Technical Report Documentation Page

| 1. Report No.<br><br>FHWA/TX-03/4080-4 | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|

| 4. Title and Subtitle<br><br>ACTIVITY-BASED TRAVEL-DEMAND MODELING FOR METROPOLITAN AREAS IN TEXAS: A MICRO-SIMULATION FRAMEWORK FOR FORECASTING | 5. Report Date<br>February 2003 |
|---|---|
| 7. Author(s)<br><br>Chandra R. Bhat, Sivaramakrishnan Srinivasan, Jessica Y. Guo, Aruna Sivakumar | 6. Performing Organization Code<br><br>8. Performing Organization Report No.<br>4080-4 |
| 9. Performing Organization Name and Address<br>Center for Transportation Research<br>The University of Texas at Austin<br>3208 Red River, Suite 200<br>Austin, TX 78705-2650 | 10. Work Unit No. (TRAIS)<br><br>11. Contract or Grant No.<br><br>Project 0-4080 |
| 12. Sponsoring Agency Name and Address<br>Texas Department of Transportation<br>Research and Technology Implementation Office<br>P.O. Box 5080<br>Austin, TX 78763-5080 | 13. Type of Report and Period Covered<br>Research Report (9/02 – 2/03)<br><br>14. Sponsoring Agency Code |

15. Supplementary Notes

Project conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration, and the Texas Department of Transportation.

16. Abstract

The project aims to comprehensively model the activity-travel patterns of workers as well as nonworkers in a household. The activity-travel system will take as input various land-use, sociodemographic, activity system, and transportation level-of-service attributes. It will provide as output the complete daily activity-travel patterns for each individual in the household. This report presents a methodology to apply the developed model system for forecasting. A micro-simulation framework is developed and details of the proposed implementation scheme are discussed.

| 17. Key Words<br>Activity-Based Analysis, Modeling Framework, Micro-Simulation, Forecasting | 18. Distribution Statement<br>No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161. |
|---|---|

| 19. Security Classif. (of report)<br>Unclassified | 20. Security Classif. (of this page)<br>Unclassified | 21. No. of pages<br>56 | 22. Price |
|---|---|---|---|

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

# Activity-Based Travel-Demand Modeling for Metropolitan Areas in Texas: A Micro-Simulation Framework for Forecasting

Chandra R. Bhat
Sivaramakrishnan Srinivasan
Jessica Y. Guo
Aruna Sivakumar

Research Report 4080-4

Research Project 0-4080
*Activity-Based Travel-Demand Modeling for Metropolitan Areas in Texas*

Conducted for the
Texas Department of Transportation
in cooperation with the
U.S. Department of Transportation
Federal Highway Administration
by the
Center for Transportation Research
The University of Texas at Austin

February 2003

# Disclaimers



Chandra R. Bhat
Research Supervisor

# Acknowledgments

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

# Table of Contents

# List of Figures

# List of Tables

# 1. Introduction

Transportation planners and engineers have to be able to forecast the response of transportation demand to changes in the attributes of the transportation system and changes in the sociodemographics of the people using the transportation system in order to make informed transportation infrastructure planning decisions. Travel-demand models are used for this purpose; specifically, travel-demand models are used to predict travel characteristics and usage of transport services under alternative socioeconomic scenarios, and for alternative transport service and land-use configurations.

The need for realistic representations of behavior in travel-demand modeling is well acknowledged in the literature. This need is particularly acute today as emphasis shifts from evaluating long-term, investment-based capital improvement strategies to understanding travel behavior responses to shorter-term, congestion management policies such as alternate work schedules, telecommuting, and congestion pricing. The limitations of the traditional *statistically oriented*, trip-based approach in evaluating demand management policies (Gordon et al., 1988; Lockwood and Demetsky, 1994; Hanson, 1980) has led to the emergence of a more *behaviorally oriented*, activity-based approach to demand analysis.

The activity-based approach to travel-demand analysis views travel as a derived demand, derived from the need to pursue activities distributed in space (Jones et al., 1990; Axhausen and Gärling, 1992). The approach adopts a holistic framework that recognizes the complex interactions in activity and travel behavior. The conceptual appeal of this approach originates from the realization that the need and desire to participate in activities is more basic than the travel that some of these participations may entail. Activity-based travel analysis has seen considerable progress in the past couple of decades (see Guo and Bhat (2001) for a detailed review of the state of the art in activity-based research).

The current project aims to advance the state of the art in daily activity-travel modeling. It represents one of the first attempts to comprehensively model the activity-travel patterns of both workers and nonworkers in a household. The activity-travel system will take as input various land-use, sociodemographic, activity system, and transportation level-of-service attributes. It will provide as output the complete daily activity-travel patterns for each individual in the household within a continuous time domain.

A previous research report (Bhat et al., 2002) presented the conceptual and analysis frameworks to model the different attributes to completely characterize the daily activity-travel patterns of both workers and nonworkers. Travel survey data from Dallas-Fort Worth was used to estimate the different model components. Results of model estimations were also provided in the previous report.

The focus of this report is to present a framework for using such an estimated modeling system to predict the activity-travel patterns of individuals. Methodologies for the prediction of individual choice instances and the integration of all the different choices into the complete activity-travel pattern of individuals are presented. As part of this research project, a simulation software called the "Comprehensive Econometric Micro-simulator for Daily Activity-travel Patterns" (CEMDAP) is being developed to implement this prediction mechanism. Proposed implementation details of the software are also discussed in this report.

The rest of this report is organized as follows: Chapter 2 discusses representations for the daily activity-travel patterns for both workers and nonworkers and the modeling framework to be implemented within CEMDAP. Chapter 3 provides details of a micro-simulation framework developed for applying the modeling system to predict activity-travel patterns of individuals. Chapter 4 discusses implementation details and input data requirements. Chapter 5 summarizes the contents of the report.

# 2. An Econometric Framework for Modeling Activity-Travel Patterns

This chapter describes the frameworks embedded within CEMDAP (Comprehensive Econometric Micro-simulator for Daily Activity-travel Patterns) to model the daily activity-travel patterns of individuals. Representation frameworks for worker and nonworker activity-travel patterns are first presented. The different model components that constitute the overall modeling system are then listed. The econometric model structure of each of these models is also discussed. To use CEMDAP for activity-travel predictions, a planner needs to estimate all these different model components using travel survey data from the city under study.

## 2.1 Representing Activity-Travel Patterns

Individuals make choices about different activities to be pursued during a day. Travel may be required to participate in these activities in a desired sequence. This sequence of activities and travel that a person undertakes is defined as the individual's "activity-travel pattern" for the day. The activity pattern of workers rests on the regularity and the fixity of the work activity. No such obvious fixity is present in the case of nonworkers (retired people and homemakers). This critical difference motivated development of separate representations for worker and nonworker activity-travel patterns. For both the worker and nonworker representations, we consider 3 a.m. as the beginning of the day and assume that the individual is at home at this time. The following discussion on activity-travel representations for workers and nonworkers is drawn from earlier works of Bhat and Singh (2000) and Bhat and Misra (2002).

### 2.1.1 A representation framework for workers' activity-travel patterns

The daily pattern of workers is characterized by four different (sub-) patterns: a) Before-work pattern, which represents the activity-travel undertaken before leaving home to work; b) Commute pattern, which represents the activity-travel pursued during the home-to-work and work-to-home commutes; c) Work-based pattern, which includes all activity and travel undertaken from work; and d) After-work pattern, which comprises the activity and travel behavior of individuals after arriving home at the end of the work-to-home commute. Within each of before-work, work-based and after-work patterns, there might be several tours. A tour is a circuit that begins and ends at home for the before-work and after-work patterns and is a circuit that begins and ends at work for the work-based pattern. Each tour, the home-to-work commute and the work-to-home commute may comprise several out-of-home activity episodes (referred to as "stops" in the rest of the report). A stop is characterized by the type of activity undertaken, in addition to spatial and temporal attributes. Figure 2.1 provides a diagrammatic representation of the worker activity-travel pattern.

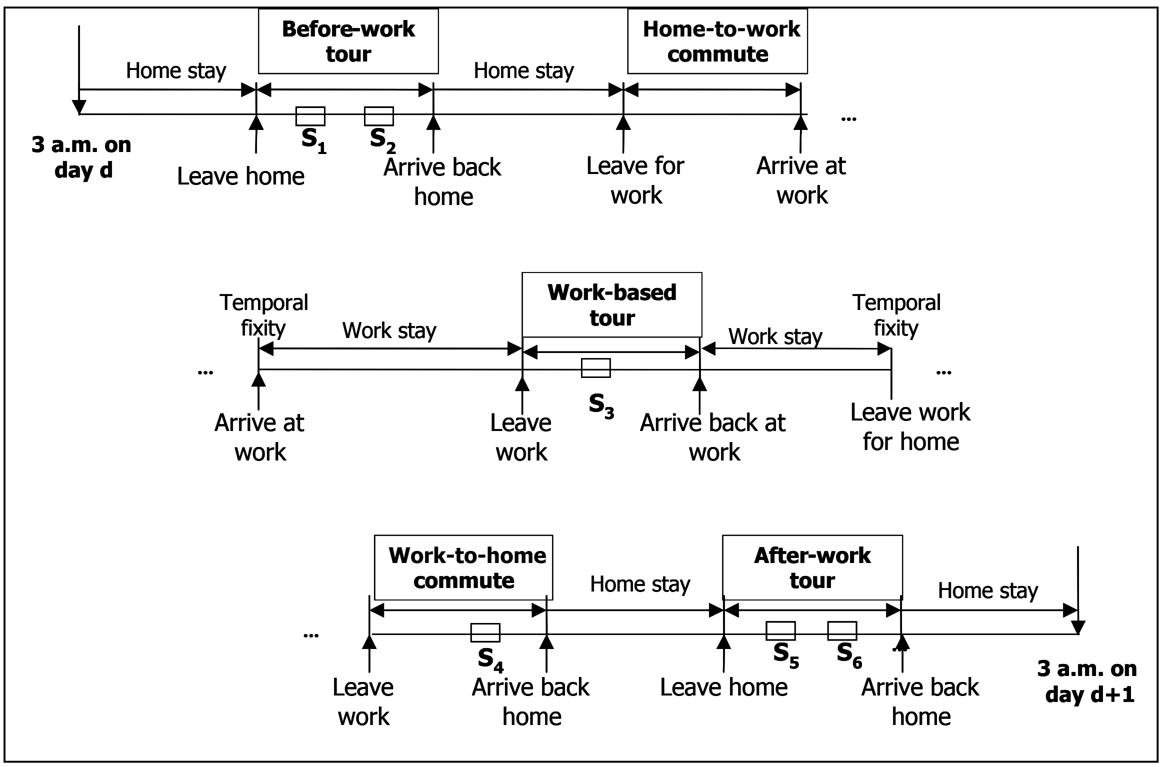


*Figure 2.1 A representation framework for workers' activity-travel pattern*

The characterization of the complete workday activity-travel pattern is accomplished by identifying a number of different attributes. These attributes may be classified based on the level of representation they are associated with; that is, whether they are associated with a pattern, a tour, or a stop. Pattern-level attributes include the travel mode, number of stops and the duration for each of the work-to-home and home-to-work commutes in addition to the number of tours that the worker undertakes during each of before-work, work-based and after-work periods. Tour-level attributes include travel mode, number of stops, tour duration and home-stay duration (or work-stay duration, in the case of the work-based tour) before the tour. Stop-level attributes include activity type, duration of the activity, travel time to stop, location, and sequence of stop in a tour/commute.

### 2.1.2 A representation framework for nonworkers' activity-travel patterns

In the case of nonworkers, the activity-travel pattern is considered as a set of out-of-home activity episodes (or "stops") of different types interspersed with in-home activity stays. The chain of stops between two in-home activity episodes is referred to as a tour. The pattern is represented diagrammatically in Figure 2.2.

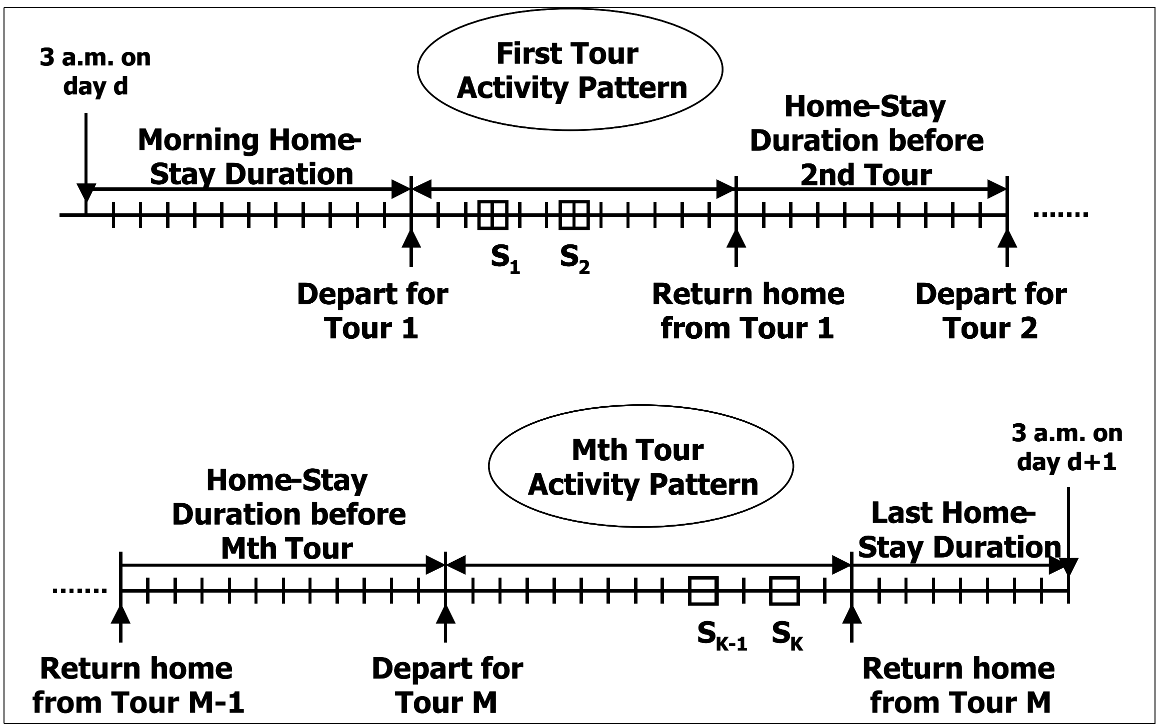


*Figure 2.2 A representation framework for nonworkers' activity-travel pattern*

A nonworker's daily activity-travel pattern is characterized by several attributes, which can again be classified into pattern-, tour-, and stop-level attributes. The only pattern-level attribute is the total number of tours that the person decides to undertake during the day. The tour-level attributes are the travel mode, number of stops in the tour, tour duration, the home-stay duration before the tour and the sequence of the tour in the day. Stop-level attributes include activity type, duration of the activity, travel time to stop, location, and the sequence of the stop in a tour.

The representations described above for workers and nonworkers are generic and can be used to describe any complex activity-travel pattern (i.e. any number of stops sequenced into any number of tours). Considering practical implementation constraints, certain restrictions are imposed on the maximum number of tours and the maximum number of stops in any tour. In the case of nonworkers, CEMDAP is designed to handle up to a total of four tours and up to four stops during each tour. In the case of workers, the implementation is capable of handling one tour during each of before-work, work-based and after-work periods and up to four stops during any tour or commute.

## 2.2 Model Components

This section identifies all model components that constitute the overall modeling framework and their econometric model structures as implemented within CEMDAP. These models have to be calibrated and the parameters provided as inputs in order to use CEMDAP for activity-travel predictions. The model components presented here are primarily based on the analysis framework developed for the Dallas-Fort Worth (DFW) area (details are presented in Bhat et al. (2002)). The econometric structures implemented for some of the model components are, however, more generic (i.e. the model structures estimated for DFW may be viewed as simplifications of these more generic structures). It is also recognized that, in some cases, it may not be possible to estimate the prescribed model type due to data limitations. Hence, for each model component, a "simple" model type is also presented that may be easily estimated when it is not possible to estimate the "prescribed" model type. The model components are listed under three major categories: (1) the generation-allocation models, (2) the scheduling models for workers, and (3) the scheduling models for nonworkers.

### 2.2.1 The Generation-Allocation Models

The eleven different model components of the generation-allocation model system are listed in Table 2.1. Models for decisions to go to work and school are estimated for employed persons and students respectively. This requires that data be available to classify adults into three mutually exclusive and collectively exhaustive types: employed, students, and non-employed. Models for work-based (school-based) duration and the work start-time (school start time) are estimated only for employed persons (students) who chose to go to work (school). The work/school start times are modeled as duration in minutes from 3 AM until the start of work/school episode.

The household activity generation model determines the decision of the household to undertake none, one, or more of the following activity types for the day: shopping, social/recreational, and personal business. Models 8, 9, and 10 (i.e., the activity allocation models) are estimated only for households with multiple adults that have decided to undertake the corresponding activity for the day. These model the decision of an individual to participate in an activity type, given that the household has decided to participate in it. The last model component is the "other" activity participation model, estimated for all adults. "Other" activities comprise of activity types such as eat out, serve-passenger and any other miscellaneous types such as volunteer work and community service etc.

*Table 2.1 Components of the generation-allocation model system*

| S.No | Model Description | Prescribed Model Type | Simple Model Type |
|---|---|---|---|
| 1 | Decision to go to work | Binary logit | Constant only |
| 2 | Work-based duration | Hazard-duration[1] | Simple Probabilistic |
| 3 | Work start time | Hazard-duration[1] | Simple Probabilistic |
| 4 | Decision to go to school | Binary logit | Constant only |
| 5 | School-based duration | Linear-regression | Simple Probabilistic |
| 6 | School start time | Linear-regression | Simple Probabilistic |
| 7 | HH activity generation | Multinomial logit | Constants only |
| 8 | Shopping activity allocation | Binary logit | Constant only |
| 9 | Social/Recreational activity allocation | Binary logit | Constant only |
| 10 | Personal business activity allocation | Binary logit | Constant only |
| 11 | "Other" activity participation | Binary logit | Constant only |

[1] proportional hazard function with non-parametric baseline hazard and gamma heterogeneity

The last column of Table 2.1 provides a "simple" model type that is supported by the simulator, in case the "prescribed" model type (provided in the third column of Table 2.1) cannot be estimated. For all unordered discrete choice models (binary and multinomial logit), a simple model that consists of only the constant(s) may be specified if a more elaborate model with explanatory variables could not be estimated. The constant(s) can be determined from the observed sample shares in the data. For models of time duration (models 2 3 5 and 6), a simple probabilistic model may be specified assuming that the logarithm of the duration is normally distributed. Such models are specified using the mean and variance of the observed distribution in the sample.

### 2.2.2 The Scheduling Models for Workers

The components of the scheduling model system (which in turn comprises the pattern-, tour- and the stop-level model systems) for workers are listed in Table 2.2. “Workers” are defined as employed persons who decided to go to work and students who decided to go to school.

The first two models in the pattern-level system characterize the work-to-home commute for the workers. CEMDAP allows for the specification of the mode choice and the number of stops for this commute as either a joint model or as independent models. In the case of a joint model specification, two different models are required: (1) a joint model of mode and number of stops, for workers who decided to participate in one or more activities other than work and (2) A mode choice model for workers who decided to participate only in the work activity during the day (and consequently the number of commute stops is known to be zero). In the case of independent model specifications, separate models are needed for the mode choice and the number of stops (again, only for workers that decided to undertake activities other than work). Therefore a model system will either have models 12 and 13 or 12B and 13B. The implementation is capable of handling the following modes: drive alone, shared-ride (or shared-ride 2 and shared-ride 3+), Transit, Walk/bike and DA-SR (drive alone and shared ride modes).

The home-to-work commute mode choice may be assumed to be significantly dependent on the work-to-home commute mode choice. Consequently, the modeling framework allows specification of the home-to-work commute mode choice model to be completely segmented based on the work-to-home commute mode.

The last model of the pattern-level system determines the decision of workers to undertake tours. The model is estimated only for workers who decided to participate in activities other than work. Since there are three different periods (before-work (BW), work-based (WB) and after-work (AW) periods) in which the worker may undertake a tour, there are a total of eight different choice alternatives: no tours at all, tours during one of the three periods (3 choices), tours during two of the three periods (3 choices), and tours during all the three periods.

*Table 2.2 Components of the scheduling model system for workers*

| S.No. | Model Description | Prescribed Model Type | Simple Model Type |
|---|---|---|---|
| **The pattern-level model system** | | | |
| 12 | WH commute mode and stops | Joint unordered-ordered | Independent models |
| 13 | WH commute mode and no stops | Multinomial logit | Constants only |
| 12(B) | WH commute mode | Multinomial logit | Constants only |
| 13(B) | WH commute stops | Ordered probit | Thresholds only |
| 14 | WH commute duration | Linear-regression | Simple Probabilistic |
| 15 | HW commute mode[1] | Multinomial logit | Constants only |
| 16 | HW commute stops | Ordered probit | Thresholds only |
| 17 | HW commute duration | Linear-regression | Simple Probabilistic |
| 18 | Decision to make a tour in each period | Multinomial logit | Constants only |
| **The tour-level model system[2]** | | | |
| 19 | Mode and stops | Joint unordered-ordered | Independent models |
| 19(m) | Mode | Multinomial logit | Constants only |
| 19(s) | Stops | Ordered probit | Thresholds only |
| 20 | Tour duration | Linear-regression | Simple Probabilistic |
| 21 | Home-stay duration before tour | Linear-regression | Simple Probabilistic |
| **The stop-level model system[3]** | | | |
| 22 | Activity type | Multinomial logit | Constants only |
| 23 | Activity duration and travel time | Simultaneous equations | Independent models |
| 23(a) | Activity duration | Linear-regression | Simple Probabilistic |
| 23(t) | Travel time | Linear-regression | Simple Probabilistic |
| 24 | Location | Spatial location choice | Multinomial logit |

[1] Separate models for each of the possible work-to-home modes

[2] Separate models for each of the BW, WB and AW tours

[3] Separate models for stops in each of WH and HW commutes and BW, WB, and AW tours

The tour-level model system determines the mode, number of stops, duration and the home-stay duration for each of the before-work (BW), work-based (WB) and after-work (AW) tours (if any). The implementation allows for the specification of a separate model for each of the BW, WB and AW tours. The tour mode and the number of stops may be specified jointly (model 19 in Table 2.2) or by independent models (models 19(m) and 19(s) in Table 2.2).

The stop-level model system determines the activity type, duration, travel time to stop and the location for each of the stops made during the home-to-work commute (HWC), work-to-home commute (WHC) and as a part of any tour (BW, WB, or AW). The implementation allows for separate models to be specified for stops made during each of the HWC, WHC, and BW, WB and AW tours.

While modeling the generation of activities, four different activity types are supported by CEMDAP, viz., shopping, social/recreational, personal-business and "other". While modeling the activity type of the stop, the "other" activity type is further classified into eat-out, serve-passenger and miscellaneous types. Activity duration and the travel time to a stop may be specified as joint (model 23) or independent models (models 23(a) and 23(t)).

The last column of Table 2.2 provides the simpler modeling methods. Ordered probit models may be replaced by simpler thresholds-only models. Spatial location choice models (which use a probabilistic choice set generation method) may be replaced by a simpler location choice model that uses a random set of destinations as the choice set.

### 2.2.3 The Scheduling Models for Nonworkers

The components of the scheduling model system for nonworkers are listed in Table 2.3. "Nonworkers" are defined to include workers who did not decide to go to work and students who did not decide to go to school, in addition to the persons who are not employed. The only pattern-level model is the one that determines the number of tours and this is estimated for nonworkers who decided to participate in one or more activity types for the day.

The components of the nonworker tour- and stop-level model system are identical to that of the tour- and stop-level model system for workers respectively. Since the implementation is designed to handle up to four different tours for the nonworkers, one may specify a separate model for each of the four tours. Similarly, one may also specify separate stop-level models for stops in each of the four tours.

*Table 2.3 Components of the scheduling model system for nonworkers*

| S.No. | Model Description | Prescribed Model Type | Simple Model Type |
|---|---|---|---|
| **The pattern-level model system** | | | |
| 25 | Number of tours | Ordered probit | Thresholds only |
| **The tour-level model system**[1] | | | |
| 26 | Mode and stops | Joint unordered-ordered | Independent models |
| 26(m) | Mode | Multinomial logit | Constants only |
| 26(s) | Stops | Ordered probit | Ordered probit |
| 27 | Tour duration | Linear-regression | Simple Probabilistic |
| 28 | Home-stay duration before tour | Linear-regression | Simple Probabilistic |
| **The stop-level model system**[2] | | | |
| 29 | Activity type | Multinomial logit | Constants only |
| 30 | Activity duration and travel time | Simultaneous equations | Independent models |
| 30(a) | Activity duration | Linear-regression | Simple Probabilistic |
| 30(t) | Travel time | Linear-regression | Simple Probabilistic |
| 31 | Location | Spatial location choice | Multinomial logit |

[1] Separate models for tours 1 2 3 and 4

[2] Separate models for stops in each of tours 1 2 3 and 4

Again, the last column of the tables lists the simple model types. All multinomial logit models may be replaced with constants-only models and all ordered probit models may be replaced with thresholds-only models. Linear regression models for durations may be replaced with simple probabilistic models by specifying the mean and variance. The spatial location choice model may be replaced with a simpler location choice model that uses a random sample for candidate destinations.

# 3. A Micro-Simulation Framework for Predicting Activity-Travel Patterns

This chapter of the report presents a micro-simulation-based framework for a predictive mechanism that uses the calibrated model system to predict individuals' daily activity-travel patterns. First, the general details of an econometric micro-simulation approach are discussed and compared against other predictive mechanisms. The implementation of the micro-simulation platform consists of two dimensions: the mechanism of micro-simulating individual decision instances and the mechanism of integrating the outputs obtained from the different decision instances to determine the final predicted activity-travel pattern. Section 2 of this chapter deals with the prediction of each individual choice instance and Section 3 presents an overall framework, which integrates the different choices to completely determine the activity-travel pattern of individuals. The last section of this chapter deals with spatial and temporal consistency checks that will be performed to ensure that the predicted choices are not unreasonable.

## 3.1 Alternate Predictive Mechanisms

This section discusses alternate methodologies that can be used to predict or forecast a set of choices using a calibrated model system that comprises of a sequence of individual model components, with the possibility that outcomes from any model may be an input to one or more of the subsequent models. The individual choice instances may be discrete (such as the mode for a tour) or continuous outcomes (such as the duration of a tour).

A simple approach would be to select the alternative with the highest utility for each of the model components with discrete outcomes consistent with the theory of utility maximization. Continuous choice variables may be assigned the expected value predicted by the model. The disadvantage of this methodology is that it introduces systematic bias in the outcome of each stage (see Bhat and Misra (2001) for a detailed discussion). Consequently, the cumulative prediction errors for large modeling systems comprising of several model components, such as the one implemented in CEMDAP, can be quite significant.

An alternate approach is to develop a full decision tree where the probabilities of all the alternatives are carried over to the root node of the decision tree. The chosen set of alternatives can hence be determined by extracting the path with the highest path probability in the decision tree (Bhat and Misra, 2001). Since the probabilities for all the alternatives for all choice instances need to be carried till the end, this approach can get computationally intensive for a large tree (i.e., a model system with several models). Further, decision trees require discrete choice instances and cannot handle models with continuous choice outcomes.

The micro-simulation approach adopted in the implementation of CEMDAP eliminates the bias of the simplistic approach while avoiding the computational complexity of the decision-theoretic approach. Unlike the decision-theoretic approach, in the micro-simulation approach the choice outcome from each model is uniquely determined and carried over to the next model component. In the case of discrete choices, the chosen alternative is determined through a random draw from a pseudo-sample containing all the alternatives in proportion to their predicted probabilities. This ensures an unbiased selection of an alternative at any choice instance since each alternative appears in the pseudo-sample in proportion to its probability of being chosen (Bhat and Misra, 2001). In the case of continuous choice instances, the choice is determined by a random draw from the predicted distribution of the choice variable. Thus, it is ensured that the chosen continuous outcome is not the same for all observationally similar decision makers.

## 3.2 Disaggregate Predictions Using Individual Component Models

This section of the chapter presents algorithms for disaggregate predictions of the different individual choices. Since the methodology depends primarily on the econometric model type, the algorithms are presented separately for each of the *eight* different econometric model types embedded within CEMDAP.

### 3.2.1 Simple Probabilistic Models

Simple probabilistic models may be used to model continuous choice variables when data does not permit the development of more elaborate models with explanatory variables. In this case, one needs to specify the type of distribution along with the parameters (such as mean, variance, etc.) defining the distribution. The prediction method involves the single step of drawing a random number from the specified distribution. The implementation tool allows for all linear regression and hazard duration models (used for predicting continuous time durations) to be replaced with simple probabilistic models.

### 3.2.2 Multinomial Logit Models

The multinomial logit model is generally used to model choice outcomes when the alternatives are discrete and unordered. See, for example, Ben-Akiva and Lerman (1985) for details on the econometric model structure. The major steps involved in predicting the choice outcome using a multinomial logit model are:

1. Determine the available choice alternatives for the decision maker from among the universal set of choices.
2. If the person has only one alternative available, then this is the chosen alternative. STOP.
3. If the person has multiple alternatives available ($A_1, A_2 ... A_K$), compute the probability ($P_1, P_2 ... P_K$) for each of the different choice alternatives using the calibrated model parameters and the values of exogenous variables specific to the decision maker under consideration.
4. Generate a uniformly distributed random number ($U$) between 0 and 1.
5. The chosen alternative is determined using the computed choice probabilities and the uniform random number drawn as follows:

   If $0 <= U < P_1$, chosen alternative is $A_1$.
   If $P_1 <= U < P_1+P_2$, chosen alternative is $A_2$.

   If $P_1+P_2+..P_{J-1} <= U < P_1+P_2+..P_J$, chosen alternative is $A_J$.
   If $P_1+P_2+..P_{K-1} <= U < 1$, chosen alternative is $A_K$.

The household activity generation model, tour and commute mode choice models and the choice of activity type at the stop location are formulated as multinomial logit models in CEMDAP. The binary logit model is a special case of the multinomial logit model with exactly two choice alternatives. Therefore the above-described algorithm is much simplified in the context of a binary logit model. Decisions to go to work/school and to participate in different activity types for the day are formulated as binary logit models.

### 3.2.3 Ordered Probit Models

The ordered probit model is generally used to model choice outcomes when the alternatives are discrete and ordered. See for example Maddala (1999) for details on the econometric model structure. The major steps involved in predicting the choice outcome are (assuming $K$ ordered alternatives, ($0,1,2,\ldots K\text{-}1$) and that the model has been estimated without a constant term) :

1. Compute the propensity ($V$) using the calibrated model parameters and the values of exogenous variables specific to the decision maker under consideration.
2. Compute the probability ($P_1, P_2 \ldots P_K$) for each of the different choice alternatives using the computed propensity ($V$) and the threshold values ($T_1, T_2 \ldots T_{K-1}$).
3. Generate a uniformly distributed random number ($U$) between 0 and 1.
4. The choice alternative is determined using the computed choice probabilities and the uniform random number drawn as follows:

If $0 <= U < P_1$, chosen alternative is $0$.

If $P_1 <= U < P_1+P_2$, chosen alternative is $1$..

If $P_1+P_2+..P_{J-1} <= U < P_1+P_2+..P_J$, chosen alternative is $J\text{-}1$.

If $P_1+P_2+..P_{K-1} <= U < 1$, chosen alternative is $K\text{-}1$.

In CEMDAP, the number of stops in a tour or during the commute and the number of tours during the day (for nonworker) are formulated as ordered probit models.

### 3.2.4 Linear Regression Models

Linear regression models are used to model continuous choice variables. See for example Greene (2000) for details on the econometric model structure. In the modeling framework, most of the temporal choices (such as commute duration, tour durations, home-stay duration before a tour, activity duration at a stop, travel time to a stop, etc.) are formulated as linear regression models. In all the above cases, the logarithm of the time duration has been chosen as the explanatory variable and the error term assumed to be normally distributed. The major steps involved in predicting the time duration from a linear regression model are:

1. Compute the expected value of log-duration ($L$) using the calibrated model parameters and the values of explanatory variables specific to the decision maker under consideration.
2. Draw from a normally distributed random variable with mean equal to the above computed value ($L$) and known variance ($\sigma^2$) from the calibration.

### 3.2.5 Hazard-Duration Models

Hazard-duration models are used to model time durations recognizing that the likelihood of termination of a duration depends on the length of the elapsed time since the start of the duration. See for example Hensher and Mannering (1994) and Kiefer (1988) for a detailed discussion on duration models. In CEMDAP work-based duration and work-start time are modeled as hazard-duration models. A proportional hazard function with a non-parametric baseline hazard specification is used. A gamma-distributed error term is used to account for the unobserved heterogeneity. Bhat (1996) provides a detailed discussion of this specification. The methodology for determining the choice of duration (on a continuous scale) using the above-described model structure is as follows (assuming $K$ discrete periods for the baseline hazard distribution):

1. Compute the probability ($P_1, P_2 ... P_K$) of each of the discrete periods using the computed parameters, values of exogenous variables specific to the decision maker under consideration, estimated variance of the heterogeneity term ($\sigma^2$) and threshold values ($T_1, T_2 ... T_{K-1}$).
2. Generate a uniformly distributed random number ($U$) between 0 and 1.
3. The duration chosen (in discrete time scale) is determined using the computed probabilities and the uniform random number drawn as:

   If $0 <= U < P_1$, chosen discrete period is $1$.

   If $P_1 <= U < P_1+P_2$, chosen discrete period is $2$.

If $P_1+P_2+..P_{J-1} <= U < P_1+P_2+..P_J$, chosen discrete period is $J$.

If $P_1+P_2+..P_{K-1} <= U < 1$, chosen discrete period is $K$.

4. In the final step, the continuous time duration is determined from the discrete time interval chosen in the above step assuming that the hazard is constant over the discrete time duration:

   Draw another random number ($U_2$) from a uniform distribution over (0,1)

   If $t_L$ and $t_H$ are respectively the lower and higher bounds of the discrete time interval chosen, determine the chosen duration as: $t_L + (t_H - t_L)* U_2$.

### 3.2.6 Spatial Location Choice Models

The spatial location choice models, as the name suggests, are used to predict the location of out-of-home activity stops. The methodology employs a probabilistic choice set generation method that uses the predicted distribution of travel time to the stop in the determination of the candidate locations for the stop. See for example Misra (1999) for a detailed discussion of the mathematical formulation of this model. The steps involved in the disaggregate prediction using this model are summarized below:

1. Determine $N$-$1$ thresholds ($T_1, T_2...T_{N-1}$) on the logarithmic time scale and classify the candidate destinations into $N$ different choice sets, which are mutually exclusive and collectively exhaustive.
2. Compute the likelihood ($L_1, L_2...L_N$) that each of these $N$ different sets is considered by the individual using the thresholds ($T_1, T_2...T_{N-1}$) and the predicted distribution for the travel time to the stop.
3. Compute the conditional probability ($\pi_1, \pi_2... \pi_K$) for each of the different $K$ candidate locations using the calibrated model parameters and the values of exogenous variables specific to the decision maker under consideration.
4. Compute the probability ($P_1, P_2...P_K$) that each of the zones are chosen as the product of its conditional probability and the likelihood that the choice set containing the corresponding zone is selected.

5. Generate a uniformly distributed random number ($U$) between 0 and 1.
6. The chosen alternative is determined using the computed choice probabilities and the uniform random number drawn as follows:

   If $0 <= U < P_1$, chosen alternative is $A_1$.
   If $P_1 <= U < P_1+P_2$, chosen alternative is $A_2$.

   If $P_1+P_2+..P_{J-1} <= U < P_1+P_2+..P_J$, chosen alternative is $A_J$.
   If $P_1+P_2+..P_{K-1} <= U < 1$, chosen alternative is $A_K$.

### 3.2.7 Joint Unordered-Ordered Discrete Choice Models

A joint unordered-ordered discrete choice model may be used in the joint determination of two choice outcomes, in which, the alternatives are unordered for one of the choices and are ordered for the other. See Bhat (1997) for details on the econometric model structure. The modeling system implemented in CEMDAP allows for the mode choice for a tour/commute and the number of activity stops in the same tour/commute to be estimated jointly using this methodology. This section presents a method to predict the mode and number of stops using a joint unordered-ordered discrete choice model system.

1. Compute the joint probability of choosing mode $j$ (from among $K$ alternatives) and $n$ stops (from among $Nj$ alternatives) for each of the mode-stops combinations.
2. Generate a uniformly distributed random number ($U$) between 0 and 1.
3. The chosen combination of mode and number of stops is determined using the computed choice probabilities and the uniform random number drawn as follows in a manner similar to that described for multinomial logit models.

### 3.2.8 Simultaneous Equations Models

Simultaneous equations models can be used to model cases in which multiple continuous choice outcomes are jointly determined. The modeling system in CEMDAP allows for the joint determination of activity duration and the travel time to the stop using a simultaneous equations system. See for example Misra (1999) for a discussion of simultaneous equations models in this context. The logarithm of the time duration has been chosen as the explanatory variable. The methodology for determining activity duration and travel time using this methodology is discussed below:

1. Compute the expected value of (the log of) activity duration ($L_a$) using the calibrated model parameters and the values of explanatory variables specific to the decision maker under consideration.
2. Compute the expected value of (the log of) travel time ($L_t$) using the calibrated model parameters and the values of explanatory variables specific to the decision maker under consideration.
3. Draw from a bivariate normal distribution with means equal to the above computed values ($L_a$ *and* $L_t$) and known variance and covariance $(\sigma_a^2, \sigma_t^2, \rho_{at})$ from the calibration.

## 3.3 The Overall Micro-Simulation Platform for Predicting Activity-Travel Patterns

The previous section discussed the methods to predict individual choice outcomes. This section develops a framework for systematically and sequentially applying the different model components in order to construct the complete activity-string for individuals. The proposed modeling framework identifies individuals and households as the primary decision making units. Consequently the micro-simulation platform simulates one household at a time (Figure 3.1). The generation-allocation model system is first applied to the household. The scheduling model systems are then applied to each of the household adults. The activity-travel patterns of workers are simulated first, followed by the activity-travel patterns of nonworkers. The scheduling model system itself can be further subdivided into three major sequential model systems: the pattern-level model system, the tour-level model system and the stop-level model system.

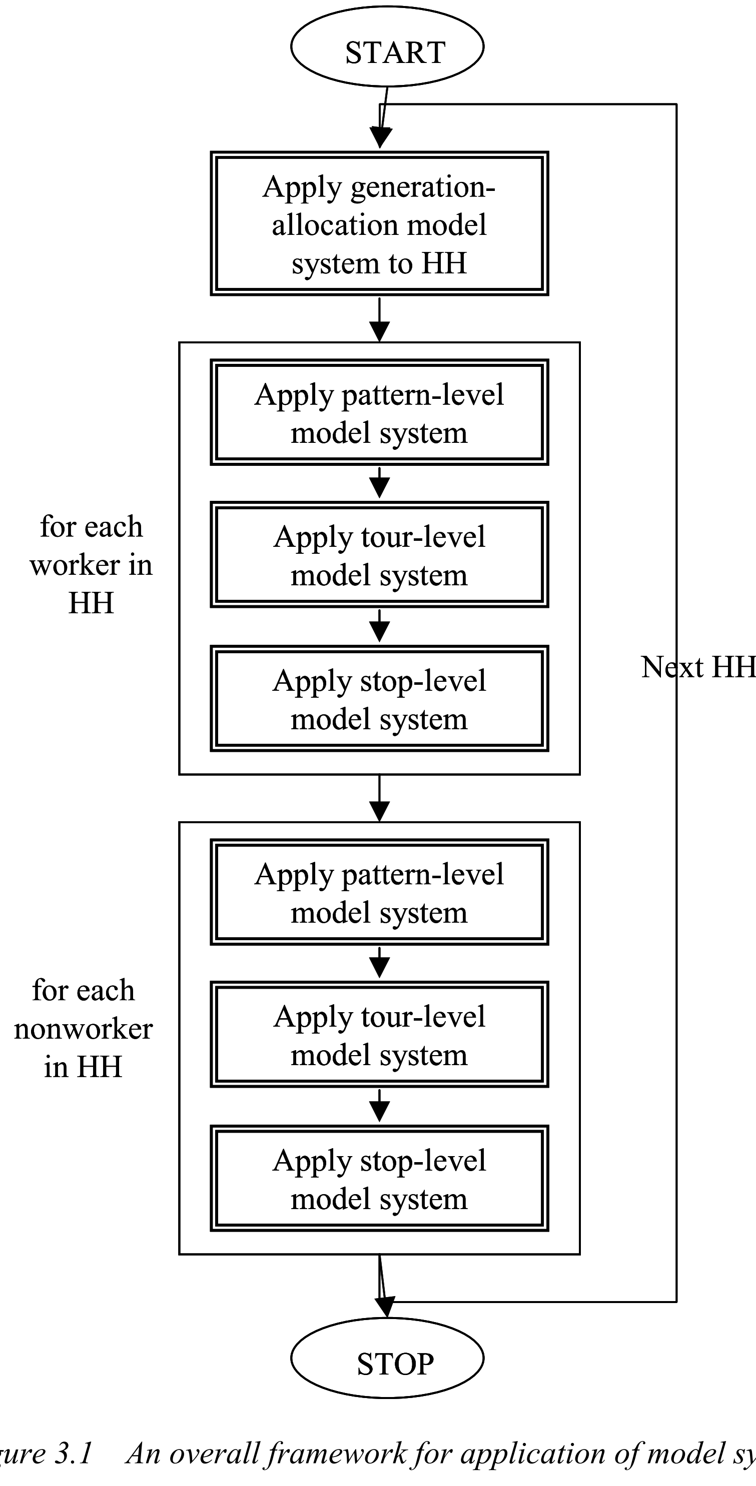


*Figure 3.1 An overall framework for application of model system*

### 3.3.1 Micro-Simulation Framework for Applying the Generation-Allocation Model System

The generation-allocation system models the decisions of the household adults to participate in different activity types during the day. The framework for applying this model system in forecasting is presented in Figure 3.2.

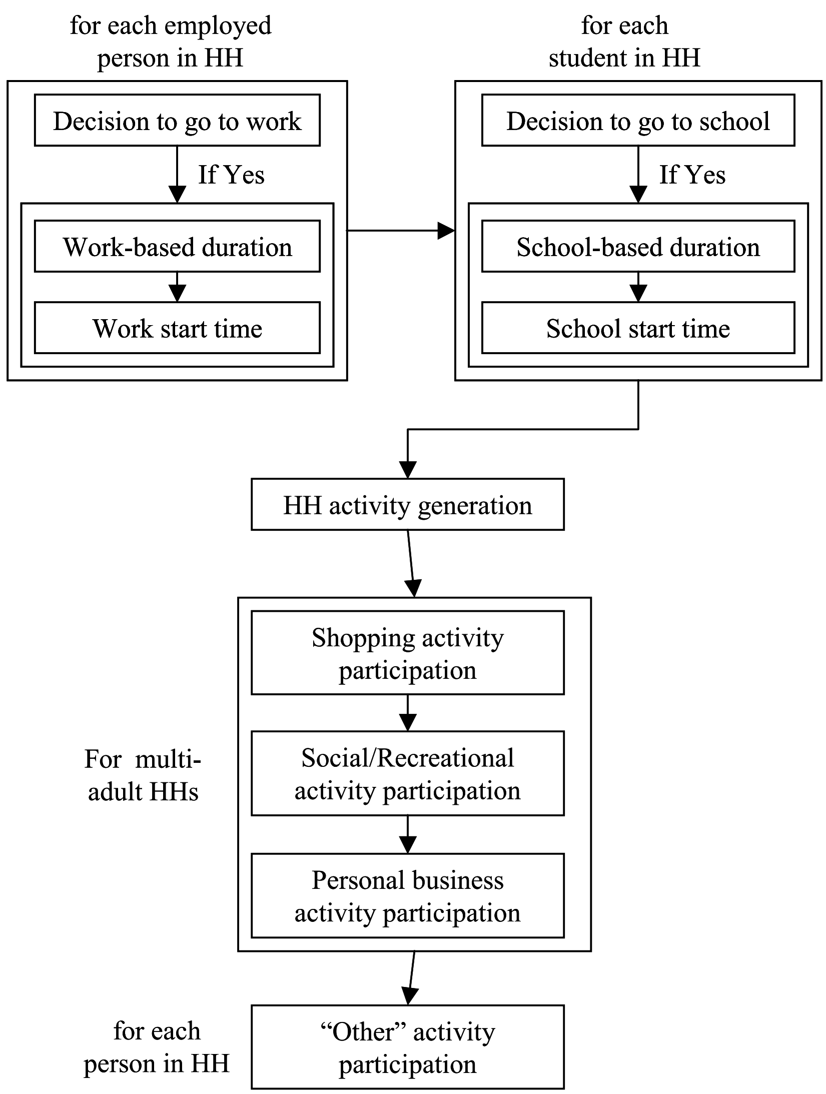


*Figure 3.2 Framework for application of the generation-allocation model system*

Decisions about subsistence activities form the highest level of models. For each employed adult in the household, the decision to go to work is first determined. If the person decides to travel to work, the work-based duration and the work start times are determined. The decision of students to go to school is then determined. Again, if the students do decide to go to school, the school-based duration and the school start-time are determined. For all scheduling models, employed persons who chose to go to work and students who chose to go to school are classified as "workers". The rest (adults who are neither students nor employed, employed persons who chose not to go to work and students who chose not to go to school) are classified as "nonworkers".

The household activity generation model is then applied to determine the decision of the household to undertake shopping, personal business and social/recreational activities for the day. If the household has only a single adult, activity allocation is trivial: this one adult in the household has to perform all the different activities generated. In the case of multi-adult households, activity-allocation models are applied for each of shopping, social/recreational and personal business activities. The models are applied conditional on the household deciding to undertake the activity during the day. The last model in the generation-allocation model system is the "other" activity participation model. This is applied to each adult in the household.

Thus, at the end of the application of the generation-allocation model system, the decision of each household adult to participate in different activities such as work (only for employed persons), school (only for students), shopping, social/recreational activities, personal business and "other" activities will be determined. In addition, the work start and end times (for employed persons who decided to go to work) and school start and end times (for students who decided to go to school) will also be determined.

### 3.3.2 Micro-Simulation Framework for Applying the Scheduling Model System for Workers

The scheduling model system for workers can be subdivided into three sequential model systems: the pattern-level model system, the tour-level model system and the stop-level model system. Frameworks for applying each of these three different model systems are discussed in detail subsequently.

The framework for the pattern-level model system for workers is presented in Figure 3.3. The work-to-home commute mode choice and the number of stops during this commute are determined first, either using a joint model or two independent models. A model for number of stops is applied only if the worker had decided to participate in activities other than work (or school in the case of students) during the day. The work-to-home commute duration is then determined. Next, the home-to-work commute is characterized. The travel mode is first determined, followed by the determination of number of stops during the commute to work (again, only if the worker had decided to participate in activities other than work), and finally the home-to-work commute duration is determined. If work is the worker's only activity for the day, the characterization of the worker's activity-travel pattern for the day is complete at this point.  However, if the worker had also decided to participate in other activities, the final pattern-level model is applied to predict the worker's decision to undertake tours during one or more of before-work, work-based and after-work periods.

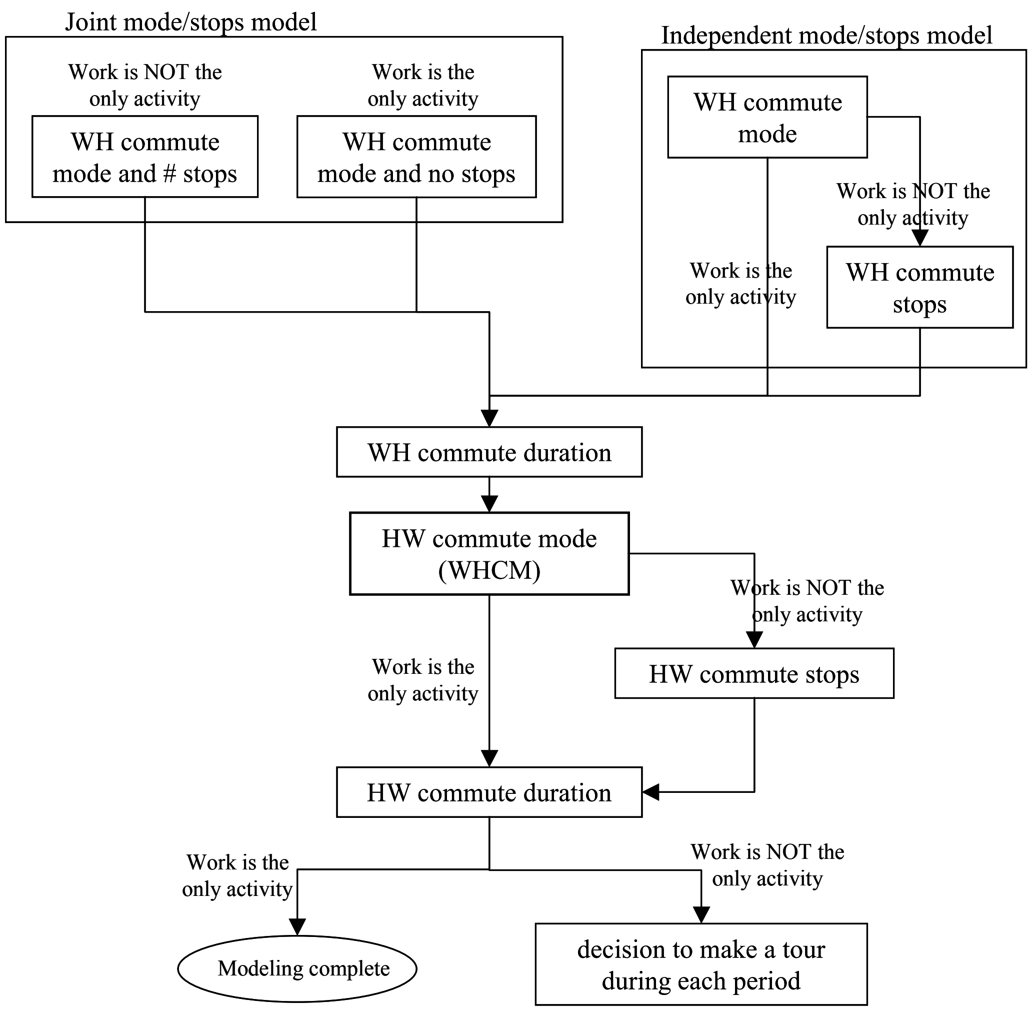


*Figure 3.3    Framework for application of the pattern-level model system for workers*

The framework for applying the tour-level model system for prediction is presented in Figure 3.4. The characteristics of a before-work (BW) tour are determined first, followed by the characterization of work-based (WB) and after-work (AW) tours (if one is made during each of these periods). The tour-level model system determines, sequentially, the travel mode, number of stops, total tour duration and the home-stay duration (or work-stay duration in the case of work-based tours) for each of the tours. The tour mode and number of stops may be determined simultaneously using a joint model or sequentially using independent models.

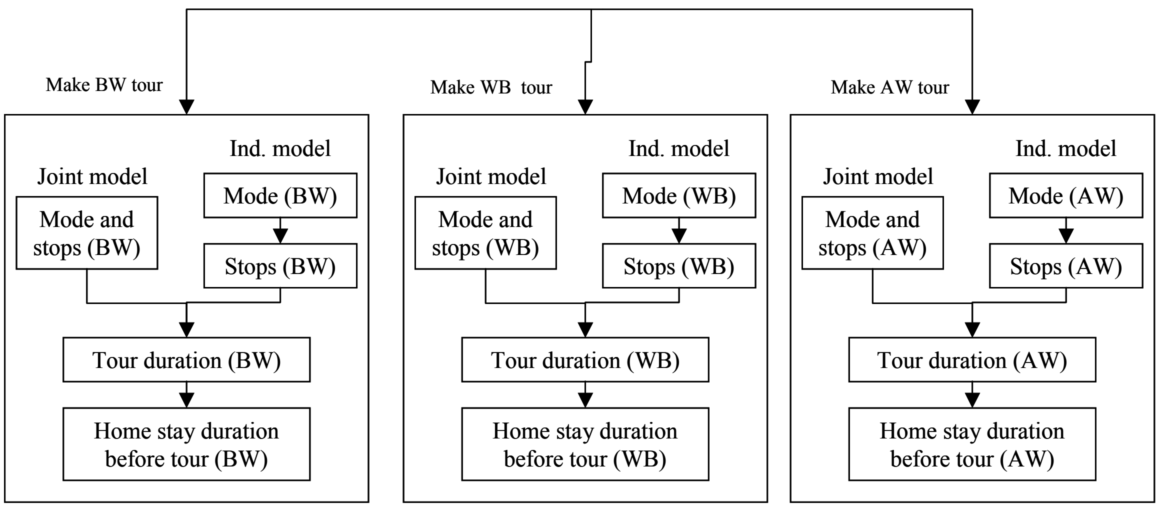


*Figure 3.4 Framework for application of tour-level model system for workers*

The framework for the stop-level model system is presented in Figure 3.5. Stops made during the work-to-home (WHC) and home-to-work (HWC) commutes are characterized first, followed by stops made as a part of any other tour (before-work, work-based and after-work). Within any tour or commute, the characteristics of stops are determined sequentially from the first to the last stop. For any stop, the activity type is first determined. Next, the activity duration and travel time to the stop are determined. The simulation platform allows for either a joint model or two independent models to determine activity duration and travel time. Finally, the location of the stop is determined using a spatial location choice model.

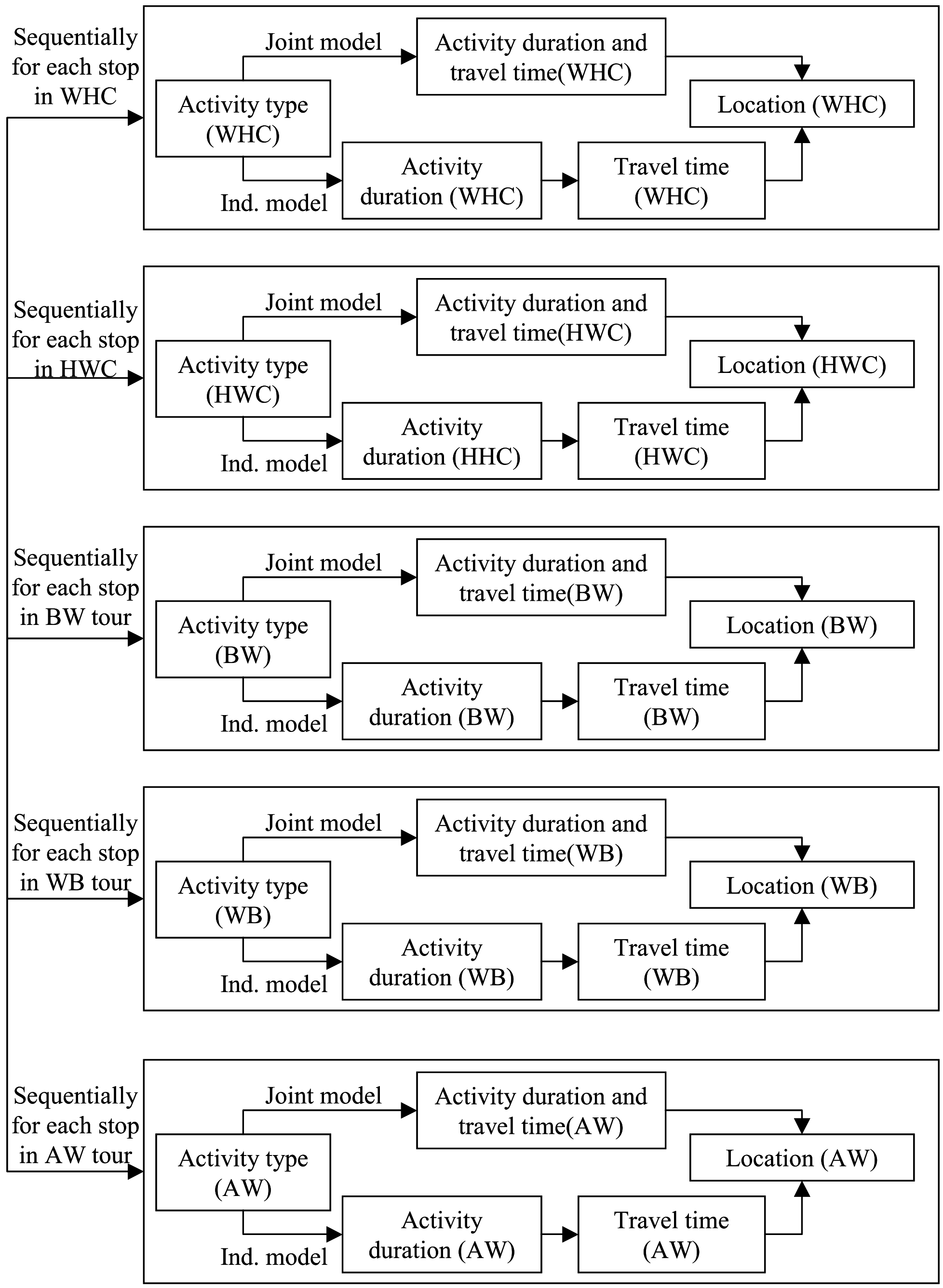


*Figure 3.5 Framework for application of stop-level model system for workers*

### 3.3.3 Micro-Simulation Framework for Applying the Scheduling Model System for Nonworkers

Analogous to the scheduling model system for workers, the scheduling model system for nonworkers can also be subdivided into three sequential model systems: the pattern-level model system, the tour-level model system and the stop-level model system.

Frameworks for applying each of these three different model systems are discussed in detail subsequently.

If the nonworker did not decide to participate in any activity (shopping, social/recreational, personal business or other) during the day, there are no scheduling decisions to be modeled. Hence the characterization of this person's activity-travel pattern is complete by noting that the person stays home all day. However, if the nonworker decided to participate in one or more activity types for the day, the total number of tours is determined. This is the only model in the pattern-level model system for nonworkers.

The framework for the applying the tour-level model system for nonworkers is presented in Figure 3.6. The characteristics of the different tours are determined sequentially from the first tour to the last. For any tour, the attributes determined are the travel mode, number of stops, total tour duration and the home-stay duration. The tour mode and number of stops may be determined simultaneously either using a joint model or sequentially using independent models.

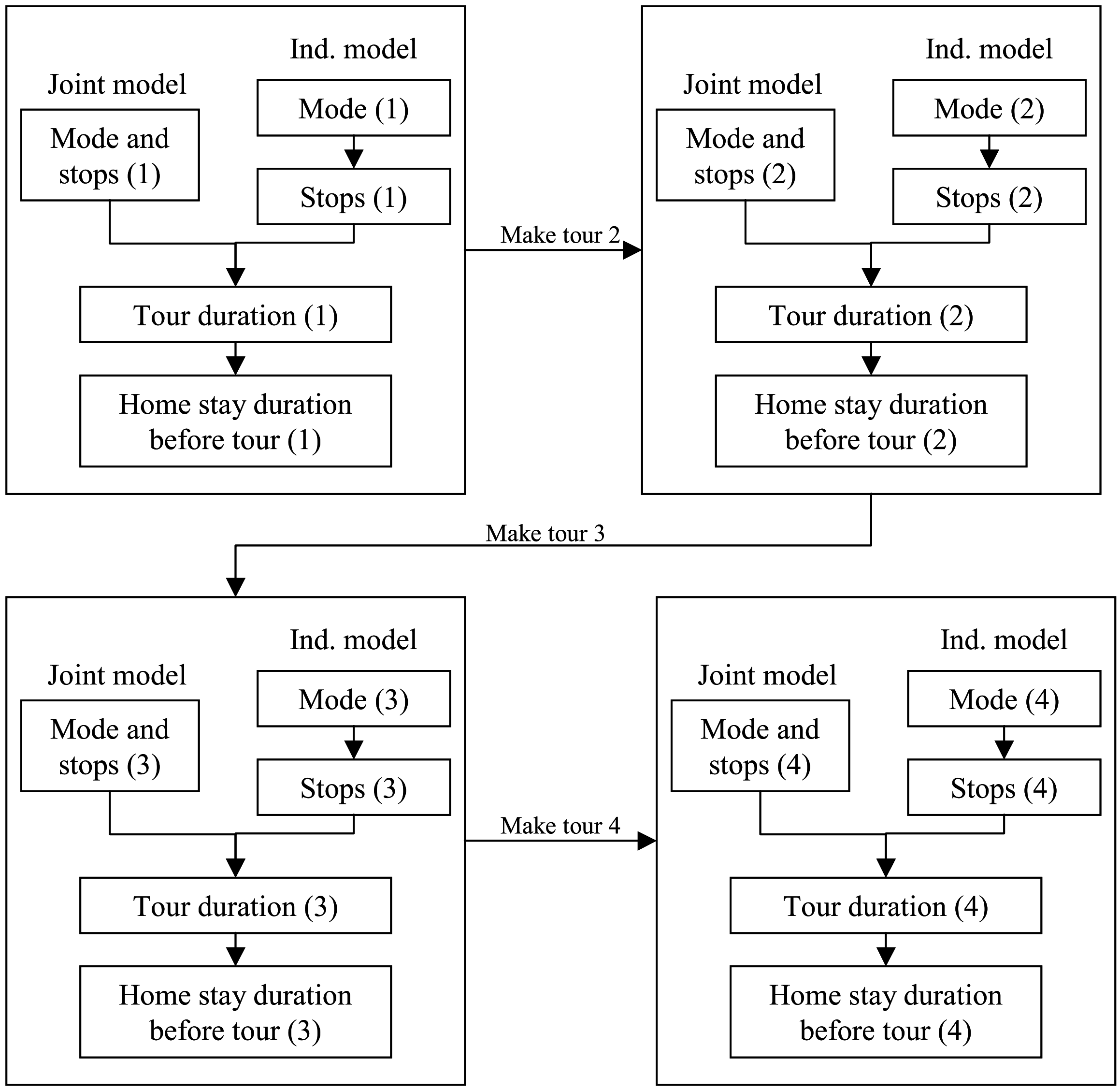


*Figure 3.6 Framework for applying the tour-level model system for nonworkers*

The framework for applying the nonworkers' stop-level model system for prediction is presented in Figure 3.7. Stops in the first tour are modeled initially, followed by stops in second, third and fourth tours (if any). Within any tour, the characteristics of stops are determined sequentially from the first to the last stop. The different attributes modeled are similar to the stop-level attributes modeled for the workers. The activity type is determined first. Activity duration and travel time to the activity are determined next, either using a joint model or two independent models. Finally, the location of the stop is determined using a spatial location choice model.

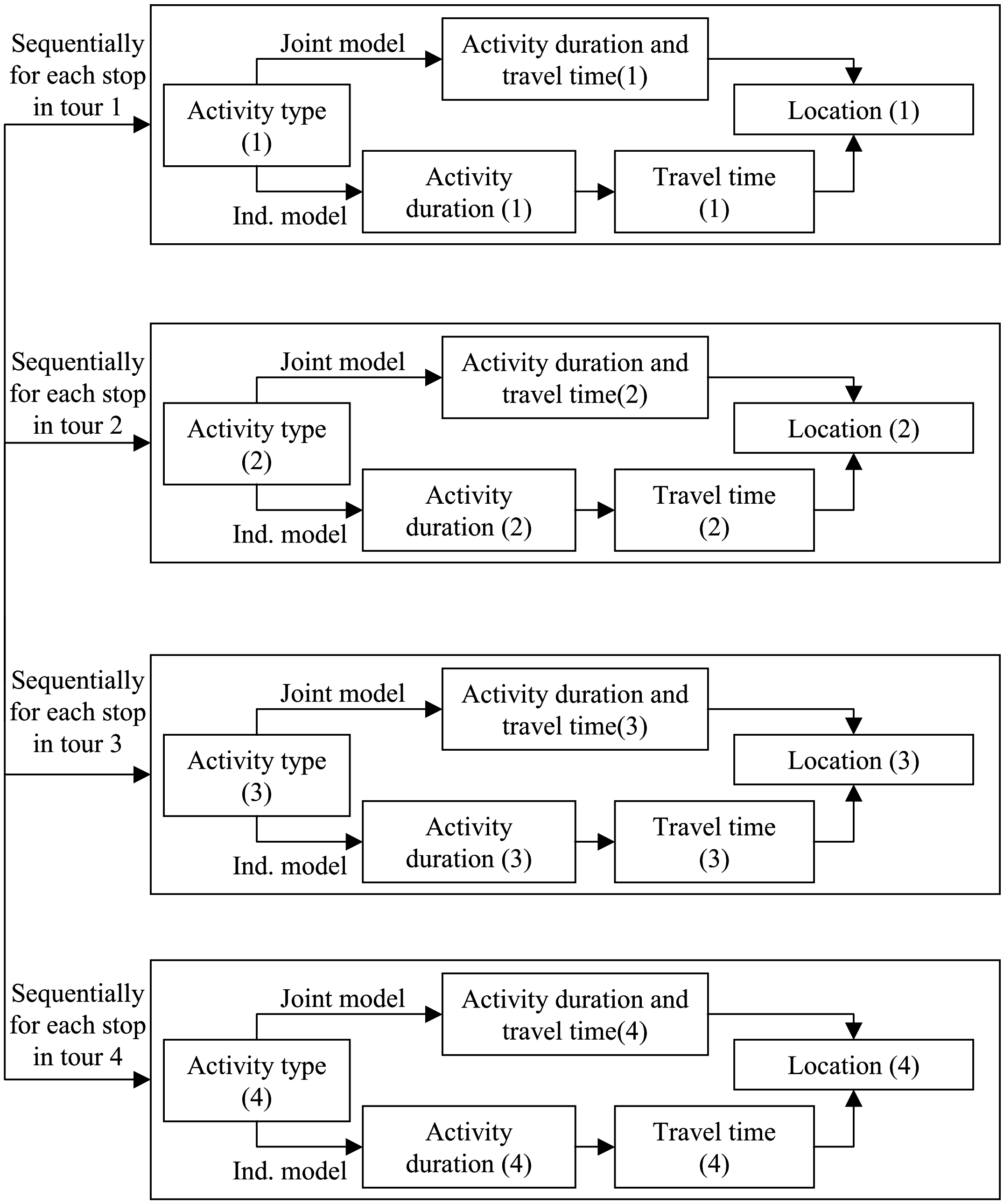


*Figure 3.7 Framework for applying the stop-level model system for nonworkers*

## 3.4 Consistency Checks

Several spatial and temporal consistency checks will be implemented in CEMDAP to ensure that the simulation process does not result in unreasonable or impossible activity patterns. Rules to ensure temporal consistency are discussed below. Rules to ensure spatial consistency are being developed and are not included in this report.

Most of the temporal choices (such as commute and tour durations, home-stay durations before tours, activity durations and travel times to stops) are determined using linear regression models. Since the chosen duration is determined by a random draw from a normal distribution, there is a small (but non-zero) possibility that the duration determined is either very high or very low. This may lead to situations in which the total predicted duration for a person exceeds 24 hours or the predicted end time of an activity falls after the predicted start time of the next activity. Rules for temporal consistency have been developed to handle cases in which the predicted duration is unreasonably high or low.

These rules are defined in terms of lower and upper bounds for each of the different durations that will be determined by the model system. If the predicted value of the duration falls below the lower bound, it is set to the lower bound; if it falls above the upper bound; it is set to the upper bound. The values were determined based on an empirical examination of data from the Dallas-Fort Worth area (DFW). In most cases, the 5-percentile value of the duration in the sample is chosen as the lower bound and the 95-percentile value chosen as the upper bound. Finally, it is also to be noted that models, in general, mimic properties of the sample which is being used to estimate them. Therefore, temporal consistency should first be ensured within the data sample. This was done in the case of the data from DFW.

The lower bound for work-based duration is set at 270 minutes (4.5 hours) and the upper bound at 720 minutes (12 hours). The work start-time shall be no earlier than 6 a.m. (or 180 minutes from 3 a.m.) and no later than 12:30 p.m. (or 570 minutes from 3 a.m.). Although not explicitly modeled, it will also be ensured that the work end-time is no later than 9 p.m. (or 1080 minutes from 3 a.m.). This ensured availability of time for after-work tours. The lower bound for school-based duration is set at 90 minutes and the upper bound at 600 minutes. The school start time is assumed to be no earlier than 7 a.m. (or 240 minutes from 3 a.m.) and no later than 5 p.m. (or 840 minutes from 3 a.m.). Again, an additional check ensures that the school end-time is no later than 9 p.m.

Table 3.1 presents bounds for the work-to-home and home-to-work commute durations. The bounds are presented in terms of percentages of available time, as opposed to absolute values. This will ensure, for example, that the predicted work-to-home commute duration is always less than the time duration from the end of work to the end of day (this is defined as the available time for work-to-home commute). The available time for home-to-work commute is defined as the time from 3 a.m. to the work start-time.

*Table 3.1 Bounds for WH and HW commute durations*

| Work-to-home (WH) commute | | | | |
|---|---|---|---|---|
| | Number of stops | | | |
| | 0 | 1 | 2 | >=3 |
| Lower bound | 0.92 | 3.45 | 5.04 | 7.01 |
| Upper bound | 11.11 | 44.46 | 52.57 | 68.71 |

| Home-to-work (HW) commute | | | |
|---|---|---|---|
| | Number of stops | | |
| | 0 | 1 | >=2 |
| Lower bound | 1.72 | 4.15 | 8.50 |
| Upper bound | 21.74 | 34.97 | 56.79 |

One observes an increase in both the lower and upper bounds with the increase in the number of stops. This is consistent with our expectation that with a greater number of stops, more time is required for the commute.

Table 3.2 provides the bounds on the tour durations and home-stay (or work-stay) periods before the tour for workers. Bounds are defined in terms of the percentage of available time. Available time for a before-work tour is the time from 3 a.m. until the departure to work. Available time for a work-based tour is the entire work-based duration and the available time for an after-work tour is the time from the arrival at home after-work to the end of day (3 a.m.). Available time for home-stay (or work-stay) before a tour is defined as the difference between the available time for the corresponding tour and the tour duration.

*Table 3.2 Bounds on tour and home-stay durations for workers*

| Before-work (BW) period | | | | |
|---|---|---|---|---|
| | Tour duration | Home-stay before tour | | |
| Lower bound | 1.35 | 44.70 | | |
| Upper bound | 49.67 | 95.73 | | |
| **Work-based (WB) period** | | | | |
| | 1 stop | | 2 or more stops | |
| | Tour duration | Work-stay before tour | Tour duration | Work-stay before tour |
| Lower bound | 3.40 | 32.90 | 6.24 | 6.15 |
| Upper bound | 26.47 | 74.91 | 57.29 | 89.60 |
| **After-work (AW) period** | | | | |
| | 1 stop | | 2 or more stops | |
| | Tour duration | Home-stay before tour | Tour duration | Home-stay before tour |
| Lower bound | 2.85 | 1.62 | 6.66 | 2.99 |
| Upper bound | 37.50 | 43.35 | 59.75 | 44.77 |

Bounds are provided separately for the before-work, work-based and after-work periods. Further, for work-based and after-work periods, separate bounds are provided depending on the number of stops in the tour. The bounds for the tour duration are observed to increase with an increase in the number of stops, consistent with expectations.

Table 3.3 presents bounds on activity duration and travel time to the different activity stops for workers. Again, the bounds are presented in terms of percentages of available time. For the first stop in a tour/commute, the available time is defined as the total duration of the tour/commute. For subsequent stops, available time is defined as the time from the end of the previous activity to the end of the tour/commute. Available time for travel to a stop is defined as the difference in the available time for the activity and the activity duration. The bounds are developed separately for each of the tours and commutes and are also based on the number of stops in the tour/commute.

*Table 3.3    Bounds for activity duration and travel time to the activity for workers*

Stops in before-work (BW) tour

| | Activity duration | Travel time to activity |
|---|---|---|
| Lower bound | 0.00 | 6.13 |
| Upper bound | 91.02 | 75.00 |

Stops in work-based (WB) tour

| | 1 stop | | 2 or more stops | |
|---|---|---|---|---|
| | Activity duration | Travel time to activity | Activity duration | Travel time to activity |
| Lower bound | 15.72 | 25.42 | 1.76 | 9.87 |
| Upper bound | 89.29 | 62.50 | 77.00 | 67.15 |

Stops in after-work (AW) tour

| | 1 stop | | 2 stops | | 3 or more stops | |
|---|---|---|---|---|---|---|
| | Activity duration | Travel time to activity | Activity duration | Travel time to activity | Activity duration | Travel time to activity |
| Lower bound | 10.00 | 33.33 | 3.88 | 3.83 | 0.71 | 1.72 |
| Upper bound | 90.63 | 60.00 | 85.84 | 66.67 | 90.35 | 73.68 |

Stops in work-to-home (WH) commute

| | 1 stop | | 2 stops | | 3 or more stops | |
|---|---|---|---|---|---|---|
| | Activity duration | Travel time to activity | Activity duration | Travel time to activity | Activity duration | Travel time to activity |
| Lower bound | 2.94 | 25.00 | 0.37 | 8.31 | 0.45 | 1.43 |
| Upper bound | 84.61 | 85.71 | 77.56 | 79.66 | 66.69 | 81.44 |

Stops in home-to-work (HW) commute

| | 1 stop | | 2 or more stops | |
|---|---|---|---|---|
| | Activity duration | Travel time to activity | Activity duration | Travel time to activity |
| Lower bound | 0.00 | 14.74 | 1.47 | 6.07 |
| Upper bound | 66.19 | 81.19 | 67.14 | 71.31 |

Table 3.4 presents the bounds on the duration for the first tour for nonworkers. The bounds are presented in terms of absolute duration  since, by definition, the entire day is available for this tour. Bounds are developed based on both the number of tours for the day and the number of stops in the first tour. One observes that the bounds are higher with an increasing number of stops in the tour.

*Table 3.4 Bounds on the duration for the first tour for nonworkers*

| | | 1 tour | 2 or more tours |
|---|---|---|---|
| 1 stop in tour | Lower bound | 20 | 13 |
| | Upper bound | 610.5 | 360.75 |
| 2 stops in tour | Lower bound | 24.4 | 43.75 |
| | Upper bound | 629 | 371.25 |
| 3 or more stops in tour | Lower bound | 85 | 66.2 |
| | Upper bound | 786 | 555 |

Table 3.5 presents the bounds on the home-stay duration before the first tour. This is presented in terms of percentage of available time, where available time is defined as the difference between the duration for the entire day (1440 minutes) and the duration for the first tour. Bounds on the home-stay duration are found to be lesser when the nonworker makes multiple tours in the day when compared to the bounds when the nonworker makes only a single tour.

*Table 3.5 Bounds on home-stay duration before tour 1 for nonworkers*

| | | 1 tour | 2 or more tours |
|---|---|---|---|
| 1 stop in tour | Lower bound | 37.14 | 30.72 |
| | Upper bound | 88.93 | 64.62 |
| 2 stops in tour | Lower bound | 38.00 | 33.72 |
| | Upper bound | 91.22 | 69.66 |
| 3 or more stops in tour | Lower bound | 43.79 | 37.29 |
| | Upper bound | 83.60 | 64.46 |

Table 3.6 presents the bounds on tour duration and home-stay duration as percentages of available time for the second tour for nonworkers. Available time for a tour is defined as the time from the end of the previous tour until the end of day. Available time for home-stay before a tour is defined as the difference between the available time for the corresponding tour and the actual tour duration.

*Table 3.6 Bounds on tour and home-stay duration for the second tour, nonworkers*

| | | Tour duration | Home-stay before tour |
|---|---|---|---|
| 1 stop in tour | Lower bound | 1.71 | 3.86 |
| | Upper bound | 39.80 | 73.77 |
| 2 stops in tour | Lower bound | 4.78 | 0.69 |
| | Upper bound | 54.65 | 71.17 |
| 3 stops in tour | Lower bound | 11.99 | 1.85 |
| | Upper bound | 75.11 | 71.25 |

The lower bound for the tour duration for tours 3 and 4 is set at 2.35% of the available time. The upper bound is set at 39.62%. In the case of home-stay durations before tours 3 and 4, the lower bound is 18.19% of available time and the upper bound is 88.54%.

Table 3.7 presents the bounds on the activity duration and the travel time to activity stops for nonworkers as percentage of available time. The definitions of available time are identical to that presented in the case for workers.

*Table 3.7 Bounds on activity duration and travel time to stops for nonworkers*

| Stops in tour 1 | | | | | | |
|---|---|---|---|---|---|---|
| | 1 stop | | 2 stops | | 3 or more stops | |
| | Activity duration | Travel time to activity | Activity duration | Travel time to activity | Activity duration | Travel time to activity |
| Lower bound | 4.89 | 31.19 | 1.34 | 4.66 | 2.89 | 3.10 |
| Upper bound | 92.32 | 66.67 | 84.87 | 75.86 | 85.47 | 64.38 |
| **Stops in tour 2** | | | | | | |
| | 1 stop | | 2 stops | | 3 or more stops | |
| | Activity duration | Travel time to activity | Activity duration | Travel time to activity | Activity duration | Travel time to activity |
| Lower bound | 6.01 | 39.45 | 4.26 | 6.41 | 0.20 | 3.69 |
| Upper bound | 92.28 | 66.15 | 81.63 | 75.36 | 81.40 | 73.75 |
| **Stops in tours 3 and 4** | | | | | | |
| | Activity duration | Travel time to activity | | | | |
| Lower bound | 4.86 | 6.14 | | | | |
| Upper bound | 90.23 | 60.46 | | | | |

# 4. Implementation Details and Input Data Requirements

The Comprehensive Econometric Micro-simulator for Daily Activity-travel Patterns (CEMDAP) is designed to serve as a user-friendly modeling platform for activity-based travel demand analysis. This chapter describes implementation details of this simulation software. The implementation platform is first identified, followed by a description of the system architecture. The different input data required are then identified and the methodology to assemble the inputs in the required format is finally discussed.

## 4.1 Implementation Platform

CEMDAP is designed using the object-oriented paradigm, which offers the advantage of code reuse and software extensibility. The current version of the software is implemented using the C++ language. More specifically, the software is written in Visual C++ using the Microsoft Visual Studio .NET development tool. CEMDAP will therefore be compatible with various MS Windows operating systems, including NT, 2000 and XP. Such Windows environments will have the ODBC (Open Database Connectivity) driver that provides a means for data access in CEMDAP.

## 4.2 System Architecture

CEMDAP provides a user-friendly interface to aid in configuring the mathematical model components in the system. This is achieved by allowing the user to specify the model structure and relevant parameters. The software then applies the model components in the predefined sequence to simulate the various choices of the individuals for whom data are provided as input. The output from CEMDAP is the activity travel patterns, which can be subsequently translated into trip sequences, predicted for these individuals. The software architecture of CEMDAP is shown in Figure 4.1. The various components in the system are described in more detail below.

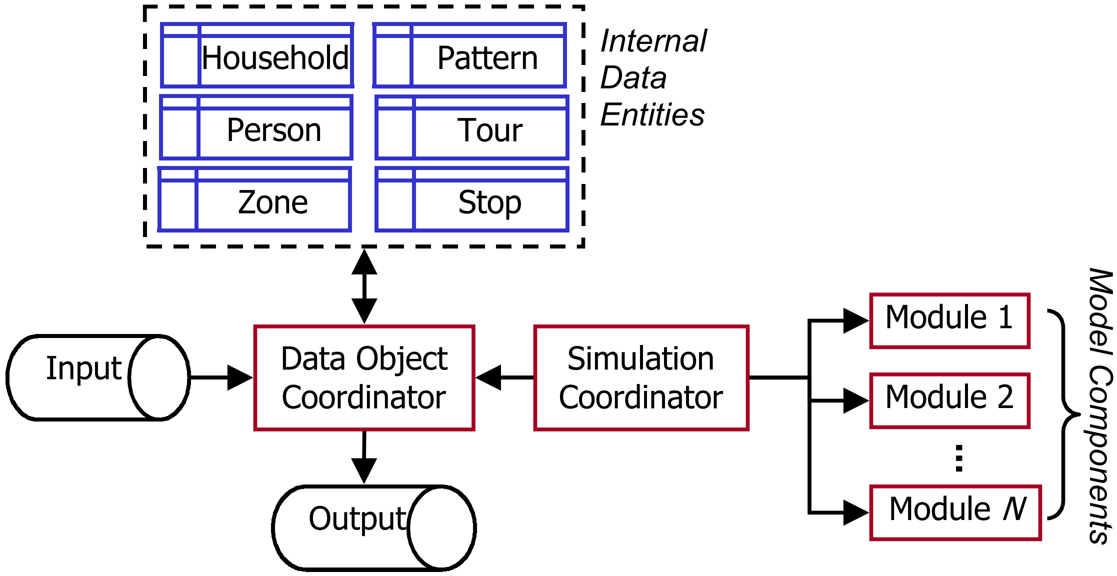


*Figure 4.1 Software architecture of CEMDAP*

### 4.2.1 Inputs and Outputs

Input data required for the simulation process may be classified into (1) disaggregate socio-economic characteristics of the population, (2) aggregate zonal-level land-use and demographic characteristics and (3) zone-to-zone transportation system level-of-service characteristics by time of day. All these required input data are organized as relational databases in MS Access format. CEMDAP produces the activity-travel patterns of individuals as output in a relational database in MS Access format.

### 4.2.2 Data Entities

These are the main data structures that the simulator operates upon internally. Entities such as household, person and zone are produced from the input data files. The other entities (pattern, tour, and stop) are created during the simulation process.

### 4.2.3 Modeling Modules

Each modeling module in the system corresponds to a model component in the activity-based travel analysis framework. Once a module is configured by the user via the user interface, it possesses the knowledge about the structure and all the relevant parameters required to produce the desired endogenous variable. Once called upon, the module executes a forecasting algorithm to predict the corresponding choice and this predicted choice is returned as a result.

### 4.2.4 Data Object Coordinator

In CEMDAP, the models do not communicate directly with each other; rather, they communicate via shared data entities that are managed by the data object coordinator. The data object coordinator is responsible for establishing the connection with external databases. It extracts the content and structural information of the data tables to create the data entities and also to supply such information to the user front end for the purpose of model configuration. During the simulation process, the coordinator keeps track of the current household and person records for which the activity travel pattern is to be simulated. Any updates to existing data entities are performed via the coordinator.

### 4.2.5 Simulation Coordinator

The simulation coordinator is responsible for controlling the flow of the simulation. It coordinates the sequence and logic in which the modeling modules are called upon. It also performs any consistency checks as required. Once a choice outcome is predicted and checked for consistency, the simulation coordinator updates corresponding data entities through the data object coordinator.

## 4.3 Preparing Input Data

As described in the system architecture, the simulator requires data to be provided in a pre-determined format. Zonal-level land use and demographic data (such as total population, total number of households, median income, number of jobs in basic, service and retail industries, and the percentage of area covered by each land-use type) and the network level-of-service data (such as distance between zonal pairs, travel time by mode and time of day, transit availability) are generally available at the desired level of aggregation. Hence not much pre-processing is required in preparing these data tables. These only need to be converted into relational databases and imported into MS Access.

As the various activity-travel choices of individuals and households are simulated in CEMDAP, detailed disaggregate socio-economic characteristics of the population are also required as inputs. Household characteristics required may include number of adults and children, number of vehicles, household structure, etc.; and person-level characteristics may include age, gender, education level, employment status, marital status, etc. In general, the explanatory variables used in the model components embedded within the micro-simulator dictate the sociodemographic data requirements.

Typically, detailed individual and household demographic data are not available for the entire population. Hence, a population synthesis procedure is required to prepare this data for input. One of the methods of creating synthetic baseline populations using Iterative Proportional Fitting (IPF) is discussed in Beckman et al. (1996). The basic technique is to use summary tables from the Census data in conjunction with the Public Use Microdata Sample (PUMS) to estimate the proportion of households in a census tract with a desired combination of demographics. Households are generated by selection of households from the associated PUMS according to these proportions. This is just one of many possible approaches to population synthesis (Miller, 1996). For our purposes, we use the IPF methodology to generate a synthetic population. Separate software is being developed for this purpose. For the purpose of demonstrating the application of the micro-simulator, the research group will synthesize population using the developed software and data from the DFW area.

# 5. Summary

There has been an increasing realization in the travel-demand modeling field that the conventional trip-based approach needs to be replaced with an activity-based approach that is behaviorally oriented. Several comprehensive activity-based systems have been developed in the recent past. The current research aims at advancing the state of the art in activity-based modeling by addressing the activity patterns of both workers and nonworkers within a household. A comprehensive modeling framework for determining the complete activity-travel patterns of individuals on a continuous time domain has been developed. The different model components have been calibrated using travel survey data from the Dallas-Fort Worth area. These have been described in detail in previous research reports.

The objective of this report is to present a micro-simulation-based methodology that uses a calibrated model system to predict activity-travel patterns of individuals. The report begins with a detailed description of the representation frameworks developed for workers and nonworkers. The different components in the overall modeling system and their econometric structures are then listed. Chapter 3 describes the philosophy of a micro-simulation-based approach to predicting choice outcomes using a calibrated model system. Algorithms for predicting individual choice instances using the various types of econometric models embedded within the simulator are then described. This chapter also provides the flowchart for an overall procedure that would systematically simulate the different individual choice instances and integrate these individual decisions into the complete activity-travel string of an individual. Temporal consistency checks to be performed are also discussed in detail. In addition, the simulator will also perform checks on spatial consistency. The rules for these checks are being developed.

Chapter 4 describes the implementation details of the simulation software, called CEMDAP. The software will be developed in Visual C++ using the object-oriented programming methodology. MS Access has been chosen as the data base software. The design architecture for the software is discussed in this chapter. The input data required is identified and methods to assemble this data in the required format are also discussed. In particular, the methodology of synthetic population generation to determine detailed disaggregate socio-economic characteristics is described.

Development of the software is underway. The software will be developed to be user friendly and adequate help features will be provided. User and reference manuals will also be written. Assembly of input data for the Dallas-Fort Worth area is also underway. This data will be used in the testing of the code and also to demonstrate the use of the software for TCM evaluations.

# References


Axhausen, K. and Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models and research problems. Transport Reviews, 12, 324-341.

Beckman, R.J., Baggerly, K.A., and McKay, M.D. (1996). Creating synthetic baseline populations. Transportation Research Part A, 30(6), 415-429.

Ben-Akiva, M. and Lerman, S. (1985). Discrete-choice analysis: theory and application to travel demand. MIT Press, Cambridge, MA.

Bhat, C.R. (1996). A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity. Transportation Research Part B, 30, 189-207.

Bhat, C.R. (1997). Work travel mode choice and the number of non-work commute stops. Transportation Research Part B, 31, 41-54.

Bhat, C.R. and Misra, R. (2001). A comprehensive and operational econometric modeling framework for analysis of the of the activity-travel patterns of nonworkers. 9th World Conference on Transport Research (WCTR), Seoul, Korea.

Bhat, C.R. and Misra, R. (2002). Comprehensive activity-travel pattern modeling system for nonworkers with empirical focus on the organization of activity episodes. Transportation Research Record, 1777, 16-24.

Bhat, C.R. and Singh, S.K. (2000). A comprehensive daily activity-travel generation model system for workers. Transportation Research Part A, 34, 1-22.

Bhat, C.R, Srinivasan, S., and Guo, J.Y. (2002). Activity-based travel-demand modeling for metropolitan areas in Texas: Data sources, sample formation and estimation results. Research Report 4080-3, Center for Transportation Research, Austin, Texas.

Gordon, P., Kumar, A., and Richardson, H.W. (1988). Beyond the journey to work. Transportation Research Part A, 21(6), 419-426.

Greene, W.H. (2000). Econometric Analysis. Prentice Hall, New Jersey.

Guo, J.Y. and Bhat, C.R. (2001). Representation and analysis plan and data needs analysis for the activity-travel system. Research Report 4080-1, Center for Transportation Research, Austin, Texas.

Hanson, S. (1980). Spatial diversification and multipurpose travel: implications for choice theory. Geographical Analysis, 12, 245-257.

Hensher, D.A., and Mannering, F.L. (1994). Hazard-based duration models and their application to transport analysis. Transport Reviews, 14(1), 63-82.

Jones, P.M., Koppelman, F.S., and Orfeuil, J.P. (1990). Activity analysis: state of the art and future directions. In Developments in Dynamic and Activity-Based Approaches to Travel Analysis, 34-55, Gower, Aldershot, England.

Kiefer, N.M., (1988). Economic duration data and hazard functions. Journal of Economic Literature, 26(2), 646-679.

Lockwood, P.B. and Demetsky, M.J. (1994). Nonwork travel - A study of changing behavior. Presented at the 73rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Madalla, G.S. (1999). Limited-dependent and qualitative variables in econometrics. Econometric Society Monographs No. 3, Cambridge University Press, Cambridge.

Miller, E.J. (1996). Microsimulation and activity-based forecasting. Activity-Based Travel Forecasting Conference Proceedings, June 2-5, 1996.

Misra, R. (1999). Toward a comprehensive representation and analysis framework for the nonworker activity-travel modeling. Ph.D. dissertation, Department of Civil Engineering, The University of Texas at Austin.