NISS

Generation of Synthetic Daily Activity-Travel Patterns: Outline of the Approach

Ryuichi Kitamura

Technical Report Number 39
October, 1995

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

GENERATION OF SYNTHETIC DAILY ACTIVITY-TRAVEL PATTERNS: OUTLINE OF THE APPROACH[1]

Ryuichi Kitamura
August, 1995
Revised October, 1995

1. INTRODUCTION

Achieving desired levels of accuracy in the outcome of travel demand forecasts produced by micro-simulation of household behavior may require a large sample of households. This may happen when: high levels of spatial or temporal resolution are required of the outcome; sample households do not have a desirable geographical distribution; demand by a small population segments is desired; or a high level of accuracy is desired. In such instances the number of households available in the data set at hand may not be sufficiently large and generation of synthetic households may be required. When the micro-simulation expects daily travel patterns of household members as input data, it calls for generation of synthetic daily travel patterns.

An approach to the problem of synthetic travel pattern generation is proposed in this report. The approach adopted here is sequential. The proposed model system can be decomposed into components to which certain aspects of observed activity-travel behavior correspond. This establishes a link between the mathematical models and observational data. The model components are each relatively simple and can be estimated using commonly used estimation methods and existing data sets.

The problem of synthetic travel pattern generation is first formulated and presented formally in Section 2. The knowledge that has been accumulated on the characteristics of daily travel patterns is briefly reviewed in Section 3. Following this, Section 4 is devoted to the discussion of the relative advantages of sequential and simultaneous modeling approaches. The formulation of the model system and its components is described in Section 5. Significant portions of Section 5 are dedicated to the discussion of behavioral and statistical issues associated

---

[1]Funding for this research was provided by the National Science Foundation through a grant, DMS 9313013, to the National Institute for Statistical Sciences, and by the Los Alamos National Laboratory as part of the Federal Highway Administration's Travel Model Improvement Program.

with modeling daily activity-travel patterns. Section 6 offers an explorative analysis of history dependence in activity engagement. Section 7 is a conclusion.

2. PROBLEM DESCRIPTION

Consider a household member, i, whose daily activity-travel pattern can be characterized as

$$(\mathbf{X}_i, \mathbf{T}_i, \mathbf{L}_i) = (X_{i1}, X_{i2}, \ldots, X_{in}; T_{i1}, T_{i2}, \ldots, T_{in}; L_{i1}, L_{i2}, \ldots, L_{in}) \quad (1)$$

where

X_{ij} = the type of the j-th activity pursued by i,

T_{ij} = the duration of the j-th activity pursued by i,

L_{ij} = the location of the j-th activity pursued by i (if the activity is travel, then L_{ij} refers to the destination of trip j; in this case $L_{ij} = X_{i,j+1}$), and

n = the number of activities involved in i's daily activity-travel pattern.

Note that travel is included here as one of the activity types. Also note that travel mode, which may be stored in another vector, say $\mathbf{M}_i$, is not included in the system here. The discussions of this report will be limited to general discussions of travel mode choice in daily travel behavior. A framework to introduce travel mode into the stochastic process mode system described in this report will be presented separately in another report.

The generation of a synthetic daily activity-travel pattern implies generating vectors $\mathbf{X}_i$, $\mathbf{T}_i$ and $\mathbf{L}_i$ given:

1. attributes of the individual i,
2. attributes of the household to which i belongs,
3. residence and work location of i,
4. demographic and socio-economics characteristics of the region,
5. land use characteristics of the region, and
6. transportation network and travel time characteristics of the region.

Since it is most likely that synthetic activity-travel patterns will be generated for synthetic individuals and households, items 1 through 3 will comprise synthetic data. Generating synthetic individuals and households is, however, outside the scope of this report; it is assumed here that all personal and household attributes, as well as work location, are known for i. The latter three items will consist of projected values in cases where synthetic activity-travel patterns are generated for forecasting.

3. Accumulated Knowledge

The following discussions offer a brief summary of what is know about n, which also is a variable to be determined, and each of the three vectors, $\mathbf{X}_i$, $\mathbf{T}_i$, and $\mathbf{L}_i$.

n: It is often said that the total number of activity episodes captured in time use surveys tends to be 20 to 25 per person per day, including travel activities. More detailed results, e.g., how this total varies across sample sub-groups, need to be obtained through a survey of the literature in the time use analysis field.

In the transportation field, the number of trips, which makes up a part of n, is known to be 4 to 5, per person per day. It is known that the number of trips captured varies greatly depending on the survey instruments and survey administration. It has been well established that the total number of trips is associated with age with mid-age individuals as a group making the largest number of trips per day.

$\mathbf{X}_i$: There are certain regularities in the sequence with which individuals engage in different types of activities. For example, one may anticipate that the sequence of activities performed before leaving home for work or after coming back home from work, is fairly uniform across individuals. Again, the literature in time use analysis needs to be explored to determine what tendencies have been found for activity sequences involving both in-home and out-of-home activities.

Kitamura (1983) examined the sequence of trip purposes using standard trip diary data from Detroit. The trip purpose was used to identify the primary out-of-home activity type at each destination location. No information on in-home activities was available in the data set. The analysis examined how out-of-home activities were sequenced in a home-based trip chain, i.e., the home-to-home series of trips which involves one or more stops to pursue activities. The results indicated the tendency that *activities of more mandatory nature tend to be pursued first in a trip chain.* The sequencing tendencies indicated the following hierarchy:

- work and school, work-related
- chauffeuring
- personal business (e.g., banking, dental and medical)
- shopping
- social and recreational

The presence of the same sequencing hierarchy was later found for activities throughout the day (Kitamura & Kermanshah, 1983, 1984). Another important tendency is that the activities pursued in the same trip chain tend to be similar (Kitamura, 1983).

$\mathbf{T}_i$: Several studies investigated the duration of activity. In a semi-Markov process model of trip chaining, Lerman (1979) used gamma distributions for the duration of sojourns at destination locations. Survival models have recently been applied to the time dimension in activity-travel patterns (e.g., Mannering et al., 1993; Niemeier & Morita, 1995). These studies are based on the assumption that durations of successive activities are independent.

In addition, it is expected that a substantial body of literature is available from the time use research community on the duration of activity episodes. Again, this literature needs to be explored in the course of the project.

Activity durations have been examined from the viewpoint of resource allocation. A theoretical model of activity duration can be found in Kitamura et al. (1995) where the duration of an activity episode was analytically derived while assuming that the total daily activity pattern is optimized and that each activity episode has a logarithmic utility function. The model was estimated using a time use data set from the United States. Although the model is based on the assumption that the daily time use is optimized as a whole, the resulting model applies to each activity episode. Golob and McNally (1995) examined the allocation of time to different activity types using a structural-equations model system. This approach facilitates the inference of causal relationships that exist among activities of different types.

Critical in the analysis of activity durations is the correlations across the durations of respective activity episodes. Since the total amount of time available is fixed at 24 hours a day, in general negative associations can be expected. At the same time, the duration of each episode is also a function of n, the total number of episodes. The inter-relationships among durations of different types of activities and the number of activities, n, need to be explored in the future.

$\mathbf{L_i}$: Non-home activity locations have traditionally been estimated using the gravity model of spatial interaction, formulated after the law of gravitation. The multinomial logit model of destination choice can be viewed as a special case of the gravity model family. In principle, these models depict that, *ceteris paribus*, more intense interaction exists between a pair of locations that are closer to each other, and the intensity of the interaction is positively related to the attraction level of the destination and the number of trips originating from the origin.

One important issue is how to characterize destination choice for non-home-based trips, i.e., trips whose origin and destination are both non-home. For home-based destination choice underlying a simple trip chain involving only one stop (i.e., home-activity-home), the only spatial element to be considered is the separation between the destination and the home base. In case of non-home-based choice, this is not the case. For example consider the choice of a shopping opportunity on the way home from work; in this case the home location and the deviation from the regular commute route would be important considerations. Kitamura and Kermanshah (1984) constructed a non-home-based destination choice model which included both the usual origin-to-destination travel time, t_{ij}, and the destination-to-home travel time, t_{jh}, in a multinomial logit choice model. Their estimation results offered a clear indication that t_{ij} and t_{jh} are equally important for non-home-based destination choice. This finding is readily applicable to the generation of synthetic activity-travel patterns.

There are numerous studies on travel mode choice. Most studies, however, are seriously limited because they are trip-based, i.e., they analyzed each trip separately in isolation from other trips. For example, one may choose to commute by car because a car is needed for work. Then this mode choice behavior cannot be explained by just looking at the home-to-work commute trip and comparing the attributes of the travel modes available for the trip.

One of the critical requirements in the generation of synthetic activity-travel patterns is to observe the constraints imposed on the transition between travel modes. For example, transition from public transit to driving alone is in general not possible unless the transition takes place at the home or work base where a private car had been placed. For a trip chain that originates and terminates at the home base, the sequence of travel modes tends to be governed by the boundary condition that the mode of the first trip from home is identical to that of the last trip to home. These regularities and tendencies serve as a set of constraints in the generation of activity-travel patterns.

4. SEQUENTIAL VS. SIMULTANEOUS APPROACHES

There are two broad classes of approaches to the generation of synthetic activity-travel patterns: sequential (incremental) approaches vs. simultaneous (holistic) approaches. The former approaches adopt rules to generate, one by one, the activity that will immediately follow, given the history of activity generation so far. The latter approaches, on the other hand, deploy behavioral paradigms that are each concerned with the entire daily activity-travel pattern.

One paradigm for the latter approaches is that an individual of given attributes has a probability vector that depicts the likelihoods with which he or she will exhibit respective activity-travel patterns. A study by Pas (1983) is readily applicable to operationalize this paradigm. Another paradigm is utility maximization: the individual chooses that activity-travel pattern, from among a set of all feasible patterns, which has the maximum utility. Studies based on this assumption include Adler and Ben-Akiva (1979), Recker et al. (1986), and Recker (1995). The two paradigms can be integrated to produce probabilities for alternative daily activity-travel patterns.

One important advantage of sequential approaches is the ease of implementation they offer. The size of the problem to be handled at a time is much smaller because a daily pattern is synthesized incrementally. Simultaneous approaches, on the other hand, have theoretical elegance. They can be expected to be more sensitive to parameters that characterize the travel environment. In addition, simultaneous approaches can better reflect individuals' planning effort.

Despite the advantages offered by simultaneous modeling approaches, it is proposed that a sequential approach be taken in this study. There are three major reasons.

Practicality: The first is practicality. When viewed as an optimization problem, daily activity-travel behavior is very complex (Pas, 1990). Exact formulation of this behavior produces an overwhelmingly complex mathematical problem. Even when the behavior is viewed as a discrete choice behavior with a choice set of feasible activity-travel patterns, the number of possible patterns would be astronomical because of the time dimensions brought into the problem by the inclusion of activity durations and activity starting times (theoretically speaking, each activity offers infinitely many possible starting times and durations, as time is continuous).

Behavioral Basis: The second reason is behavioral. As Simon noted when he proposed the paradigm of satisficing, a person is not capable of enumerating all possible alternatives or discerning minute differences among them. Furthermore, more often than not the person will not have all the information associated with the alternatives. It rests on a very dubious behavioral basis to assume that an observed activity-travel pattern is an optimum pattern selected from among a set of theoretically infinitely many alternative patterns.

Irrelevance of Policy Sensitivity: The third reason concerns what exactly is needed from synthetic activity-travel patterns. As noted earlier, simultaneous approaches can better represent travelers' planning efforts and therefore can be more sensitive to changes in the travel environment, leading to more policy-sensitive models. The superior policy sensitivity offered by simultaneous approaches, however, is not of much importance in the context of this study. Indeed what is desired is that synthetic activity-travel patterns generated truthfully represent associations among all pertinent variables, not policy sensitivity.

5. FORMULATION OF A SYNTHETIC ACTIVITY-TRAVEL PATTERN GENERATION PROCEDURE

The sequential approach is based on the identity that, given n,

$$\begin{aligned}\Pr[\mathbf{X}_i,\mathbf{T}_i,\mathbf{L}_i] &= \Pr[X_{i1}, X_{i2}, \ldots\ldots, X_{in}; T_{i1}, T_{i2}, \ldots\ldots, T_{in}; L_{i1}, L_{i2}, \ldots\ldots, L_{in}] \\ &= \Pr[X_{in}, T_{in}, L_{in} | X_{i1}, \ldots\ldots, X_{i,n-1}; T_{i1}, \ldots\ldots, T_{i,n-1}; L_{i1}, \ldots\ldots, L_{i,n-1}] \\ &\quad \Pr[X_{i,n-1}, T_{i,n-1}, L_{i,n-1} | X_{i1}, \ldots\ldots, X_{i,n-2}; T_{i1}, \ldots\ldots, T_{i,n-2}; L_{i1}, \ldots\ldots, L_{i,n-2}]\end{aligned}$$

$$..... \Pr[X_{i1}, T_{i1}, L_{i1}]. \quad (2)$$

Each probability on the right-hand side can be formulated as a model for activity type, location and duration, given the past history of activity and travel. As noted above, knowledge has been accumulated on characteristics of activity-travel behavior, which can be incorporated in this modeling effort.

5.1. Modeling Issues

There are several major modeling issues to be addressed in the effort.

Decomposing the X-T-L Triple: Each activity (or travel) episode is characterized by X_{ij}, T_{ij} and L_{ij}. These three may be treated simultaneously, which however would lead to the same enumeration problem discussed earlier because T_{ij} is continuous. Thus, one may use one of the following permutations:

$$\begin{aligned}
&\Pr[X_{ij}, T_{ij}, L_{ij}|\underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \\
&\quad = \Pr[L_{ij}|X_{ij}, T_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]\Pr[T_{ij}|X_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]\Pr[X_{ij}| \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \\
&\quad = \Pr[T_{ij}|X_{ij}, L_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]\Pr[L_{ij}|X_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]\Pr[X_{ij}| \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \\
&\quad .....
\end{aligned} \quad (3)$$

etc., where $\underline{\mathbf{X}}_{i,j-1} = (X_{i1}, X_{i2}, ..., X_{i,j-1})$, etc.

Since all permutations of X_{ij}, T_{ij} and L_{ij} lead to the same joint probability, the model's replication capability should not depend on which permutation is adopted. Therefore that permutation which can be theoretically supported and/or which offers most modeling flexibility and sensitivity can be selected.

Exploration of History Dependence: This is a critical task for the development of a model system that accurately represent observed behavioral characteristics. History dependence is likely for X_{ij}. For example, it is highly unlikely that an eat-meal activity is followed by another eat-meal activity (although it is not impossible as one may have a meal in a restaurant then visit a coffee shop for dessert). The sequencing pattern discussed earlier that more mandatory activities tend to be pursued earlier also implies history dependence. If a less mandatory activity has been pursued in the past, then the probability will be smaller that a more mandatory activity will be pursued in the future. In fact Kitamura (1983) demonstrated the presence of history dependence in trip chains by estimating alternative models that assumed different levels of stationarity and history dependence. In the same study, it was also shown that the series of trip purposes in a trip chain is a Markov chain of at least the second-order of history dependence. Even though a large data set was used in the study, it was not possible to show the presence of higher order of history dependence due to the scarcity of trip chains with many stops. Furthermore, available results (Kitamura, 1995) suggest that different tendencies of history dependence are associated with different types of activities. Further discussions on this issue can be found in Section 6.

History dependence is likely also for T_{ij}. One may anticipate that if a large amount of time has already been spent for a type of activity, then the probability that more time will be allocated for that type of activity in the future is smaller. Yet the knowledge of the history dependence in T_{ij} appears to be extremely limited.

Nor is the nature of history dependence in L_{ij} well explored. There is ample evidence that movement patterns are at least first-order history dependent. Namely the destination location (j) depends on the origin location (i). There have been models that described travelers' movement patterns in an urban area as a Markov chain of the first order (e.g., Horton and Shuldiner, 1967; Nystuen, 1967; Hanson & Marble, 1971; Sasaki, 1972; Kondo, 1974). Whether the degree of dependence is of a higher order, however, has not been explored. In addition, L_{ij} may exhibit future dependence because of the aforementioned effect of the home location on destination choice.

Time-of-Day Dependence: Activity engagement is strongly dependent on the time of day. Tabulations of time use data (e.g., Robinson, et al., 1992) show surprising homogeneity in activity engagement across individuals. This is partly institutional (e.g., work and school) and partly physiological (e.g., meals, sleeping). The comparative analysis of Dutch and Californian time use data sets shows that over 98% of the Dutch respondents were asleep at 4:00 A.M., and at 9:00 P.M. about 40% of respondents in California watched TV. Eating meals shows three sharp peaks, with the peak for supper most concentrated.

The time-of-day dependence of activity engagement can be represented by formulating engagement probabilities as time-dependent functions. An example can be found in Kitamura and Kermanshah (1983), where time of day, t, and its logarithm, ln(t), are used in multinomial logit models of activity type choice by time of day. Work activities are excluded from the analysis. The results using a 1977 travel diary data set from Baltimore show clear tendencies that personal-business activities tend to be pursued earlier in the day; social-recreational activities toward the end of the day; and shopping engagement peaks in the early afternoon. Similar approaches can be adopted for the development of a synthetic activity-travel pattern generator.

Inter-Dependence across Activity Episodes: As noted earlier, activity episodes are inter-related due to time budget constraints and possibly for many other reasons. It is anticipated that this inter-dependence can be accounted for by properly representing history and time-of-day dependence of activity engagement.

Spatial and Temporal Fixities: Different activities have different levels of fixity in terms of (i) engagement, (ii) duration, (iii) location, and (iv) timing. Typically associated with work and school activities are higher levels of engagement and duration fixities. It is logical to assume that other, more flexible activities are organized around these activities. Chauffeuring people often involves fixities in location and timing. Typical example would be taking a child to school, for which the destination location is fixed, and the timing is restricted to a relatively narrow time window. Or consider a carpool member picking up other members; again, the location is fixed, and the timing has to be precise for a successful carpool. These activities also have high levels of engagement fixity, although they may be performed by other individuals, e.g., other family members.

Some types of activities have large degrees of flexibility in terms of engagement, duration, location and timing. For example banking may be performed at any ATM in an urban area, and grocery shopping at any grocery store. These opportunities are often available 24 hours a day, providing high levels of timing flexibility. Some recreational activities (e.g., going to a concert) may have least fixity in engagement, but once engagement decision is made, then location, timing, and duration are all fixed (unless one chooses to be late for a concert or decides to leave early). In this case, high levels of fixities may be associated with engagement as well, especially once some commitment (e.g., purchasing the tickets) has been made.

Many types of activities may have high levels of fixity when they are pursued with prior commitment, e.g., a medical appointment or meeting a friend at a certain time at a certain location. This is the case where fixities are determined by how activity engagement was committed and arranged beforehand, not by activity type. Likewise engagement flexibilities may be determined by the needs for the activity, not necessarily by activity type alone. For example, performing grocery shopping may be absolutely necessary when there is no food in the house, and there happen to be no restaurants open in town or no neighbors kind enough to offer to share a meal.

These examples illustrate that fixities associated with activity engagement vary significantly depending on institutional and situational factors (e.g., store hours), prior arrangement and commitment, as well as the type of activity. Many contributing factors are situational for which adequate levels of detail may not be available. An issue for the development of a procedure for activity-travel pattern generation is whether fixities associated with each activity should be explicitly considered and modeled, or they should be treated as random elements. Considering data availability, only the latter approach seems to be feasible. Attention should be given, however, to fixities of events that recur with regularity such as work and school starting times and lunch hours.

Planned vs. Unplanned Activities: Some activities are routine, some are planned ahead, yet some are unplanned and are pursued in response to unanticipated events. For example, a business meeting may end earlier than expected, allowing the business-person to visit a nearby café for quick lunch. Also conceivable is unplanned disengagement; the business meeting may take longer than anticipated, forcing the business-person to give up a visit to a gym he had intended to make.

It is desirable that the degree of plannedness can be represented when synthesizing an individual's travel pattern. In particular, the ability to describe how a daily activity-travel pattern is planned implies the ability to represent the future dependence of activity-travel decision which arises through the individual's act of planning. This ability is important when analyzing how transportation policy measures affect travel behavior. In the context of generating synthetic activity-travel patterns, however, representing the level of plannedness in activity engagement is of lesser importance, given the fixities associated with activities are well understood. Based on these considerations, it is decided not to incorporate the degree of plannedness into the model system of this study.

Travel Time Budget: History dependence in $\mathbf{L}_i$ as well as in $\mathbf{T}_i$ would arise if a traveler allocates a certain amount of time for traveling. This leads to the notion of travel time budgets (e.g., see Beckmann et al., 1983). There have been disputes about the notion proposed earlier that individuals have a fixed time budget that is invariant across individuals. However, more recent results offer evidence that when the duration of a trip is reduced by, say, reduced congestion, then a portion of the time saving tends to be used to travel more (RDC, 1993; Golob & McNally, 1995).

Modal Continuity, Permissible Transitions and Time-of-Day Dependence: Despite the voluminous studies on travel mode choice, little is known of history dependence and time-of-day dependence of travel mode choice. This is largely because most past studies are trip-based, i.e., the choice of travel mode for a particular trip is studied in isolation without considering the preceding or subsequent trips. Consequently modal continuity and modal transition have rarely been addressed in the literature (a rare example can be found in Kondo, 1974).

The travel modes used by an individual in a series of trips tend to be governed by the condition that certain modal transitions are not permissible. As noted earlier, if a traveler leaves home by public transit, then in general the drive-alone mode cannot be adopted in subsequent trips before he returns home. This implies strong history dependence in modal transition. Exceptions, however, exist. For example, suppose a traveler leaves home driving his car, parks the car at the work place, goes out for lunch on foot, returns to the work place on foot, then drives the car back home after work. The series of modes used in this case are: drive alone - walk - walk - drive alone. What facilitated the transition from walk to drive alone is the fact that the private car had been parked at the work base. This condition must be appropriately accounted for in analyzing modal transition. It is also noted that theoretically impermissible transitions can in fact occur, e.g., a transit traveler meets and travels with a friend who is driving a car.

Both transit and highway levels of service vary along the time of day. Most obviously public transit is often unavailable during certain parts of the day, while highway travel time increases substantially during peak periods due to traffic congestion. It is conceivable that such variations in transit and highway levels of service produce apparent time-of-day dependence in mode choice. Testing whether this time-of-day dependence is genuine or spurious, is difficult because time-of-day variations in network service levels are rarely well measured. One will probably have to resort to the position of accepting observed time-of-day dependence even if it is spurious.

Another challenge is to develop a method of assigning travel modes to a series of trips made by a traveler, which is sensitive to policy measures. If one only wishes to replicate and does not seek policy sensitivity in the resulting model system, then one may determine travel mode in a sequential manner. If one wishes to account for the interplay, in response to policy measures, of mode choice, destination choice and activity duration choice associated with a series of trips, it would require a much more elaborate approach. This issue will be addressed in a separate report.

5.2. PROPOSED FORMULATION

For X_{ij} which is not travel, the following decomposition of X-T-L triple will be adopted in the study:

$$
\begin{aligned}
&\Pr[X_{ij}, T_{ij}, L_{ij} | \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \\
&\quad = \Pr[L_{ij} | X_{ij}, T_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \Pr[T_{ij} | X_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \Pr[X_{ij} | \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] \qquad (4)
\end{aligned}
$$

Namely an activity type is first selected; given the type, its duration is determined; and finally, a location is chosen for activity engagement given the type and duration. Each of these decision elements is assumed to be dependent on the past history of behavior.

This formulation is based on the view that activity engagement is the most fundamental decision that drives duration and location choice. Given the activity type, the time needed to perform the activity can be assessed and an activity duration can be determined. Also given the activity type, the suitability of respective opportunities for the activity can be evaluated and a location can be selected.

Of course a daily activity-travel pattern may not always evolve in this manner. For example, an individual may choose to pursue a certain type of activity because he or she happened to be close to a suitable opportunity (L_{ij} conditioning X_{ij}). Or one may select a type of activity and pursue it because it fits into a time window which happened to be available (T_{ij} conditioning X_{ij}). The proposed decomposition, however, can be considered to be most representative of activity engagement decision.

When X_{ij} is travel, the following decomposition would be more appropriate:

$$\Pr[X_{ij}, T_{ij}, L_{ij}|\mathbf{\underline{X}}_{i,j-1}, \mathbf{\underline{T}}_{i,j-1}, \mathbf{\underline{L}}_{i,j-1}]$$
$$= \Pr[T_{ij}|X_{ij}, L_{ij}, \mathbf{\underline{X}}_{i,j-1}, \mathbf{\underline{T}}_{i,j-1}, \mathbf{\underline{L}}_{i,j-1}]\Pr[L_{ij}|X_{ij}, \mathbf{\underline{X}}_{i,j-1}, \mathbf{\underline{T}}_{i,j-1}, \mathbf{\underline{L}}_{i,j-1}]\Pr[X_{ij}| \mathbf{\underline{X}}_{i,j-1}, \mathbf{\underline{T}}_{i,j-1}, \mathbf{\underline{L}}_{i,j-1}] \qquad (4')$$

Namely, the destination, L_{ij}, is determined before travel time, T_{ij}. This reflects the view that travel time cannot be determined before destination and mode are determined.

So far the discussion of this section has not touched on the travel mode. An approach that can be most readily integrated into the model system discussed here is that (i) a model is developed to determine the travel mode for the first trip of each home-based series of trips, and (ii) a travel mode transition matrix is developed and applied to subsequent trips on a trip-by-trip basis. As noted earlier, the issue will be further explored in a separate report. The modeling approach for each of the three probability components is presented below.

5.2.1. $Pr[X_{ij}| \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]$

The history dependence of activity type transition will be represented by formulating the probability of an activity type as a function of the series of activities so far engaged, $\underline{\mathbf{X}}_{i,j-1}$, and the time that has been allocated to them, $\underline{\mathbf{T}}_{i,j-1}$. In the discussion here, the transition probability is assumed to be dependent on the current location, $L_{i,j-1}$, but independent of the prior history of activity locations, $\underline{\mathbf{L}}_{i,j-2}$, namely,

$$Pr[X_{ij}| \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] = Pr[X_{ij}| \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, L_{i,j-1}]. \quad (5)$$

The adequacy of this assumption needs to be examined through statistical analysis of empirical data.

A simplified representation of the past history of activity engagement may be adopted. For example, let k be the number of activity categories and let

$$\underline{\mathbf{D}}_{i,j-1} = (D_{i1,j-1}, D_{i2,j-1}, \ldots.., D_{ik,j-1}) \quad (6)$$

where

$D_{im,j-1}$ = 1, if activity type m is included in $\underline{\mathbf{X}}_{i,j-1}$,

= 0, otherwise.

Likewise, let

$$\underline{\mathbf{S}}_{i,j-1} = (S_{i1,j-1}, S_{i2,j-1}, \ldots.., S_{ik,j-1}) \quad (7)$$

where

$S_{im,j-1}$ = total activity duration for type m in $\underline{\mathbf{T}}_{i,j-1}$.

Then, letting t be the time of the day when the (j - 1)th activity ended,

$$Pr[X_{ij} = m|\underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] = F(m: t, L_{i,j-1}, \underline{\mathbf{D}}_{i,j-1}, \underline{\mathbf{S}}_{i,j-1}, \mathbf{Z}_i), \; m = 1, 2, \ldots, k \quad (8)$$

where $\mathbf{Z}_i$ is the vector of person attributes and other explanatory variables. Of course $\underline{\mathbf{X}}_{i,j-1}$ and $\underline{\mathbf{T}}_{i,j-1}$ themselves may be used in function F. Discrete choice models such as multinomial logit models and nested logit models will be suitable for F.

5.2.2. $\Pr[T_{ij}|X_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]$

Given that $X_{ij} = m$, T_{ij} will have a probability distribution function whose parameters will be determined as a function of t, $\underline{\mathbf{X}}_{i,j-1}$, and $\underline{\mathbf{T}}_{i,j-1}$:

$$\Pr[T_{ij} \leq q|X_{ij} = m, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] = G_m(q\text{: } t, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}, \mathbf{Z}_i),\ \ q \geq 0, m = 1, 2, ..., k \quad (9)$$

where G_m is a distribution function.

Some distribution functions may be preferred over others for activity durations. For example, suppose an activity comprises the completion of n tasks, and suppose task completion times are identically and independently distributed (i.i.d.) with a negative exponential distribution for all tasks. Then the distribution of the duration of this activity is a type-n Erlang distribution. Other distributions, including negative exponential, Weibull, and log-normal distributions, have geneses that offer interpretations that are suitable for activity durations; see Appendix to this report.

5.2.3. $\Pr[L_{ij}|X_{ij}, T_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}]$

The problem here is to determine the probability that the destination of the j-th activity is g, given the type and duration of the activity, the completion time of (j - 1)th activity, t, and $\underline{\mathbf{X}}_{i,j-1}$, $\underline{\mathbf{T}}_{i,j-1}$, $\underline{\mathbf{L}}_{i,j-1}$. It is proposed as an initial assumption of the model development effort that this probability be formulated as conditionally independent of $\underline{\mathbf{X}}_{i,j-1}$, $\underline{\mathbf{T}}_{i,j-1}$, and $\underline{\mathbf{L}}_{i,j-2}$, given t, X_{ij} (= m), T_{ij} (= q), L_{ij} (= f) and home location, h. Namely,

$$\Pr[L_{ij} = g|\ t, h, X_{ij}, T_{ij}, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] = \Pr[L_{ij} = g|\ t, h, X_{ij} = m, T_{ij} = q, L_{ij-1} = f]$$
$$= H_m(g\text{: } t, h, f, q, \mathbf{A}, \mathbf{S}, \mathbf{Z}_i). \quad (10)$$

where **A** is a vector of attractiveness measures of alternative destinations, and **S** is a matrix of origin-to-destination travel times.

Measures of spatial separation between the current activity location and potential next activity locations, such as distance and travel time, are considered to be the true determinants of the location to visit. Another reason for the above formulation is that future activity locations are conditioned on the location of the home base, simply because the individual in most cases eventually returns to the home base. In comparison, a typical model of spatial interaction takes on the form, $H_m'(g: f, \mathbf{A}, \mathbf{S}, \mathbf{Z}_i)$.

The conditional history independence of course implies that, given t, h, f and q, whether a zone has been visited in the past does not affect the selection of the next destination location in any way. This may be the case when the model is formulated using geographical zones. Since a zone is an artificial construct which contains many opportunities, whether the next destination location falls in a particular zone is to some extent random. Likewise, the set of locations visited in the past is represented by a set of zones which again are designated rather randomly; e.g., three locations visited in the past may happen to fall in one zone, or in three separate zones. This randomness, or ambiguity, in zone assignment is one reason for the assumption of conditional history dependence.

However there are reasons to believe that $\mathbf{L}_i$ is history dependent. For example, if a worker makes a work-based trip chain, then he or she will for sure return to the zone in which the work base lies, generating history dependence. On the other hand, a worker who is not making a work-based chain may never return to the work zone after the work is over. This creates history dependence in the opposite direction. Thus different causal relations are conceivable for the nature of history dependence in $\mathbf{L}_i$. As noted earlier, little effort has been made on this subject in the past. Exploring the presence and nature of history dependence in $\mathbf{L}_i$ is one of the challenges of the present project. A few additional modeling issues are introduced below.

Prism Constraints. The spatial expanse that is accessible to an individual for activity engagement is determined by the speed of movement and the amount of time available.

Hagerstrand (1970) defined this expanse in the time-space dimension and called it time-space "prism." The prism contains all possible locations where activities can be engaged, and defines the amount of time available for activities at each location within it. The latter varies from location to location depending on the amount of time spent for traveling. Kondo and Kitamura (1987) adopted the prism concept in the analysis of trip chaining behavior. A similar concept can be found in Beckmann et al. (1983) which is concerned with the definition of accessibility measures. In the present context, the prism concept is important because it defines the state space for the transition that defines the evolution of $\mathbf{L}_i$.

Trade-off between Activity Duration and Travel Time: Another aspect which deserves attention is the trade-off between the duration of activity and the time spent to reach the activity location. One may choose to visit a nearby opportunity and spend more time on the activity there, or visit a farther, but better opportunity and spend less time there. In general, a farther destination opportunity that is visited can be considered to offer a larger utility than opportunities that are closer because otherwise the former opportunity would not be visited. This consideration is adopted by Kitamura et al. (1995) in the formulation of time-utility functions. The formulation of this study accounts for this by making the probability of L_{ij} conditional on T_{ij}. Whether this is adequate is a subject for future investigation.

5.3. Statistical Issues

Exploring these issues statistically imposes difficulties because the models at hand involve endogenous variables. Suppose

$$\Pr[X_{ij}|\, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] = \Pr[X_{ij}|\, X_{i,j-1}, T_{i,j-1}, L_{i,j-1}],\ j = 1,2, ..., n. \tag{11}$$

Then X_{ij} is said to be conditionally independent of the past history, given $X_{i,j-1}$, $T_{i,j-1}$, $L_{i,j-1}$. If

$$\Pr[X_{ij}|\, \underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}] = \Pr[X_{ij}],\ j = 1, 2, ..., n \tag{12}$$

then X_{ij} is purely random. Unless the process is purely random, a model to describe the evolution of X_{ij} would involve variables that represent past states, $X_{i,j-1}$, $X_{i,j-2}$, ..., which are endogenous, lagged-dependent variables.

The presence of lagged-dependent variables in a model creates estimation problems when serial correlation is present. Suppose X_{ij} is discrete and let $\{1, 2, ..., J\} = \Omega$ be the state space. Let Z_{ijq} be a latent variable associated with state q and individual i ($-\infty < Z_{ijq} < \infty$). Suppose

$$X_{ij} = q \ \text{ iff } Z_{ijq} > Z_{ijr}, \ q \in \Omega, \ \forall r \in \Omega, \ r \neq q \tag{13}$$

$$Z_{ijr} = F(\underline{\mathbf{X}}_{i,j-1}, \underline{\mathbf{T}}_{i,j-1}, \underline{\mathbf{L}}_{i,j-1}, W_{ijr}: \Theta) + \epsilon_{ijr}, \ \forall r \in \Omega \tag{14}$$

where Θ is unknown parameter vectors, W_{ijr} is a vector of explanatory variables measured for state r and traveler i, and ϵ_{ijr} is a random error term. This formulation leads to a discrete choice model for the next state, X_{ij}.

Consider a series of random error terms, (ϵ_{i1q}, ϵ_{i2q}, ..., ϵ_{ijq}). Since all these error terms are associated with state q, it is probable that they are serially correlated. If the traveler prefers activity type q, then ϵ_{i1q}, ϵ_{i2q}, ..., ϵ_{ijq} will all tend to be positive and are serially positively correlated. When the model is linear, i.e., $Y_q = \beta' V_q + \tau Y_{q-1} + \epsilon_q$, then the simultaneous presence of lagged-dependent variables and serially correlated errors will lead to the loss of efficiency, unless a proper method is used to account for serial correlation. Yet, the ordinary least squares estimator will retain its consistency. This is not the case when the model is non-linear, which discrete choice models are. In this case lagged-dependent variables combined with serially correlated errors lead to inconsistent estimates, again unless a proper method is taken to account for serial correlation. Accounting for serial correlation in a system of discrete choice models is not a trivial task, however. An example can be found in Kitamura and Bunch (1988) where error components are used in a series of ordered-response probit models. This, however, is a relatively simple problem. Accounting for serial correlation in general discrete choice models when Ω contains several states (say, five or more), is difficult.

State Dependence versus Heterogeneity: One critical issue in developing longitudinal models is discerning true state dependence from spurious state dependence due to heterogeneity. The former refers to the case where the probability of being in a state depends on the past history, especially on previous visits at the state itself. The latter, on the other hand, refers to the case that a particular state is visited more often because the individual is predisposed to do so.

Heterogeneity may be classified into observed heterogeneity and unobserved heterogeneity. The former refers to the case where the individual's predisposition to visit a state more, or less, frequently, can be accounted for by measured variables. The latter, no the other hand, refers to the case where this disposition is associated with unmeasured variables, or random effects. Observed heterogeneity can therefore be accounted for by introducing pertinent explanatory variables into the model system to explain differences in the probability of visiting respective states. Unobserved heterogeneity can be represented by introducing individual specific error terms. For example, Kitamura and Bunch (1988) used an error component to account for unexplained differences in the orientation toward vehicle ownership across households. An alternative in longitudinal analysis is the use of serially correlated error terms. With non-linear models with limited-dependent variables such as discrete choice models, incorporation of correlated errors often impose estimation difficulties.

5.4. Data and Validation

The time-use portion of the AMOS survey conducted in the Metropolitan Washington Council of Governments (MWCOG) area will be used in aspatial analysis and model development. The sample of this time-use data set, however, contains only commuters. Furthermore this data set has a relatively small sample size. Consequently possible use of the time use data sets from the U.S. and California that are available to the research team, will be considered. It is anticipated that the use of these time-use data that offer information on in-home activities as well as out-of-home activities, will enhance the analysis of activity engagement and duration.

AMOS survey results have not been geo-coded, and therefore the data file is not usable for spatial analysis. Because the availability of geo-coded time-use data is questionable, existing trip diary data and accompanying land use and network data collected and maintained by metropolitan planning organizations (MPOs) will be used in spatial analyses. Recently collected trip diary data are available from the San Francisco-San Jose metropolitan area (collected in 1990) and Washington, D.C., metropolitan area (collected in 1994). Currently data collection efforts are ongoing in Portland, Oregon, Dallas-Fort Worth, Texas, and San Francisco-San Jose, California. The survey instruments used in these current efforts involve time-use elements and offer information on in-home activities at varying levels of detail.

Availability of these data sets offers the possibility of rigorously validating synthetic activity-travel patterns. For example, marginal distributions of activity durations or transition frequencies of trip purposes can be compared between observed and synthetic activity-travel patterns. Whether synthetic patterns replicate more complex relationships, e.g., daily time allocation, can be examined by estimating a model system on a data set of observed patterns and on a data set of synthetic patterns, then comparing resulting coefficients vectors.

6. An Exploration of History Dependence

Kitamura (1995) explored the issue of history dependence using the 1990 National Person Travel Survey (NPTS) data set and showed that the nature of history dependence in activity choice varies depending on the activity type. This section offers salient results from the 1995 report.

The question addressed is how the fact of engaging in out-of-home activities of a given type affects the probability of engaging again in the same type of activity during the same day. Different types of history dependence are conceivable for different activities. Certain activities may be engaged just once during a day, thus past engagement would almost certainly preclude recurrent engagement in the future. Having lunch is an example. Other types of activities, on the other hand, may have the tendency that past engagement leads to higher probabilities of engagement in the future. Comparison shopping for a durable good is an example.

The 1990 NPTS data set is used to probe these issues. Three activity types -- shopping, other family and personal business, and social or recreational activity -- are used in the analysis. Conditional probabilities of engaging in these activities in the future, given the engagement in activities of the same type in the past, are evaluated at three time points of the day, 12:00 noon, 3:00 p.m. and 6:00 p.m. Results are summarized in Table 1, where each entry represents the conditional probability of engaging (E) or not engaging (N) in the activity of the same type given past engagement.

The conditional probabilities, evaluated at three time points of the day, display the clear tendency that the probability of engaging in an out-of-home activity decreases as the day progresses, irrespective of past activity engagement. This is not at all surprising as the chance of pursuing an activity will decrease as the time that remains during the day decreases. Social and recreational activities, which are dominant activities during the evening period, show the weakest tendency of this type.

The conditional probabilities evaluated for shopping show that past engagement in shopping does not affect future engagement. The conditional probabilities shown in the first row (given past engagement) and those in the second row (given non-engagement) are surprisingly similar. *Shopping engagement appears to be history independent.* Its engagement probability, however, is dependent on the time of day with its value decreasing from over 0.25 at 12:00 noon to less than 0.08 at 6:00 p.m.

Conditional probabilities for both family or personal business and social or recreational activity indicate strong history dependence, with engagement probabilities much greater with past engagement than without engagement. This is more pronounced for family or personal business. For example, as of 12:00 noon, the probability of engaging in this activity in the future is 0.486, given that family or personal business has been pursued by then, but the probability is only 0.205 given that no such activity has been engaged. The corresponding values evaluated as of 3:00 p.m. are 0.302 versus 0.131, and at 6:00 p.m. 0.134 versus 0.054.

The result found for family or personal business and social or recreational activity that conditional engagement probabilities are greater given that activities of the same type have been engaged in the past, implies that individuals tend to be split into two groups, one of which consisting of those who do not engage in these activities at all, and the other consisting of those who engage in them multiple times in the course of the day. Obviously properly capturing these history dependencies is critically important for model development.

Table 1
Conditional Probabilities of Activity Engagement by Past Engagement and Time of Day

a. Shopping

Past Engagement	At 12:00 Noon			At 3:00 P.M.			At 6:00 P.M.		
	E	N	Total	E	N	Total	E	N	Total
Engaged (E)	0.257	0.743	1.000	0.157	0.843	1.000	0.069	0.931	1.000
Not Engaged (N)	0.256	0.744	1.000	0.177	0.823	1.000	0.079	0.921	1.000
Total	0.256	0.744	1.000	0.173	0.827	1.000	0.076	0.924	1.000

b. Other Family or Personal Business

Past Engagement	At 12:00 Noon			At 3:00 P.M.			At 6:00 P.M.		
	E	N	Total	E	N	Total	E	N	Total
Engaged (E)	0.486	0.514	1.000	0.302	0.698	1.000	0.134	0.866	1.000
Not Engaged (N)	0.205	0.795	1.000	0.131	0.869	1.000	0.054	0.946	1.000
Total	0.244	0.756	1.000	0.167	0.823	1.000	0.076	0.924	1.000

c. Other Social or Recreational

Past Engagement	At 12:00 Noon			At 3:00 P.M.			At 6:00 P.M.		
	E	N	Total	E	N	Total	E	N	Total
Engaged (E)	0.375	0.625	1.000	0.267	0.733	1.000	0.167	0.833	1.000
Not Engaged (N)	0.200	0.800	1.000	0.154	0.846	1.000	0.092	0.908	1.000
Total	0.211	0.789	1.000	0.167	0.833	1.000	0.105	0.895	1.000

Before closing this section, it is important to note that the analysis here represents an initial cursory exploration of the data set regarding the history dependence of activity engagement. Only the frequency of trips by time of day is considered in the analysis and the attributes of individuals and other possible contributing factors are not incorporated into the analysis. In particular, the issue of history dependence versus heterogeneity noted earlier remains to be explored in the future. Furthermore, history dependence is examined only within the same type of activity while dependencies across different types of activities have not been examined. Nonetheless, this initial analysis has made evident that the dependence of activity engagement on the time of day and on its own history must be explicitly incorporated into the generation of synthetic travel patterns.

7. Conclusion

An analytical framework has been proposed in this report for the development of a procedure for generation of synthetic activity-travel patterns. Attempts have been made to include a broad range of analytical issues and to develop a rationale for the proposed approach. It is hoped that the report aided in paving the way for the development of a synthetic activity-travel pattern generator.

References

Adler, T., and M. Ben-Akiva (1979) A theoretical and empirical model of trip chaining behavior. *Transportation Research B*, **13B**, 243-57.

Beckmann, M.J., T.F. Golob and Y. Zahavi (1983) Travel probability fields and urban spatial structure: 1. Theory. *Environment and Planning A*, **15**, 593-606.

Golob, T.F., R. Kitamura, D. Brownstone, D.S. Bunch (1995) Forecasting Electric Vehicle Ownership and Use in the California South Coast Air Basin. Draft final report prepared for Southern California Edison Company and California Energy Commission, Institute of Transportation Studies, University of California, Irvine, CA, August.

Golob, T.F. and M.G. McNally (1995) A model of household interactions in activity participation and the derived demand for travel. Presented at the EIRASS Conference on Activity Based Approaches, Eindhoven, The Netherlands.

Hagerstrand, T. (1970) What about people in regional science.

Hanson, S. and D.F. Marble (1971) A preliminary typology of urban trip linkages. *The East Lakes Geographer*, **7**, 49-59.

Horton, F.E. and P.W. Shuldiner (1967) The analysis of lank use linkages. *Highway Research Record*, **165**, 96-107.

Kitamura, R. (1983) A sequential, history dependent approach to trip-chaining behavior. *Transportation Research Record*, **944**, 13-22.

Kitamura, R. (1995) Time-of-day characteristics of travel: an analysis of 1990 NPTS data. In Special Reports on Trip and Vehicle Attributes, 1990 NPTS Report Series, Publication No. FHWA-PL-95-033, U.S. Department of Transportation, Federal Highway Administration, Washington, D.C., February, pp. 4-1 - 4-56.

Kitamura, R. and D.S. Bunch (1990) Heterogeneity and state dependence in household car ownership: A panel analysis using ordered-response probit models with error components. In M. Koshi (ed.), *Transportation and Traffic Theory*, Elsevier, New York, pp. 477 - 496.

Kitamura, R. and M. Kermanshah (1983) Identifying Time and history dependencies of activity choice. *Transportation Research Record*, **944**, 22-30.

Kitamura, R. and M. Kermanshah (1984) A Sequential model of interdependent activity and destination choices. *Transportation Research Record*, **987**, 81-89.

Kitamura, T. van der Hoorn and F. van Wijk (1995) A comparative analysis of daily time use and the development of an activity-based traveler benefit measure. Presented at the EIRASS Conference on Activity Based Approaches, Eindhoven, The Netherlands.

Kondo, K. (1974) Estimation of person trip patterns and modal split. In D.J. Buckley (ed.), *Transportation and Traffic Theory*, Elsevier, New York, pp. 715-742.

Kondo, K. and R. Kitamura (1987) Time-space constraints and the formation of trip chains. *Regional Science and Urban Economics*, **17**, 49-65.

Lerman, S.R. (1979) The use of disaggregate choice models in semi-Markov process models of trip chaining behavior. *Transportation Science*, **13**(4), 273-91.

Mannering, F., E. Murakami and S.-G. Kim (1993) Temporal stability of travelers' activity choice and home-stay duration: Some empirical evidence. Mimeograph.

Niemeier, D.A. and J.G. Morita (1995) Duration of trip-making activities by men and women: A survival analysis. *Transportation* (forthcoming).

Nystuen, J.D. (1967) A theory and simulation of intra-urban travel. In W.L. Garrison and D.F. Marble (eds.), *Quantitative Geography -- Part I: Economic and Cultural Topics*. Studies in Geography 13, Department of Geography, Northwestern University, Evanston, Illinois, pp. 54-83.

Pas, E.I. (1983) A flexible and integrated methodology for analytical classification of daily travel-activity behavior. *Transportation Science*, **17**(4), 405-29.

Pas, E.I. (1990) Is travel demand analysis and modeling in the doldrums? In P.M. Jones (ed.), *New Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Avebury, Aldershot, United Kingdom, pp. 3-27.

Robinson, J.P., R. Kitamura and T.F. Golob (1992) Daily Travel in the Netherlands and California: A Time-diary Perspective. Report prepared for Rijkswaterstaat, Dutch Ministry of Transport, The Hague.

Sasaki, T. (1972) Estimation of person trip patterns through Markov chains. In G.F. Newell (ed.), *Traffic Flow and Transportation*. Elsevier, New York, pp. 119-130.

RDC, Inc. (1993) Further Comparative Analysis of Daily Activity and Travel Patterns & Development of a Time-Activity-Based Traveler Benefit Measure. Report prepared for Rijkswaterstaat, Dutch Ministry of Transport, the Hague.

Recker, W.W. (1995) The household activity pattern problem: General formulation and solution. *Transportation Research B*, **29B**(1), 61-77.

Recker, W.W., M.G. McNally, and G.S. Root (1986) A model of complex travel behavior: Part I--Theoretical development, Part II--An operational model. *Transportation Research A*, **20A**(4), 307-318, 319-330.

APPENDIX
ALTERNATIVE DISTRIBUTION FUNCTIONSFOR ACTIVITY DURATIONS[2]

With the objective of determining the most suitable distribution function for activity durations, alternative distribution functions are reviewed in this Appendix. The analysis here assumes that the expected duration varies from individual to individual depending on their attributes. The distributions examined here are: negative exponential distribution, Weibull distribution, log-normal distribution, log-logistic distribution, and generalized gamma distribution.

a. Negative Exponential Distribution
Let the probability density function (pdf) of T be

$$f_T(t) = \alpha \exp(-\alpha t), \quad t > 0 \tag{A.1}$$

where $\alpha = e^{-\mu}$. This distribution function has a mean of $1/\alpha = e^{\mu}$. Let $W = \ln T$. Then the pdf of W is

$$g_W(w) = f_T(e^w)|dT/dW| = f_T(e^w)e^w = e^{(w-\mu)}\exp(-e^{(w-\mu)}). \tag{A.2}$$

This distribution may be used to estimate the parameters of the distribution function with $\mu = \beta'X$, where β is the vector of coefficients and X is the vector of explanatory variables.

The negative exponential distribution is associated with the stochastic process of purely random events, i.e., events that occur with an invariant probability over time and whose occurrence neither depends on the past event history nor affects the occurrence of future events. The distribution represents the elapsed time between two successive events that are purely random, while the frequency of such events counted over intervals of a fixed length will have a Poisson distribution.

In survival analysis, the negative exponential distribution represents the basic distribution for which the hazard function, h(t), is constant:

$$h(t) = f_T(t)/[1 - F_T(t)] = \alpha, \tag{A.3}$$

where $F_T(t)$ is the cumulative distribution function (CDF) of T. When durations have a negative exponential distribution, therefore a constant hazard function, they are duration independent, i.e., the fact that the duration in a state has reached a value, t_0, does not in any way influence the probability of staying in that state for another duration of time, say x. In this sense, these durations are "memoriless." More formally,

$$\Pr[T \le t_0 + x | T > t_0] = \Pr[t_0 < T \le t_0 + x]/\Pr[T > t_0]$$

[2]The review of the statistical distributions presented here was originally conducted for Southern California Edison (SCE) Company and California Energy Commission in a project to develop a model system for electric vehicle demand forecasting in the SCE service area. See Golob et al. (1995).

$$= \{F_T(t_0+x) - F_T(t_0)\}/(1 - F_T(t_0)) = \{-\exp(-(t_0+x)) + \exp(-t_0)\}/\exp(-t_0)$$

$$= 1 - \exp(-x) = F(x), \quad x > 0, \; t_0 \geq 0 \qquad (A.4)$$

which is independent of t_0.

b. Weibull Distribution

T has a Weibull distribution if there exist γ (> 0), α (> 0) and ξ_0 such that $Y = [(T - \xi_0)/\alpha]^\gamma$ has the standard negative exponential distribution, $f_Y(y) = e^{-y}$, y > 0. The pdf of T is

$$f_T(t) = (\gamma/\alpha)[(t - \xi_0)/\alpha]^{\gamma-1}\exp\{-[(t - \xi_0)/\alpha]^\gamma\}, \quad t > \xi_0. \qquad (A.5)$$

For duration models we may assume $\xi_0 = 0$. Then

$$f_T(t) = (\gamma/\alpha)(t/\alpha)^{\gamma-1}\exp[-(t/\alpha)^\gamma], \quad t > 0. \qquad (A.5')$$

$W = ln T$ has

$$g_W(w) = (\gamma/\alpha)e^{\gamma w}\exp[-e^{\gamma w}/\alpha^\gamma], \quad -\infty < w < \infty. \qquad (A.6)$$

Letting $\sigma = 1/\gamma$ and $\mu = ln\alpha$, therefore $\alpha = e^\mu$ and $\alpha^\gamma = e^{\gamma\mu}$, we may rewrite this as

$$g_W(w) = \sigma^{-1}\exp[(w - \mu)/\sigma]\exp\{-\exp[(w - \mu)/\sigma]\}, \quad -\infty < w < \infty. \qquad (A.6')$$

The mean and variance of T are, with $\xi_0 = 0$, given as

$$E[T] = \alpha\Gamma(1 + \gamma^{-1}) = e^\mu\Gamma(1 + \sigma), \text{ and}$$

$$Var(T) = \alpha^2\{\Gamma(1 + 2\gamma^{-1}) - [\Gamma(1 + \gamma^{-1})]^2\} = e^{2\mu}\{\Gamma(1 + 2\sigma) - [\Gamma(1 + \sigma)]^2\}. \qquad (A.7)$$

The distribution was used in 1939 by a Swedish physicist Waloddi Weibull to represent the distribution of the breaking strength of materials (Johnson & Kotz, 1970a). It can be argued that a distribution of this form arises when we consider the limiting distribution of the failure time of a system when it consists of n elements, each of which consists of γ redundant components. Each element fails when all γ components fail, while the system fails if at least one element fails. Because of this linkage to failure times, the Weibull distribution is often used in survival analysis.

c. Log-normal Distribution

If there is a number, ξ_0, such that $Z = ln(T - \xi_0)$ has a normal distribution, then T is said to have a log-normal distribution. Let μ and σ be the expected value and standard deviation of Z, respectively. Let

$$U = \{ln(T - \xi_0) - \mu\}/\sigma \qquad (A.8)$$

have the unit normal distribution. Then the pdf of T is

$$f_T(t) = [(t - \xi_0)(2\pi)^{-1/2}\sigma]^{-1}\exp\{-(ln(t - \xi_0) - \mu)^2/\sigma^2\}, \quad t > \xi_0. \tag{A.9}$$

With $\xi_0 = 0$, therefore $T > 0$, we have

$$f_T(t) = \exp\{-(lnt - \mu)^2/\sigma^2\}/\{(2\pi)^{1/2}\sigma t\}, \quad t > 0 \tag{A.9'}$$

The pdf of $W = lnT$ is

$$g_W(w) = \exp[-(w - \mu)^2/2\sigma^2]/\{(2\pi)^{1/2}\sigma\}, \quad -\infty < w < \infty. \tag{A.10}$$

And $E[W] = \mu$ and $Var(W) = \sigma^2$.

Consider n independent random variables, $X_1, X_2, ..., X_n$ (> 0), and let T_n be the product of the X's. Then lnT_n tends to be normal as n tends to be infinity. The limiting distribution of T_n would then be log-normal. Thus a log-normal distribution is a theoretical distribution for durations if they are determined as a product of independent positive random variables.

d. Log-logistic Distribution

Let the CDF of T be

$$F_T(t) = 1 - 1/(1 + \alpha t^\gamma) \tag{A.11}$$

and its pdf be

$$f_T(t) = \alpha\gamma t^{\gamma-1}/(1 + \alpha t^\gamma)^2, \quad t > 0. \tag{A.12}$$

Then the pdf of $W = lnT$ is

$$g_W(w) = \alpha\gamma e^{\gamma w}/(1 + \alpha e^{\gamma w})^2, \quad -\infty < w < \infty. \tag{A.13}$$

Letting $\gamma = 1/\sigma$ and $\alpha = e^{-\mu/\sigma}$,

$$g_W(w) = \sigma^{-1}\exp[(w - \mu)/\sigma]/\{1 + \exp[(w - \mu)/\sigma]\}^2, \quad -\infty < w < \infty, \tag{A.13'}$$

and $E[W] = \mu$ and $Var(W) = (\pi^2/3)\sigma^2$.

The logistic function has often been used as a growth curve based on the differential equation,

$$dH/dx = c[H(x) - A][B - H(x)] \tag{A.14}$$

where c (> 0), A and B (> A) are constant parameters. Function H may be viewed to represent growth from a lower asymptote A to an upper asymptote B, and the rate of growth is proportional to the product of the distances from the two asymptotes. The above H is a CDF when A = 0 and B = 1. The logistic distribution has been shown to be the limiting distribution (as $n \to \infty$) of the standardized mid-range (average of largest and smallest sample values) of random samples of size n (Johnson and Kotz, 1970b).

e. Generalized Gamma Distribution

The standardized forms (μ = 0, σ = 1) of the generalized gamma distributions used in this analysis can be written as:

$$f_T(t) = \delta(t^\delta/\kappa)^{1/\kappa}\exp(-t^\delta/\kappa)/t\Gamma(1/\kappa), \quad t > 0, \tag{A.15}$$

and

$$g_W(w) = \delta(e^{\delta w}/\kappa)^{1/\kappa}\exp(-e^{\delta w}/\kappa)/\Gamma(1/\kappa), \quad -\infty < w < \infty, \tag{A.16}$$

where $\kappa = \delta^2$. The mean and variance of T are:

$$E[T] = \delta^{2/\delta}\Gamma(1/\kappa + 1/\delta)/\Gamma(1/\kappa), \text{ and}$$

$$Var(T) = \delta^{4/\delta}\{\Gamma(1/\kappa + 2/\delta) - [\Gamma(1/\kappa + 1/\delta)]^2/\Gamma(1/\kappa)\}/\Gamma(1/\kappa). \tag{A.17}$$

The gamma distribution includes as its special case the chi-square distribution. The latter is the distribution of the sum of squares of independent unit normal random variables. Namely, if $U_1, U_2, ..., U_\upsilon$ are independent unit normal random variables, then $X = U_1^2 + U_2^2 + ... + U_\upsilon^2$ has a χ^2 distribution with degrees of freedom υ. It also contains as its special case the Erlang distribution, which is the distribution of the sum of α (= 1, 2, ...) independent negative exponentially distributed random variables. Gamma distributions have been used as approximate distributions for chi-square distributions. Johnson and Kotz (1970a) note that "In applied work, gamma distributions give useful representations of many physical situations. They have been used to make realistic adjustments to exponential distributions in representing lifetimes in 'life-testing' situations. Of recent years, Weibull distributions have been more popular for this purpose, but this may not be permanent. The fact that sum of independent exponentially distributed random variables has a gamma distribution ... leads to the appearance of gamma distribution in the theory of random counters and other topics associated with random processes in time ...."

References

Golob, T.F., D. Brownstone, D.S. Bunch and R. Kitamura (1995) Forecasting Electric Vehicle Ownership and Use in the California South Coast Air Basin. Draft final report prepared for the Southern California Edison Company. University of California, Irvine, August.

Johnson, N.L. and S. Kotz (1970a) *Distributions in Statistics, Continuous Univariate Distributions -- 1.* John Wiley & Sons, New York.

Johnson, N.L. and S. Kotz (1970b) *Distributions in Statistics, Continuous Univariate Distributions -- 2.* John Wiley & Sons, New York.