the convenience of available estimation procedures.

We first discuss the nature of transportation behavior data and the requirements of the transportation planner. Next, we present the general theory of population demand, and develop a series of specific demand models for binary and multiple alternatives. The independence and separability properties of these models are discussed in the following sections. Finally, several theoretical issues related to applications of estimated aggregate demand are discussed.

## 4.2. Trip tables and choice probabilities

The "trip" of interest to a transportation policy analyst is in general distinguished by mode, time of day, origin, destination, purpose, and socioeconomic characteristics of the trip-takers. The aggregate demand for such a trip is simply the number of journeys with these specifications taken by the relevant urban subpopulation over a given period of time. The array of demands for various trips is termed a *trip table.*

Transportation data ordinarily consist of observations on the trips taken by a sample of individuals, using household surveys, on-board surveys, cordon counts, etc. Using the survey design and the demographic characteristics of the population, the aggregate trip demand of the population can be inferred from the observed sample trip demand. Hence both in theory and in practice, the aggregate demand for a trip is obtained by aggregating over individual choices.

It is often convenient to express aggregate demand for a trip as a *frequency* by dividing by population size. Thus one can interpret a trip frequency as the probability that an individual drawn at random from the population will choose to make this trip.[1] These *choice probabilities* play a fundamental role in our analysis. If the functional dependence of the choice probabilities on transportation policy variables is known, then the planner can construct trip tables predicting the effect of alternative actions. On the other hand, observations on the travel choices of a sample from the population can be interpreted as drawings from a statistical distribution with these probabilities, making possible statistical inference on the functional dependence of the probabilities on policy variables.

## *4.3. Determination of population choice probabilities*

In principle, the theory of individual utility maximization provides a complete model of individual choice. However, within the framework of economic rationality and the postulated structure of utility maximization, there will be unobserved characteristics, such as tastes and unmeasured attributes of alternatives, which vary over the population. These variations may induce variations in observed choice among individuals facing the same measured alternatives. A specification of a distribution for the unobserved factors then generates a distribution of choices in the population.

To clarify the conceptual issues involved in this construction, we consider the textbook model of economic consumer behavior. The

[1] We assume the period of observation is sufficiently short to make the probability of an individual making the same trip more than once negligible.

individual has a utility function $u = U(x, s, \varepsilon)$, representing tastes, where $x$ is the vector of observed attributes of an alternative, $s$ is a vector of observed socioeconomic characteristics, such as sex, education, and age, and $\varepsilon$ is a vector of unobserved characteristics of alternatives and unobserved factors, such as intelligence, experience, childhood training and other variables determining tastes. The utility function is maximized subject to a "budget constraint" $x \in B$ at a value $x$ given by a system of demand functions,[2]

$$x = h(B, s; \varepsilon). \tag{4.1}$$

The econometrician typically observes the budget constraint $B_n$, socioeconomic characteristics $s_n$, and chosen alternative $x_n$ for a cross-section of consumers $n = 1, \ldots, N$. He wishes to test hypotheses about the behavioral model (4.1), ranging from specific structural features of parametric demand functions to the general revealed preference hypothesis that the observed data are generated by utility-maximizing consumers.

The unobserved vector $\varepsilon_n$ will vary in the sample due to variation in tastes and unmeasured attributes of alternatives, and will induce a variation in observed demands which will be influenced by the structure of tastes. The procedure of most empirical demand studies is to ignore the possibility of taste variations in the sample, and make the assumption that the cross-section of consumers has observed demands which are distributed randomly about the exact values for some *common* tastes.[3] In the conventional demand study, where quantities vary continuously, it is reasonable to expect errors in measurement of the chosen alternative to be important, and perhaps dominate the effect of taste variations. Hence in this case, this specification is fairly realistic. On the other hand,

[2] The set $B$ may be the conventional budget hyperplane, or it may be a more general set of available alternatives.

[3] This specification will continue to hold in conventional consumer demand models in the presence of some types of taste variation. Suppose one can postulate that consumer tastes are identical up to a vector of parameters that appear linearly in the demand function. (An example would be individuals with log-linear utility functions who face conventional budget constraints, with variation in the parameters of the utility function across individuals.) Then the demand functions can be estimated using a random coefficients econometric model. What is important is that except for refinements in estimation of the error structure, this approach will lead to the same models and estimates as are obtained under the identical tastes postulate. This "robustness" property of the conventional demand model does not, however, carry over to the case in which the consumer faces discrete alternatives.

we argue below that in the alternative case of a finite set of choices this is not a plausible model.

Under the conventional specification described above, the relation of observed aggregated demand to individual demand is straightforward. In a population of consumers who are homogeneous with respect to budgets faced, aggregate demand will equal individual demand "writ large", and all systematic variations in aggregate demand are interpreted as having been generated by a common variation at the *intensive margin* of the identical individual demands, where each consumer is choosing to buy more or less of a commodity. In the absence of unobserved variations in tastes or budgets, there is no *extensive* margin, where individuals are choosing to buy or not to buy, affecting aggregate demand.

We must re-examine the conventional demand specification in the case that the set of alternatives is finite. A utility maximum exists under conventional conditions, and generates the demand equation (4.1). This equation predicts a single chosen $x$ when tastes and unobserved attributes of alternatives are assumed uniform across the population. The conventional statistical specification above implies that all observed variation in demand $x_t$, over the finite set of alternatives, is the result of errors in measurement. The argument that measurement error is sufficiently serious to confound discrete alternatives is implausible on the face of it. Further, we must question the relevance of this behavioral model in which a substantial proportion of the observed variation in choice is attributed to aspects of behavior described only by the ad hoc error specification.

The effect of the discreteness of an individual's alternatives on the aggregate demand for a "lumpy" commodity is often negligible in a large population, and a continuous approximation is justified. For example, the rate at which an urban population takes a particular trip can be treated as a continuous variable, even though the individual's decision to make this journey is discrete. However, systematic variations in the aggregate demand for the lumpy commodity are all due to shifts at the *extensive* margin, where individuals are switching from one alternative to another, and not at the *intensive* margin as in the divisible commodity, identical individual case. Thus, it is falacious to apply the latter model to obtain specifications of aggregate demand for discrete alternatives. What is needed is a formulation of the demand model in which the effects of individual differences in tastes on the error structure in eq. (4.1) are made explicit.

We next describe in more detail how the probabilities of choices are deduced from the theory of utility maximization. Suppose an individual has $J$ alternative choices, indexed $j = 1, 2, \ldots, J$. In our study of travel demand, each of these choices will represent travel along a particular link by a particular mode at a particular time; we have compressed the link–mode–time characterization of a "trip" into a single index $j$ to simplify notation. One of these alternatives will generally be the option of not taking a trip. (It is from consideration of this option that we will be able to draw conclusions on the total demand for trips. It is important to note that estimation of the choice probability function for this option will require data on non-trip-takers.)[4]

In some applications, the alternatives will always be ranked (i.e., $j = 1$ will be the peak-hour transit work-trip option, $j = 2$ will be the peak-hour automobile work-trip option, etc.). In others, the alternatives will be unranked, and will not be paired in observations across individuals. For example, the shopping destinations of one individual may have no natural pairing with the shopping destinations of a second. Further, the number of options available may vary from individual to individual. The following analysis will apply in all these cases.

For the individual we are considering, each alternative $j = 1, \ldots, J$ has a vector of observed attributes $x^j$. Then the "budget constraint" $B$ entering (4.1) is composed of these vectors, $B = \{x^1, x^2, \ldots, x^J\}$. The observed socioeconomic characteristics of the individual are summarized in a vector $s$. As mentioned earlier, we assume that this individual has a *utility function* measuring the desirability of an option with a vector of attributes $x$, which we have written in eq. (4.1) as $u = U(x, s, \varepsilon)$, where $\varepsilon$ is an unobserved vector containing all the attributes of the alternatives and characteristics of the individual which we are unable to measure. Provided our transportation survey samples randomly from the population of individuals with common socioeconomic character-

[4] By an appropriate definition of the alternatives available to the individual, we can always take his options to be mutually exclusive. For example, multiple-link trips, such as triangular trips, can be labeled as distinct alternatives. In the following analysis, we assume this has been done, and the individual can choose one and only one alternative. This condition simplifies our analysis, and provides the most convenient format for analyzing available transportation survey data.

istics $s$ and the same alternatives, the vector $\varepsilon$ will be random, and as a consequence the values of the utility function will be stochastic. To simplify notation, we will suppress the random effect $\varepsilon$ as an explicit argument in the utility function, and define

$$u = U(x, s) \tag{4.2}$$

to be a random function whose value at any argument is a random variable depending on exactly which individual we have drawn from the subpopulation of persons with the same observed characteristics and alternatives.

The individual will choose option $i$ if this is the alternative which maximizes his utility; i.e., the individual will choose $i$ if

$$U(x^i, s) > U(x^j, s) \quad \text{for } j \neq i,\ j = 1, \dots, J. \tag{4.3}$$

Since these utility values are stochastic, the event that the condition in eq. (4.2) holds will occur with some probability, which we can denote by

$$\begin{aligned} P_i &= H(B, s, i) \\ &= \text{Prob}\left[U(x^i, s) > U(x^j, s) \quad \text{for } j \neq i,\ j = 1, \dots, J\right]. \end{aligned} \tag{4.4}$$

(We assume the probability of a "tie" is zero.) Note that this is precisely the *choice* probability introduced in the preceding section. With complete generality it is always possible to write the stochastic utility function $U(x, s)$ in the form

$$U(x, s) = V(x, s) + \eta(x, s), \tag{4.5}$$

where $V$ is non-stochastic and reflects the "representative" tastes of the population, while $\eta$ is stochastic (with mean independent of $x$) and reflects the effect of individual idiosyncrasies in tastes or unobserved attributes for alternatives in $B$. Then eq. (4.3) can be written as

$$\begin{aligned} P_i = \text{Prob}\big[&\eta(x^j, s) - \eta(x^i, s) < V(x^i, s) - V(x^j, s) \\ &\text{for } j \neq i,\ j = 1, \dots, J\big]. \end{aligned} \tag{4.6}$$

Let $\psi(t_1, \dots, t_J)$ denote the cumulative joint distribution function of $(\eta(x^1, s), \dots, \eta(x^J, s))$. Let $\psi_i$ denote the derivative of $\psi$ with respect to its $i$th argument, and let $V_j = V(x^j, s)$. Then, eq. (4.6) becomes

$$P_i = \int_{-\infty}^{+\infty} \psi_i(t + V_i - V_1, \dots, t + V_i - V_J)\mathrm{d}t. \tag{4.7}$$

Any specified joint probability distribution, such as joint normal, will yield a family of probabilities depending on the unknown parameters of the distribution and of the functions $V_j$.

Derivation of choice probabilities from an (intra-individual) stochastic utility function was first suggested for a particular case by the psychologist Thurstone (1927a). This model of the determination of choice probabilities forms the theoretical basis for both classical psychophysical laws, such as Fechner's law, and modern individual choice theories, such as the axiomatic theory of Luce (1959). A second line of development has been concerned with the effect of taste variation in a population of consumers, and has deduced specific forms of eq. (4.7). The studies of travel demand by Quandt (1968, 1970, 1972) provide an excellent statement of the foundations of this approach. Our analysis will combine these two approaches to obtain a broad class of functional forms for the choice probabilities.

To complete our task of specifying the probability functions up to a small number of parameters, we must specify an explicit functional form and probability distribution for the stochastic utility function $U(x, s)$, and then use eq. (4.7) to obtain the desired conclusion. We shall now consider several alternative specifications, beginning with an important family of functional forms, which lead in the case of binary choice to what are termed "linear", "logit", or "probit" probability functions. Following this investigation of forms for choice between two alternatives, we shall consider forms appropriate to multiple-choice settings.

## *4.4. Probability functions for binary choice*

We consider an individual with a choice between two alternatives (indexed $j = 1, 2$), with vectors of attributes $x^1$ and $x^2$, respectively. The choice probability for the first alternative is then given by eq. (4.7) as

$$P_1 = \int_{-\infty}^{+\infty} \psi_1(t, t + V(x^1, s) - V(x^2, s))\mathrm{d}t, \tag{4.8}$$

where $\psi$ is the cumulative joint distribution function of the random components $\eta(x^1, s)$ and $\eta(x^2, s)$ of the stochastic utility function. A more transparent form for this probability is obtained by introducing the cumulative distribution function $G$ of the difference of the random components, $\eta(x^2, s) - \eta(x^1, s)$. Then from eq. (4.6),

$$P_1 = G(V(x^1, s) - V(x^2, s)). \tag{4.9}$$

The form of the functions $V$ and $G$ will be influenced both by the implications of our theory of individual choice behavior and by the constraints of computational practicality.

For the purposes of this discussion (and in our empirical analysis), we assume that $V$ has the general form

$$\begin{aligned} V(x, s) &= Z^1(x, s)\beta_1 + \ldots + Z^k(x, s)\beta_k \\ &= Z(x, s)'\beta, \end{aligned} \tag{4.10}$$

where $Z^k(x, s)$ are empirical functions with no unknown parameters, $Z' = (Z^1, \ldots, Z^k)$ is a row vector of these functions, and $\beta = (\beta_1, \ldots, \beta_k)'$ is a column vector of unknown parameters. This assumption makes $V$ a linear function of the parameter vector $\beta$, a fact which greatly facilitates its estimation and statistical interpretation. The variables $Z^1, \ldots, Z^k$ may be complex transformations of the raw data (e.g., logs, reciprocals, ratios, or empirical functions) and may incorporate interactions between socioeconomic characteristics and attributes of alternatives. Further, these variables can be defined to incorporate either generic or non-generic attributes of alternatives. For example, if on-vehicle travel time is an attribute of each alternative, one may introduce this variable generically by defining $Z^1(x^j, s)$ to equal on-vehicle travel time on both alternatives $j = 1, 2$. Then $\beta_1$ is a generic weight for on-vehicle travel time. Alternately, one may define $Z^1(x^j, s)$ to equal on-vehicle travel time if $j = 1$, and zero if $j = 2$, and define $Z^2(x^j, s)$ to equal on-vehicle travel time if $j = 2$, and zero if $j = 1$. Then $\beta_1$ is a non-generic weight for travel time on the first alternative. Thus, the form in eq. (4.10) is sufficiently general to accommodate almost any phenomenon that can be usefully examined with available data.

The cumulative distribution function $G$ in eq. (4.9) is an increasing function of one variable which translates the range of $V$ into the probability scale (a number between zero and one). The parameters of this distribution are, in general, functions of $x^1$, $x^2$, and $s$. For example, systematic variations in tastes due to unmeasured socioeconomic characteristics, which are themselves correlated with observed socioeconomic characteristics, may cause the mean of $G$ to shift with $s$. Similarly, if tastes in some socioeconomic subpopulations tend to be more homogeneous than in others, this may be reflected in a dependence of the variance of $G$ on $s$. Finally, if the underlying stochastic variables $\eta(x^1, s)$ and $\eta(x^2, s)$ are non-independent, with a correlation depending on the "similarity" of $x^1$ and $x^2$ in some components, then the variance of $G$ may depend on this similarity. At this point in the discussion we shall assume that $G$ is independent of $x^1$, $x^2$, and $s$. This will be the case in particular if $\eta(x^1, s)$ and $\eta(x^2, s)$ are independent of each other and do not depend on the values of $x$ and $s$. This assumption greatly facilitates empirical analysis, and is plausible as an initial working hypothesis.[5] However, we shall see later that this assumption may have implausible implications in some applications, particularly for multiple choice.

If the distribution function $G$ is linear over the range of $V$, then eqs. (4.9) and (4.10) yield

$$P_1 = (Z(x^1, s) - Z(x^2, s))'\beta, \tag{4.11}$$

which is termed the *linear probability function*. For example, suppose $P_1$ is the probability of choosing auto when faced with a binary choice between auto and transit modes, and there are four variables: $Z^1$ = auto travel time ($T_a$) for the auto mode and zero for the transit mode; $Z^2$ = transit travel time ($T_b$) for the transit mode and zero for the auto mode; $Z^3$ = income ($I$) for the auto mode and zero for the transit mode; and $Z^4$ = one for the auto mode and zero for the transit mode. Then eq. (4.11) becomes

$$P_1 = \beta_1 T_a - \beta_2 T_b + \beta_3 I + \beta_4. \tag{4.12}$$

The $\beta_i$'s measure the effect on the probability of the auto choice of a one minute change in either auto or transit time (we expect $\beta_1, \beta_2 < 0$) or a one dollar change in income. We note that the constant term $\beta_4$ in eq. (4.12) is actually the coefficient of an "auto-mode" dummy variable, while $\beta_3$ is the coefficient of an interaction variable formed by the product of income and an "auto-mode" dummy. It should be clear that socioeconomic variables can enter eq. (4.11) only via interaction with variables which are not constant over alternatives.

It is essential for the validity of the specification (4.11) that the values

[5] In the case of binary choice, it is often impossible to identify the influence of $x^1$, $x^2$, and $s$ via $V$ and via $G$, in the sense that one can find a "standard" model with a more complex $V$ function and a $G$ distribution independent of these variables which yields the same choice probabilities for all $x^1$, $x^2$, and $s$. Then the standard model can be singled out on *grounds of convenience*.

of the function not range outside the unit interval. We shall show below that this restriction can substantially bias the statistical estimates obtained by applying straightforward regression methods to this model.

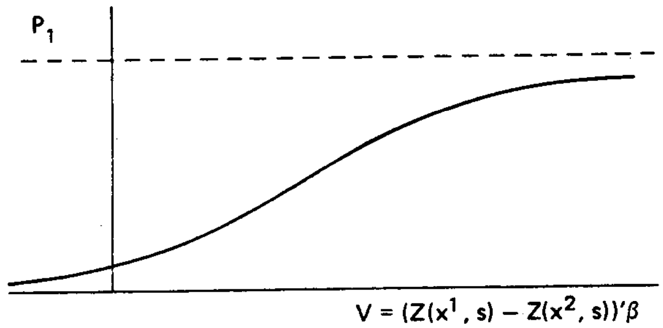


Fig. 4.1. Cumulative probability distributions giving a two-tailed ogive curve.

| | |
|---|---|
| Normal: | $G(V) = \Phi(V) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{V} e^{-t^2/2}\, dt$ |
| Logistic: | $G(V) = \frac{1}{1 + e^{-V}}$ |
| Arctan: | $G(V) = \frac{1}{\pi} \tan^{-1}(V) + \frac{1}{2}$ |

Instead of a linear function, the function $G$ may be specified to be an ogive (see fig. 4.1) which maps the real line into the zero-one interval; any cumulative distribution function without jumps gives this basic shape. Three commonly used ogives are the cumulative normal, logistic, and Cauchy distributions. The normal distribution gives the probability function

$$P_1 = \Phi(\beta' Z(x^1, s) - \beta' Z(x^2, s)), \tag{4.13}$$

where $\Phi$ is the standard cumulative normal distribution. This equation is termed the *binary probit probability model.* The Cauchy distribution gives the probability function

$$P_1 = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(\beta' Z(x^2, s) - \beta' Z(x^1, s)), \tag{4.14}$$

and is known as the *arctan probability model.* The logistic distribution gives the probability function

$$P_1 = \frac{1}{1 + \exp\left[\beta' Z(x^2, s) - \beta' Z(x^1, s)\right]}, \tag{4.15}$$

and is termed the *binary logit probability model.*

We use the logit model to continue the example of a binary mode choice probability. Again let $P_1$ be the probability of choosing the auto mode, and let $T_a$, $T_b$, and $I$ denote auto travel time, transit travel time, and income, respectively. Using the inverse transformation of the cumulative logistic distribution, the model can be written

$$\log\left[\frac{P_1}{1 - P_1}\right] = \beta_1 T_a - \beta_2 T_b + \beta_3 I + \beta_4. \tag{4.16}$$

Here, $\beta_1$ and $\beta_2$ measure the change in the log of the *odds* of choosing

Table 4.1

| Argument | Logit[a] | Probit | Arctan[b] | Logit-Probit | Logit-Arctan |
|---|---|---|---|---|---|
| 0.0 | 0.5 | 0.5 | 0.5 | 0.0 | 0.0 |
| 0.2 | 0.5791 | 0.5792 | 0.5781 | −0.0001 | 0.0009 |
| 0.4 | 0.6543 | 0.6554 | 0.6479 | −0.0010 | 0.0064 |
| 0.6 | 0.7226 | 0.7257 | 0.7052 | −0.0031 | 0.0173 |
| 0.8 | 0.7818 | 0.7881 | 0.7504 | −0.0062 | 0.0314 |
| 1.0 | 0.8314 | 0.8413 | 0.7856 | −0.0099 | 0.0457 |
| 1.2 | 0.8715 | 0.8849 | 0.8132 | −0.0133 | 0.0583 |
| 1.4 | 0.9032 | 0.9192 | 0.8351 | −0.0159 | 0.0681 |
| 1.6 | 0.9277 | 0.9451 | 0.8527 | −0.0174 | 0.0750 |
| 1.8 | 0.9464 | 0.9640 | 0.8671 | −0.0176 | 0.0792 |
| 2.0 | 0.9605 | 0.9772 | 0.8791 | −0.0167 | 0.0813 |
| 2.2 | 0.9709 | 0.9860 | 0.8892 | −0.0151 | 0.0817 |
| 2.4 | 0.9787 | 0.9918 | 0.8978 | −0.0130 | 0.0809 |
| 2.6 | 0.9844 | 0.9953 | 0.9052 | −0.0108 | 0.0792 |
| 2.8 | 0.9886 | 0.9974 | 0.9116 | −0.0087 | 0.0770 |
| 3.0 | 0.9917 | 0.9986 | 0.9172 | −0.0069 | 0.0744 |
| 4.0 | 0.9983 | 1.0 | 0.9373 | −0.0016 | 0.0609 |
| 5.0 | 0.9996 | 1.0 | 0.9496 | −0.0003 | 0.0500 |
| 6.0 | 0.9999 | 1.0 | 0.9579 | 0.0 | 0.0420 |
| 7.0 | 1.0 | 1.0 | 0.9638 | 0.0 | 0.0361 |
| 8.0 | 1.0 | 1.0 | 0.9683 | 0.0 | 0.0316 |
| 9.0 | 1.0 | 1.0 | 0.9718 | 0.0 | 0.0281 |
| 10.0 | 1.0 | 1.0 | 0.9746 | 0.0 | 0.0253 |

[a] The logit formula (normalized to have the same slope at zero as the standard normal) is $P = 1/\{1 + \exp[-2x\sqrt{(2/\pi)}]\}$.

[b] The arctan formula (normalized to have the same slope at zero as the standard normal) is $P = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}[x\sqrt{(\pi/2)}]$.

auto as a result of a one-minute change in auto or transit travel time, and $\beta_3$ measures the effect on the odds of a unit change in income.

The binary probit, arctan, and logit models are virtually indistinguishable except at arguments yielding probabilities extremely close to zero or one, where the probit model approaches the extreme values most rapidly and the arctan model least rapidly. Table 4.1 gives selected values of the choice probabilities for these three ogives. (The curves are symmetric about zero and are normalized to have the same slope at zero.) The maximum deviation in probability between the logit and probit curves is 0.018, with the result that these curves are virtually equivalent for empirical purposes. The arctan curve approaches the asymptotes considerably less rapidly than the probit and logit curves, with a maximum deviation in probability from the logit curve of 0.082. In the following chapter we discuss the consequences for parameter estimates of misspecifying the functional form of the ogive curve, or, for example, calibrating the logit model using data generated by the arctan model. Within the range of most data, the models presented above provide essentially equivalent probability functions, and except for computational reasons, there is little to choose among them. The logit model has computational advantages since it is a closed (explicit) functional form with convenient curvature properties for numerical optimization. The probit model, on the other hand, has its argument as the limit of an integral which cannot be expressed in closed form.

There are a number of other functional forms for ogives which we shall not use in our empirical analysis, but which could be relevant in certain applications. Fig. 4.2 and 4.3 illustrate two ogives, based on the negative and positive exponential distributions, which have the property that the limiting probability on one side or the other is attained. Fig. 4.4 illustrates the ogive generated by a uniform distribution over an interval; this might appropriately be called the *truncated linear probability model* because it coincides with the specification of eq. (4.11) when the restriction on the range of the linear probability function to the zero-one interval is valid. We shall return to a discussion of these forms in the context of statistical estimation procedures.

We turn next to the question of the relationship of the (truncated) linear, probit, logit and arctan models to the underlying theory of a population of utility-maximizing consumers. We shall demonstrate that each of these concrete probability models is consistent with the general

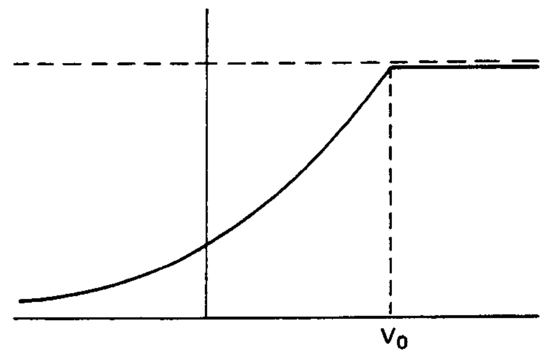


Fig. 4.2. Cumulative exponential probability distribution giving a one-tailed ogive curve. $G(V) = e^{-(V_0 - V)}$ for $V < V_0$; $G(V) = 1$ for $V \geqslant V_0$.

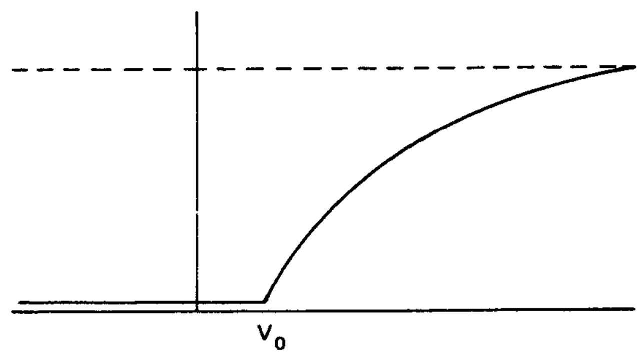


Fig. 4.3. Cumulative exponential probability distribution giving a one-tailed ogive curve. $G(V) = 1 - e^{-(V - V_0)}$ for $V > V_0$; $G(V) = 0$ for $V \leqslant V_0$.

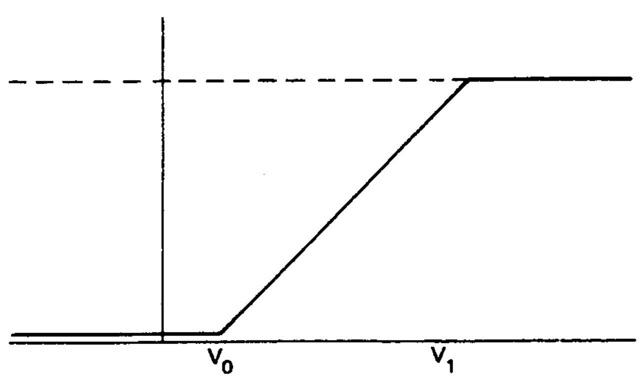


Fig. 4.4. Uniform distribution giving a "truncated linear model". $G(V) = 1$ for $V \geqslant V_1$; $G(V) = (V - V_0)/(V_1 - V_0)$ for $V_0 < V < V_1$; $G(V) = 0$ for $V < V_0$.

formulas, (4.8) or (4.9), for choice probabilities, provided that we assume specific distributions for the stochastic components of utility, $\eta_1 = \eta(x^1,s)$ and $\eta_2 = \eta(x^2,s)$. In particular, in the binary case we can assume that $\eta_1$ is always zero and $\eta_2$ has any specified cumulative distribution function $G$ in eq. (4.8). The cases of uniform, normal, logistic and Cauchy distributed $\eta_2$ yield, respectively, the (truncated) linear, probit, logit and arctan models.

The assumption above treats $\eta_1$ and $\eta_2$ asymmetrically, making its generalization to multiple-choice cases difficult. It is of interest to know whether these concrete models can also be obtained under the alternative condition that $\eta_1$ and $\eta_2$ are independently identically distributed. Letting $\psi(\eta)$ denote the cumulative distribution function of $\eta_1$ and $\eta_2$ under this assumption, we obtain from eq. (4.8) the convolution formula

$$G(v) = \int_{-\infty}^{+\infty} \psi'(t)\psi(v+t)\,\mathrm{d}t. \tag{4.17}$$

A textbook exercise in probability theory shows that the uniform distribution cannot be obtained in this manner by convoluting independent, identically distributed distributions.[6] The normal, logistic, and Cauchy distributions can be obtained in this fashion, as we now show.

If $\eta_1$ and $\eta_2$ are jointly normally distributed, then $\eta_2 - \eta_1$ is normal, and a probit model results. This is true particularly when $\eta_1$ and $\eta_2$ are independent with identical means and variances. However, it also holds more generally when $\eta_1$ and $\eta_2$ are dependent. Suppose, for example, the covariances of $\eta_1$ and $\eta_2$ are determined by the "psychometric" proximity of $x_1$ and $x_2$, i.e., their perceived similarity along the dimensions they are being perceived. Then this effect is absorbed into the binary probit functional form. (See the argument below leading to eq. (4.36).) As a consequence, the probit model is somewhat "robust" with respect to changes in the structure of normally distributed stochastic components of utility.

If $\eta_1$ and $\eta_2$ have independent Cauchy distributions, not necessarily

[6] The characteristic function of a uniform distribution on the interval $[-1,1]$ is $(\sin t)/t$, which is negative at $t = 3\pi/2$. If the uniformly distributed random variable could be written as the difference $\eta_1 - \eta_2$ of two identically independently distributed random variables, each with characteristic function $\phi(t)$, then one would have

$$(\sin t)/t = \phi(t)\phi(-t) = (E_\eta \cos t\eta)^2 + (E_\eta \sin t\eta)^2 \geqq 0,$$

for a contradiction.

identical, then $\eta_2 - \eta_1$ has a Cauchy distribution, and one obtains the arctan model.

The final case we will consider leads to the logit model. This model and the construction described in the following paragraphs will also play a substantial role in our analysis of multiple choice. We shall see that this model is computationally tractable and in many applications corresponds to a plausible stochastic specification. It is virtually the only multinomial choice model known to date satisfying both these criteria. As in the constructions above, the key to our analysis is the specification of a statistical distribution with the property that the difference of independent random variables having this distribution is a logistically-distributed random variable. We first introduce a distribution with this property and discuss its characteristics.

A random variable $\eta_i$ has a Weibull (extreme value, Gnedenko) distribution if

$$\text{Prob}\,[\eta_i \leqq \eta] = e^{-e^{-(\eta+\alpha)}}, \tag{4.18}$$

where $\alpha$ is a parameter. The associated frequency function is $\psi(\eta) = e^{-(\eta+\alpha)}\exp[-e^{-(\eta+\alpha)}]$. Fig. 4.5 plots the Weibull frequency function (with $\alpha = 0$) along with a normal frequency function, with unit variance and mean 0.5. One sees that the Weibull frequency has the same general bell shape as the normal frequency, but is skewed, with a thinner left tail than the normal distribution and a thicker right tail. (The right tail behaves like the tail of an exponential distribution. Thus, it is easy to show that the Weibull distribution has all positive moments.) The parameter $\alpha$ determines the mode of the Weibull distribution; hence changing $\alpha$ shifts the location of the mode and mean, but not the shape of the distribution. We shall next establish a lemma summarizing the properties of this distribution which are important for our purposes. The first significant property is that the Weibull distribution is *stable* under maximization, in the sense that the maximum of two independent Weibull random variables is again a Weibull random variable. Compare this property with the property that the sum of two normal variables is again normal, so that the normal family is stable under addition. In our problem, where maximization of utility is the critical operation, this stability property of the Weibull distribution makes it a natural distribution with which to work, just as the normal distribution is natural for problems involving addition of random variables. The second significant

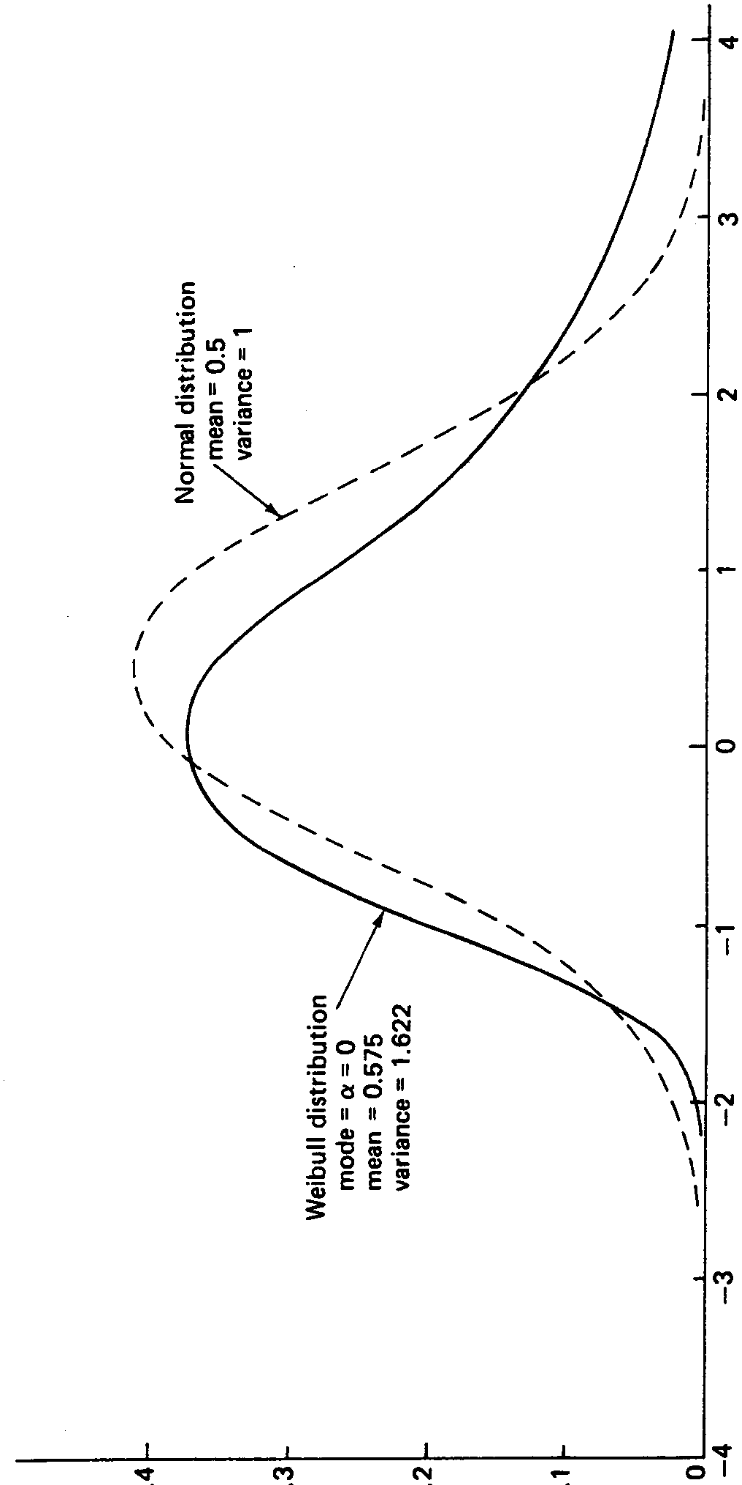


Fig. 4.5. Frequency functions of normal and Weibull distributions.

property of the Weibull distribution is that the difference of Weibull distributed variables has a binary logistic distribution. One of the authors has shown elsewhere [McFadden (1973)] that this property characterizes the Weibull distribution in the sense that it is essentially the only distribution with this property in the multinomial case.

*Lemma.* If random variables $\eta_i$ have independent Weibull distributions with parameters $\alpha_i$ for $i = 1, \ldots, n$, then:

(a) $\eta_i + v$ has a Weibull distribution with parameter $\alpha_i - v$ for any real $v$;

(b) $\max_{i=1,\ldots,n} \eta_i$ has a Weibull distribution with parameter $-\log \sum_{i=1}^{n} e^{-\alpha_i}$;

(c) $$\operatorname{Prob}[v_1 + \eta_1 \geqq v_2 + \eta_2] = \frac{e^{v_1 - \alpha_1}}{e^{v_1-\alpha_1} + e^{v_2 - \alpha_2}}; \tag{4.19}$$

(d) $$\operatorname{Prob}[v_1 + \eta_1 \geqq v_i + \eta_i \quad \text{for } i = 2, \ldots, n] = \frac{e^{v_1 - \alpha_1}}{\sum_{i=1}^{n} e^{v_i - \alpha_i}} \tag{4.20}$$

Verification of this result is straightforward; we outline the steps for the sake of completeness. To show (a), note that

$$\operatorname{Prob}[\eta_i + v \leqq \eta] = \operatorname{Prob}[\eta_i \leqq \eta - v],$$

and substitute the argument $\eta - v$ in the cumulative distribution function of $\eta_i$. To show (b), note that

$$\begin{aligned}\operatorname{Prob}[\max_{i=1,\ldots,n} \eta_i \leqq \eta] &= \operatorname{Prob}[\eta_1 \leqq \eta] \cdot \ldots \cdot \operatorname{Prob}[\eta_n \leqq \eta] \\ &= \exp\left[-\sum_{i=1}^{n} e^{-(\eta+\alpha_i)}\right] \\ &= \exp\left[-e^{-\eta} \cdot \sum_{i=1}^{n} e^{-\alpha_i}\right].\end{aligned} \tag{4.21}$$

Setting

$$e^{-\alpha} = \sum_{i=1}^{n} e^{-\alpha_i}$$

establishes that the maximum value is distributed Weibull with parameter $\alpha$.

To establish (c), we use the convolution formula

$$\text{Prob}\,[v_1 + \eta_1 \geqq v_2 + \eta_2] = \int_{-\infty}^{+\infty} \psi_1'(\eta)\psi_2(v_1 - v_2 + \eta)\mathrm{d}\eta, \tag{4.22}$$

where $\psi_i$ is the cumulative distribution function of $\eta_i$. In this case

$$\psi_1(\eta) = \exp\left(-\mathrm{e}^{-(\eta+\alpha_1)}\right),$$
$$\psi_1'(\eta) = \mathrm{e}^{-(n+\alpha_1)}\exp(-\mathrm{e}^{-(n+\alpha_1)}).$$

Then eq. (4.22) becomes

$$\begin{aligned}
&\text{Prob}\,[v_1 + \eta_1 \geqq v_2 + \eta_2] \\
&\quad = \int_{-\infty}^{+\infty} \mathrm{e}^{-(\eta+\alpha_1)}\exp(-\mathrm{e}^{-(\eta+\alpha_1)})\exp(-\mathrm{e}^{-(\eta+v_1-v_2+\alpha_2)})\mathrm{d}\eta \\
&\quad = \int_{-\infty}^{+\infty} \mathrm{e}^{-(\eta+\alpha_1)}\exp(-\mathrm{e}^{-\eta}(\mathrm{e}^{-\alpha_1} + \mathrm{e}^{-v_1+v_2-\alpha_2}))\mathrm{d}\eta \\
&\quad = \frac{\mathrm{e}^{-\alpha_1}}{[\mathrm{e}^{-\alpha_1} + \mathrm{e}^{-v_1+v_2-\alpha_2}]} \\
&\qquad \cdot \int_0^1 \mathrm{d}\{\exp(-\mathrm{e}^{-\eta}(\mathrm{e}^{-\alpha_1} + \mathrm{e}^{-v_1+v_2-\alpha_2}))\} \\
&\quad = \frac{\mathrm{e}^{-v_1-\alpha_1}}{\mathrm{e}^{v_1-\alpha_1} + \mathrm{e}^{v_2-\alpha_2}}.
\end{aligned} \tag{4.23}$$

The argument for the general case (d) can be made simply by extending the number of terms in the demonstration of condition (c). Alternately, we can combine results (b) and (c) directly to reach the desired conclusion:

$$\text{Prob}\,[\eta_1 + v_1 \geqq \eta_i + v_i \quad \text{for } i = 1, \ldots, n] =$$

$$\text{Prob}\,[\eta_1 + v_1 \geqq \underset{i=2,\ldots,n}{\text{Max}}(\eta_i + v_i)] = \frac{\mathrm{e}^{v_1-\alpha_1}}{\mathrm{e}^{v_1-\alpha_1} + \mathrm{e}^{-\alpha}}, \tag{4.24}$$

by (c), where

$$\alpha = -\log \sum_{i=2}^{n} \mathrm{e}^{v_i - \alpha_i}$$

is the parameter of the Weibull distributed variable $\text{Max}_{i=2,\ldots,n}(\eta_i + v_i)$

by condition (b). This formula simplifies directly to the formula of (d). This completes the demonstration of the lemma.

As indicated earlier, the importance of this lemma lies in the fact, established in condition (c), that the difference of two independent Weibull distributed random variables has a binary logit distribution

$$\operatorname{Prob}[\eta_2 - \eta_1 \leqq v_1 - v_2] \equiv G(v_1 - v_2) = \frac{e^{v_1 - \alpha_1}}{e^{v_1 - \alpha_1} + e^{v_2 - \alpha_2}}. \tag{4.25}$$

Hence, independent Weibull distributed stochastic components of utility lead to the logit model. When $v_1 = V(x^i, s) = Z(x^i, s)'\beta$, as defined in eq. (4.10), and the "location" parameters $\alpha_j$ in the underlying Weibull distributions are the same, this equation coincides with logit eq. (4.15). (We note that in general one can absorb the parameters $\alpha_j$ in the definition of $V(x^j, s)$, with $\alpha_j$ interpreted as an effect specific to this alternative. Hence, an assumption that all $\alpha_j$ are zero involves no loss of generality.)

From the analysis above we conclude that in the case of binary choice, a wide variety of functional forms for the probability function are consistent with the underlying model of individual utility maximization, where random elements are introduced because the tastes of specific individuals cannot be determined completely from available data. In particular, the linear, probit and logit probability models are consistent with this theory. Noting that the probit and logit models are virtually equivalent, we can choose between them on grounds of computational convenience. The consistency of the linear probability model requires the addition of truncation conditions to its linear form. Further comparisons of these models will be developed in chapter 5.

## *4.5. Probability functions for multiple choice*

We now wish to generalize the derivations of probability functions for binary choice situations to the multiple choice case with $J$ alternatives. Again let $B = \{x^1, \ldots, x^J\}$ denote the set of available alternatives $(1, \ldots, J)$, identified by their attribute vectors $x^j$, and let $s$ denote the vector of socioeconomic characteristics of the individual. From eqs. (4.4) and (4.7), the choice probabilities satisfy

$$P_i = H(B, s, i) = \text{Prob}\left[U(x^i, s) > U(x^j, s) \quad \text{for } j \neq i, j = 1, \ldots, J\right]$$
$$= \int_{-\infty}^{+\infty} \psi_i(t + V_i - V_1, \ldots, t + V_i - V_J)\mathrm{d}t, \tag{4.26}$$

where $V_j = V(x^j, s)$ is a non-stochastic "representative" utility, $\psi$ is the joint cumulative distribution function of the stochastic components of utility $(\eta(x^1, s), \ldots, \eta(x^J, s))$, and $\psi^1$ is the derivative of $\psi$ with respect to its $i$th argument. Eq. (4.26) provides a formula that can in principal be applied to concrete joint probability distributions to obtain specific functional forms for the probability functions. In practice, it is extremely difficult to specify joint distributions for which the expression in eq. (4.26) can be evaluated without numerical multivariate integration. Of the distributions considered for binary choice, only the Weibull distribution yields a convenient functional form; the multiple-choice generalizations of the probit and arctan models are computationally intractable. The possibilities and difficulties of each of these alternatives are outlined below.

We will first consider the case in which $\psi$ is multivariate normal. We derive the formula for the probability $P_1$; the remaining probabilities have analogous expressions. In this case the convolution property of the normal distribution can be used to simplify eq. (4.26) to

$$P_1 = \int_{r_2=-\infty}^{V_1-V_2} \cdots \int_{r_J=-\infty}^{V_1-V_J} n(r; 0; \Omega)\mathrm{d}r_2 \ldots \mathrm{d}r_J, \tag{4.27}$$

where $n(r; 0; \Omega)$ is the multivariate normal frequency function with mean vector 0 and covariance matrix $\Omega$ evaluated at argument $r$, and where the elements $\omega_{ij}$ of $\Omega$ satisfy

$$\omega_{ij} = E\eta_i\eta_j + E\eta_1^2 - E\eta_i\eta_1 - E\eta_j\eta_1, \tag{4.28}$$

with $\eta_i = U(x_i, s) - V(x_i, s)$ and $E\eta_i = 0$. A further simplification of this formula can be obtained by defining

$$\lambda_{2j} = \omega_{2j}, \qquad t_2 = r_2 \tag{4.29}$$

and recursively, for $i = 3, \ldots, J$,

$$\lambda_{ik} = \omega_{ik} - \sum_{j=2}^{i-1} \lambda_{ji}\lambda_{jk}/\lambda_{jj}, \quad k = 2, \dots, J, \tag{4.30}$$

$$t_i = r_i - \sum_{j=2}^{i-1} \lambda_{ji} r_j/\lambda_{jj}. \tag{4.31}$$

Then the $t_i$ are independent normal with zero mean and variance $\lambda_{ii}$, and eq. (4.27) becomes, with a transformation to the variables $t_i$,

$$P_1 = \int_{t_2=-\infty}^{V_1-V_2} \phi\left[\frac{t_2}{\sqrt{\lambda_{22}}}\right] \int_{t_3=-\infty}^{V_1-V_3-\frac{\lambda_{23}}{\lambda_{22}}t_2} \phi\left[\frac{t_3}{\sqrt{\lambda_{33}}}\right]$$

$$\int_{t_4=-\infty}^{V_1-V_4-\frac{\lambda_{24}}{\lambda_{22}}t_2-\frac{\lambda_{34}}{\lambda_{33}}t_3} \phi\left[\frac{t_4}{\sqrt{\lambda_{44}}}\right] \cdots \int_{t_J=-\infty}^{V_1-V_J-\sum_{j=2}^{J-1}\frac{\lambda_{jJ}}{\lambda_{jj}}t_j} \phi\left[\frac{t_J}{\sqrt{\lambda_{JJ}}}\right] \mathrm{d}t_J \dots \mathrm{d}t_2, \tag{4.32}$$

where $\phi$ is the univariate standard normal distribution. While this expression can be evaluated by straightforward repeated numerical integration, the expression is too cumbersome for efficient use in iterative statistical procedures.

It is instructive to consider how the multivariate normal model might arise. Suppose the random utility function $U(x, s)$ has the form

$$U(x, s) = \sum_{k=1}^{K} \alpha_k Z^k(x, s) + \varepsilon(x, s), \tag{4.33}$$

where the $\alpha_k$ are taste parameters which vary randomly in the population, but are independent of $(x, s)$, and $\varepsilon(x, s)$ is a random component in utility which is a function of $(x, s)$. The parameters $\alpha_k$ can be interpreted as weights associated with particular components of the attribute vector of alternatives. Write $\varepsilon_j = \varepsilon(x^j, s)$, and suppose $\alpha_1, \dots, \alpha_K$ and $\varepsilon_1, \dots, \varepsilon_J$ are multivariate normal with $E\alpha_k = \beta_k$, $E\varepsilon_j = 0$, $\mathrm{cov}(\alpha_k, \alpha_l) = 0$ for $l \neq k$, $\mathrm{cov}(\alpha_k, \varepsilon_j) = 0$, and $\mathrm{cov}(\varepsilon_j, \varepsilon_l)$ equal to a positive non-increasing function of the "distance" between $x^j$ and $x^l$. Then in the terminology used previously,

$$\eta_j = \eta(x^j, s) = \sum_{k=1}^{K} (\alpha_k - \beta_k) Z^k(x^j, s) + \varepsilon_j, \tag{4.34}$$

implying $E\eta_j = 0$, and

$$\omega_{jl} = \operatorname{cov}(\eta_j, \eta_l) = \operatorname{cov}(\varepsilon_j, \varepsilon_l) + \sum_{k=1}^{K} \operatorname{var}(\alpha_k) Z^k(x^j, s)^2. \tag{4.35}$$

In the special case that $\operatorname{var}(\alpha_k) = 0$ and $\operatorname{cov}(\varepsilon_j, \varepsilon_l)$ is a constant independent of $x^j$, $x^l$, one obtains the direct multivariate generalization of the binary probit model that was discussed above. Alternately, in the special case that $\operatorname{var}(\varepsilon_j) = 0$, one obtains a multiple-choice generalization of a binary choice model used by Quandt (1968, 1970, 1972) in a series of transportation demand studies. In general, for the binary case, one obtains an analogue of the previous probit formula of eq. (4.13) with the argument modified by the inclusion of the covariance effects as functions of the $x^j$ and $s$; e.g.,

$$P_1 = \Phi\left[\frac{\beta' Z(x^1, s) - \beta' Z(x^2, s)}{\sqrt{(\omega_{11} + \omega_{22} - 2\omega_{12})}}\right]. \tag{4.36}$$

We next consider the case in which the stochastic components of utility have independent Cauchy distributions. Then,

$$\psi(t_1, \ldots, t_J) = \prod_{j=1}^{J} \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1}(t_j/a_j)\right), \tag{4.37}$$

where the $a_j$ are positive constants, and eq. (4.26) becomes, for $P_1$,

$$P_1 = \int_{t=-\infty}^{+\infty} \frac{a_1}{a_1^2 + t^2} \prod_{j=2}^{J} \left[\frac{1}{2} + \frac{1}{\pi} \tan^{-1}\left(\frac{V_1 - V_j + t}{a_j}\right)\right] \mathrm{d}t. \tag{4.38}$$

This probability can again be obtained by straightforward, but costly, numerical integration.

## 4.6. The multinomial logit model

The case of independent Weibull distributed stochastic components of utility provides a fortunate contrast to the mathematical complexity of the cases above. Eq. (4.20), proved earlier in the lemma, establishes

$$P_i = \frac{e^{V_i - \alpha_i}}{\sum_{j=1}^{J} e^{V_j - \alpha_j}}, \tag{4.39}$$

where $\alpha_j$ is a parameter of the Weibull distribution and $V_i = V(x^i, s)$ is the non-stochastic component of utility. Absorbing the parameter $\alpha_j$ into the definition of $V(x^i, s)$, we can rewrite this as

$$P_i = \frac{e^{V(x^i,s)}}{\sum_{j=1}^{J} e^{V(x^j,s)}}. \tag{4.39a}$$

This is the formulation of multiple-choice selection probabilities which will provide a computationally practical basis for our empirical analysis.

## 4.7. *Independence of irrelevant alternatives and demand for a new transportation mode*

An important implication of the above result is that the *odds* $P_i/P_j$ of choosing alternative $i$ over alternative $j$ are independent of the presence or absence of third alternatives, satisfying the equation

$$\operatorname{Log}(P_i/P_j) = V(x^i, s) - V(x^j, s). \tag{4.40}$$

The conclusion stated in eq. (4.40) is consistent with the "independence of irrelevant alternatives" axiom [Luce (1959)]. This axiom requires that the relative odds of two options being chosen be independent of the presence or absence of non-chosen third alternatives. The equivalence of this axiom and the assumption that the stochastic utility function has a Weibull distribution has been established by Marschak (1960). This result is strengthened by Holman and Marley [in Luce and Suppes (1965)] and McFadden (1968, 1973a).

The function $e^{V(x,s)}$ in eq. (4.39a) is known as a "strict utility function" in the literature of psychology [e.g., Luce and Suppes (1965)]. From this equation we draw the conclusion that the probability of an alternative being chosen is proportional to its strict utility, with the proportion determined by the condition that exactly one alternative must be chosen. This implies that the probabilities $P_j$ must sum to one over the available alternatives.

The independence of irrelevant alternatives condition in eq. (4.40) is both the principal strength and the principal weakness of the "strict utility" probability model in eq. (4.39). It is a strength because this restriction allows the introduction of new alternatives, such as new or additional modes or new destinations, without re-estimation of the model, once a numerical functional form for $V(x, s)$ has been established. This is done by simply adding the new term to the denominator of eq. (4.39) for each alternative and expanding the list of probability functions to include the new alternative. This procedure is possible because the addition of an alternative cannot change the *relative* odds with which the previous alternatives are selected. It is a weakness because it requires that the alternatives be perceived as completely distinct and independent. This will be discussed at greater length below.

We illustrate the advantages of the strict utility model by showing how one would utilize transportation survey data on currently existing alternatives to forecast the effects on modal choice of introducing a new transportation mode. Suppose data are collected on individuals who have two options, auto trip ($j = 1$) and bus trip ($j = 2$), each described by a vector of attributes $x^j$. Choosing a specific functional form for $V(x^j, s)$ which depends on a small number of unknown parameters, we apply one of the statistical procedures described in the following chapter to obtain from these data an estimated numerical $V$ function. Suppose that we have specified the form of $V$ such that it is not mode-specific (i.e., in eq. (4.39) the parameters $\alpha_i$ are zero, and $V$ contains no pure "mode" effect, but rather evaluates a mode solely in terms of its generic attributes).

Thus $V$ might be a function of mode attribute variables such as waiting time, line-haul time, walking time, fares and tolls, and an index of vehicle comfort, as well as variables describing the individual socioeconomic characteristics of the trip-makers. $V$ would not depend on mode-specific variables, such as a dummy variable which is one for the auto mode and zero otherwise, a specific mode attribute such as "automobile tolls" (a variable which records auto toll charges for this mode and is zero otherwise), or a mode-specific individual characteristic such as a "pure income-linked auto preference" (a variable which equals family income for this mode and is zero otherwise). The difficulty with variables such as these is that they confound the effects of generic attributes with effects which are specific to the designated modes. (To the extent that there are mode-specific effects which cannot be captured by a full description of the generic attributes of the mode, one cannot hope to forecast the impact of a new mode without additional empirical evidence on *its* mode-specific effect.)

Given the fitted function $V(x^j, s)$ for $j = 1, 2$, we have the probability function

$$P_i = \frac{e^{V(x^i,s)}}{e^{V(x^1,s)} + e^{V(x^2,s)}} \quad \text{for } i = 1, 2. \tag{4.41}$$

These functions provide estimates of the frequencies with which the population chooses the current modes. We can now introduce a new mode (for example, rail rapid transit ($j = 3$)) for which we can estimate a vector of generic attributes $x^3$ from engineering considerations. Substituting this vector into the estimated numerical function $V(x^3, s)$ provides a forecast of the "strict utility" attached to this new mode. The probability functions for the individuals with this new alternative now change from the values given by eq. (4.41) to

$$P_i = \frac{e^{V(x^i,s)}}{e^{V(x^1,s)} + e^{V(x^2,s)} + e^{V(x^3,s)}}, \tag{4.42}$$

for $i = 1, 2, 3$. This formula implies that the introduction of the new mode will cause the probability that each old mode will be chosen to decrease because the relative odds between the existing modes cannot change.

For example, we can derive a formula for the change in the probability of choosing the first mode,

$$\begin{aligned}
\begin{bmatrix}\text{Percentage} \\ \text{change in } P_1\end{bmatrix} &= \frac{P_1(3 \text{ modes}) - P_1(2 \text{ modes})}{P_1(2 \text{ modes})} \\
&= \frac{e^{V_1}/(e^{V_1} + e^{V_2} + e^{V_3})}{e^{V_1}/(e^{V_1} + e^{V_2})} - 1 \\
&= \frac{e^{V_1} + e^{V_2}}{e^{V_1} + e^{V_2} + e^{V_3}} - 1 \\
&= \frac{e^{V_3}}{e^{V_1} + e^{V_2} + e^{V_3}} \\
&= -P_3, \quad \text{where } V_i = V(x^i, s).
\end{aligned} \tag{4.43}$$

This formula demonstrates the ease with which the effect of a new alternative can be evaluated in the "strict utility" model. However, it also demonstrates a potential drawback in that the cross-elasticity of demand for each old mode, with respect to an attribute of the new mode, is uniform across all the old modes. This precludes the possibility of postulating a pattern of differential substitutability and complementarity between modes within the "strict utility" framework. This drawback is not unique to the transportation demand problem. It is shared by many of the empirically convenient functional forms used in consumption and production theory, e.g., Cobb–Douglas or constant elasticity of substitution (CES) functional forms.

The independence of irrelevant alternatives property also gives a "separability of decisions" property consistent with the assumption in chapter 3 of the additive separability of utility used to factor the simultaneous travel decisions of an individual into a series of separate choice models. To illustrate this link, we use the example of the compound decision of what time to take a trip and what mode to use. By the laws of conditional probability, we can always write

$$\begin{bmatrix}\text{Probability of}\\ \text{choosing time } t,\\ \text{mode } m\end{bmatrix} = \begin{bmatrix}\text{Probability of choosing}\\ \text{mode } m\text{, conditioned}\\ \text{on the event that time } t\\ \text{is chosen}\end{bmatrix} \cdot \begin{bmatrix}\text{Probability of}\\ \text{choosing time } t,\\ \text{any mode}\end{bmatrix}. \tag{4.44}$$

But by the independence of irrelevant alternatives assumption, the conditional probability above will not depend on whether or not the individual has the option of times other than $t$, and hence

$$\begin{bmatrix}\text{Probability of choosing}\\ \text{mode } m\text{, conditioned}\\ \text{on the event that time } t\\ \text{is chosen}\end{bmatrix} = \begin{bmatrix}\text{Probability of choosing}\\ \text{mode } m \text{ when the set of}\\ \text{alternatives is the set of}\\ \text{modes available at time } t\end{bmatrix}. \tag{4.45}$$

Further,

$$\begin{bmatrix}\text{Probability}\\ \text{of choosing}\\ \text{time } t\text{, any mode}\end{bmatrix} = \begin{matrix}\text{Sum over modes}\\ m' \text{ available at}\\ \text{time } t\end{matrix} \begin{bmatrix}\text{Probability of}\\ \text{choosing available}\\ \text{mode } m' \text{ at time } t\end{bmatrix}. \tag{4.46}$$

In algebraic terms, if $P_i$ in eq. (4.39) is the probability of choosing a specific mode and time from the list $j = (1, \ldots, J)$ of all possible modes and times, and if the set of alternatives $k = (1, \ldots, i)$ denotes those corresponding to the different modes available at the specified time, then

$$P_i = \mathrm{e}^{V_i} \Big/ \sum_{j=1}^{J} \mathrm{e}^{V_j}$$
$$= \left[\mathrm{e}^{V_i} \Big/ \sum_{j=1}^{i} \mathrm{e}^{V_j}\right] \left[\sum_{j=1}^{i} \mathrm{e}^{V_j} \Big/ \sum_{l=1}^{J} \mathrm{e}^{V_l}\right], \tag{4.47}$$

where $V_i = V(x^i, s)$ and the terms in brackets correspond to the terms in eq. (4.44). Thus we can reconstruct the complete probability function of simultaneous time and mode choices from estimates of probabilities of choosing among various modes at a given time and a second set of estimates of choice among times (ignoring mode choice within this decision).

To show that this result is consistent with the structure of individual utilities assumed earlier, consider the additively separable utility function specified in eq. (3.9). With a suitable monotone transformation, this function can be written

$$U(x, s) = \sum_{i=1}^{7} \phi^i(x_{(i)}, s), \tag{4.48}$$

where the $x_{(i)}$ are subvectors of $x$ corresponding to mode choice, time of travel, attributes of the destination, attributes of the no-trip choice, attributes of locational choices, and attributes of all other consumer choices, respectively. Assume that in the urban population the effect of individual differences in tastes is to add a stochastic component $\eta(x, s)$ to this "representative" utility function, yielding

$$U(x, s) = \sum_{i=1}^{7} \phi^i(x_{(i)}, s) + \eta(x, s). \tag{4.49}$$

The probability of choosing mode $m$, conditioned on the event that time of travel $t$ and other attributes of the consumer's environment are held fixed, is

$$\begin{aligned} P_{m|t} &= \mathrm{Prob}\left[U(x^m, s) > U(x^j, s) \quad \text{for all } j \neq m\right] \\ &= \mathrm{Prob}\left[\phi^1(x^m_{(1)}, s) + \eta(x^m, s)\right. \\ &\qquad \left. > \phi^1(x^j_{(1)}, s) + \eta(x^j, s) \quad \text{for } j \neq m\right], \end{aligned} \tag{4.50}$$

since $x^j_{(k)} = x^m_{(k)}$ for $j \neq m$. If we now assume that the $\eta(x^j, s)$ have independent Weibull distributions, we obtain by our earlier reasoning the result

$$P_{m|t} = \frac{e^{\phi^1(x_{(1)}^{mt}, s) - \alpha_{mt}}}{\sum_j e^{\phi^1(x_{(1)}^{mt}, s) - \alpha_{jt}}}, \tag{4.51}$$

where the $\alpha_{jt}$ are parameters of the Weibull distributions, and the modal split choice probability is given by the multinomial logit model.

Next consider the choice of time of travel when the most desired mode is used. It is now convenient to use a double index $(j, p)$ for the alternatives, where $j$ indicates mode and $p$ indicates time of travel. The probability that time $t$ will be chosen by a randomly selected individual is, analogously to eq. (3.8),

$$\begin{aligned} P_t &= \text{Prob}\,[\operatorname*{Max}_j U(x^{jt}, s) > \operatorname*{Max}_j U(x^{jp}, s) \quad \text{for } p \neq t] \\ &= \text{Prob}\,[\phi^2(x_{(2)}^t, s) + \operatorname*{Max}_j \{\phi^1(x_{(1)}^{jt}, s) + \eta(x^{jt}, s)\} > \\ &\qquad \phi^2(x_{(2)}^p, s) + \operatorname*{Max}_j \{\phi^1(x_{(1)}^{jp}, s) + \eta(x^{jp}, s)\} \quad \text{for } p \neq t]. \end{aligned} \tag{4.52}$$

Provided that the $\eta(x^{jp}, s)$ have independent Weibull distributions for each index $(j, p)$, we conclude that $\eta_p$, defined by

$$\operatorname*{Max}_j \phi^1(x_{(1)}^{jp}, s) + \eta_p = \operatorname*{Max}_j \{\phi^1(x_{(1)}^{jp}, s) + \eta(x^{jp}, s)\}, \tag{4.53}$$

is again distributed independently Weibull. Therefore, eq. (4.52) can be written in the multinomial logit form

$$\begin{aligned} P_t &= \text{Prob}\,[\phi^2(x_{(2)}^t, s) + \operatorname*{Max}_j \phi^1(x_{(1)}^{jt}, s) + \eta_t > \\ &\qquad \phi^2(x_{(2)}^p, s) + \operatorname*{Max}_j \phi^1(x_{(1)}^{jp}, s) + \eta_p \quad \text{for } p \neq t] \end{aligned} \tag{4.54}$$

$$= \frac{\exp\,[\phi^2(x_{(2)}^t, s) + \operatorname*{Max}_j \phi^1(x_{(1)}^{jt}, s) - a_t]}{\sum_p \exp\,[\phi^2(x_{(2)}^p, s) + \operatorname*{Max}_j \phi^1(x_{(1)}^{jp}, s) - a_p]}, \tag{4.55}$$

where the Weibull distribution "parameters" $a_p$ satisfy (using Lemma conclusions (a) and (b))

$$a_p = \operatorname*{Max}_p \phi^1(x_{(1)}^{jp}, s) - \log \sum_j \exp\,[-\alpha_{jp} + \phi^1(x_{(1)}^{jp}, s)]. \tag{4.56}$$

Substituting this expression in eq. (4.55) then yields

$$P_t = \frac{\exp\left[\phi^2(x^t_{(2)}, s)\right] \sum_j \exp\left[-\alpha_{jt} + \phi^1(x^{jt}_{(1)}, s)\right]}{\sum_p \exp\left[\phi^2(x^p_{(2)}, s)\right] \sum_j \exp\left[-\alpha_{jt} + \phi^1(x^{jt}_{(1)}, s)\right]}$$

$$= \frac{\sum_j \exp\left[\phi^1(x^{jt}_{(1)}, s) + \phi^2(x^t_{(2)}, s) - \alpha_{jt}\right]}{\sum_p \sum_j \exp\left[\phi^1(x^{jp}_{(1)}, s) + \phi^2(x^p_{(2)}, s) - \alpha_{jp}\right]}. \tag{4.57}$$

Combining eqs. (4.51) and (4.57) in the formula (4.44) yields the choice probability $P_{mt}$ for the simultaneous choice of mode $m$ and time $t$,

$$P_{mt} = P_{m|t} P_t$$

$$= \frac{e^{\phi^1(x^{mt}_{(1)}, s) - \alpha_{mt}}}{\sum_j e^{\phi^1(x^{mt}_{(1)}, s) - \alpha_{jt}}} \frac{\sum_j e^{\phi^1(x^{jt}_{(1)}, s) + \phi^2(x^t_{(2)}, s) - \alpha_{jt}}}{\sum_p \sum_j e^{\phi^1(x^{jp}_{(1)}, s) + \phi^2(x^p_{(2)}, s) - \alpha_{jp}}}$$

$$= \frac{e^{\phi^1(x^{mt}_{(1)}, s) + \phi^2(x^t_{(2)}, s) - \alpha_{mt}}}{\sum_p \sum_j e^{\phi^1(x^{jp}_{(1)}, s) + \phi^2(x^p_{(2)}, s) - \alpha_{jp}}}. \tag{4.58}$$

But this is precisely the formula for the simultaneous choice probability for mode $m$ and time $t$ resulting from the assumption of independent Weibull distributed stochastic components in eq. (4.49). We conclude that when the strong assumptions of independent Weibull distributed stochastic components of utility and an additively separable utility structure are valid, the independence of irrelevant alternatives property can be exploited to greatly simplify the magnitude and complexity of the demand estimation task. In particular, for "marginal" choice probabilities, such as $P_t$ in eq. (4.57), terms of the form $\sum_j \exp\left[-\alpha_{jp} + \phi^1(x^{jp}_{(1)}, s)\right]$ can be treated as a single "inclusive" index of the desirability of travel at time $p$, taking into account the attributes of alternative modes at this time. Define

$$y_p = -\log \sum_j \exp\left[-\alpha_{jp} + \phi^1(x^{jp}_{(1)}, s)\right] \tag{4.59}$$

to be the "inclusive cost" of traveling at time $p$, and suppose $\phi^1$ has the linear-in-parameters functional form

$$\phi^1(x_{(1)}^{jp}, s) = \sum_{k=1}^{K} \beta_k Z^k(x_{(1)}^{jp}, s). \tag{4.60}$$

Define

$$q_j = -\alpha_{jp} + \sum_{k=1}^{K} \beta_k Z^k(x_{(1)}^{jp}, s),$$

and

$$y_p = y(q_1, \ldots, q_J) = -\log \sum_j e^{q_j}$$

Let $\bar{q}$ denote the average of the $q_j$, and make a first-order Taylor's expansion of $y(\bar{q}, \ldots, \bar{q})$ about the vector of values $(q_1, \ldots, q_J)$:

$$\begin{aligned} y(\bar{q}, \ldots, \bar{q}) &\equiv -\log J - \bar{q} \\ &= y(q_1, \ldots, q_J) + \sum_j \frac{\partial y}{\partial q_j}\bigg|_{(q_1, \ldots, q_J)} \cdot (\bar{q} - q_j) \\ &\quad + \text{higher-order terms in } (\bar{q} - q_j), \end{aligned} \tag{4.61}$$

but

$$\frac{\partial y}{\partial q_i} = \frac{-e^{q_i}}{\sum_j e^{q_j}}.$$

Comparing this formula with eq. (4.51),

$$\frac{\partial y}{\partial q_i} = -P_{i|p}.$$

Hence, from eq. (4.61),

$$\begin{aligned} y_p = y(q_1, \ldots, q_J) &= y(\bar{q}, \ldots, \bar{q}) - \sum_j \frac{\partial y}{\partial q_j}\bigg|_{(q_1, \ldots, q_J)} \cdot (\bar{q} - q_j) \\ &\quad - \text{higher-order terms} \\ &= -\log J - \bar{q} + \sum_j P_{j|p} \cdot \bar{q} - \sum_j P_{j|p} q_j \\ &\quad - \text{higher-order terms} \\ &= -\log J + \sum_j P_{j|p} \alpha_{jp} - \sum_{k=1}^{K} \beta_k \sum_j P_{j|p} Z^k(x_{(1)}^{jp}, s) \\ &\quad - \text{higher-order terms in } (q_j - \bar{q}). \end{aligned}$$

When the higher-order terms can be neglected, this approximation establishes that the inclusive cost of a trip at time $p$ by the best available mode can be measured by weighting the variables for the alternative modes by the corresponding modal split probabilities and summing over these weighted variables, using the coefficients estimated from empirical modal splits at a fixed travel time. Under the assumption that all mode-specific effects are included among the $z^k$ variables, we can, without loss of generality, take the $\alpha_{jp}$ above to be zero. Hence the constant term in this expression for $y$ can be ignored, and we can write

$$y_p = -\sum_{k=1}^{K} \beta_k \left[ \sum_{j=1}^{J} P_{j|p} Z^k(x_{(1)}^{jp}, s) \right], \tag{4.62}$$

and

$$P_t = \frac{\exp\left[\phi^2(x_{(2)}^t, s) - y_t\right]}{\sum_p \exp\left[\phi^2(x_{(2)}^p, s) - y_p\right]}. \tag{4.63}$$

This structure allows very substantial savings in the number of parameters and alternatives which must be treated in estimation of the model. We shall assume throughout our empirical analysis that the approximation above is valid, and we will use formula (4.62) to calculate inclusive prices. Note that $y_p$ can also be interpreted as a measure of accessibility and used as such in analyzing the impacts of transportation improvement. The above formula can be applied to each stage of the decision tree.

The weakness of the "strict utility" probability model is that the independence of irrelevant alternatives property may be implausibly strong in some applications. An example illustrates this point. Suppose a population faces the alternatives of one auto mode and one bus mode, and chooses the auto mode with probability $\frac{2}{3}$. Now suppose a second bus mode is introduced which follows a different route, but has essentially the same attributes as the first bus mode. Intuitively, we believe that individuals will still choose the auto mode with probability $\frac{2}{3}$, and will choose either of the bus modes which one-half the probability $\frac{1}{3}$ of choosing some bus mode, or $\frac{1}{6}$. However, the independence of irrelevant alternatives condition requires that the relative odds of choosing the auto mode over *either* of the bus modes be two to one, implying that the probability of choosing the auto mode drops to $\frac{1}{2}$, and the probability of choosing each bus mode is $\frac{1}{4}$. The reason this result is counterintuitive is that we expect the individual to lump the two bus modes together rather than treating them as "independent" alternatives.

This example suggests that application of the "strict utility" model should be limited to multiple-choice situations where the alternatives can plausibly be assumed to be distinct and independent in the eyes of the decision-maker. Thus care must be taken in specifying the available alternatives and decision-making structure when using this multiple-choice model. In empirical analysis this can often be done by postulating sequential choice structures in which the "inclusive prices" associated with first-level choices among close substitutes reflect the fact that these alternatives are not independent. An example illustrates how this can be done. Suppose individuals choose between an auto mode and several transit modes. Suppose the transit modes are perceived similarly, in the sense that an individual drawn randomly from the population who is positively inclined to one of the transit modes is likely to be positively inclined to the others. The individual may then be modeled as first choosing between transit modes under the hypothetical condition that he chooses transit. In this comparison, a common stochastic component in the utilities of the alternatives arising from auto versus overall transit taste variations will tend to cancel out, and the logit model is likely to provide a satisfactory explanation of choice among transit alternatives. The second stage of the individual's decision process is to choose between auto and the "best" transit mode. The interdependence of the stochastic components of the utilities of the transit alternatives now influences the level of "mean" utility and the distribution of utility associated with the "best" transit alternative. Suppose, for example, that the auto versus overall transit component of variations in tastes, which was assumed to cancel out in intra-transit mode choices, looms sufficiently large in auto versus transit choices to make the effects of further variations in tastes among transit alternatives negligible. Then, the "mean" utility of the "best" transit alternative is almost exactly the maximum of the mean utilities of the set of transit alternatives. A logit model of the choice between auto and the best transit alternative valued in this manner can then provide a satisfactory explanation of the auto—transit choice. This example illustrates a case in which the logit framework can be adapted to describe a tree decision structure in which the independence of irrelevant alternatives does not hold. It should be noted, however, that the implications drawn from the stochastic specification hold only approximately. It is computationally difficult to obtain bounds on the accuracy of such an approximation, or to develop an exact model. A further discussion of this approach and some empirical evidence on the extent to which the independence of irrelevant alternatives assumption, inappropriately applied, may bias results is given in McFadden (1973b).

A second empirical procedure for alleviating the difficulties raised by the independence property is to include among the attributes of alternatives variables describing the range of available options. For example, a multinomial logit model of mode choice between one auto alternative and several similar bus alternatives may include as an attribute of each bus alternative the total number of bus alternatives available. This procedure is consistent with the underlying theory of individual utility maximization, provided we allow the possibility that the range of available alternatives enters as an externality in the utility function. While this is an effective method of reintroducing patterns of differential substitutability ruled out by the independence assumption, no argument has been supplied to justify this model of individual behavior.

The advantages offered by the "strict utility" model in allowing separation and simplification of empirical studies of the decision process, plus some empirical evidence from psychological studies that it provides a satisfactory model of laboratory choice behavior [Luce (1959)] seem to outweigh the drawbacks, provided the model is used judiciously so that "independent" alternatives are identified. All the empirical multiple-choice models which appear in the literature have the "strict utility" structure, and hence are subject to the above caveat.

Specification of explicit probability functions for the "strict utility" specification in eq. (4.39) can be completed by specifying parametric forms for the functions $V(x, s)$. We shall consider several cases. First, suppose this function is log-linear in unknown parameters; i.e.,

$$V(x^i, s) = \log\left[\sum_{k=1}^{K} \beta_k Z^k(x^i, s)\right]. \tag{4.64}$$

Suppose further that

$$0 \leqq \sum_{k=1}^{K} \beta_k Z^k(x^i, s) \leqq 1 \quad \text{for } i = 1, \ldots, J,$$

and

$$\sum_{i=1}^{J} \sum_{k=1}^{K} \beta_k Z^k(x^i, s) = 1. \tag{4.65}$$

Then the probability functions have the form

$$P_i = h(B, s, i) = \sum_{k=1}^{K} \beta_k Z^k(x^i, s). \tag{4.66}$$

Note that this is just a multinomial extention of the binary linear probability model. Its linear form is appealing for empirical applications. However, it should be noted that consistent application of the model requires inclusion of the inequality restrictions in eq. (4.65), destroying its simple linear structure.

The second concrete specification we consider for the $V(x, s)$ function directly incorporates the restrictions above, and leads to a multinomial extension of the logit model. Suppose $V$ has the linear-in-parameters form

$$V(x^i, s) = \sum_{k=1}^{K} \beta_k Z^k(x^i, s). \tag{4.67}$$

Then, from eq. (4.39), the probability functions have the form

$$P_i = h(B, s, i) = \frac{1}{\sum_{l=1}^{J} \exp \sum_{k=1}^{K} \beta_k [Z^k(x^l, s) - Z^k(x^i, s)]}. \tag{4.68}$$

This model is termed the conditional, multinomial, or polychotomous logit model. It was first developed systematically by Gurland, Lee, and Doland (1960). A more general formulation and application to transportation problems was made by McFadden (1968). Other applications of the model have been made by Theil (1969), and an adaptation of Theil's model was made by Rassam, Ellis, and Bennet (1971). An extensive survey of statistical properties of this and related models has been made by McFadden (1973a).

Of the models for multiple choice developed above, the multinomial logit model proves to be the most useful for the demand analysis of available transportation survey data. This model is empirically tractable and has a satisfactory theoretical justification in terms of the underlying behavior of individual decision makers.

## 4.8. Applications of behavioral travel demand models

Once the parameters of a concrete behavioral travel demand model, such as the one given in eq. (4.68), are calibrated, this model can be used to predict directly the behavior of an individual selected randomly from the population. The conventional goodness of fit statistical tests discussed in the next chapter can be used to assess the accuracy of the model. However, to go beyond the prediction of individual behavior to the construction of aggregate trip tables and patronage forecasts of interest to transportation planners, it is necessary to consider the process of aggregation.

An urban population will consist of a large number of individuals who differ in their socioeconomic characteristics and the set of alternatives they face. For the subpopulation with a common vector of socioeconomic characteristics $s^i$ and a common set of vectors of attributes $\{x^{1i}, \ldots, x^{J_i i}\}$, the calibrated behavioral model gives numerical probabilities $P^i_j$ that an individual drawn at random from this subpopulation will choose alternatives $j = 1, \ldots, J_i$. The expected distribution of the subpopulation among the alternatives is given by these probabilities. Letting $i = 1, \ldots, I$ index all the subpopulations characterized by socioeconomic characteristics and available alternatives, and letting $N_i$ denote the size of subpopulation $i$, the expected population demand for alternative $j$ is

$$D_j = \sum_{i=1}^{I} N_i P^i_j. \tag{4.69}$$

In practice, the planner will have available to him either a homogeneous or stratified random sample of individuals from the population, with observations $x^{ji}$ and $s^i$ for each sampled individual, or a list of summary statistics defining the distribution of attributes of alternatives and socioeconomic characteristics in the population. In the former case, the sample average probability, weighted to the dimensions of the population, provides the best estimate of aggregate demand,

$$D_j = \sum_{i=1}^{I} P^i_j \theta_i, \tag{4.70}$$

where $i = 1, \ldots, I$ denotes the individuals in the sample, $\theta_i$ is the reciprocal of the probability that an individual would be drawn in the sample from the "strata" containing $i$, and $P^i_j$ is the numerical probability calculated from the observations $x^{ji}$ and $s^i$. A random sample used to calibrate the model may also be used to predict aggregate demand. However, it should be noted that calibration is often carried out using non-random or special-purpose samples which are not a satisfactory base for forecasting aggregate demand.

A second method of estimating aggregate demand is to utilize summary statistics from the population on the distribution of the $x$ and $s$ vectors. With sufficient assumptions to identify the distribution from the summary statistics, one can generate a hypothetical random sample of individuals from the population and then proceed as in the first case to construct the estimate of aggregate demand from the sample average. Alternately, one can form, numerically or analytically, the expectation of $P_j$ with respect to the distribution of the $x$ and $s$ vectors.[7]

To illustrate this procedure, we consider the case of a binary mode choice between auto and transit, and assume that there is a binary probit model of the form given in eq. (4.13) which fits the observed data. Suppose that the vector of explanatory variables,

$$z = Z(x^1, s) - Z(x^2, s),$$

has a multivariate normal distribution with a vector of means $\bar{z}$ and a covariance matrix $\Omega$. (The values of $\bar{z}$ and $\Omega$ might be obtained, for example, from census statistics or existing transportation surveys.) The calibrated selection probability for alternative 1 is $P_1 = \Phi(\beta' z)$. The assumption above implies $\beta' z$ is distributed normally with mean $\beta'\bar{z}$ and variance $\beta'\Omega\beta$. Hence, the expected demand for the first alternative in a population of size $N$ is

$$D_1 = \frac{N}{\sqrt{(\beta'\Omega\beta)}} \int_{-\infty}^{+\infty} \Phi(t)\, \phi\left[\frac{t - \beta'\bar{z}}{\sqrt{(\beta'\Omega\beta)}}\right] \mathrm{d}t. \tag{4.71}$$

If $X_k$ is normal with mean $\mu_k$ and variance $\sigma_k^2$ for $k = 1, 2$, then, utilizing the convolution property of two normal random variables,

$$\begin{aligned} \text{Prob}\,[X_1 - X_2 \leqq x] &= \Phi\left[\frac{x - \mu_1 + \mu_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)}}\right] \\ &= \frac{1}{\sigma_2} \int_{-\infty}^{+\infty} \Phi\left[\frac{t + x - \mu_1}{\sigma_1}\right] \phi\left[\frac{t - \mu_2}{\sigma_2}\right] \mathrm{d}t. \end{aligned} \tag{4.72}$$

Comparing this formula with eq. (4.71), we set $\sigma_1^2 = 1$, $\sigma_2^2 = \beta'\Omega\beta$, $x = \mu_1 = 0$, and $\mu_2 = \beta'\bar{z}$ to obtain

[7] The implications of this approach are developed in greater detail in McFadden and Reid (1974).

$$D_1 = N\Phi\left[\frac{\beta'\bar{z}}{\sqrt{(1 + \beta'\Omega\beta)}}\right]. \tag{4.73}$$

Thus, this example yields a simple closed formula for aggregate demand. Because of the similarity of the normal and logistic curves, the formula in eq. (4.73) can be used to give good approximations to the aggregate demand resulting from a binary logit response curve and normally distributed explanatory variables. More generally, one can apply numerical integration to the expression for the expectation of the selection probabilities with respect to the distribution of the explanatory variables. The result, scaled by population size, gives the aggregate demand.

One can analyze the effects of transport policy using the aggregate demand measure constructed by one of the methods suggested in the preceding paragraphs. We will first consider policy changes which affect the conditions of travel by existing modes but do not introduce new alternatives. Examples of this kind of change are changes in transit fares, headways, running times, auto tolls, taxes on parking or gasoline, restrictions on parking availability, or auto driving times. To analyze the effect of such changes, we first compute the values of $x^{ji}$ and $s^i$ which would prevail after the policy change for a real or hypothetical random sample of the population. Some policies can be analyzed by considering a percentage change in one of the explanatory variables. For example, we may investigate the effect of a ten percent increase in the excise tax on parking tariffs, where we assume that this increase is completely shifted forward to consumers. Provided that these changes are incremental, the impact is summarized in the market elasticities of demand for the alternative choices with respect to each of the affected explanatory variables. It should be noted that these market demand elasticities are given by a weighted average of the various elasticities of the response curve at the values of the explanatory variables for each homogeneous subpopulation. They may differ significantly from corresponding elasticities calculated at the population mean of the explanatory variable. The elasticity formulae derived in the following paragraph show this averaging effect for the individual elasticities.

Consider individuals of the "type" $i$ which have characteristics $s^i$ and face alternatives indexed $1, \ldots, J_i$ with attributes $x^{ji}$. Suppose the probability $P^i_j$ that individuals of this type choose alternative $j$ is given by the multinomial logit model,

$$P_j^i = \frac{e^{\beta' z^{ji}}}{\sum_{l=1}^{J_i} e^{\beta' z^{li}}}, \tag{4.73a}$$

where $z^{ji} = Z(x^{ji}, s^i)$ is a $K$-vector of numerical functions of the observations and $\beta$ is a commensurate vector of parameters. With $N_i$ individuals in the population of type $i$, the expected demand for alternative $j$ by this group is $N_i P_j^i$. The change in this demand caused by a one-unit change in the value of component $k$ in the vector of independent variables for alternative $j$ is, differentiating (4.73a),

$$\frac{\partial(N_i P_j^i)}{\partial z_k^{ji}} = \beta_k N_i P_j^i (1 - P_j^i). \tag{4.73b}$$

Similarly, the change in this demand caused by a one-unit change in component $k$ in the vector of independent variables for an alternative $l \neq j$ is

$$\frac{\partial(N_i P_j^i)}{\partial z_k^{li}} = -\beta_k N_i P_j^i P_l^i, \quad l \neq j. \tag{4.73c}$$

Converting these expressions to elasticity terms, we obtain first the type $i$ elasticity of demand for alternative $j$ with respect to own variable $z_k^{ij}$,

$$E_j^i(j, k) \equiv \frac{z_k^{ji}}{N_i P_j^i} \frac{\partial(N_i P_j^i)}{\partial z_k^{ji}} = \beta_k z_k^{ji}(1 - P_j^i). \tag{4.73d}$$

Second, we obtain the type $i$ elasticity of demand for alternative $j$ with respect to a "cross" variable $z_k^{il}$,

$$E_j^i(l, k) = \frac{z_k^{li}}{N_i P_j^i} \frac{\partial(N_i P_j^i)}{\partial z_k^{li}} = -\beta_k z_k^{li} P_l^i. \tag{4.73e}$$

We next compare these individual elasticities with the market demand elasticities for the population. From eq. (4.69), the market demand for alternative $j$ is

$$D_j = \sum_j N_i P_j^i, \tag{4.73f}$$

where the index $i$ extends over the types in the population who have alternative $j$ available. Suppose $\bar{z}_k^{li}$ denotes the initial value of a variable $z_k^{li}$. Then, a uniform percentage change in this variable for each type $i$ can be defined by writing $z_k^{li} = t\bar{z}_k^{li}$, where $t$ is a scalar. The elasticity of market demand with respect to such a uniform percentage change is then defined as the elasticity with respect to $t$, evaluated at $t = 1$. Using this definition, the elasticity of market demand for alternative $j$, with respect to a uniform one percent increase in own variable $z_k^{ji}$, is

$$\begin{aligned} E_j(j, k) &\equiv \frac{t}{D_j} \frac{\partial D_j}{\partial t}\bigg|_{t=1} = \left[ \frac{t}{\sum_i N_i P_j^i} \sum_i \frac{\partial (N_i P_j^i)}{\partial (\bar{z}_k^{ji} t)} \frac{\partial (\bar{z}_k^{ji} t)}{\partial t} \right]\bigg|_{t=1} \\ &= \sum_i w_i \frac{\partial (N_i P_j^i)}{\partial (\bar{z}_k^{ji} t)}\bigg|_{t=1} \frac{\bar{z}_k^{ji}}{N_i P_j^i} \\ &= \sum_i w_i E_j^i(j, k), \end{aligned} \tag{4.73g}$$

where

$$w_i = N_i P_j^i \Big/ \sum_l N_l P_j^l$$

is a weight giving the proportion of the total demand for alternative $j$ originating from individuals of type $i$. Similarly, the elasticity of market demand for alternative $j$, with respect to the "cross" variable $z_k^{li}$ with $l \neq j$, is

$$E_j(l, k) = \sum_i w_i E_j^i(l, k).$$

The qualitative effect of "averaging" over the population is clear from these formulae. In many applications, market demand for an alternative will be comprised of a large group of individuals whose choice of the alternative is clear-cut, and a much smaller group whose choice is sensitive to small changes in the independent variable. Then, the contribution of the large group to the market demand elasticity is small, because the individual demand elasticities are small, while the contribution of the smaller group is small because of their numerical size. The result is a market demand elasticity which is typically smaller in magnitude than the value of the individual elasticity formula evaluated at the population mean of the independent variable.

Many policy changes affect different segments of the population differently, and their impact cannot be determined by simply considering market demand elasticities. For example, a change in the structure of transit fares or in headways on particular transit routes requires recalculation of the relevant components of the $x^{ij}$ vectors on an individual-by-individual basis. The costs of obtaining such data can be substantial;

nevertheless, it is clear that any demand forecasting model providing accurate measures of the impact of particular local transport policy changes will require detailed data of this sort.

If the impact of a policy is localized, the real or hypothetical random sample of the population used to carry out the demand forecast should reflect the distribution of the impact. Thus, the impact of changing headways on a particular bus route would be assessed using a stratified random sample taken predominantly from the population of the corridor serviced by the bus route. It is important to emphasize that the criteria for selection of a population sample differ according to whether the sample is to be used for calibration of the behavioral model or for forecasting. Furthermore, provided that the calibration process is successful in establishing behavioral models which are valid across subpopulations, the fitted models can be applied to the forecasting samples without further calibration. This has two implications. First, the population sample used for the calibration can be designed to optimize the statistical properties of the estimates. It is desirable to retain the property of having a (stratified) random sample of the population in order to make inferences on the accuracy of the calibration as a representation of population behavioral parameters. However, the simple stratification or complete design which would be required for easy use of the sample as a forecasting base are unnecessary. Second, the specialized sample used as a forecasting base need not correspond to a full household survey. Only the values of the explanatory variables $x^{ij}$ and $s^i$ are needed, and it may be feasible to use a variety of data sources, such as census block statistics, transport grid calculations, and general transportation survey statistics, to construct $x^{ij}$ and $s^i$ for hypothetical individuals, rather than obtaining these figures from a real sample survey. Because of this it may be possible to provide accurate forecasts of the impact of transportation policy without making costly and time-consuming household surveys for each impact analysis.

The transportation planner is concerned with the impact of major policy changes which involve the introduction of new transportation alternatives as well as the impact of these incremental policy changes, and the procedure outlined above can also be used to forecast the impact of such major policy changes. Earlier in this chapter, we described the procedure for determining the selection probabilities after the introduction of a new alternative for the case in which alternatives are assessed in terms of generic attributes, and the assumptions underlying the multinomial logit model are met. This method can be used if the model has been calibrated for the present alternatives, and the attributes of the new alternative for each individual in the forecasting data base are known. Data on these attributes can be derived from the design specifications of the new transportation mode.

Forecasting the impacts of major changes in the transportation system involves two problems which are not analyzed in this study. First, major changes in transportation are likely to have significant impacts on related consumer decisions, particularly automobile ownership and residential and work locations. These decisions, in turn, may have a significant impact on transportation demand, and a fully successful forecasting model must take into account the structural interrelationships between these decisions. Second, major changes in transportation policy can sometimes be expected to result in shifts in demand which are large enough to affect the attributes of transport alternatives. For example, the introduction of a major new mass transit mode will increase transit patronage and may reduce highway congestion and travel time, while this, in turn, may lessen the increase in transit patronage. Accurate forecasting requires explicit consideration of the process of equilibration, which takes place, for example, as the transit patronage and highway congestion mentioned in the preceding example reach a balance, and analysis of the equilibration process requires explicit models of the relationship between transport mode service levels and attributes. For example, we may need functions relating auto and bus travel times to congestion levels in the system, or transit service levels to patronage rates for transit agencies operating under budget constraints. This study represents only the demand side of the modeling effort. Fully equilibrated forecasts require a parallel model of the supply of transportation services. This supply model must determine trip attributes as functions of system loads. It is also necessary to provide methods for carrying out the equilibration process.

## 4.9. *Marginal and conditional trip tables*

The construction of aggregate demand forecasts described in the preceding section provides the transportation planner with a method of constructing detailed policy-sensitive trip tables. In their most detailed form, these tables would distinguish all the components of a "trip" relevant to the planner: mode, time of day, origin, destination, purpose, and the socioeconomic characteristics of the trip-takers. However, specific policy questions can normally be answered by examining choices over just one of these dimensions; for example, aggregate modal split or the modal split for a particular corridor and socioeconomic group. For this analysis it is convenient to work with various reduced trip tables. Exploiting the relationship between these tables and the selection probabilities, we adopt the terminology of statistics and speak of *marginal* and *conditional* trip tables. In this section, we explain the relationship between these concepts and the trip generation, trip distribution, and modal split tables familiar to transportation planners.

The components of the trip will be denoted by subscripts: $m$ for mode, $t$ for time-of-day, $o$ and $d$ for origin and destination, $p$ for purpose, and $s$ for socioeconomic characteristics of the subpopulation being considered. Let $T_{mtodps}$ denote the number of trips, with the specified characteristics, made by an urban population in a given period. The array of values of $T$ for all possible values of the vector of subscripts defines the basic *detailed trip table.*

We first relate this trip table to our behavioral models of individual travel demand. Suppose, for each individual (or individual type) in the population, we enumerate in complete detail all the possible daily travel patterns. For example, "auto trip, residence to 4th and Main, at 8:15 a.m., work purpose; followed by walk trip, 4th and Main to 6th and Main, at 12:15 p.m., shopping purpose; followed by ..." is one alternative, while "bus trip, residence to 4th and Main, at 8:00 a.m., work purpose; followed by bus trip, 4th and Main to 12th and Main at 12:15 p.m., shopping purpose; ..." is a second, and "no trips from residence all day" is a third. These descriptions are chosen to be mutually exclusive and exhaustive, so that the individual selects exactly one daily travel pattern. The theory of individual choice behavior described in chapter 3 is assumed to describe this selection process. The distribution of tastes in the subpopulation facing the same objective environment then yields selection probabilities for the daily travel patterns. If we let $s$ denote the vector of observed characteristics of the individuals in this homogeneous subpopulation, index daily travel patterns by $j \in J$, and let $x = (x^j)$ denote the vector whose subvector $x^j$ gives the observed attributes of alternative $j$, then we can define $P_j(x, s)$ to be the selection probability for pattern $j$, or the proportion of this homogeneous subpopulation choosing $j$. It should be emphasized that at this level of complexity and detail, actual observation of $x$ and calibration of a model (such as the multinomial logit model) for $P_j(x, s)$ is impractical. The separability conditions on utility introduced in chapter 3 are utilized to break apart the travel pattern choice into manageable components, and the overall $P_j(x, s)$ can be visualized in principle as being built up from these components.

We let $N(x, s)$ denote the number of individuals in the homogeneous subpopulation described in the preceding paragraph. The size of the total urban population is then

$$N = \sum_x \sum_s N(x, s) \tag{4.74}$$

We let $n(x, s) = N(x, s)/N$ denote the distribution of "types" of subpopulations in the urban area.

Now consider a "trip" as defined in a detailed trip table, identified by mode $m$, time of day $t$, origin and destination (or link) $o$ and $d$, and purpose $p$. For a particular homogeneous subpopulation identified by a set of daily travel patterns $J$ and "environment" $(x, s)$, a particular travel pattern $j$ may result in zero, one, or more than one "trip" $mtodp$. We shall assume the classification of time-of-day, purpose, and destination of trips is sufficiently fine so that we can neglect the possibility that some daily travel patterns may result in more than one trip of a specified type $mtodp$. Then, we can identify the set $J_{mtodp}$ of daily travel patterns which result in one trip of type $mtodp$, with the remaining daily travel patterns resulting in no trips of this type. Then, the detailed trip table satisfies

$$T_{mtodps} = \sum_x N(x, s) \sum_{j \in J_{mtodp}} P_j(x, s). \tag{4.75}$$

It is convenient to also distinguish the total number of trips of type $mtodp$ taken by each homogeneous subpopulation,

$$T_{mtodp}(x, s) = N(x, s) \sum_{j \in J_{mtodp}} P_j(x, s), \tag{4.76}$$

so that

$$T_{mtodps} = \sum_x T_{mtodp}(x, s). \tag{4.77}$$

Eq. (4.75) can be interpreted as giving an estimate of the detailed trip table as a function of the calibrated selection probabilities. These selection probabilities are themselves functions of estimates of the behavioral parameters of the model and specified values of the attributes describing the alternatives.

As noted earlier, both policy objectives and the practicalities of data collection and calibration require aggregation of the detailed trip table into marginal and conditional trip tables. Aggregating over some components of the trip description yields *marginal trip tables.* For example, letting $J_{todp}$ be the set of all daily travel patterns containing a "trip" on any mode on *todp*, we define

$$\begin{aligned} T_{todp}(x, s) &= N(x, s) \sum_{j \in J_{todp}} P_j(x, s) \\ &= N(x, s) \sum_m \sum_{j \in J_{mtodp}} P_j(x, s), \end{aligned} \tag{4.78}$$

and

$$T_{todps} = \sum_x T_{todp}(x, s). \tag{4.79}$$

This is the marginal trip table specifying the total number of trips on all modes at time $t$ for origin—destination pair $o$—$d$, for purpose $p$, by individuals with characteristics $s$. In a similar fashion, we can define the marginal number of trips from origin $o$ at time-of-day $t$ for purpose $p$ by group $s$:

$$T_{top}(x, s) = N(x, s) \sum_{j \in J_{top}} P_j(x, s), \tag{4.80}$$

$$T_{tops} = \sum_x T_{top}(x, s). \tag{4.81}$$

Suppose we include the "no-trip" option at time $t$ in our accounting by treating it as a "trip" from the origin $o$ to the same point $o$. Then the $(m, d, p)$ triples represent an exhaustive list of mutually exclusive options for the individuals whose daily travel pattern places them at location $o$ at time $t$. Then,

$$T_{to}(x, s) = N(x, s) \sum_{j \in J_{to}} P_j(x, s) \tag{4.82}$$

gives the total number of individuals, equal to the total number of potential trips, at *to* for the subpopulation characterized by $(x, s)$. In the long run, the numbers $T_{to}$ are influenced by residential and work location decisions and by life-style decisions influencing daily travel patterns, and can themselves be estimated from the calibration of an overall choice model. In the short run, they are essentially fixed by the location and demography of the sample.

As a notational shorthand, write

$$Q_{mtodp} \equiv Q_{mtodp}(x, s) \equiv \sum_{j \in J_{mtodp}} P_j(x, s), \tag{4.83}$$

with analogous definitions for other subscripts. Then

$$\begin{aligned} T_{mtodp}(x, s) &= N(x, s)Q_{mtodp}(x, s) \\ &= \left[\frac{Q_{mtodp}(x, s)}{Q_{todp}(x, s)}\right] \cdot \left[\frac{Q_{todp}(x, s)}{Q_{top}(x, s)}\right] \cdot \left[\frac{Q_{top}(x, s)}{Q_{to}(x, s)}\right] \cdot N(x, s)Q_{to}(x, s) \\ &\equiv Q_{m|todp}(x, s) \cdot Q_{d|top}(x, s) \cdot Q_{p|to}(x, s) \cdot T_{to}(x, s), \end{aligned} \tag{4.84}$$

where we have defined

$$Q_{m|todp}(x, s) = Q_{mtodp}(x, s)/Q_{todp}(x, s), \tag{4.85}$$

with analogous definitions for the remaining terms in eq. (4.84). Since $Q_{todp}(x, s)$ is the probability that a randomly drawn member of the subpopulation facing $(x, s)$ will make a trip by any mode on *todp*, and $Q_{mtodp}(x, s)$ is the probability of this trip on mode $m$, we see that $Q_{m|todp}(x, s)$ is the conditional modal split distribution for the trip *todp* and this subpopulation. Similarly, $Q_{d|top}$ is the conditional distribution of destinations of trips starting at *to* for purpose $p$, and $Q_{p|to}$ is the conditional distribution of trip purposes for trips starting at *to*.

From eq. (4.84), we obtain the marginal trip tables for the subpopulation,

$$T_{todp}(x, s) = Q_{d|top}(x, s)Q_{p|to}(x, s)T_{to}(x, s), \tag{4.86}$$

$$T_{top}(x, s) = Q_{p|to}(x, s)T_{to}(x, s). \tag{4.87}$$

Eq. (4.86) is interpreted as providing a trip generation and distribution table for the subpopulation (by time-of-day $t$, origin $o$, and purpose $p$). Since $T_{toop}$ corresponds to the "no-trip" option, the magnitude of $Q_{o|top}(x, s)$ determines the total number of trips generated at *top*, and

$Q_{d|top}(x, s)$ for $d \neq o$ determines the distribution of these trips to alternative destinations.[8] Eq. (4.87) provides a breakdown of trips by purpose.

Aggregating the formulae (4.84), (4.86) and (4.87) over $x$ provides the basic trip tables of interest:

$$T_{mtodps} = \sum_x Q_{m|todp}(x, s) Q_{d|top}(x, s) Q_{p|to}(x, s) T_{to}(x, s), \tag{4.88}$$

$$T_{todps} = \sum_x Q_{d|top}(x, s) Q_{p|to}(x, s) T_{to}(x, s), \tag{4.89}$$

$$T_{tops} = \sum_x Q_{p|to}(x, s) T_{to}(x, s). \tag{4.90}$$

The procedure for analyzing the response of these trip tables to policy changes is first to obtain expressions for the $Q$ probabilities as functions of the calibrated individual choice models, and then to calculate the effect of the policy on the $Q$ values for each subpopulation. Suppose we assume the multinomial logit model developed earlier in this chapter, along with the separability assumptions on individual utility given in chapter 3. Then $P_j(x, s)$ is proportional to $e^{V(x^j, s)}$, where the "mean" utility $V$ is additively separable across distinct trips in the daily travel pattern, and across characteristics of the time-of-travel, destination, etc. within each trip. Consider the probability

$$Q_{mtodp}(x, s) = \sum_{j \in J_{mtodp}} P_j(x, s),$$

for various $m$, with $todp$ fixed. Under our separability assumptions, for each $j \in J_{mtodp}$, $V(x^j, s)$ can be written as the sum of a term, $\beta'_{(1)} z_{(1)}^{mtodp}$, involving the attributes of the trip $mtodp$ contained in the daily travel pattern $j$, and a series of terms which are independent of the mode choice $m$ on the trip $todp$. Then $Q_{mtodp}(x, s)$, for various $m$, is proportional to $\exp[\beta'_{(1)} z_{(1)}^{mtodp}]$, implying

$$Q_{m|todp}(x, s) = \frac{\exp[\beta'_{(1)} z_{(1)}^{mtodp}]}{\sum_{m'} \exp[\beta'_{(1)} z_{(1)}^{m'todp}]}. \tag{4.91}$$

[8] The imputation of "purpose" to "non-trips" is arbitrary. We shall by convention take it to have the same distribution as the purposes of actual trips. Then $Q_{p|to}$ can be interpreted either as the distribution of purposes for actual trips or as the distribution of purposes prior to the trip–no-trip decision.

Hence, calibration of the parameter vector $\beta_{(1)}$ from data on mode choices alone allows us to obtain fitted values of $Q_{m|todp}$. Proceding in the same way, define terms $\beta'_{(2)}z^{todp}_{(2)}$ and $\beta'_{(3)}z^{top}_{(3)}$ of $V$ giving, respectively, attributes of destination $d$ for trips from $top$, and attributes of trip purpose $p$ for trips from $to$. Then,

$$Q_{d|top} = \frac{\sum_m \exp[\beta'_{(1)}z^{mtodp}_{(1)} + \beta'_{(2)}z^{todp}_{(2)}]}{\sum_{d'}\sum_m \exp[\beta'_{(1)}z^{mtod'p}_{(1)} + \beta'_{(2)}z^{tod'p}_{(2)}]}, \tag{4.92}$$

and

$$Q_{p|to} = \frac{\sum_d\sum_m \exp[\beta'_{(1)}z^{mtodp}_{(1)} + \beta'_{(2)}z^{todp}_{(2)} + \beta'_{(3)}z^{top}_{(3)}]}{\sum_{p'}\sum_d\sum_m \exp[\beta'_{(1)}z^{mtodp'}_{(1t} + \beta'_{(2)}z^{todp'}_{(2)} + \beta'_{(3)}z^{top'}_{(3)}]}. \tag{4.93}$$

The parameters $\beta_{(2)}$ and $\beta_{(3)}$ in these formulae can be calibrated from data on trip—no-trip and destination choices in the first case, and on trip purposes in the second case. Substitution of these calibrated formulae for the $Q$ probabilities into eqs. (4.88)–(4.90), as a final step, provides the policy-sensitive trip tables required for policy analysis.

An important characteristic of the detailed trip tables constructed from the disaggregated behavioral model is the interdependence of the modal split, generation, and distribution formulae. For example, a decrease in transit time for one trip $todp$ will, in general, affect not only modal split on this trip, but will also affect the probabilities $Q_{d|top}$ of trip generation and distribution, and the probabilities for trip purposes $Q_{p|to}$. Thus, this approach provides a logically consistent, behaviorally plausible basis for a joint policy analysis of modal split, generation and distribution.

Conventional methods of constructing trip tables from aggregate data employ a variety of assumptions on the structure and independence properties of the $Q$ probabilities, which in turn can be viewed as hypotheses on the separability of utility, the distribution of independent variables, and homogeneity of the environments facing subpopulations. It is possible to formulate statistical tests of these hypotheses for specified data sets, providing information on the validity of alternative "skim tree" methods for forming trip tables. This topic has been explored further by Brand (1972a) and Ben-Akiva (1972). We note in conclusion, however, that direct construction of trip tables from behavioral principles, along the lines suggested in this section, would appear to be the most useful direction from the standpoint of policy applications.

## 4.10. Population travel demand and cost–benefit analysis

We have seen that consumer choice behavior among discrete transportation alternatives can be conveniently described as the result of individual utility maximization, with a distribution of tastes in which the extensive margin determines the response of aggregate demand to transportation policy. We will next examine the consistency of this model of market demand with the measure of benefits ordinarily employed in cost–benefit analysis.[9]

The consumer surplus arguments which form the foundation for conventional cost–benefit analysis assume a population of identical consumers (e.g., both tastes and environments are identical) with continuous demand for the commodity whose supply is subject to public policy. Then, all demand variation occurs at the intensive margin, and the satisfaction of any consumer can be taken as an index of social welfare. Consider the further assumption that the "marginal utility of the numeraire is constant", i.e., that individual utility is additively separable into the numeraire good and the group of all remaining goods. The conventional analysis then concludes that the "income effect" on the commodity in question is zero, implying that the area under the market demand function (which coincides in this case to the Hicksian compensated demand function) gives a correct measure of consumer benefits, in the sense that this measure exceeds the cost of making the commodity available if and only if individual satisfaction increases. A presentation of the details of this theory is given in Diamond and McFadden (1974).

We wish to consider the status of this conclusion when there is a distribution of tastes in the population and the commodity demanded is lumpy, with each consumer demanding either zero or one units. We show that despite the major differences in the structure of demand in the conventional and qualitative choice formulations, essentially the same conclusion obtains. To make this statement meaningful, we first need to define a measure of social welfare when tastes vary in the population. We do this by assuming that individual utility is additively separable in the numeraire commodity, and that this commodity is transferable across individuals. If we call the numeraire "money", then this assumption is

[9] We are indebted to Eytan Sheshinski for posing the question of whether the demand models specified in this study are consistent with conventional cost–benefit calculations.

equivalent to assuming that a given amount of money will yield the same level of social welfare no matter how it is distributed over the population and no matter what the prices of non-numeraire commodities are. This is clearly an extremely strong assumption. It will hold approximately, in general, if the society always maintains an optimal distribution of income under any policy for the commodity in question, but is unlikely to hold in the absence of optimal income distribution. We note that in the absence of this assumption, we would confront the same problems of the measurement of social welfare and "second best" analysis that make conventional consumer surplus arguments insufficient.

Under the assumption on the structure of individual utility and the definition of social welfare given in the previous paragraph, we conclude exactly as in the conventional case of identical consumers with continuous demand that the area under the market demand curve provides a correct measure of benefits in cost–benefit analysis.

As a specific example, consider bus patronage as a function of fares. Assuming, for the moment, that all consumers face the same fare, this demand curve has the form illustrated in fig. 4.6. Suppose that a change in transport policy leads to a change in fare from level $p_1$ to level $p_2$, resulting in a demand shift from $D_1$ to $D_2$. Let $C$ denote the net cost of meeting this demand shift, including both operating and capital costs. We express both benefits and costs in per capita terms. Thus, $D$ is demand per capita, or *frequency* of transit use, estimated by the selection probability. Benefits are measured by the Marshallian consumer surplus $S$, given by the shaded area under the demand curve in fig. 4.6. The policy is judged desirable if $S > C$.

To demonstrate this result, we consider a model of individual utility maximization among discrete transportation alternatives, where tastes vary within the population. Assume that individual utility can be written in the form

$$u(y, x, s) = y + \phi(x, s) + \eta(x, s), \tag{4.94}$$

where $y$ is the quantity consumed of a numeraire commodity, $x$ is the vector of all other attributes of the alternative, $s$ is a vector of socioeconomic characteristics, $V(y, x, s) \equiv y + \phi(x, s)$ is the non-stochastic component of utility, and $\eta(x, s)$ is the stochastic component. Suppose each consumer has a binary choice between a non-transit alternative with attributes $x^0$ and a transit alternative with attributes $x^1$. Let $p_1$ denote the initial transit fare and $y_1$ the quantity of the numeraire commodity consumed at this fare if the non-transit alternative is chosen.

Consider a change in transportation policy leading to a new transit fare $p$, and let $C(p)$ equal the net total cost of this policy change. The quantity of the numeraire commodity available at price $p$ will differ from $y_1$ by the net per capita cost of the policy change, or

$$y(p) = y_1 + (pD(p) - p_1D_1) - C(p), \tag{4.95}$$

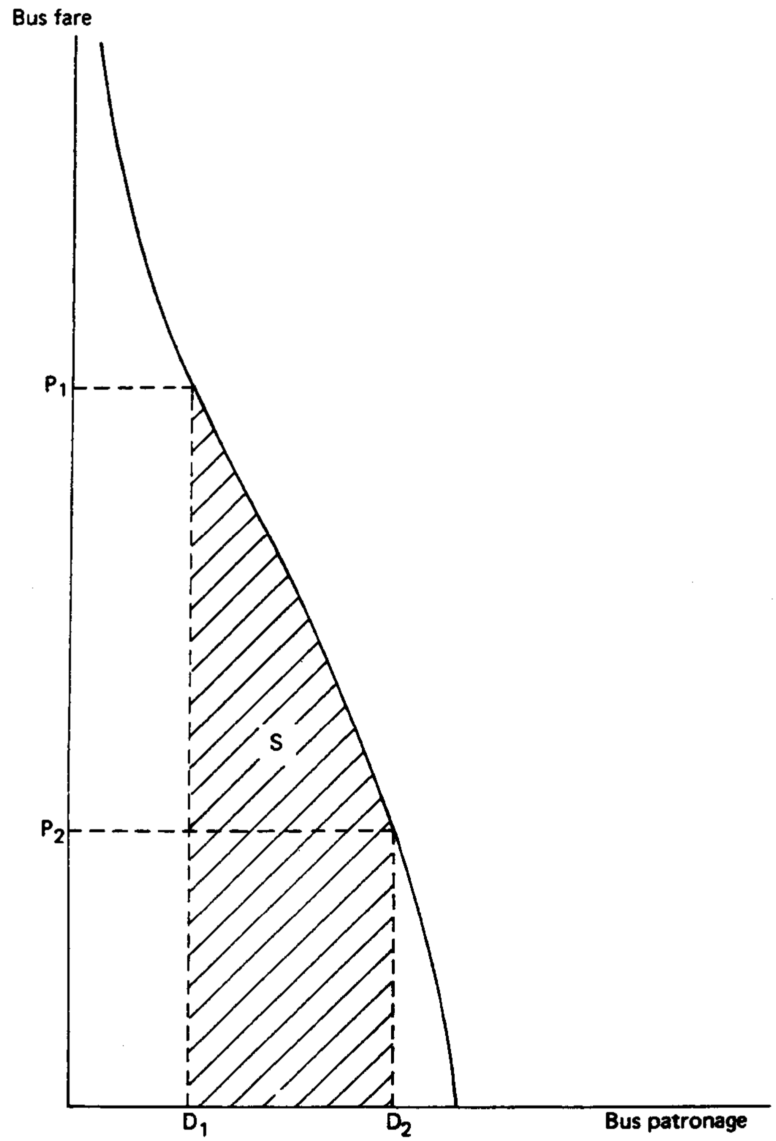


Fig. 4.6.

where $D(p)$ is market demand for transit at fare $p$ and $(pD(p) - p_1 D_1)$ equals the change in transit revenues. Then $y(p)$ or $y(p) - p$ of the numeraire commodity will be consumed when the non-transit and transit alternatives are chosen, respectively. The consumer will choose transit at fare $p$ if

$$\eta(x^0, s) - \eta(x^1, s) < \phi(x^1, s) - \phi(x^0, s) - p. \tag{4.96}$$

Letting $G$ denote the cumulative distribution function of $\eta(x^0, s) - \eta(x^1, s)$, the demand for transit is

$$D(p) = G(\phi(x^1, s) - \phi(x^0, s) - p). \tag{4.97}$$

Again consider the change in transit fare from $p_1$ to $p_2$ in fig. 4.6. The Marshallian consumer's surplus $S$, the shaded area in this figure, can be calculated as the area to the left of the demand curve between $p_1$ and $p_2$, plus the area $p_2 D_2$, less the area $p_1 D_1$, or

$$S = \int_{p_2}^{p_1} G(\phi(x^1, s) - \phi(x^0, s) - p)\mathrm{d}p + p_2 D_2 - p_1 D_1. \tag{4.98}$$

The utility-maximizing consumer will have a utility level

$$\begin{aligned} u(p) &= \mathrm{Max}[y(p) + \phi(x^0, s) + \eta(x^0, s), \\ &\qquad y(p) - p + \phi(x^1, s) + \eta(x^1, s)] \\ &= \phi(x^0, s) + \eta(x^1, s) + y(p) \\ &\qquad + \mathrm{Max}[\eta, \phi(x^1, s) - \phi(x^0, s) - p] \end{aligned} \tag{4.99}$$

for fare $p$, where $\eta = \eta(x^0, s) - \eta(x^1, s)$.

Define social welfare to be the sum of individual utility levels, i.e., the expectation of $u(p)$ with respect to $\eta(x^0, s)$ and $\eta(x^1, s)$. Since only the last two terms in eq. (4.99) depend on $p$, it is sufficient to index social welfare by the expectation of the sum of these terms, or

$$\begin{aligned} W &= y(p) + \int_{-\infty}^{+\infty} \mathrm{Max}[\eta, \phi(x^1, s) - \phi(x^0, s) - p]\, G'(\eta)\mathrm{d}\eta \\ &= y(p) + \lambda G(\lambda) + \int_{\lambda}^{\infty} \eta G'(\eta)\mathrm{d}\eta, \end{aligned} \tag{4.100}$$

where $\lambda = \phi(x^1, s) - \phi(x^0, s) - p$. Under the assumptions $E\eta = 0$ and

$$\int_{-\infty}^{0} G(\eta)\mathrm{d}\eta < +\infty,$$

satisfied, for example, by the logit and probit response curves, eq. (4.100) simplifies to

$$\begin{aligned} W &= y(p) + \lambda G(\lambda) + E\eta - \textstyle\int_{-\infty}^{\lambda} \eta G'(\eta)\mathrm{d}\eta \\ &= y(p) + \lambda G(\lambda) - \eta G(\eta)\big|_{-\infty}^{\lambda} + \textstyle\int_{-\infty}^{\lambda} G(\eta)\mathrm{d}\eta \\ &= y(p) + \textstyle\int_{-\infty}^{\lambda} G(\eta)\mathrm{d}\eta. \end{aligned} \tag{4.101}$$

Considering the change in welfare resulting from a fare change from $p_1$ to $p_2$, we obtain from either eq. (4.100) or eq. (4.101) the result:

$$\Delta W = y(p_2) - y(p_1) + \textstyle\int_{\lambda_1}^{\lambda_2} G(\eta)\mathrm{d}\eta, \tag{4.102}$$

where

$$\lambda_i = \phi(x^1, s) - \phi(x^0, s) - p_i,$$

or

$$\begin{aligned} \Delta W = y_1 &+ (p_2 D_2 - p_1 D_1) - C - y_1 \\ &+ \textstyle\int_{p_2}^{p_1} G(\phi(x^1, s) - \phi(x^0, s) - p)\mathrm{d}p = S - C. \end{aligned} \tag{4.103}$$

In conventional consumer theory the assumption that utility is linear in the numeraire commodity provides the theoretical justification for the measurement of benefits by consumer surplus. Under this same assumption, it follows from eq. (4.103) that this cost–benefit calculation provides a correct index of social welfare changes for the model of individual choice among discrete alternatives. Furthermore, this conclusion can be extended to the case of multiple markets. Hence these arguments can be applied to the transportation demand case in which different subpopulations face different fares (or differing vectors of attributes $x^1, x^2$), since they can be treated as engaging in distinct submarkets, and the consumer's surplus can be summed over the populations.

When the condition

$$\textstyle\int_{-\infty}^{0} G(\eta)\mathrm{d}\eta < +\infty$$

is met, as it is in logit analysis for example, the case of the introduction of a new alternative can be treated in the above framework by setting $p_1 = +\infty$, making this choice unavailable in the initial case.

The cost–benefit analysis for policy changes affecting aspects of travel other than transit price can be converted into the terms of the above analysis by treating the expression

$$\lambda = \phi(x^1, s) - \phi(x^0, s) - p$$

as the relative inclusive price of the non-transit trip. Any policy resulting in a change in $x^1$ and $x^0$ can be converted into an equivalent change in fare using this formula. When this is done, the previous calculation of the net benefit due to this price change provides a correct index of the corresponding welfare change.

The arguments of this section establish that when utility is linear in a numeraire commodity, the usual calculation of net benefits gives a valid index of potential welfare gains when all individuals are treated "equally". The linearity requirement has unrealistic implications that cast serious doubts on the validity of all cost–benefit calculations. However, it is of interest to note that all the empirical demand systems fitted in this study satisfy the linearity assumption. Thus, to the extent that these models are successful in depicting behavior, they provide a consistent starting point for cost–benefit analysis.