CHAPTER 2

ALTERNATIVE STRUCTURES FOR THE ESTIMATION AND FORECASTING OF URBAN TRAVEL DEMAND

Introduction

The preceding chapter introduced a model of the "comprehensively rational" decision-making household, where location of residence and job, auto ownership, frequency of trips for various purposes, destination, time-of-day of travel, and mode were all the result of choice of an overall lifetime plan. This model is unrealistic in two respects, one practical and one behavioral. The practical difficulty is that description and forecasting of lifetime household decisions would require data and a system structure of a level and complexity greatly exceeding current modeling capacity. The behavioral difficulty is that everyday experience and experimental evidence indicate that individuals find it impossible or impractical to process the information required to form comprehensive overall decisions. They instead use "boundedly rational" rules-of-thumb for decision-making on various pieces of their overall decision problem, without full integration of the pieces into a whole, or feedback to ensure complete compatibility of various decisions. Similarly, the multiple-person household typically fails to operate as an "organic" decision-maker, instead decentralizing many decisions.

Behavioral realities may, on one hand, help simplify the problem of the transportation demand modeler. To the extent that decisions such as residential location, auto ownership, and mode choice are made in isolation, or with weak feedback, the analyst can hope to model these aspects of choice separately, reducing the complexity of the model system to manageable proportions. On the other hand, there is no comprehensive theory of "boundedly rational" behavior to guide the development of demand models. In particular, the important role of expectations and perceptions in a world where information and calculating capacity are limited is poorly understood.

The demand analysis in this volume takes an intermediate position between the classical model of full rationality and thorough consideration of decision-making with limited information and processing. The overall decision problem of the household is broken into pieces that have a plausible correspondence to divisions of decisions that households utilize. On the other hand, choice within a particular sphere is assumed to conform to the classical rational model with decision-makers reacting to the objective attributes of the available alternatives. This approach has several important practical consequences. First, the complexity and range of a particular decision is reduced to a level compatible with the limitations of data collection, statistical analysis, and forecasting. Second, model structure is relatively straightforward, leading to a definition of utility that allows "trade-offs" between attributes of alternatives, and hence model sensitivity to the degree to which individuals will, in response to transportation policy shifts, be willing to substitute one attribute for another. Third, the approach provides a direct "causal" link from objective environmental variables to behavior, without requiring the definition or direct measurement of intervening variables such as expectations or perceptions. This is particularly important for practical forecasting, where the presence of any variables in the model that require direct measurement or monitoring at the level of the individual make the prediction process much more expensive, cumbersome, and time-consuming.

A final remark on the modeling strategy adopted here is that the final demand models are consistent with more broadly defined, and less "rational" behavioral rules than the classical decision theory used to motivate their structure would suggest. In particular, although we use the classical economic model of the utility-maximizing consumer to derive choice models, the resulting models are also consistent with "random utility" models from psychology, in which considerable scope for "irrational" behavior in the classical economic sense exists.

Components of Travel Demand

The range of transportation decisions identified in the preceding chapter included the locations of residence and job, automobile ownership, frequency of trips for various purposes, destination of trips, scheduling of trips, and mode of travel. These components of travel demand can be pictured in a structure such as that depicted in Figure 2. When the decisions in the figure are made jointly, the "feedbacks" denoted by dotted lines are strong. The more "boundedly rational" is behavior, and the less well integrated different decisions, the weaker the feedbacks.

The pattern of primary and secondary links may be different from that described in Figure 2. For example, mode-choice may "precede" destination-choice and trip-scheduling, in the sense that the mode-choice decision is largely determined by higher level decisions, and destination and trip-scheduling decisions are made conditional on mode-choice, with little feedback to the level of the mode-choice decision. Or, these decisions may be truly joint. The question of an appropriate model structure is empirical--the presence of feedback can be tested, and the goodness-of-fit of alternative structures to observations provides a way of eliminating unrealistic structures. However, model structure must also be guided by the practicalities of available data.

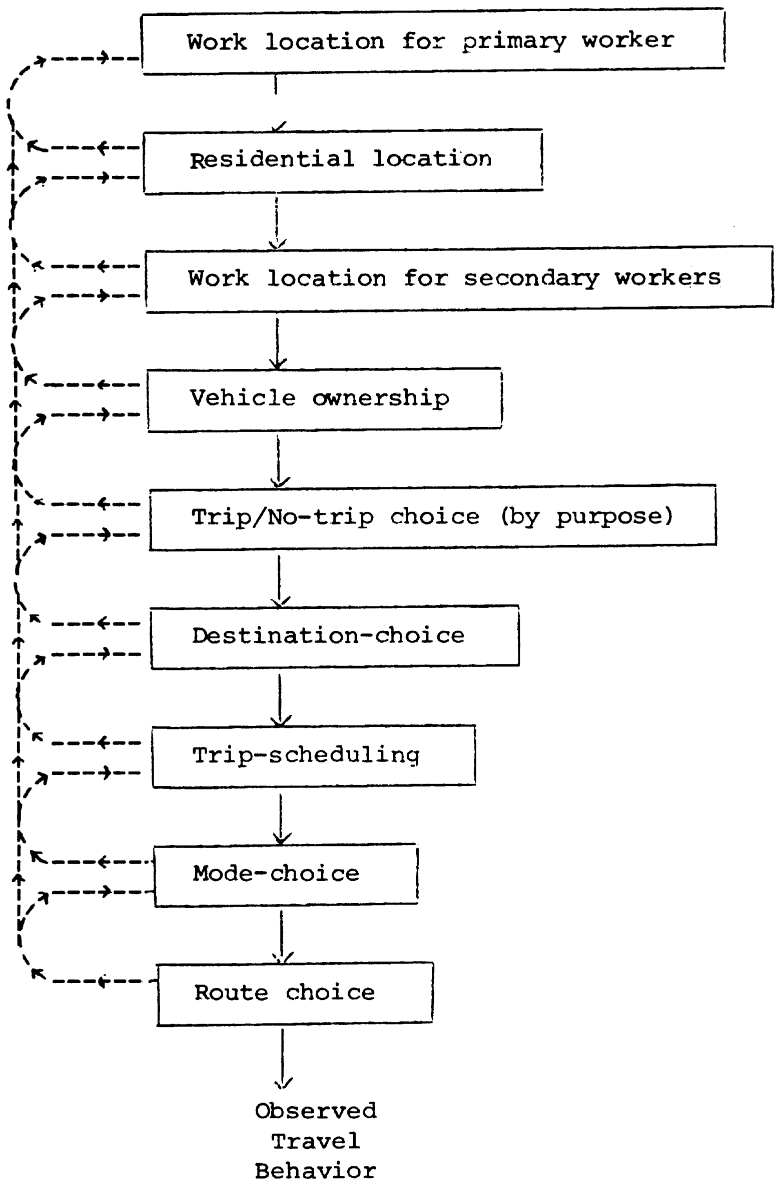
Work location for primary worker
Residential location
Work location for secondary workers
Vehicle ownership
Trip/No-trip choice (by purpose)
Destination-choice
Trip-scheduling
Mode-choice
Route choice
Observed
Travel
Behavior

Figure 2

Factorization of the Demand System

We have stressed "bounded rationality" as a motivation for the division of the overall transportation decision into manageable sub-choices. However, "tree" decision structures can also be derived within the classical framework of utility maximization by postulating a separable utility structure. Domencich and McFadden (1975, p. 39) discuss the method of factoring the transportation demand function into manageable components by postulating separable entities; the following paragraphs are extracted from their book:

“To examine this method, let us consider the hypothetical question: 'If you were making a round trip from home to γ at time β on day π, what mode would you use?' Our hypothesis is that the answer to this question depends only on the attributes of the different available modes and is independent of the time of day the trip is made, the destination of the trip or the number of trips that are made for this purpose each day. By this we mean that for the given purpose, the choice of mode depends on the comparison of times and costs of travel by the different modes (the attributes of the modes), while the evaluation of a given minute of travel time or dollar of cost is independent of when or where the trip is made or the number of trips that are made for that purpose each day. We hypothesize that it is also independent of other prices in the economy. (However, we would expect the answer to depend on socioeconomic characteristics like income, family size, and auto ownership.)

"We assume that attribute vectors of alternatives can be used to index consumption activities directly. ...

... Thus utility $U(x,s)$ is expressed as a function of a vector of socioeconomic characteristics s and a vector x of attributes of each alternative, with a set B of available alternatives. Suppose the hypothetical question above has two possible responses, modes a and b. The vector of attributes x can be partitioned into subvectors $x_{(1)}$ and $x_{(2)}$, such that $x_{(2)}$ is the same for the choices a and b, while $x_{(1)}^{a}$ and $x_{(1)}^{b}$ differ. The factorization hypothesized above requires

$$U(x_{(1)}^{a},x_{(2)},s) \geq U(x_{(1)}^{b},x_{(2)},s) \quad , \tag{3.3}$$

if and only if

$$U(x^a_{(1)},x'_{(2)},s) \geq U(x^b_{(1)},x'_{(2)},s) \quad , \tag{3.4}$$

A sufficient condition for this property is that U have the additively separable form[1]

$$U(x^a_{(1)},x_{(2)},s) = \psi(\varphi^1(x_{(1)},s) + \varphi^2(x_{(2)},s)) \quad . \tag{3.5}$$

Then, a is chosen over b if

$$\varphi^1(x^a_{(1)},s) > \varphi^1(x^b_{(1)},s) \quad . \tag{3.6}$$

Since $x_{(1)}$ is a "relatively short" vector, and the set of alternative $x_{(1)}$ subvectors for x in B is "relatively small," this structure greatly simplifies econometric analysis.

"We would not expect the additivity hypothesis to hold strictly. For example, the relative utility associated with a and b may differ depending on the commitment of family vehicles specified in $x_{(2)}$, or on levels of fatigue resulting from other trips specified in $x_{(2)}$. Further, the effects of varying transport time and cost on the consumer's expenditure and time budget constraints will, in general, introduce interaction effects in the demand functions, even when the utility function is separable. Despite these exceptions, additive separability seems a good general working hypothesis.

"The assumption that the marginal rates of substitution between modal attributes are constant for a given trip purpose, regardless of the time of day, trip destination or travel frequency seems quite plausible. One can, for example, envision a traveler evaluating the trade-off between a minute spent waiting for a bus and a minute of bus line-haul time differently in the peak as compared to the

"[1]When there are three or more subvectors of x, each with the independence property above, additive separability is also necessary. See Debreu (1960b, theorem 3, p. 21). The function ψ must be monotone increasing.

off-peak, but the difference would probably stem from the uncertainty associated with waiting and perhaps the greater need to be punctual in the peak. However, if this were the case, the list of modal attributes should also include measures of the need to be punctual (e.g., schedule delay), in which case there would no longer be reason to expect the weights assigned to each attribute to be different for the two time periods.

"With regard to the effect of overall income and time constraints, we note that the importance of these effects depends on the proportion of total income and time spent on transport. These proportions are likely to be sufficiently low to allow us to ignore the overall effects of changes in the cost and time of trips. In the special case that utility can be written as a function linear in numéraire commodity, this commodity will "absorb" all income effects and the additive separability will carry over to the demand functions, even if transport expenditures are a substantial proportion of income. ...

...

...

"A further observation on the structure of the utility function can be made by recalling that the function $U(x,s)$ of attributes of alternatives is derived from a more basic utility function defined over levels of satisfaction of fundamental wants. Consequently, it will usually be the generic attributes of an alternative that matter to the individual, and not the specific "labeling" of the alternative. For instance, walking time is walking time regardless of the principal mode used for the trip. This structure allows a further reduction in the complexity of the description of alternatives.

"By contrast to the hypothetical question posed above, in which the mode choice is indicated *ceterus paribus* for the remaining environment, one could ask a question of the form: 'If you had to make a round trip to γ on day π, at what time of day β would you go, given that you can choose your preferred mode (a or b) ?' The response to this question is conditioned on an optimal choice along another travel dimension. Thus, one would expect this choice to depend not only on the socioeconomic factors and destination attributes, such as work and child-care schedules and size of crowds, but also on the attributes of each mode at each alternative travel time.

"We now assume, analogously to the previous case, that the consumer's trade-off between such factors is not influenced by the mode of travel or day of the trip. For example, the degree of additional crowding an individual would accept at a shopping destination in exchange for lower goods prices should be independent of the mode used to reach the destination when the attributes of the trip remain the same. One can envision a traveler finding a crowded store more objectionable because he rode a crowded bus to get to the store, but we assume such instances of interdependence are rare or that their effects are negligible. This assumption holds if utility is now assumed to have the separable form

$$U(x_{(1)},x_{(2)},x_{(3)},s) = \psi\left[\sum_{i=1}^{3}\varphi^{i}(x_{(i)},s)\right] , \qquad (3.7)$$

where $x_{(1)}$ describes the attributes of the trip, $x_{(2)}$ the attributes of the destination at each time of day, and $x_{(3)}$ the remainder of the environment.

"Suppose, as before, that two modes, a and b, are available; and now suppose two choices of time of day, peak (p) and non-peak (n), are offered. The subvector $x_{(1)}$ will vary with both mode and time of day; e.g., $x_{(1)}^{ap}$ will describe the attributes of a trip by mode a at the peak. The subvector $x_{(2)}$ will vary with time of travel, $x_{(2)}^{p}$ and $x_{(2)}^{n}$, but not with mode, while $x_{(3)}$ will be the same for all choices. The response to the hypothetical question above will be a choice of a peak trip if

$$\varphi^{2}(x_{(2)}^{p},s) + \max_{i=a,b} \varphi^{1}(x_{(1)}^{ip},s) > \varphi^{2}(x_{(2)}^{n},s) + \max_{i=a,b} \varphi^{1}(x_{(1)}^{in},s) \quad .$$

The important observation to be drawn from this formula is that all the attributes of the trip are summarized in the 'index' of desirability $\varphi^{1}(x_{(1)},s)$, whose form could be determined from the first hypothetical question. Thus, the analysis of time of day of trip can proceed with the use of this previously determined 'index' rather than with the much more extensive 'raw' data on trip attributes. This saving can obviously be extremely valuable in facilitating econometric analysis."

The Short-Range-Generalized-Policy (SRGP) Model

The Urban Travel Demand Forecasting Project has, by agreement with the BART Impact Study of the Metropolitan Transportation Commission of the San Francisco Bay Area, limited its investigations to work travel, with primary emphasis on mode-choice, and decisions directly related to mode-choice, such as auto ownership and trip-scheduling. In general, policy analysis requires a more complete description of travel demand, including non-work travel. In order to carry through policy studies for the San Francisco Bay Area, this project has incorporated its work demand models in a general disaggregate demand system developed by Cambridge Systematics for the Metropolitan Transportation Commission, and called the Short-Range-Generalized-Policy (SRGP) system. This system is illustrated in Figure 3. Each block represents a constellation of decisions modeled separately from the remaining decisions, except for the transmission of the impacts of one block on another through environmental variables. For example, the auto ownership decision is influenced by general environmental variables, by work-mode-choice, and by the assessibility of non-work destinations by alternative modes, but does not depend directly on non-work trip frequency. Work-mode-choice is influenced by auto ownership decisions, and in turn influences non-work trip frequency and auto ownership.

Note that the SRGP model excludes job and household location decisions, trip-scheduling, route-choice, and some of the more complex characteristics of trips, such as multiple-stop tours. Future development of disaggregate behavioral systems, particularly for long-run policy forecasting, should integrate these aspects of demand into the forecasting model.

A variety of possible alternatives to the SRGP structure may seem plausible to the reader. For example, non-work travel decisions may directly influence work-mode-choice, or the work mode-choice decision may be made jointly with the decision of scheduling of work trips. To a considerable extent, a structure such as that in Figure 3 is sufficiently flexible to accommodate these sort of modifications, with the various feedbacks between modules changing importance. It is also possible to test empirically the validity of alternative structures, as has been done by Ben Akiva (1974). This topic has not been pursued in this study.

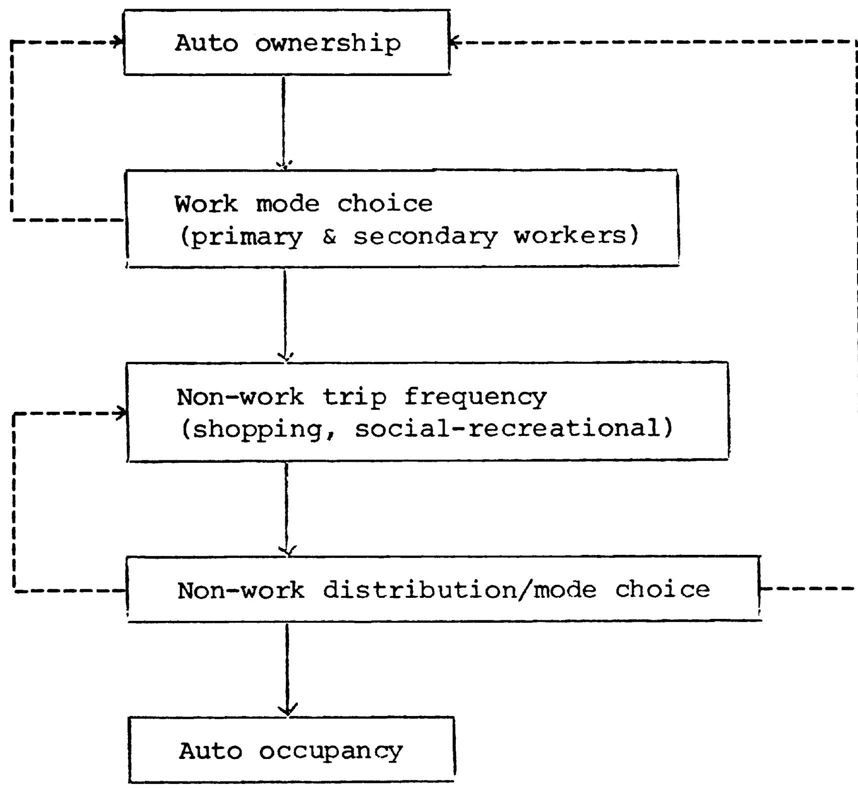

Figure 3: Structure of SRGP Model

Attitudinal and Objective Data

Transportation demand analysis has two objectives, which are complementary in principle, but often conflict in practice. The first is the desire to gain an understanding of the determinants and structure of behavior--to "explain" the nature of travel demand. The orientation of this objective is toward the inner structure of the household or individual, and the psychological mechanisms of behavior. The second objective is policy-oriented demand forecasting, where the emphasis is on the causal relationship between readily obtained objective measurements and travel behavior.

These objectives are complementary in the sense that an understanding of the mechanisms and structure of behavior is fundamental to the construction of reliable forecasting models, and to understanding the domain of validity of forecasting systems. On the other hand, forecasting validity provides an objective external test for the constructs used to explain the process of choice.

Conflict between the objectives comes when analysts try to serve both masters within a single model system, or where one objective is pursued under the guise of attacking the second. This conflict has been evident in transportation demand analysis in the perceived contest between "attitudinal" and "objective" models. On the whole, models employing attitudinal variables provide a tool for exploring the decision-making process, and identifying objective variables that are important determinants of behavior. However, such models are poorly suited for policy forecasting, due to the need for field measurement of attitudes, which can be costly and time-consuming. On the other hand, objective models treat the decision-maker as a "black box," and are often poorly suited to revealing the mechanisms of the decision process. In particular, it may be difficult to achieve environmental settings in which the mapping from stimulus to response can be used alone to identify internal structures.

To the extent that travel behavior is influenced by long-run dynamic forces of experience and habit, forecasting models based on contemporary objective data will be inaccurate. Then, measures of perceptions and attitudes may summarize the decision-maker's objective history more efficiently than any fully external measures, and their introduction--in spite of costs of measurement--may be the most effective way to improve forecasting performance.

The most satisfactory way to incorporate attitudinal variables into a forecasting system would be to make changes in the attitudinal variables endogenous to the model; i.e.,

$$\begin{matrix}\text{Current}\\ \text{Behavior}\end{matrix} = f\begin{bmatrix}\text{Objective} & \text{State of attitudes}\\ \text{environment} , & \text{and perceptions}\end{bmatrix};$$

$$\begin{matrix}\text{Change in State of}\\ \text{Attitudes and Perceptions}\end{matrix} = g\begin{bmatrix}\text{State of attitudes} & \text{Current environ-}\\ \text{and perceptions} , & \text{ment and behavior}\end{bmatrix}.$$

Base-line measurement of the state of attitudes and perceptions would be required, but then for policy forecasting the mapping g would predict attitudes as a function of predetermined data and objective variables, eliminating the need for continual monitoring and field collection of attitudinal data.

Aggregation of Alternatives

The ideal disaggregate model, with every distinguishable alternative from the standpoint of the decision-maker treated as distinct by the analyst, is usually impractical in terms of data collection and statistical analysis. Instead, the demand system is expressed in terms of reasonable aggregates of alternatives. For example, in mode-choice the alternative "bus" may contain several alternative bus lines, while the alternative "carpool" contains alternative routes and pooling arrangements.

The section of this chapter on factoring demand gives a method for describing the attributes of composite alternatives when data is available on each of the components and the form of utility governing choice among component alternatives is known. This approach is implemented in sequential logit models of the choice process, discussed in Part IV, Chapter 2. In this section, we consider the problem of dealing with composite alternatives when component data is not observed, or the utility for choice among components is unknown.

Consider the problem where data on selected component alternatives can be observed, but enumeration of all alternatives is impractical. When a multinomial logit choice model is employed, McFadden (1976e, 1977a) has shown that consistent statistical estimates of the utility function can be obtained using fixed or random subsets of the set of all available alternatives. McFadden (1976e) also discusses forecasting from subsets of the full alternative set--this problem has a less satisfactory solution and deserves further study.

Consider further the problem where component alternative data is available, at least for a sample of alternatives within a group. Suppose there are J choice groups, indexed $J = 1,...,J$, and M_j components within each group, indexed $m = 1,...,M_j$. Each component alternative jm has attributes y_j common to the group and attributes x_{jm} specific to the sub-component. The (unknown) utility function for jm is assumed to have the form $v_{jm} = \alpha' y_j + \beta' x_{jm}$. By convention, we can absorb the group means of attributes into y_j , and thus assume the x_{jm} have group means equal to zero. We shall assume that the choice probabilities for component alternatives satisfy a generalization of the multinomial logit model, the generalized extreme value model discussed in Part IV, Chapter 2:

(1) $$P_{jm} = e^{\frac{v_{jm}}{1-\sigma_j}} \left[\sum_{m=1}^{M_j} e^{\frac{v_{jm}}{1-\sigma_j}} \right]^{-\sigma_j} \Big/ \sum_{\ell=1}^{J} \left[\sum_{n=1}^{M_\ell} e^{\frac{v_{\ell n}}{1-\sigma_\ell}} \right]^{1-\sigma_\ell} ,$$

where σ_j is a measure of the degree to which component alternatives within a qroup are perceived as similar by individuals. The parameters σ_j vary between zero and one, with $\sigma_j = 0$ corresponding to the MNL case where unobserved parts of the utilities of component alternatives are uncorrelated, and $\sigma_j = 1$ corresponding to the case where they are perfectly correlated. From (1) , the choice probability for a group is

$$(2) \qquad P_j = e^{\alpha' y_j + w_j} \Big/ \sum_{\ell=1}^{J} e^{\alpha' y_\ell + w_\ell} ,$$

where

$$(3) \qquad w_j = \log \left[\sum_{m=1}^{M_j} e^{\frac{\beta' x_{jm}}{1-\sigma_j}} \right]^{1-\sigma_j} .$$

If there were no variation in the attributes of component alternatives, so that $x_{jm} = 0$, then $w_j = (1 - \sigma_j) \log M_j$, reflecting the average contribution to the utility of the group of the number of components in the group. Then, more components in the group increases the possibility that some component alternative has unobserved attributes making it attractive to the decision-maker. This particular model structure has been studied empirically and interpreted by Lerman (1975).

When the x_{jm} in (3) are not zero, the convexity of the exponential implies

$$\sum_{m=1}^{M_j} e^{\frac{\beta' x_{jm}}{1-\sigma_j}} \geq M_j ,$$

and hence $w_j \geq (1 - \sigma_j) \log M_j$, with the difference in the two sides of the inequality depending on the variances of the x_{jm} . One limiting case of (3) that is of interest occurs when the number of component alternatives within a group is large, and the x_{jm} behave as if they were independently identically normally distributed with covariance matrix Ω_j . Then,

$$(4) \qquad E\, e^{\frac{\beta' x_{jm}}{1-\sigma_j}} = e^{(1/2)\beta' \Omega_j \beta / (1-\sigma_j)^2}$$

and

$$(5) \qquad \frac{1}{M_j} \sum_{m=1}^{M_j} e^{\frac{\beta' x_{jm}}{1-\sigma_j}} \underset{\text{a.s.}}{\rightarrow} e^{(1/2)\beta'\Omega_j\beta/(1-\sigma)^2} \quad .$$

Hence,

$$w_j \approx (1-\sigma_j)\log M_j + (1/2)\beta'\Omega_j\beta/(1-\sigma) \ .$$

For example, if $\sigma_j = \sigma$, $M_j = \theta_j M$, and the number of component alternatives is large, (2) becomes

$$P_j = \frac{\exp\left\{\alpha' y_j + (1-\sigma)\log\theta_j + (1/2)\beta'\Omega_j\beta/(1-\sigma)\right\}}{\sum_{\ell=1}^{J}\exp\left\{\alpha' y_\ell + (1-\sigma)\log\theta_\ell + (1/2)\beta'\Omega_\ell\beta/(1-\sigma)\right\}} \quad .$$

When the components x_{jm} are not observed, but their distribution can be approximated or estimated, and β is known, then the model (7) can be estimated and employed for forecasting using standard MNL models. If θ_j is unobserved, then it can also be estimated using the MNL model. Note, however, that when $\alpha' y_j$ contains an alternative-specific dummy and θ_j is unobserved, the alternative-specific coefficient and the term $(1 - \sigma)\log\theta_j$ are unidentified. This suggests one interpretation of alternative-specific coefficients as indicating, in part, the number of "equivalent" component alternatives contained in the group. When β is not known, direct maximum likelihood estimation of (7) could be carried out with modification of a MNL estimation program to handle quadratic parameters in the "apparent" utility function. Alternately, writing out the quadratic form in (7) and ignoring quadratic restrictions across parameters permits consistent estimation of all the parameters in (7) with a standard MNL program.

A common practice in the disaggregate transportation demand literature when confronted with models of the form in (2) and (3) when β is unknown has been to define "composite" attribute variables

$$(8) \qquad x_j^k = \log \sum_{m=1}^{M_j} e^{x_{jm}^k} \quad ,$$

where $k = 1,\ldots,K$ indexes individual variables and the sign of x_{jm}^{k} is defined so that "more is better;" then consider the model

$$(9) \qquad P_j = e^{\alpha' y_j + \beta' x_j} \sum_{\ell=1}^{J} e^{\alpha' y_\ell + \beta' x_\ell} \quad .$$

Because (8) weighs most heavily the most attractive value of a variable among the component alternatives, this method will tend to over-represent the desirability of a group unless there tends to be a dominant component alternative. When different variables are negatively correlated (e.g., travel time and travel cost across modes for specified trips), this over-representation will be particularly strong. Hence, the approximation (8) - (9) does not seem satisfactory. However, when the variance of x_{jm} is small, the approximation error <u>may</u> be acceptable.

Model Complexity

One of the motivations for the development of disaggregate demand models was dissatisfaction with the cost and complexity of traditional modeling systems. Some of the policy applications of disaggregate analysis to sketch planning have achieved remarkable simplicity and flexibility (Cambridge Systematics, Inc., 1976). On the other hand, the desire for comprehensive demand systems capable of providing forecasts to a full range of policy scenarios has led to the development of increasingly complex model systems. The logical conclusion of this line of development will be behavioral models that simulate all the individual transportation choices of a population. The complexity and cost of such a simulation package will probably rival or exceed the cost of traditional models. The use of models of this complexity may be required for comprehensive transportation policy analysis, particularly for delicate policy choices where indirect impacts are critical. On the other hand, many policy questions are sufficiently limited in scope, or clear-cut in results, to make elaborate analysis unnecessary. For these problems, what is needed are simpler models that provide bounds on the impacts of policy. What would be desirable, then, would be a hierarchy of behavioral models, ranging from simple to complex, with an understanding (based on analysis of more complex structures) of the limits of accuracy and range of application of models at each level. In this manner, each policy issue can be analyzed within a model framework of a complexity and cost commensurate with the scope and delicacy of the problem.