Technical Report Documentation Page

1. Report No. FHWA/TX-07/0-4080-8	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Activity-Based Travel-Demand Analysis for Metropolitan Areas in Texas: CEMDAP Models, Framework, Software Architecture and Application Results		5. Report Date October 2006 6. Performing Organization Code
7. Author(s) Abdul Pinjari, Naveen Eluru, Rachel Copperman, Ipek N. Sener, Jessica Y. Guo, Sivaramakrishnan Srinivasan, Chandra R. Bhat.		8. Performing Organization Report No. 0-4080-8
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3208 Red River, Suite 200 Austin, TX 78705-2650		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 0-4080
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, TX 78763-5080		13. Type of Report and Period Covered Technical Report 14. Sponsoring Agency Code
15. Supplementary Notes Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.		
16. Abstract This report describes the modeling and software enhancements of the earlier version of CEMDAP (the activity-travel simulator that simulates the detailed activity-travel patterns of the population) and presents the application results for the Dallas-Fort Worth (DFW) region.		
17. Key Words Activity-based analysis, analysis frameworks, econometric models, empirical results		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161; www.ntis.gov.
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 212 22. Price

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

Activity-Based Travel-Demand Analysis for Metropolitan Areas in Texas: CEMDAP Models, Framework, Software Architecture, and Application Results

Abdul Pinjari
Naveen Eluru
Rachel Copperman
Ipek N. Sener
Jessica Y. Guo
Sivaramakrishnan Srinivasan
Chandra R. Bhat

CTR Technical Report:	0-4080-8
Report Date:	October 2006
Project:	0-4080
Project Title:	Second Generation Activity-Based Travel Modeling System for Metropolitan Areas in Texas Accommodating Demographic, Land Use, and Traffic Microsimulation Components
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Center for Transportation Research
The University of Texas at Austin
3208 Red River
Austin, TX 78705

www.utexas.edu/research/ctr

Disclaimers

Engineering Disclaimer

Project Engineer: Chandra R. Bhat
Professional Engineer License State and Number: Texas No. 88971
P. E. Designation: Research Supervisor

Acknowledgments

Research performed in cooperation with the Texas Department of Transportation, North Central Texas Council of Governments, and the U.S. Department of Transportation, Federal Highway Administration. The authors appreciate the help of the North Central Texas Council of Governments (NCTCOG) travel modeling staff (especially Ken Cervenka, Arash Mirzaei, Bin Chen, and Francisco Torres) for CEMDAP sensitivity testing efforts, providing the data for model estimation, and their overall support of this research effort. The authors also express appreciation to Bill Knowles, Janie Bynum, Jack Foster, and Greg Lancaster for their valuable input throughout the course of the project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Conventional wisdom has long indicated that demographics, land use, and transportation are intimately linked. While demographics represent the characteristics of decision makers and land use represents the spatial pattern of urban development and activities, transportation serves as the mechanism for spatial interaction between geographically dispersed activity sites. Recognizing these linkages among demographics, land use, and transportation is important for realistic forecasts of travel demand. To achieve this, the current research project develops a demand-forecasting approach that captures land-use and travel behavior in an integrated way, while accommodating the moderating role of individuals' demographic characteristics. This behavioral approach entails integrating activity-based travel models with disaggregate models that capture the population demographic processes, the households' long-term choice behaviors, and the economic markets in which the households act.

The proposed activity-based land-use transportation modeling system is labeled CEMDAP-II (Second Generation Comprehensive Econometric Micro-simulator of Daily Activity-Travel Patterns). As depicted in Figure 1.1, CEMDAP-II takes as input the aggregate sociodemographics and the activity-travel environment characteristics for the base year, different policy actions (scenarios) for future years, and relevant externally estimated model parameters. The aggregate sociodemographic data are first run through the Synthetic Population Generator (SPG) to create a disaggregate representation of all individuals and households in the study area. The activity-travel simulator, CEMDAP, then takes the disaggregate data as input and produces as output the detailed activity-travel characteristics for each individual. These then feed into a traffic micro-assignment simulator to determine the network link flows and speeds by time of day. The evolution of the population and the urban environment is modeled by the Comprehensive Econometric Microsimulator for Socioeconomics, Land-Use, and Transportation System (CEMSELTS). Taking as input the current sociodemographics and activity-travel characteristics, prescribed policy actions, and speed characteristics obtained from the traffic micro-assignment processor, CEMSELTS provides as output sociodemographic characteristics of the population and the attributes of the activity-travel environment for a time increment into the future (e.g.,1 year). This information feeds back into the activity-travel simulator (CEMDAP) to obtain the detailed individual activity-travel characteristics for the future year. The loop is executed until the link flows and speeds are obtained for the forecast year specified by the analyst. The effects of the prescribed policy actions can then be evaluated based on the simulated network flows and speeds for any year between the base year and the forecast year.

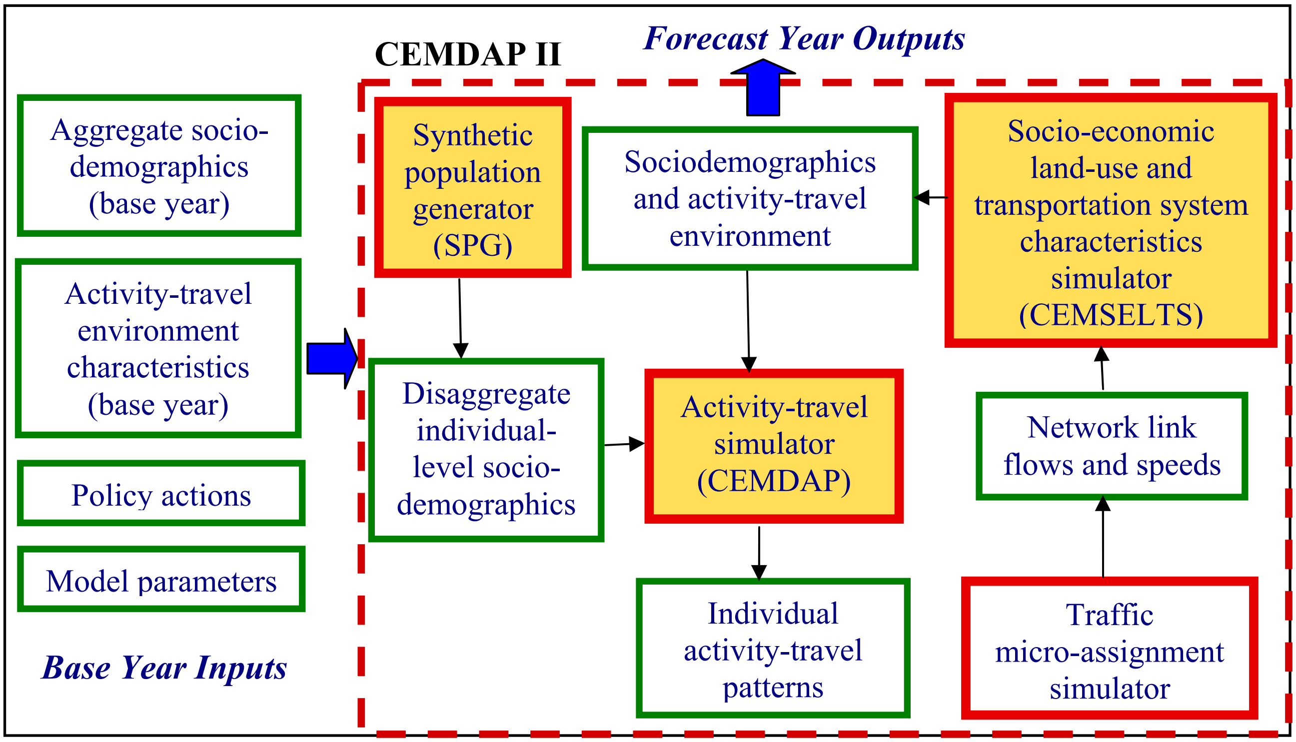

Figure 1-1 The Structure of CEMDAP II

Within the overall framework of CEMDAP-II, the focus of the current report is on the latest version of CEMDAP, the activity-travel simulator. Specifically, this report documents the following: (1) the modeling and software enhancements to CEMDAP, (2) the generation of the inputs for CEMDAP using software components SPG and CEMSELTS, and (3) the empirical validation of CEMDAP and the results of sensitivity testing carried out using CEMDAP.

The report is organized as follows. Chapter 2 describes the econometric modeling system and the microsimulation framework embedded within the latest version of CEMDAP. Chapter 3 describes the software features of CEMDAP, including the object-oriented approach, the software architecture, and the software enhancements implemented in the recent version of CEMDAP. Chapter 4 presents details of generating and verifying the synthetic population for the base year (year 2000) and forecast year (year 2025). Chapter 5 discusses the implementation of CEMSELTS to generate the disaggregate household and person level inputs required for

CEMDAP. Chapter 6 presents the empirical validation of CEMDAP and the results of sensitivity testing undertaken using CEMDAP. Chapter 7 summarizes the report.

2. ENHANCED CEMDAP SYSTEM

This chapter describes the new econometric modeling system and the microsimulation framework embedded within the latest version of CEMDAP. This new modeling system enhances the previous system in several ways. First, the new system is developed at a finer spatial resolution and applied to a 4,874-zone system for the Dallas–Fort Worth (DFW) area in Texas. Second, the activity-travel patterns of children (persons under 16 years of age) are now explicitly modeled and forecasted. Third, the interdependencies between the travel patterns of children and their parents (such as escort to and from school and joint participation in discretionary activities) are explicitly accommodated. Finally, for estimation of the models, the raw survey data obtained for the DFW area were reprocessed to create a larger sample and all the model components (over fifty in all) were re-estimated.

The reader will note here that the design and architecture of CEMDAP is generic. In particular, CEMDAP can be applied to any metropolitan area, as long as local area models are estimated to produce the appropriate sensitivity parameters. Currently, we have estimated all the CEMDAP models using the DFW data and the resulting specifications and parameters are embedded in CEMDAP as default specifications and parameters. Moreover, the user can use the graphical interface of CEMDAP to modify the specifications and parameter values if local area specifications and parameters are available (see the CEMDAP user manual by Bhat et al. (2006), for details on modifying the specifications). CEMDAP has also been designed to provide a friendly diagrammatic interface to help the user understand the logic of the system.

The remainder of this chapter is organized as follows. Section 2.1 describes the representation frameworks used to characterize the complete activity-travel patterns of individuals. Specifically, this section identifies all the choice elements that are predicted within CEMDAP to construct the activity-travel patterns of all household members, including both adults and children. Section 2.2 focuses on the econometric modeling system used for daily activity-travel prediction. Section 2.3 describes the data used in the empirical model estimations. Section 2.4 presents, in detail, the microsimulation procedure implemented within CEMDAP. Section 2.5 discusses the spatial and temporal consistency checks implemented within CEMDAP to ensure that the simulation process does not result in unreasonable or impossible activity travel patterns.

2.1 Representation Frameworks

This section describes the representation frameworks developed to describe the activity-travel patterns of individuals. These representation frameworks identify the complete set of attributes that are required to characterize an individual's daily activity-travel pattern. The simulation of an individual's activity-travel pattern then entails computing a predicted value for each of these attributes based on the underlying econometric models.

Broadly, the activity-travel pattern of an individual is defined as the sequence of activities and travel pursued during a day. Among all the different activities that an individual undertakes during the day, the work and school activities are undertaken under the greatest space-time constraints for most individuals. Also, participation in these activities significantly influences an individual's participation in all other activities during the day. Consequently, separate representations have been developed to characterize the daily activity-travel patterns of workers, students, non-workers, and non-students. The workers and students include adults (persons aged 16 years or older) who go to work or school and children (persons aged 15 years or younger) who go to school. The non-workers and non-students, on the other hand, include adults who neither go to work nor attend school during the day, as well as children who do not go to school during the day. For presentation ease, in the remainder of this section, we will use the term "workers" to represent workers and students and the term "non-workers" to represent non-workers and non-students. Similarly, the term "work" will be used generically to refer to either work or school as appropriate.

The representation frameworks for workers and non-workers are discussed in Sections 2.1.1 and 2.1.2, respectively. In both frameworks, the start of the day is defined as 3:00 a.m. and all individuals are assumed to be at home at this time.

2.1.1 Representation for the Activity-Travel Pattern of Workers

The daily pattern of workers is characterized by four different sub-patterns: (1) before-work pattern, which represents the activity-travel undertaken before leaving home to work; (2) commute pattern, which represents the activity-travel pursued during the home-to-work and work-to-home commutes; (3) work-based pattern, which includes all activity and travel undertaken from work; and (4) 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 the 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 of the tours, the home-to-work commute, and the work-to-home commute may include several activity stops. An activity 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.

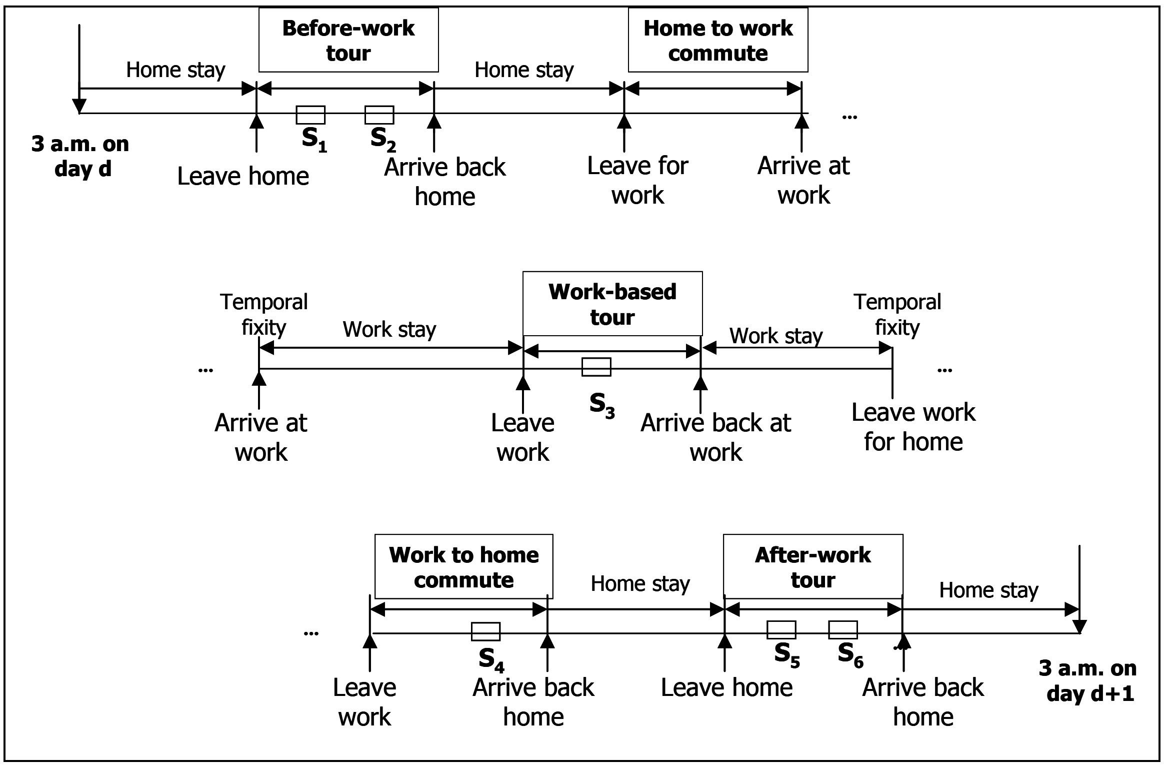

Figure 2-1 A Representation of the Activity-Travel Patterns of Workers

The characterization of the complete workday activity-travel pattern is accomplished by identifying a number of different attributes. The **primary attributes** that characterize the pattern of a worker are the start and end times of the work activity. The remaining attributes may be classified based on the level of representation that 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, as well as the number of tours that the worker undertakes during each of the before-work, work-based, and after-work periods. **Tour-level attributes** include travel mode, number of stops, home-stay duration (or work-stay duration, in the case of the work-based tour) before the tour, and the sequence number of the tour within the before-work, work-based, and after-work periods. **Stop-level attributes** include activity type pursued, whether the activity at the stop is done alone or with other household members (and with which household members), duration of the activity stop, travel time to stop, whether the travel to the stop is undertaken alone or with other household members (and with which household members), stop location, and the sequence of the stop in a tour or commute.

The representation described above is generic and can be used to describe any worker 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 development of CEMDAP. Specifically, in the case of adults who go to work or school, CEMDAP is designed to handle up to three tours during each of the before-work, work-based, and after-work periods and up to five stops during any tour or commute. In the case of school-going children, CEMDAP accommodates non-school activity participation of children only during the school-to-home commute and the after-school period. Further, only a single tour with one stop is supported for the after-school period.

2.1.2 Representation of the Activity-Travel Patterns of Non-Workers

In the case of non-workers, the activity-travel pattern is considered as a set of out-of-home activity episodes (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.

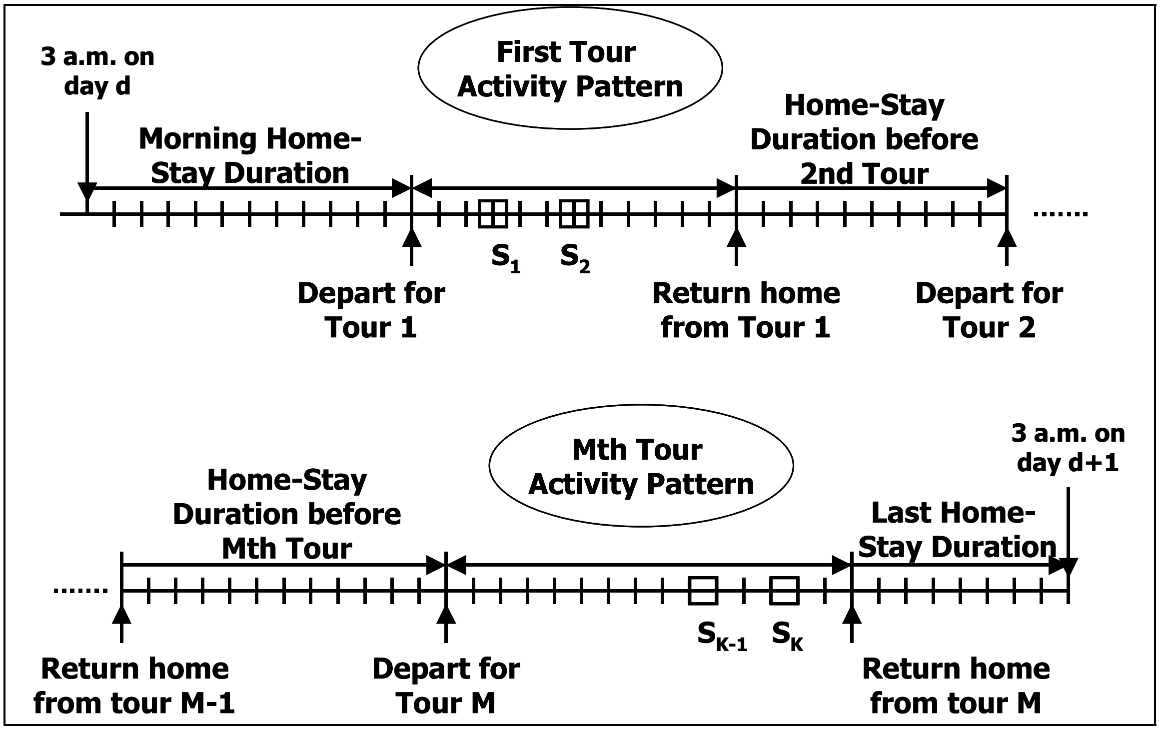

Figure 2-2 A Representation of the Activity-Travel Patterns of Non-Workers

A non-worker'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, the number of stops in the tour, the home-stay duration before the tour, and the sequence of the tour in the day. **Stop-level attributes** include activity type, whether the activity at the stop is done alone or with other household members (and with which household members), duration of the activity, travel time to stop, whether the travel to the stop is undertaken alone or with other household members (and with which household members), location, and the sequence of the stop in a tour or commute.

The representation described above is generic and can be used to describe any non-worker 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. Specifically, CEMDAP is designed to handle up to a total of four tours and up to five stops during each tour.

2.2 Econometric Modeling System

This section identifies all the model components that constitute the overall modeling system implemented within CEMDAP. Each model corresponds to the determination of one or more of the attributes characterizing the activity-travel pattern of a worker or a non-worker. Together, the set of all models identified in this section, once estimated, can be used in a systematic predictive fashion to completely characterize the activity-travel patterns of all individuals in a household. (The systematic prediction procedure is described in Section 2.4.)

The overall modeling system is broadly subdivided into the following five categories: (1) the generation-allocation model system (Table 2.1), (2) the worker scheduling model system (Table 2.2), (3) the non-worker scheduling model system (Table 2.3), (4) the joint discretionary tour scheduling model system (Table 2.4), and (5) the children scheduling model system (Table 2.5). The precise econometric structure and the choice alternatives for each of the model components are also identified in Tables 2.1 through 2.5. Further, a unique identifier is associated with each model. (For example, "GA1" identifies the first model within the "generation-allocation" category, which is the decision of a child to go to school.) To facilitate easy cross-referencing, these identifiers have also been included in the figures presented in Section 2.4 (which describe the prediction procedure), as well as in Appendix A (where the estimation results for each model component are presented). The reader will also note that not all models in the tables are applicable to all households and individuals, as we discuss further in Section 2.4.

It can be observed from Tables 2.1 through 2.5 that the econometric structure for each choice dimension being modeled in CEMDAP falls under one of the six econometric model categories: binary logit, multinomial logit, hazard-duration, regression, ordered probit, and spatial location choice. The mathematical model structures of these model categories are provided in research Report 4080-2 (Bhat et al. 2001).

Table 2.1 The Generation-Allocation Model System

Model Id	Model Name	Econometric Structure	Choice Alternatives	Comments
GA1 GA2 GA3	Children's decision to go to school Children's school start time (time from 3 a.m.) Children's school end time (time from school start time)	Binary logit Hazard-duration Hazard-duration	Yes, No Continuous time Continuous time	Applicable only to children who are students. The determination of whether or not a child is a student is made in the CEMSELTS module (see Chapter 5).
GA4 GA5 GA6	Decision to go to work Work start and end times Decision to undertake work related activities	Binary logit MNL Binary logit	Yes, No 528 discrete time period combinations Yes, No	Applicable only to individuals above the age of 12 and who are workers. The determination of whether or not an individual is a worker is made in the CEMSELTS module.
GA7 GA8 GA9	Adult's decision to go to school Adult's school start time (time from 3 a.m.) Adult's school end time (time from school start time)	Binary logit Regression Regression	Yes, No Continuous time Continuous time	Applicable only to adults who are students, as determined in CEMSELTS
GA10 GA11	Mode to school for children Mode from school for children	MNL MNL	Driven by parent, Driven by other, School bus, Walk/bike Driven by parent, Driven by other, School bus, Walk/bike	Applicable only to children who go to school
GA12 GA13	Allocation of drop off episode to parent Allocation of pick up episode to parent	Binary logit Binary logit	Father, Mother Father, Mother	Applicable only to non-single parent household with children who go to school
GA14 GA15 GA16	Decision of child to undertake discretionary activity jointly with parent Allocation of the joint discretionary episodes to one of the parents Decision of child to undertake independent discretionary activity	Binary logit Binary logit Binary logit	Yes, No Father, Mother Yes, No	Second model in this row is applicable only to non-single parent households with children who go to school.
GA17 GA18 GA19 GA20 GA21 GA22	Decision of household to undertake grocery shopping Decision of an adult to undertake grocery shopping given household undertakes it Decision of an adult to undertake household/personal business activities Decision of an adult to undertake social/recreational activities Decision of an adult to undertake eat out activities Decision of an adult to undertake other serve-passenger activities	Binary logit Binary logit Binary logit Binary logit Binary logit Binary logit	Yes, No Yes, No Yes, No Yes, No Yes, No Yes, No	Self-explanatory

General Notes: A child is a individual whose age is less than 18 years. An adult is an individual whose age is 18 years or more.
In the CEMDAP architecture, all individuals in the population have to be classified into one of the following three categories: (1) student, (2) worker, and (3) non-student, non-worker. CEMDAP, in its current form, does not accept the category of "student and worker."

Table 2.2 The Worker Scheduling Model System

Model ID	Model Name	Econometric Structure	Choice Alternatives
WSCH1	Commute mode	MNL	Solo driver, Driver with passenger, Passenger, Transit, Walk/bike
WSCH2	Number of stops in work-to-home commute	Ordered probit	0,1, or 2
WSCH3	Number of stops in home-to-work commute	Ordered probit	0,1, or 2
WSCH4	Number of after-work tours	Ordered probit	0,1, or 2
WSCH5	Number of work-based tours	Ordered probit	0,1, or 2
WSCH6	Number of before-work tours	Ordered probit	0 or 1
WSCH7	Tour mode	MNL	Solo driver, Driver with passenger, Passenger, and Walk/bike
WSCH8	Number of stops in a tour	Ordered probit	1,2,3,4, or 5
WSCH9	Home/work stay duration before a tour	Regression	continuous time
WSCH10	Activity type at stop	MNL	Work-related, Shopping, Household/personal business, Social/recreational, Eat out, and Other serve passenger
WSCH11	Activity duration at stop	Regression	Continuous time
WSCH12	Travel time to stop	Regression	Continuous time
WSCH13	Stop location	Spatial Location Choice	Choice alternatives based on estimated travel time

Table 2.3 The Non-Worker Scheduling Model System

Model ID	Model Name	Econometric Structure	Choice Alternatives
NWSCH1	Number of independent tours	Ordered probit	1, 2, 3, or 4
NWSCH2	Decision to undertake an independent tour before the pick-up or joint discretionary tour	Binary logit	Yes, No
NWSCH3	Decision to undertake an independent tour after the pick-up or joint discretionary tour	Binary logit	Yes, No
NWSCH4	Tour mode	MNL	Solo driver, Driver with passenger, Passenger, and Walk/bike
NWSCH5	Number of stops in a tour	Ordered probit	1, 2, 3 4, or 5
NWSCH6	Number of stops following a pick-up/drop-off stop in a tour	Ordered probit	0 or 1
NWSCH7	Home stay duration before a tour	Regression	Continuous time
NWSCH8	Activity type at stop	MNL	Work-related, Shopping, Household/personal business, Social/recreational, Eat out, and Other serve passenger
NWSCH9	Activity duration at stop	Regression	Continuous time
NWSCH10	Travel time to stop	Regression	Continuous time
NWSCH11	Stop location	Spatial Location Choice	Choice alternatives based on estimated travel time

Table 2.4 The Joint Discretionary Tour Scheduling Model System

Model ID	Model Name	Econometric Structure	Choice Alternatives
JSCH1	Departure time from home (time from 3 a.m.)	Regression	Continuous time
JSCH2	Activity duration at stop	Regression	Continuous time
JSCH3	Travel time to stop	Regression	Continuous time
JSCH4	Location of stop	Spatial Location Choice	Predetermined subset of the 4,874 zones

Table 2.5 The Children Scheduling Model System

Model ID	Model Name	Econometric Structure	Choice Alternatives
CSCH1	School to home commute time	Regression	Continuous time
CSCH2	Home to school commute time	Regression	Continuous time
CSCH3	Mode for independent discretionary tour	Binary logit	Drive by other, Walk/bike
CSCH4	Departure time from home for independent discretionary tour (time from 3 a.m.)	Regression	Continuous time
CSCH5	Activity duration at independent discretionary stop	Regression	Continuous time
CSCH6	Travel time to independent discretionary stop	Regression	Continuous time
CSCH7	Location of independent discretionary stop	Spatial Location Choice	Predetermined subset of the 4,874 zones

2.3 Data

This section discusses the data used for the estimation of all the model components identified in Section 2.2. Only the sources of the data are discussed in this report. The reader is referred to Guo et al. (2005) for a discussion of the data-cleaning procedure and the sample formation procedure to generate the estimation sample.

2.3.1 Data Sources

The data used in the estimation of all the model components were obtained from three main sources: (1) the 1996 DFW household activity survey, (2) the DFW zonal land-use database, and (3) the DFW interzonal transportation level of service data. All three data sets were acquired from the North Central Texas Council of Governments (NCTCOG). Each of these three major data components is described below.

2.3.1.1 1996 DFW household activity survey

The data from the 1996 DFW household activity survey are available as four separate files: (1) household file, (2) person file, (3) vehicle file, and (4) activity file. The household file contains the location of each household, housing type, housing tenure, and several household socio-economic characteristics (such as household size and household income). The person file includes socio-demographic characteristics such as age, gender, ethnicity, education level, and employment status for each person in each sampled household. For employed individuals, work location, work schedule characteristics, and income levels are also available. The vehicle file contains information on the characteristics of each vehicle owned by each sampled household. The activity file contains sequential information on all the activities pursued by the surveyed individuals on their diary day. Each data record in this file provides information for one particular activity. The available information includes the type of activity (one of thirty different categories such as home, work, school, shopping, and pick-up), location, start time, and end time. For travel activities, information on the travel mode used (e.g., driver of a vehicle, passenger in a vehicle, transit, and walk) is available.

2.3.1.2 DFW zonal land-use database

The DFW zonal land-use file provides information on several characteristics of each of the 4,874 zones (sixty-one of which are external stations) in the DFW area, including total population, number of households, median income, basic employment levels, service employment levels, retail employment levels, and the acreage by each of several land-use purposes (including water area, park land, roadway, office, and retail space). In addition, this database identifies the zones with "special" land use, such as airports, hospitals, colleges, and major shopping malls. Finally, the parking costs for zones in the Dallas and Fort Worth CBDs are also provided. In addition, the GIS layer of the zone boundaries was processed using a geographic information system (GIS) to identify the set of zones that are adjacent (i.e., share a boundary) to each of the 4,874 zones.

2.3.1.3 DFW interzonal transportation level of service data

The DFW interzonal transportation level of service (LOS) file provides information on several LOS characteristics for each of the highway and transit modes and between every pair of zones (4,874 X 4,874 zonal pair combinations in all) in the DFW region. The LOS characteristics available for the highway mode include distance and in-vehicle and out-of-vehicle travel times for each of the a.m. peak, p.m. peak, and off-peak periods. The LOS characteristics available for the transit mode include, for each of the peak and off-peak periods, the in-vehicle and out-of-vehicle travel times, accessibility to the transit stop, and the number of transfers.

2.3.2 Sample Formation

The original raw survey data provide over 119,000 activity records for 10,607 persons from 4,641 households. Each of the household, person, vehicle, and activity files were subject to preliminary cleaning and consistency checks. If critical information (such as age, employment status, work location, and school location) of one or more household members was missing, then such households were removed from further analysis. The activity records of the persons in households without any missing information were processed to generate a trip file. In this trip file, each record corresponds to a trip that is characterized by the start and end times, the start and end locations, the activity types at the origin and the destination, and the travel mode. Again, if a substantial amount of travel information was missing or inconsistent for one or more household members, then such households were removed from further analysis. The only exception to the above rule occurred when the missing information was activity locations. Specifically (and unlike in the development of models for the previous version of CEMDAP), households were not discarded if the location information was missing for one or more trips of its constituent members. Discarding such households would have resulted in a substantial reduction of the sample size. The implication of this approach is that our sample for the estimation of models for location choice decisions is smaller than the sample for the estimation of all other activity-travel decisions.

Several attributes of the activity-travel patterns (such as the commutes, the tours, and the identification of the tours to which each trip and stop belongs) that are not directly reported in the surveys were derived from the overall sequence of trip records for each person. Finally, the travel patterns of the parents and children were matched to identify (1) the discretionary activities pursued jointly and (2) the pick-up and drop-off activities undertaken by parents to escort children to and from school. There were very few joint activity and travel episodes between household adults that we could identify based on our matching procedure. Thus CEMDAP, in its current form, does not explicitly consider joint activity-travel patterns of household adults.

The final estimation data set comprises about 23,000 activity-travel records for 6,166 persons from 2,750 households. Of the 6,166 persons, 1,253 are children and 4,913 are adults. Of the 1,253 children, 939 (75 percent) are students. Of the 4,913 adults, 3,152 (64 percent) are employed, 413 are students (8.5 percent), and the rest are unemployed, retired, or homemakers.

2.4 Microsimulation Framework

This section describes the microsimulation procedure implemented within CEMDAP for predicting the complete activity-travel patterns of all individuals in a household. This procedure is repeatedly applied to each household in the input synthetic population to completely determine the activity-travel patterns of all individuals in the study area. The overall prediction procedure (for a household) can be subdivided into two major sequential steps: (1) the prediction of activity generation and allocation decisions and (2) the prediction of activity scheduling decisions. The first step predicts the decisions of household members to pursue various activities such as work, school, shopping, and escorting of children during the day. This step is described in detail in Section 2.4.1. The second step predicts the sequencing of these activities, accommodating the space-time constraints imposed by work, school, and escorting of children's activities. This step is described in detail in Section 2.4.2. The mathematical procedures used to predict the choice outcomes from various econometric models such as the multinomial logit, ordered probit, hazard duration model, and linear regression have been presented in Bhat et al,(2003).

2.4.1 Prediction of Activity Generation and Allocation Decisions

The prediction of activity generation and allocation decisions comprises the following three sequential steps: (1) the generation of work and school activity participation, (2) the generation of children's travel needs and allocation of escort responsibilities to parents, and (3) the generation of independent activities for personal and household needs. Each of these steps is discussed in further detail below.

2.4.1.1 Generation of work and school activity participation

Decisions regarding work and school activities are predicted as the first activity generation decisions because these are pursued with significant regularity and also impose constraints on participation in all other activities during the day. This prediction step is presented schematically in Figure 2-3. For each child in the household who is a student, the decision to go to school and the timing (i.e., start and end times) are first determined (note that the model numbers in the figure for each component correspond to the numbering scheme employed in Table 2.1). Next, the decision of employed adults to go to work during the day and the timing of the work activity are determined. These decisions of the adults may be influenced by the need to take care of non–school-going children at home during the day, which is the reason for modeling work participation decisions subsequent to the decisions of children to go to school. The locations of the school and work are modeled and predetermined in the CEMSELTS module discussed in Chapter 5. Employed adults may also choose to undertake work-related activities. These are different from the main work activity in that the location of these activities is not predetermined. Finally, the school participation and timing decisions of each adult who is a student are determined. (Adults are exogenously classified into one of the following three categories: employed, student, or unemployed/non-student.) Adults who decide to undertake either work or school activities during the day are classified as "workers" and the other adults are classified as "non-workers." For the rest of the prediction procedure, the term "work" will be used to refer to either a work or school activity of an adult as appropriate.

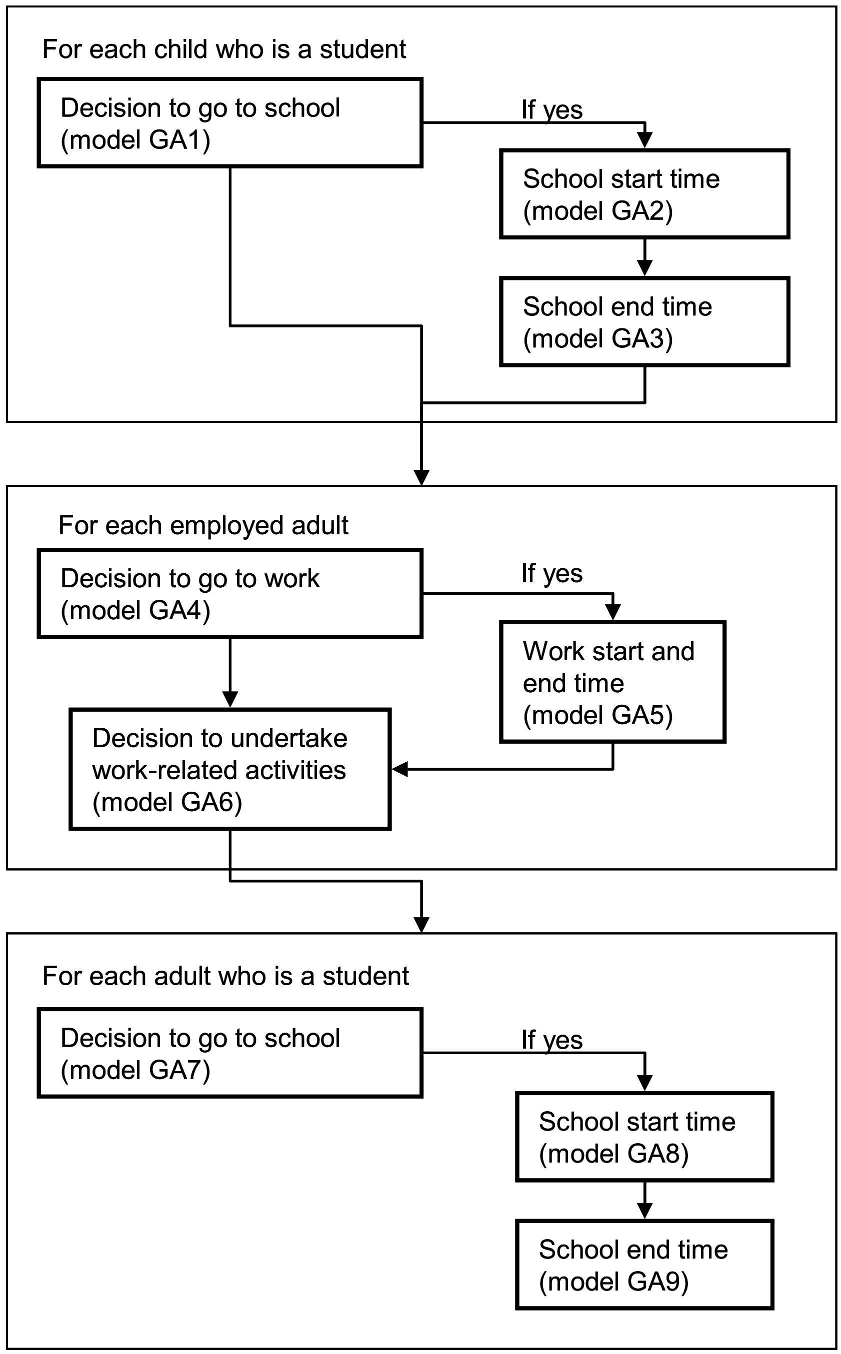

Figure 2-3 Generation of Work and School Activity Participation

2.4.1.2 Generation of children's travel needs and allocation of escort responsibilities to parents

The second major step in the prediction of the generation-allocation decisions involves the children's travel needs (Fig 2-4). In this step, the children's travel mode to and from school are first determined. The travel mode can be one of these: drive by parent, drive by other, school bus, and walk/bike. For children driven to and from school by a parent, the escort responsibilities have to be allocated to the parents. For children in single-parent households, this allocation is trivial as there is only one parent. For children in nuclear family households (i.e., a male-female couple with children), each of the pick-up and drop-off responsibilities is allocated to either the mother or the father. The reader will note that the framework assumes that there is at most one episode each of pick-up and drop-off activities. (However, multiple children may be picked up or dropped off in a single episode.) It was necessary to impose this restriction because of data limitations. Specifically, the estimation data set did not provide data to develop models to accommodate multiple pick-up and drop-off episodes (as may be required in households with many children who go to different schools). Also, the interdependencies between children and parents are not explicitly captured in complex households (i.e., households other than those of the single-parent or nuclear-family types), again owing to data limitations. Nonetheless, because single-parent and nuclear-family are the most common types of households with children, we believe that this is not a serious limitation. If any escort responsibility is allocated to a worker, then the work start and end times of this person are suitably updated to ensure feasibility of the escort activity. (Based on empirical analysis of the DFW travel survey data, we assume that escort activities undertaken by workers are pursued during the commute.)

In addition to going to school, children may also pursue discretionary activities (such as visiting friends and sports events) jointly with a parent. The next two model components in this overall second step determine these joint discretionary activity participation decisions of children, along with the parent participating in the joint discretionary activity. The chosen parent escorts the child to and from the activity and also participates in the activity jointly with the child. The reader will note two implied assumptions: (1) there is at most one joint discretionary episode (even if there are multiple children in the household) and (2) only one of the parents undertakes discretionary activities jointly with children. These assumptions can be relaxed if more data on the travel patterns of households with children are available.

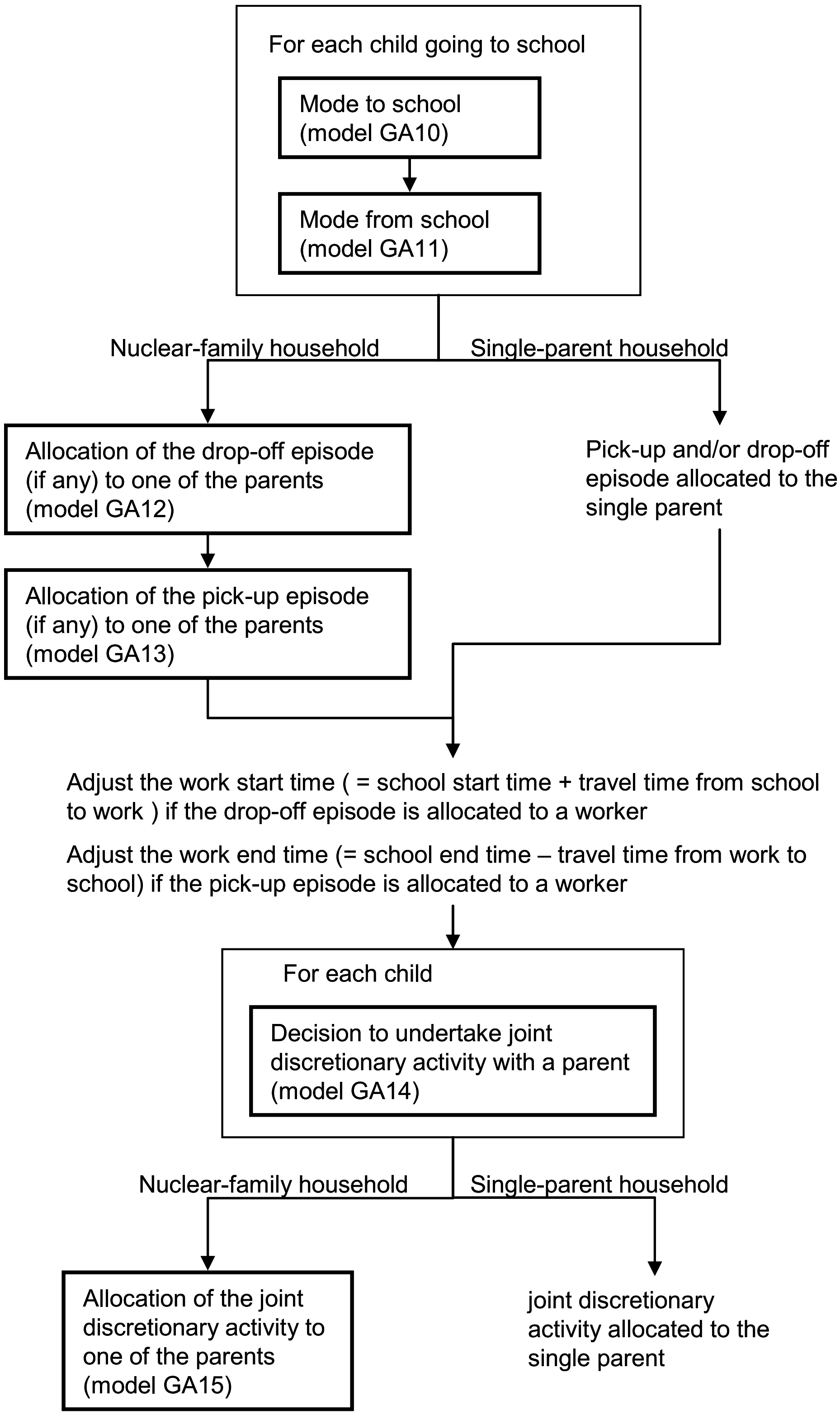

Figure 2-4 Generation of Children's Travel Needs and Allocation of Escort Responsibilities to Parents

2.4.1.3 Generation of independent activities for personal and household needs

The third and final step in the prediction of activity generation and allocation involves decisions about independent activity participation (Fig 2-5).

These independent activities may be pursued for personal needs (e.g., recreation) or for household needs (e.g., grocery shopping). Children's decisions to undertake independent discretionary activities are determined first. For these activities, the children are not escorted by household members. Next, the household's decision to undertake grocery shopping during the day is determined. Conditional on the household deciding to shop for groceries during the day, the shopping responsibility is allocated to one or more adults in the household. The next three model components in this step determine the decisions of household adults to undertake independent activities for (1) household or personal business (e.g., banking), (2) social activities or recreation (e.g., visiting friends or going to the movies), and (3) eating out. The final model component determines the decision of adults to undertake "other serve-passenger activities." These are pick-up or drop-off activities pursued by adults other than the trips for escorting children to and from school. The person(s) being served in this case may be either household members or non-members. A more rigorous treatment of these "other serve-passenger" episodes to explicitly accommodate additional interpersonal interactions is identified as a potential area of future work. Such efforts will benefit substantially from travel survey improvements that explicitly collect data about the persons being served.

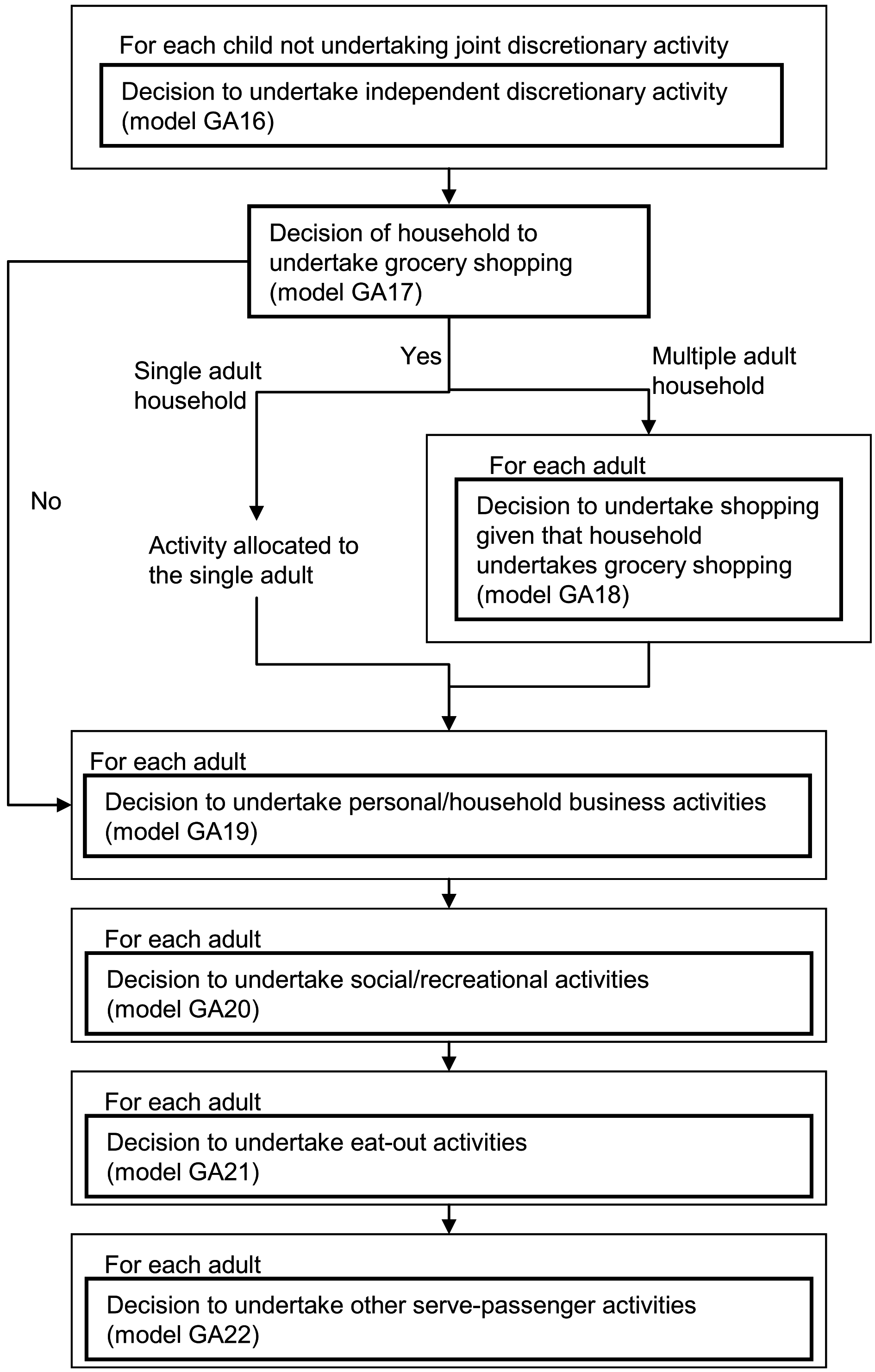

Figure 2-5 Generation of Independent Activities for Personal and Household Needs

2.4.2 Prediction of Activity Scheduling Decisions

At the end of the prediction of activity generation and allocation decisions (Section 2.4.1), the following information is available: (1) each child's decision to go to school, the school start time and end time, the modes used to travel to and from school, the decision to undertake a joint discretionary activity with a parent, and the decision to undertake an independent discretionary activity; (2) which (if either) parent undertakes the drop-off activity, the pick-up activity, and the joint discretionary activity with the children; (3) each employed adult's decision to go to work, the work start time and end time, and the decision to undertake work-related activities; (4) each adult student's decision to go to school and the school start time and end time; (5) each adult's decisions to undertake grocery shopping, personal or household business, social or recreational activities, eating out, and other serve-passenger activities.

In the next broad step of predicting activity scheduling decisions, the following sequence is adopted (see Fig 2-6): (1) scheduling the commutes for each worker in the household, (2) scheduling the drop-off tour for the non-worker escorting children to school, (3) scheduling the pick-up tour for the non-worker escorting children from school, (4) scheduling the commutes for school-going children, (5) scheduling the joint tour for the adult pursuing discretionary activity jointly with children, (6) scheduling the independent home-based tours and work-based tours for each worker in the household, (7) scheduling the independent home-based tours for each non-worker in the household, and (8) scheduling the discretionary activity tours for each child in the household. It is important to note that not all eight steps are required for each household in the population. For example, Steps (2), (3), (4), (5), and (8) are not necessary for households without children. Similarly, Steps (2) and (3) are not needed for a household if none of the school going children is escorted to or from school by his or her parents. Each of the eight steps is discussed in further detail here.

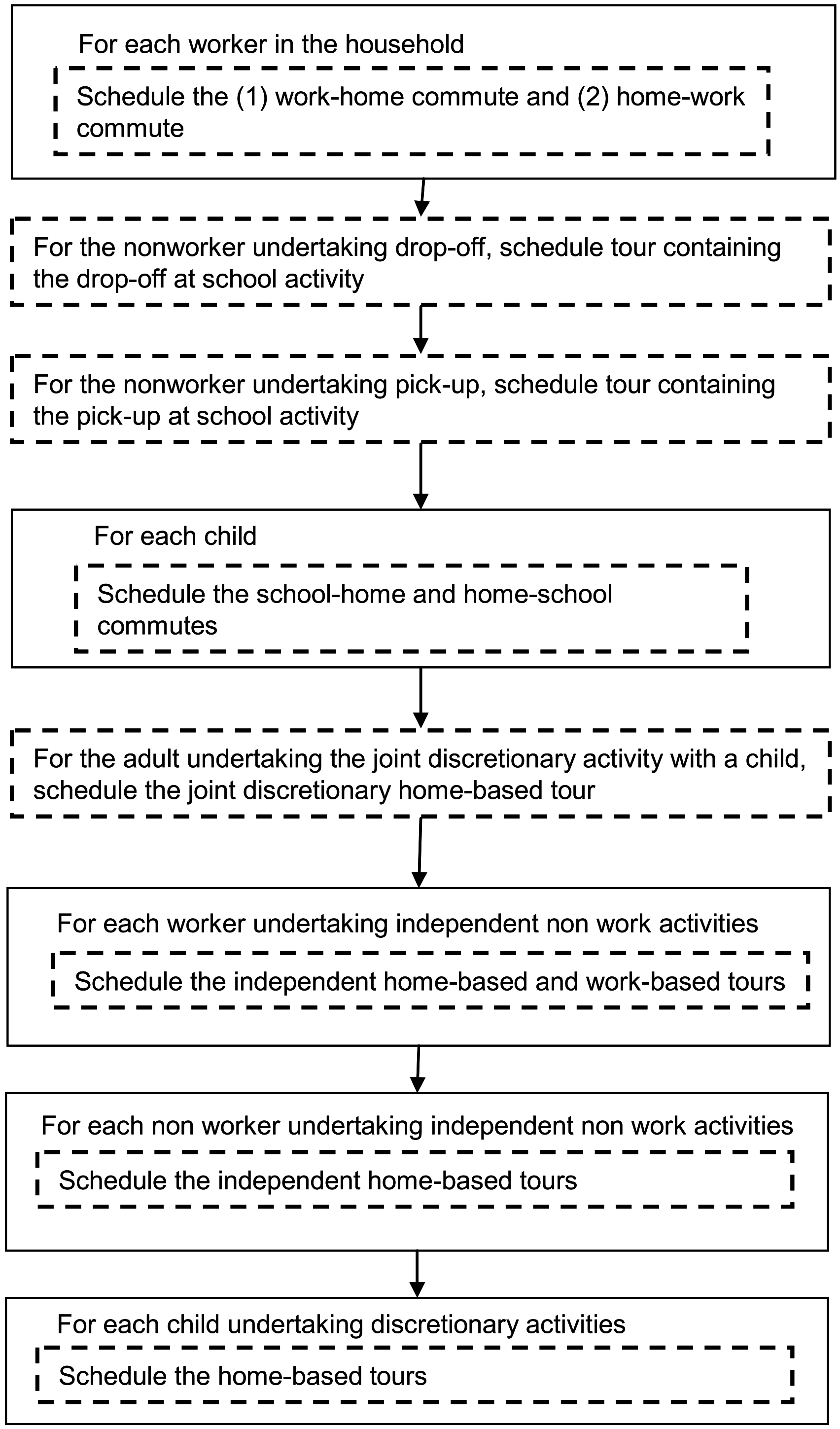

Figure 2-6 Sequence of Major Steps in the Prediction of Activity Scheduling Decisions

2.4.2.1 Scheduling the commutes for each worker in the household

Travel undertaken to and from work is arguably the most constrained in terms of space and time (because of the rather strict need to be at the work location during a certain period of the day). Further, as already indicated, if the worker escorts children to and from school, then these pick-up and drop-off episodes are assumed to be undertaken during the commutes. Hence, the scheduling decisions relating to the commute are determined first for each worker in the household. Further based on the generation of children's travel needs and allocation of child escort responsibility to parents (Section 2.4.1.2), we already know if a given worker in the household is picking up or dropping off children. If the worker is picking up a child in the evening commute but not dropping the child in the morning commute, the evening commute mode is set to "driver with passenger" and the morning commute mode is set to "driver solo." If the worker is dropping a child in the morning commute but not picking up a child in the evening commute, the morning commute mode is set to "driver with passenger" and the evening commute mode is set to "driver solo." If the worker is both dropping off and picking up the child, both the morning and evening commute modes for the worker are set to "driver with passenger."

In the rest of this section, we discuss the prediction process for the work-to-home commute activity travel pattern and the home-to-work commute pattern. The prediction begins with the work-to-home commute pattern because there is much more activity participation in this leg of the commute than in the home-to-work commute.

<u>The work-to-home-commute</u>

If the worker is picking up children from school, then this pick-up activity is assumed to be the only stop during the work-to-home commute (see Figure 2.7). The travel times from work to school and from school to home are determined as the prevailing interzonal auto travel times between the appropriate zones and at the appropriate times of day. An activity time of 5 minutes is assigned to this pick up stop.

If the worker is not picking up children from school, the first prediction is of the travel mode (see Fig 2-7). This is accomplished using a multinomial logit model with five possible choice alternatives: drive solo, drive passenger, shared ride, transit, and walk/bike. The next decision modeled is the number of stops made during the work-to-home commute. If the worker does not pursue any non-work activities during the day (as predicted earlier based on the discussion in Sections 2.4.1.2 and 2.4.1.3), then the number of work-to-home stops is set to zero. If the worker does pursue non-work activities during the day but the commute mode is transit or walk/bike, it is assumed that the worker is not making any trips during the commute (this is based on the empirical data available for estimation).

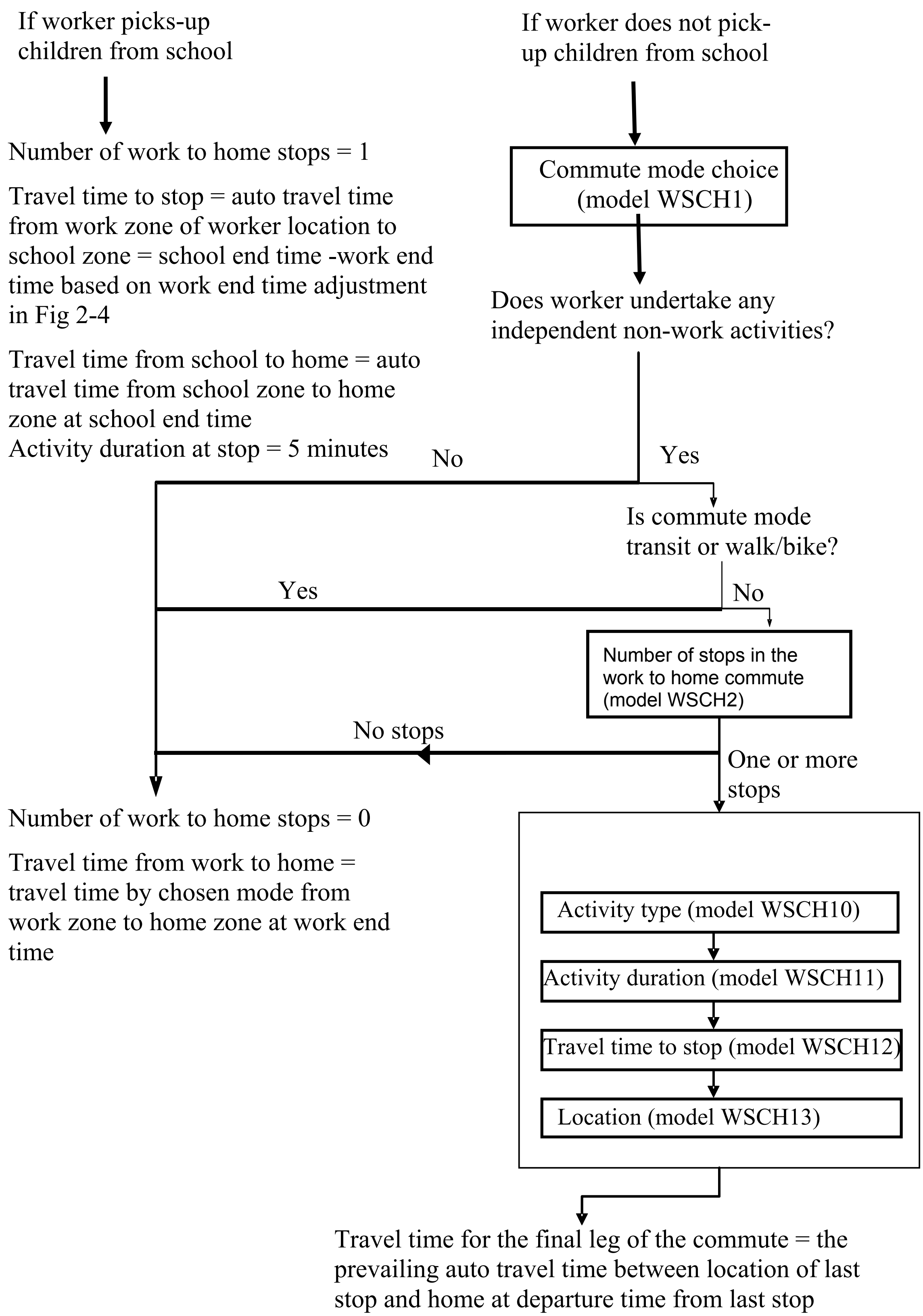

Figure 2-7 Scheduling the Work-to-Home Commute

If the worker does pursue non-work activity during the day and the commute mode is not transit and not walk/bike, the number of stops model is invoked (model WSCH2). If the number of stops predicted for the individual is zero in this model or if the worker is assigned zero stops based on earlier considerations, the work-to-home travel time is simply determined as the prevailing travel time (i.e., at work end time) by the chosen mode between the work and home locations. If one or more stops are predicted (the empirical modeling system allows a maximum of two stops during the commute), each of these stops is characterized, sequentially from the first to the last, in terms of the activity type at the stop, the duration of activity at the stop, the travel time to the stop, and the location of the stop. Once all the stops are characterized, the travel time for the last leg of the work-to-home commute (i.e., the trip ending at home) is determined as the prevailing auto travel time between the location of the last activity stop and home at the departure time from the last stop.

<u>The home-to-work commute</u>

The home-to-work commute is characterized next (see Fig 2-8).

If the worker is pursuing drop-off of children at school, then this drop-off activity is the only stop during the home-to-work commute. The travel times from home to school and from school to work are determined as the prevailing interzonal auto travel times between the appropriate zones and at the appropriate times of day. For workers not dropping off children, the scheduling of the home-to-work commute follows a procedure that is very similar to the scheduling of the work-to-home commute discussed earlier.

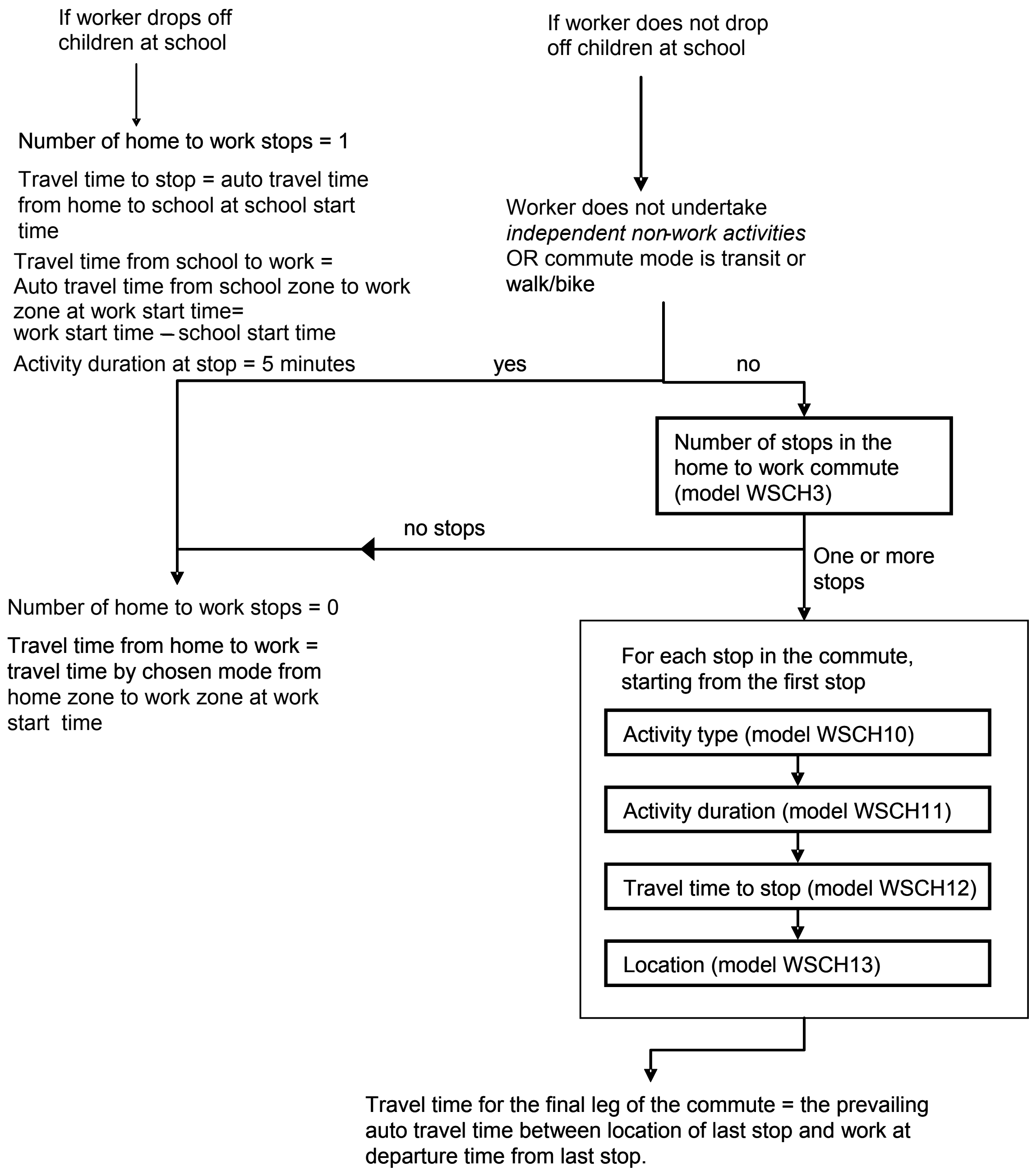

Figure 2-8 Scheduling the Home-to-Work Commute

2.4.2.2 Scheduling the drop-off tour for the non-worker escorting children to school

Among all activities and travel pursued by a non-worker, the escort of children to and from school is undertaken with perhaps the most space-time constraints. Consequently, these activities are scheduled prior to all independent activities undertaken during the day. Of the two types of escort activities, drop-off and pick-up, the scheduling of the former is undertaken first as the drop-off activities temporally precede the pick-up activities.

Non-workers dropping off children at school are assumed to undertake this activity as the first stop of their first home-based tour for the day. The scheduling of this first tour is presented in Figure 2-9. The mode for this tour is set as "driver with passenger" and the travel time is determined as the prevailing auto travel time between the home and school zones at the school start time of the children being escorted. An activity duration of 5 minutes is assigned to the drop-off stop. After dropping off the children at school, the non-worker may choose to undertake other independent activities as part of this same tour. The number of such stops in this tour is determined next. The reader will note that this is applicable only for non-workers who have decided to undertake one or more independent non-work activities (i.e., work-related activities, shopping, household or personal business, social or recreational activities, eating out, or other serve-passenger activities) during the day (as determined earlier in Section 2.4.2). If one or more stops are predicted (the empirical modeling system allows a maximum of three additional stops in a tour containing a drop-off episode), then each of these stops are characterized, sequentially from the first to the last, in terms of the activity type at the stop, the duration of activity at the stop, the travel time to the stop, and the location of the stop. Once all the stops are characterized, the travel time for the last leg of the tour (i.e., the trip ending at home) is determined as the prevailing auto travel time between the location of the last activity stop and home at the departure time from the last stop. If the non-worker is not undertaking any activity other than the drop-off as part of this tour, then the return home time is determined as the prevailing auto travel time between the school location and home at the departure time from the drop-off episode.

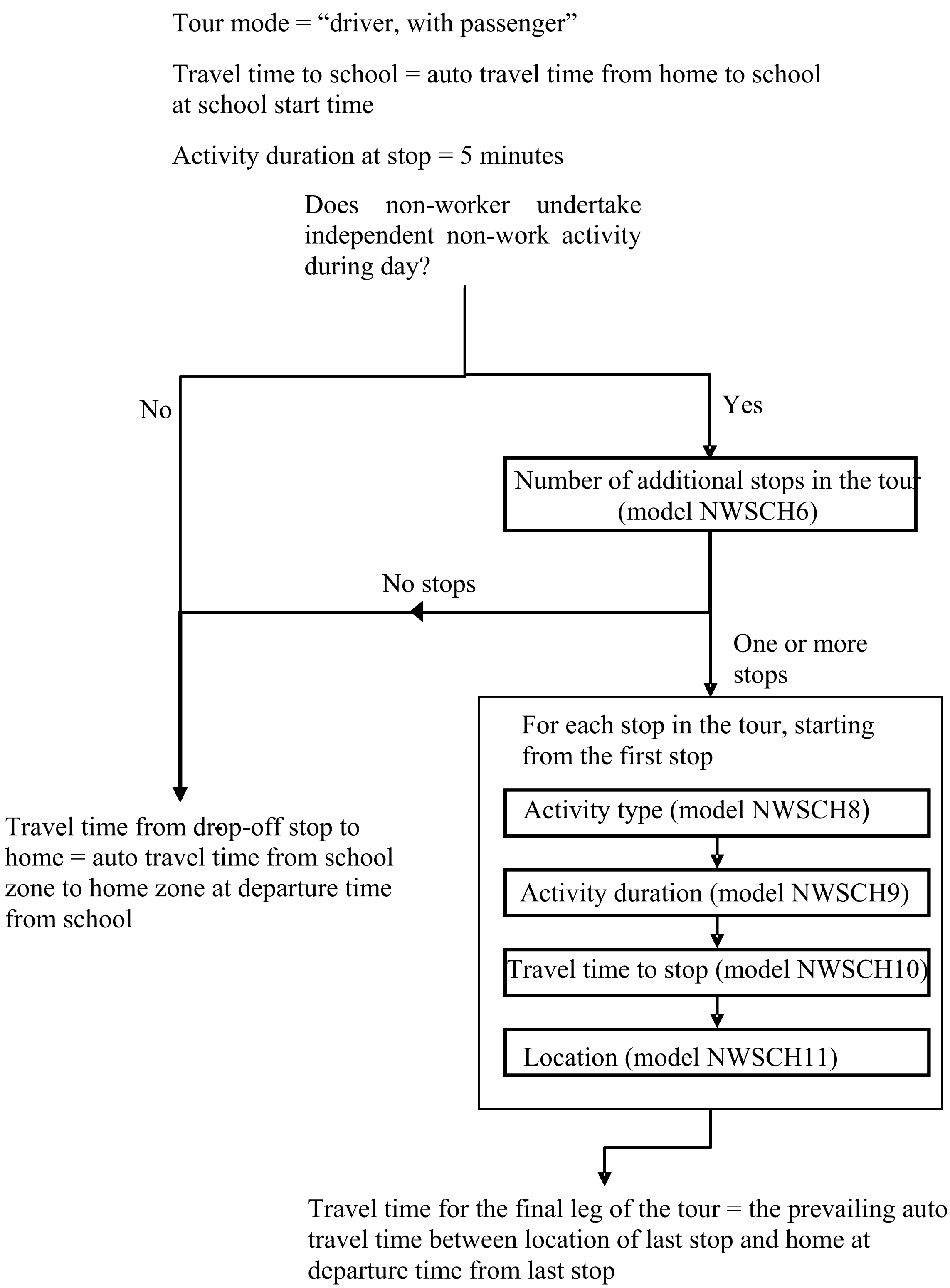

Figure 2-9 Scheduling Drop-Off Tour for Non-Worker Escorting Children to School

2.4.2.3 Scheduling the pick-up tour for the non-worker escorting children from school

Non-workers picking up children from school are assumed to be undertaking this activity as the first stop of a home-based tour. Unlike the tour containing the drop-off episode, the tour containing the pick-up episode is not necessarily the first tour of the day. In fact, it could be any (i.e., first, second, third) of the several tours made by the non-worker during the day. However, this tour would be the first tour to be scheduled if the non-worker does not undertake drop-off episodes and the second tour to be scheduled if the non-worker is also undertaking drop-off episodes. The overall scheduling of a tour containing the pick-up activity (Fig 2-10) is very similar to the procedure described for the scheduling of a drop-off tour. In this case, the tour is constrained by the school end time of the children being escorted as opposed to the school start time in the case of the drop-off tours.

Tour mode = driver

Travel time to school = auto travel time from home to school at school end time

Activity duration at stop = 5 minutes

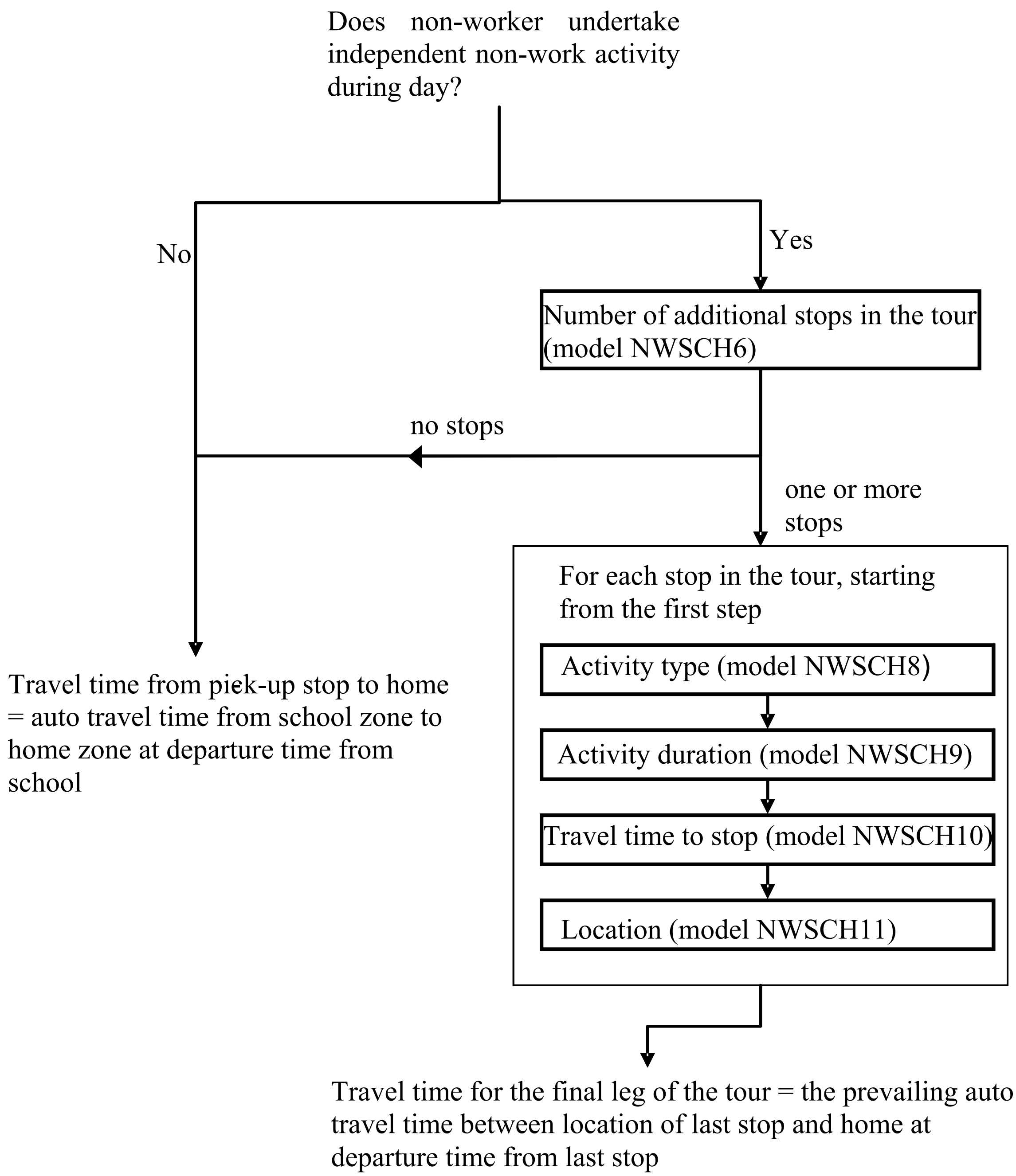

Figure 2-10 Scheduling Pick-Up Tour for the Non-Worker Escorting Children from School

2.4.2.4 Scheduling the commutes for school-going children

In the fourth major step of scheduling, the commute for each of the school-going children in the household is characterized (Fig 2-11). If a child is being escorted home from school, the school-to-home commute of this child is simply obtained as the corresponding travel pattern (i.e., the pattern from pick-up activity to arrival at home) of the escorting parent. If the child is not escorted, the travel time from school to home is determined using a regression model and the child is assumed not to make any stops during this commute. If a child is being escorted to school, the home-to-school commute of this child is simply obtained as the corresponding travel pattern (i.e., the pattern from departure from home to drop-off activity) of the escorting parent. If the child is not escorted, the travel time from home to school is determined using a regression model and the child is assumed not to make any stops during this commute.

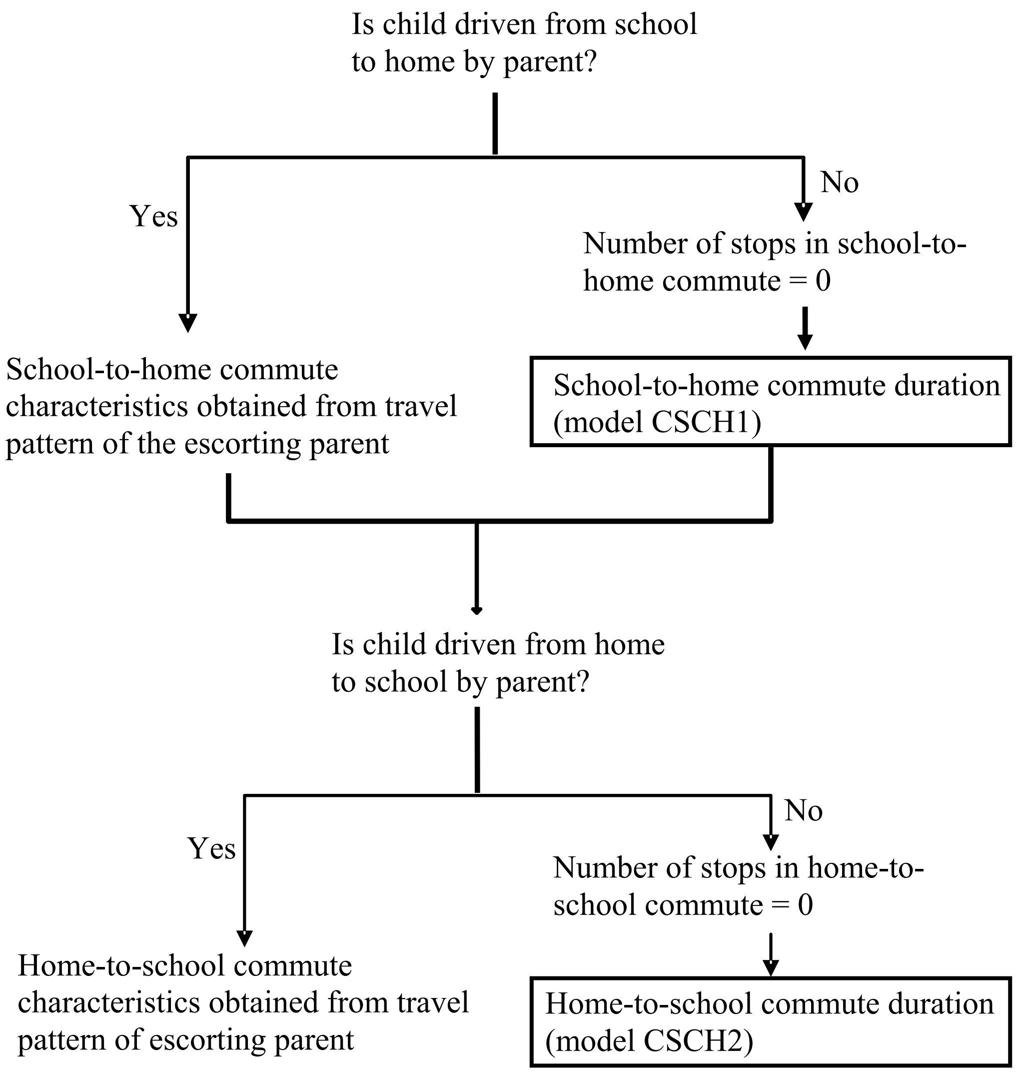

Figure 2-11 Scheduling Commutes for School-going Children

2.4.2.5 Scheduling the joint tour for the adult pursuing discretionary activity jointly with children

The next step in the scheduling procedure focuses on the discretionary activity pursued by an adult jointly with a child in the household. The scheduling procedure is illustrated in Figure 2-12. If this adult is a worker, then the joint activity episode is undertaken as the only stop in the first (and only) after-work tour of the worker. If this adult is a non-worker, then the joint discretionary activity is pursued as the only stop in a home-based tour. This tour could be any of the several tours made by the non-worker during the day. It is useful to point out here that the data sample did not provide cases in which adults undertook both escorting to and from school activities and joint discretionary activities with children. Hence, the adults undertaking joint discretionary activities are assumed not to escort children to and from school. Consequently, for a non-worker undertaking a joint discretionary activity with a child, the corresponding joint tour would be the first tour that would be scheduled. From the standpoint of the child undertaking this activity, the joint discretionary activity is assumed to be undertaken after return from school. The reader will note that the return home time from work of all the workers and the return home time from school of all the children have already been determined. The scheduling begins with the determination of the departure time for the tour and is followed by the determination of the activity duration at the stop, the travel time to the stop, and the location of the stop.

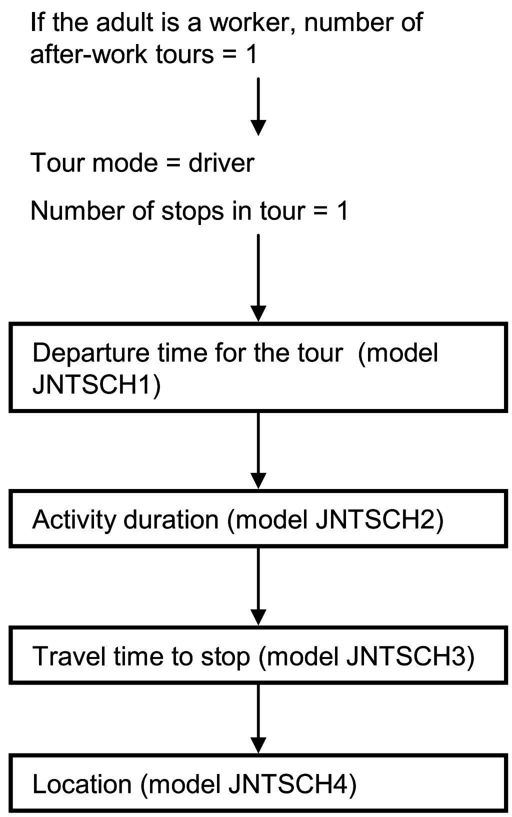

Figure 2-12 Scheduling Joint Tour for the Adult Pursuing Discretionary Activity Jointly with Children

2.4.2.6 Scheduling the independent home-based and work-based tours for each worker in the household

At this point, the scheduling of all activities that are significantly impacted by space-time constraints has been completed. The next steps in the scheduling procedure are focused on the organization of activity stops undertaken with more spatial and temporal flexibility. This sixth step (Figs 2-13 and 2-14) of the scheduling procedure is focused on the scheduling of home-based and work-based tours undertaken by workers who choose to undertake independent non-work activities during the day. For workers not undertaking joint discretionary activities with children, the number of after-work tours is first determined (Fig 2-13). If the worker chooses to undertake one or more tours (up to two after-work tours are supported by the empirical modeling system), then each of these tours is characterized (sequentially from the first after-work tour) in terms of the tour mode, number of stops in the tour, and home-stay duration prior to the tour (Fig 2-14). The reader will note that the home-stay duration before the tour determines the time of day of departure for the tour. A maximum of five stops is supported by the empirical model system in any tour. Each of the stops in the tour is characterized (sequentially from the first to the last stop) in terms of the activity type, activity duration, travel time to the stop, and location of the stop. The attributes of all the stops in a tour are completely determined before proceeding to the subsequent tour.

As shown in Figure 2-13, once the scheduling of activities during the after-work period is complete, the decision of a worker to undertake work-based tours is determined. The empirical modeling system allows up to two tours during the work-based period. The scheduling of the tours during the work-based period follows a similar procedure to the scheduling of tours during the after-work period, which has already been discussed. Finally, after the scheduling of activities during the work-based period is complete, the worker's decision to undertake tours during the before-work period is determined (a maximum of one tour is supported). Again, the scheduling of the tours during the before-work period follows a similar procedure to the scheduling of tours during the after-work and work-based periods. With this, the complete activity-travel pattern of all workers in the household has been generated.

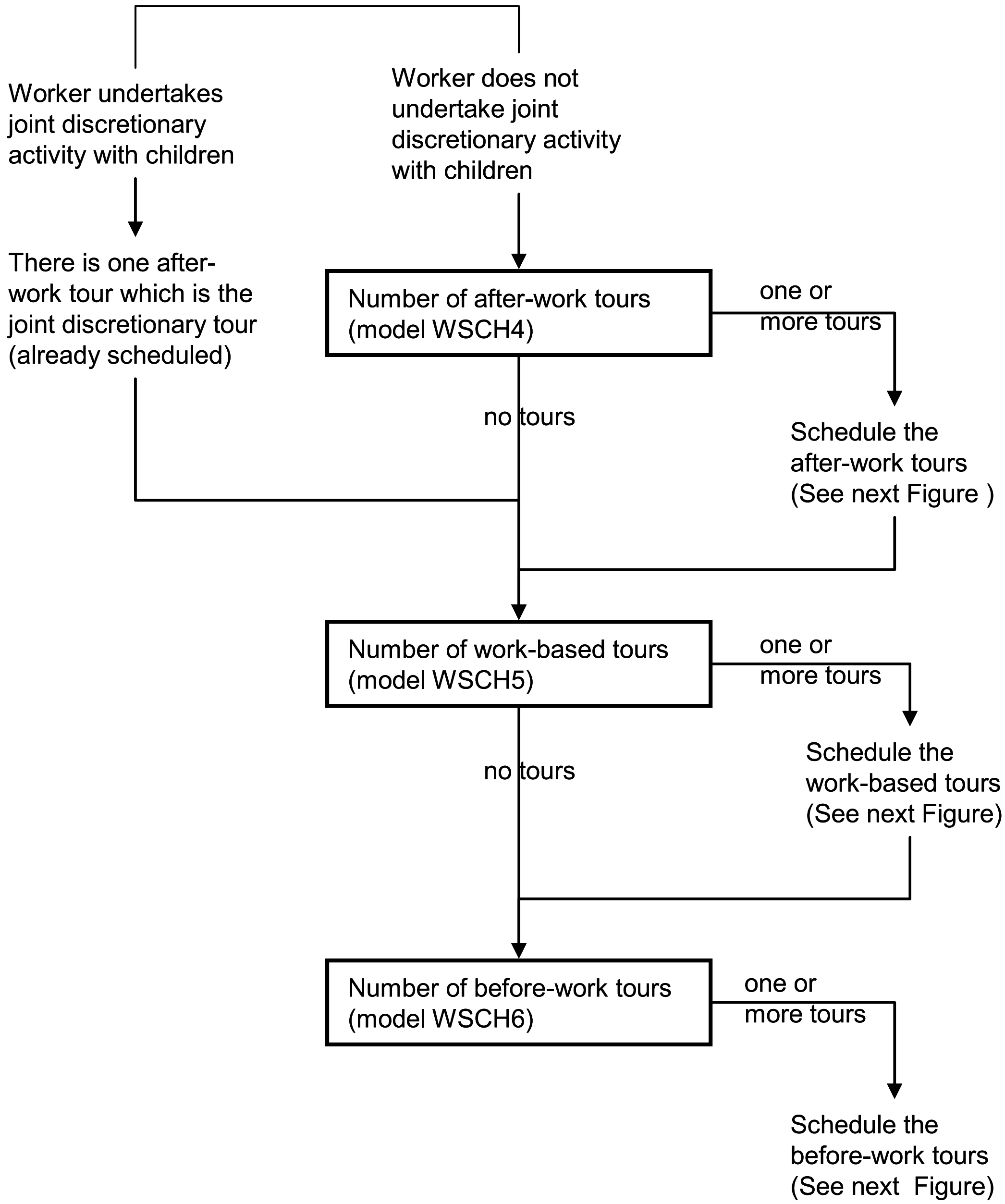

Figure 2-13 Scheduling All Independent Home-Based and Work-Based Tours for Workers

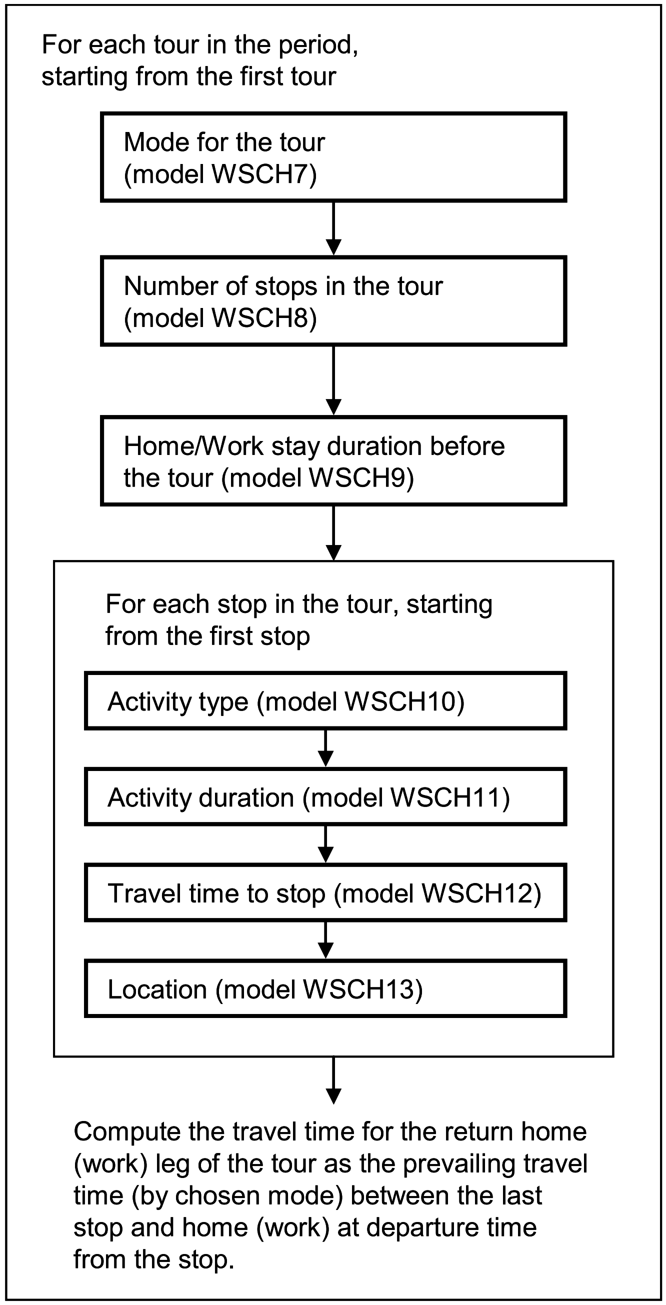

Figure 2-14 Scheduling a Single Independent Tour for Workers

2.4.2.7 Scheduling the independent home-based tours for each non-worker in the household

The penultimate step in the scheduling procedure is focused on the independent activities pursued by the non-workers in the household. If the non-worker is not pursuing pick-up or joint discretionary activities with the children, then the scheduling of independent activities begins with the determination of the total number of independent non-work tours to be undertaken by the individual. A maximum of four independent non-work tours is supported by the empirical modeling system. As depicted in Figure 2-15, each of these tours is characterized (sequentially from the first after-work tour) in terms of the tour mode, number of stops in the tour, and home-stay duration prior to the tour. Home-stay duration before the tour determines the departure time for the tour. A maximum of five stops is supported by the empirical model system in any tour. Each of the stops in the tour is characterized (sequentially from the first to the last stop) in terms of the activity type, activity duration, travel time to the stop, and location of the stop. The attributes of all the stops in a tour are completely determined before proceeding to the next tour.

If the non-worker is undertaking pick-up (joint discretionary) activities, then the decision of this person to undertake an independent tour before and after the pick-up (joint discretionary) tour is predicted (Fig 2-16). As already discussed, non-workers are assumed to undertake one escort or joint discretionary activity. This, in turn, determines the position of the pick-up (joint discretionary) tour within the overall pattern of the non-worker. For example, if a non-worker who undertakes a drop-off tour also decides to undertake an independent tour before the tour for picking up children from school, then the pick-up tour becomes the third tour in this person's overall pattern (the drop-off tour is always the first tour). Alternatively, if a non-worker who does not undertake a drop-off tour decides to undertake an independent tour before the tour for picking up children from school, then the pick-up tour becomes the second tour in this person's overall pattern. The characteristics of these tours and the stops in these tours are determined, depending on the choice to undertake a tour before and after the pick-up (joint discretionary) tour.

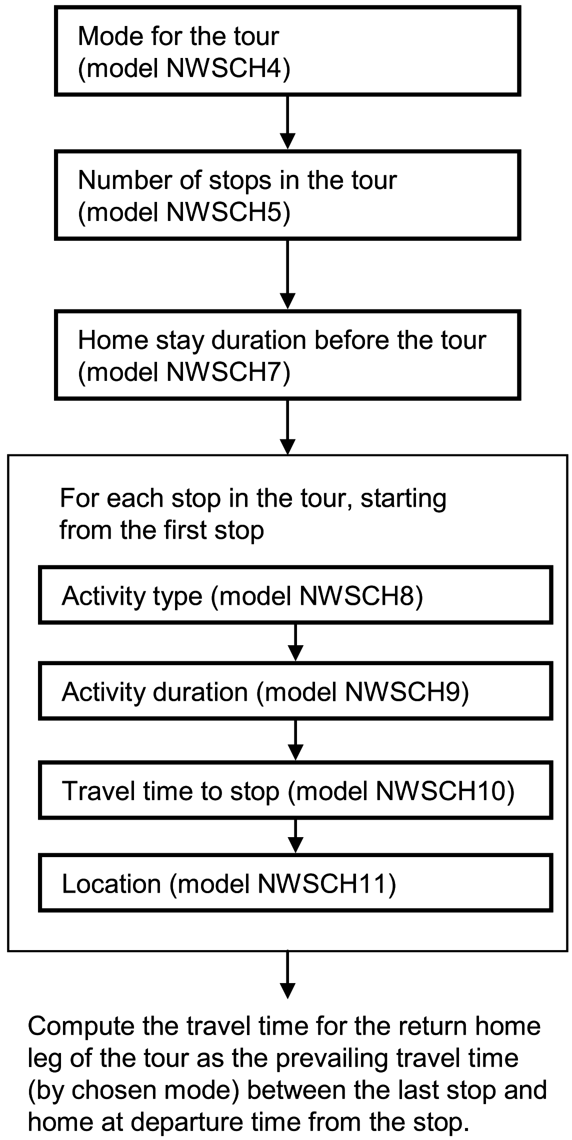

Figure 2-15 Scheduling a Single Independent Tour for Non-Workers

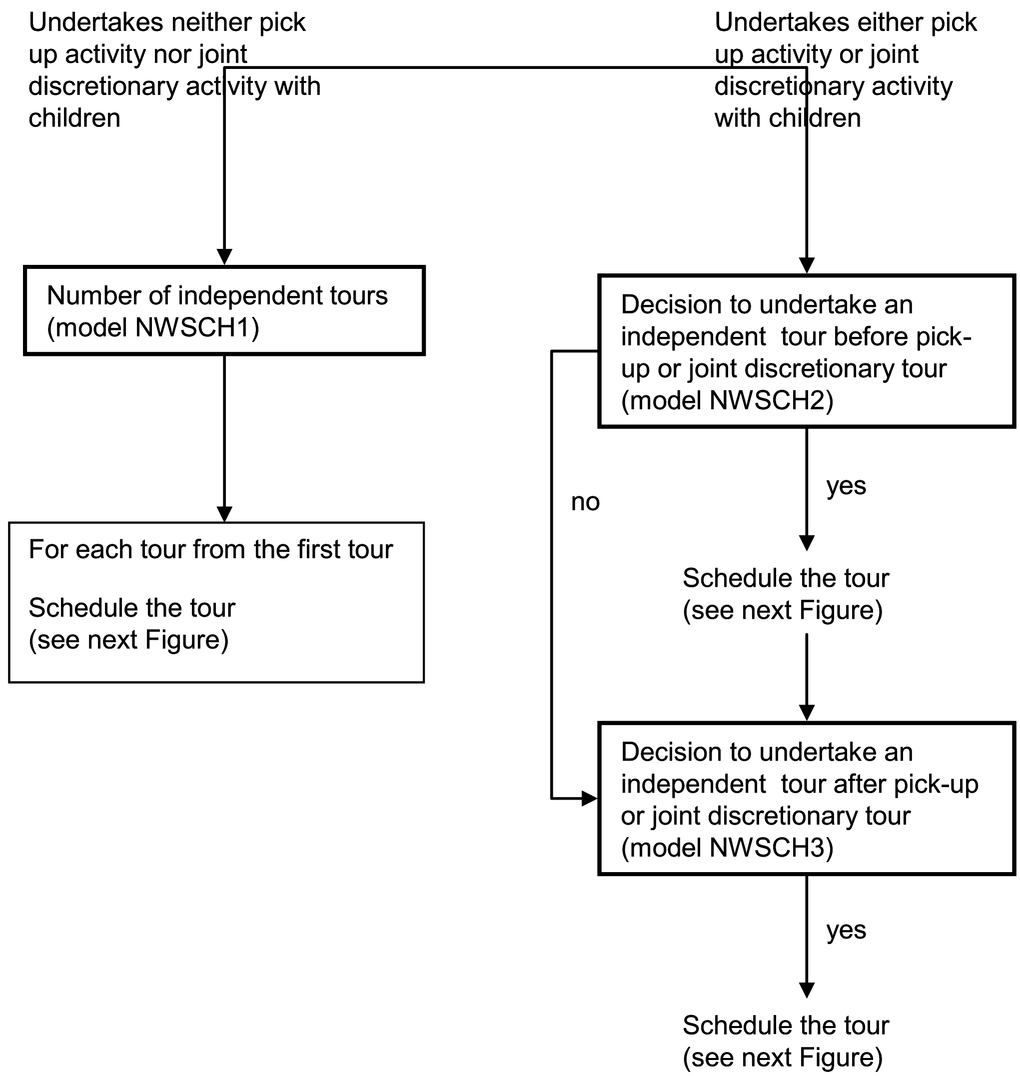

Figure 2-16 Scheduling All the Independent Home-Based Tours for Non-Workers

2.4.2.8 Scheduling the discretionary activity tours for each child in the household

In this last activity scheduling step, tours undertaken by the children for discretionary activity participation are predicted (Figure 2-17). If the discretionary activity is pursued jointly with a parent, then the characteristics of this tour are simply obtained from the corresponding tour of the parent. Otherwise, the characterization of the independent discretionary activity tour begins with the choice of the tour mode, which can be "drive by other" or "walk/bike." Next, the departure time from home for the tour is determined. If the child also goes to school, it is assumed that discretionary tours are undertaken after returning home from school. The characterization of the discretionary tour is completed by determining the activity duration at the stop, the travel time to the stop, and the location of the stop. The reader will note that there is only one stop in discretionary activity tours undertaken by children and each child undertakes at most one discretionary activity tour during the day, either independently or jointly with a parent.

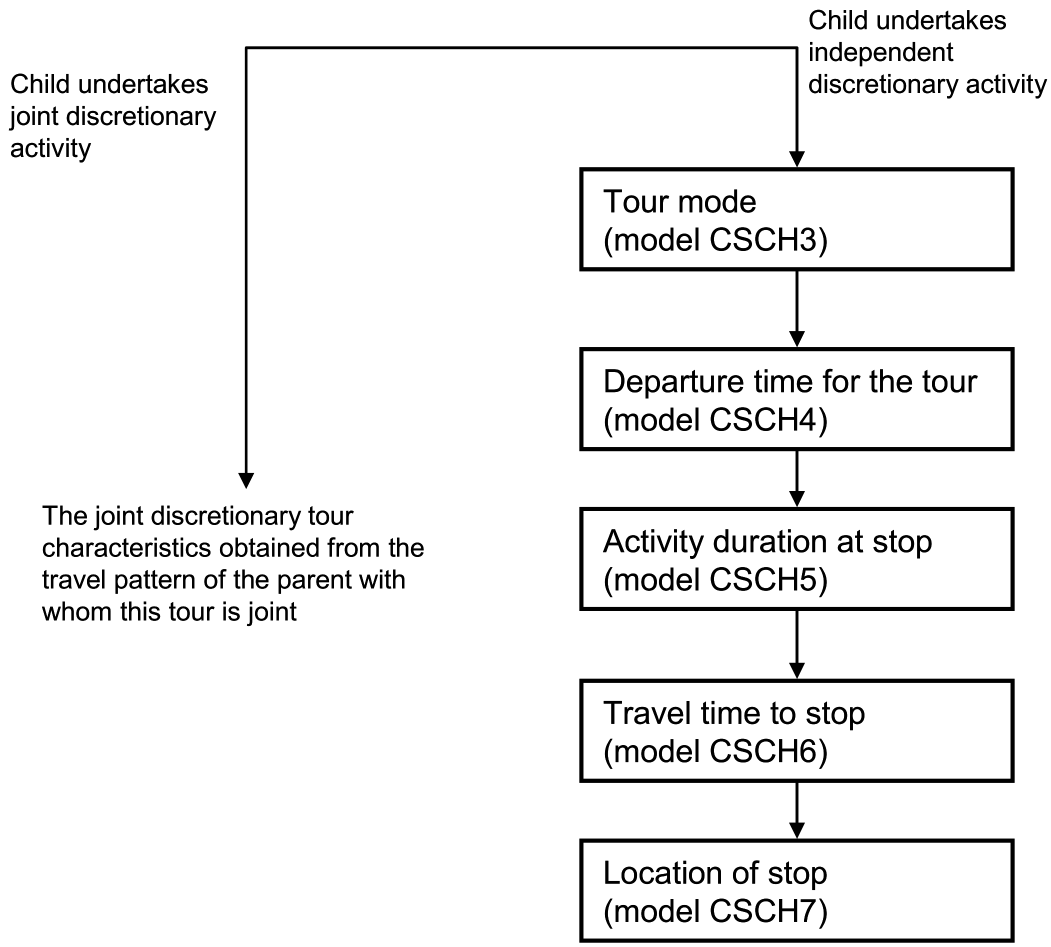

Figure 2-17 Scheduling Discretionary Activity Tours for Each Child in the Household

2.5 Spatial and Temporal Consistency Checks

Several spatial and temporal consistency checks have been implemented in CEMDAP to ensure that the simulation process does not result in unreasonable or impossible activity patterns. This section describes the spatial and temporal consistency checks used in the enhanced version of CEMDAP.

2.5.1 Spatial Consistency Checks

The spatial location choices for non-work activities are determined using the spatial location choice model. Bhat et al. (2003) describes the mathematical procedure used to apply the spatial location choice model. The methodology employs a probabilistic choice set generation method that uses the predicted travel time to the stop (from the previous stop location) in the determination of the candidate locations for the stop. Subsequently, a multinomial logit prediction procedure is used to predict the spatial location choice among the candidate locations in the choice set. It was found that the probabilistic choice set generation method was giving rise to unreasonably far (from the origin zone) spatial location choice predictions. Hence, a deterministic choice set generation method was developed to ensure the spatial consistency of the predicted activity-travel patterns. The deterministic choice set generation method and the subsequent spatial location choice prediction procedure are described below.

The deterministic choice set generation method also uses the predicted travel time to the stop (from the previous stop location) in the determination of the candidate locations for the stop. Subsequently, a multinomial logit prediction procedure is used to predict the spatial location choice among the candidate locations in the choice set.

The rationale behind using the predicted travel time to the stop in generating the location choice set is that the stop location to be predicted should be *within* a certain range of the predicted travel time to that stop. Hence, the location choice set for a stop consists of the zones that fall *within* a certain range of predicted travel times from the previous stop location. Half of the candidate zones selected into the location choice set have shorter travel times (from the previous stop location) than the predicted travel time, while the other half have travel times greater than or equal to the predicted travel time.

An important point to be noted here pertains to the definition of *predicted travel time* to the stop used in the context of spatial location choice. The travel time predicted by the "travel time to the stop" model is the *total expected travel time* that the person expects to travel for the next stop. As the "travel time to the stop" model was estimated using the reported travel times in the household travel survey data, the total expected travel time includes not only the in-vehicle-travel time, but also additional time such as the out-of-vehicle travel time. Hence, the out-of-vehicle travel time is subtracted from the *predicted total expected travel time* to obtain the *predicted travel time* on the network for spatial location choice. This predicted travel time is used to generate the location choice set. The steps involved in the disaggregate prediction (including the choice set generation) using the location choice model are summarized below:

1. Determine the predicted travel time by subtracting the out-of-vehicle travel time from the total expected travel time by using the following rules.
    a. If (activity type at the stop is personal business or shopping or serve passenger and total expected travel time >20 minutes),
    predicted travel time = total expected travel time – 8 minutes
    b. If (activity type at the stop is personal business or shopping or serve passenger and total expected travel time ≤ 20 minutes),
    predicted travel time = 0.6 X total expected travel time
    c. If (activity type at the stop none of personal business or shopping or serve passenger and total expected travel time >24 minutes),
    predicted travel time = total expected travel time – 6 minutes
    d. If (activity type at the stop none of personal business or shopping or serve passenger and total expected travel time >24 minutes),
    predicted travel time = 0.75 X total expected travel time.
2. If the predicted travel time is less than the intrazonal travel time from the previous stop location, then the chosen stop location is in the same zone as the previous stop location because this is the only choice alternative available. If the predicted travel time is greater than the intrazonal travel time, follow the steps below.
3. Arrange all the zonal locations in the ascending order of in-vehicle travel time from the previous stop.
4. Select the first spatial zone *Z,* whose in-vehicle travel time from the previous stop (t_z) is greater than the predicted travel time.
5. Select twenty-five zones with in-vehicle travel time (from the previous stop location) less than t_z and twenty-five zones with in-vehicle travel time greater than t_z. If twenty-five zones are not available on one or both sides of t_z, select the minimum number of zones available on both sides in order to maintain symmetry of travel times of the candidate zones in the choice set.

6. Compute the conditional probability ($P_1, P_2 ... P_K$) for each of the different K ($K = 50$ or less) candidate locations using the calibrated model parameters and the values of exogenous variables specific to the decision maker under consideration.
7. Generate a uniformly distributed random number (U) between 0 and 1.
8. 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.

2.5.2 Temporal Consistency Checks

Most of the temporal choices (such as home-stay durations before tours, activity durations, and travel times to stops) are determined using log-linear regression models. Because the chosen duration is determined by a random draw from a normal distribution, a small (but non-zero) possibility exists that the duration determined is either very high or very low. This may lead to temporal overlapping 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. Predictions on other temporal choice predictions, such as work start and end times and work durations, are also controlled using temporal checks, in order to avoid start and end times that are too early or late and durations that are too long.

The temporal checks 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 fifth-percentile value of the duration in the sample is chosen as the lower bound and the ninety-fifth-percentile value chosen as the upper bound. Most of the time bounds are defined as percentages of available time rather than absolute values. The concept of available time is discussed below in greater detail. (Available time is a frequently updated attribute in the CEMDAP's simulation sequence). Absolute values of time bounds are avoided to reduce the likelihood of any sort of temporal overlaps.

Table 2.6 provides the definitions for available time for various temporal attributes. The available time for a worker's home stay duration before his or her first after-work tour is given by: 1440 – arrival time at home from work; that for the subsequent after-work tours is given by: 1440 – arrival time at home from the previous after-work tour. The available time for a worker's work stay duration before the first work-based tour is given by: the work-based duration, while that for his/her subsequent work-based tours is given by: work end time – arrival time at home from the previous work-based tour. The available time for a worker's home stay duration before his or her first before-work tour is given by the departure time from home for work, while that for the subsequent before-work tours is given by: departure time from home for work – arrival time at home from the previous before-work tour.

The available time for home stay duration before a non-worker's tour depends upon whether the non-worker undertakes pick-up, drop-off, or joint discretionary activities. If the non-worker does not undertake any of the above mentioned joint activities, the available time for home stay duration before his or her first tour is 1440, while that for the subsequent tours is given by: 1440 – arrival time at home from the previous tour. If the non-worker undertakes drop-off activity, the available time for home stay duration before the first tour is given by: 1440 – arrival time at home from the drop-off tour; that for subsequent tours is given by: 1440 – arrival time from the previous tour. If the non-worker undertakes either a pick-up or joint discretionary activity, the available time for home stay duration before his or her first tour before the pick-up or joint discretionary tour is given by: time from 3 a.m. until the departure for the pick-up or joint discretionary tour; available time for the first tour after the pick-up or joint discretionary tour is given by: 1440 – arrival time at home after the pick-up or joint discretionary tour and that for all his or her subsequent tours is given by: 1440 – arrival time from the tour before.

The available time for a worker's tour (after-work, work-based, or before-work) is given by: available time for the work or home stay duration before that tour – work or home stay duration before that tour; that for the work-home commute is given by: time from 3 a.m. until the start of the work; and that for the home-work commute is given by: 1440 – work end time. The available time for a non-worker's tour is given by: available time for the home stay duration before that tour – home stay duration before that tour.

The available time for activity duration of the first stop in a tour or commute is given by: available time for the tour or commute. Available time for any subsequent stop is given by: available time for the previous stop – activity duration for the previous stop - travel duration for the previous stop. The available time for travel for any stop is given by: available time for the activity duration – activity duration at that stop.

Tables 2.7 through 2.16 provide the temporal bounds for each of the temporal choice dimensions predicted in CEMDAP. Several observations can be made from Table 2.6 and these tables. First, the available time decreases with the hierarchy of the temporal attribute (see Table 2.6). That is, the available time for home or work stay duration is greater than the available time for the corresponding tour and the available time for a tour (a tour-level attribute) is greater than the available time for activity duration and travel duration of stops (stop-level attributes) in that tour. Second, the upper and lower bounds for home or work stay duration decrease with an increase in the number of stops or an increase in the number of tours (see Tables 2.7 and 2.8). For non-workers, earlier tours in the pattern have wider time bounds on home stay (see Table 2.8). Third, the upper and lower bounds on activity durations and travel durations decrease with the increase in the number of stops. Fourth, the temporal bounds on home or work-stay, activity duration, and travel duration are in terms of percentages of available time, whereas those of other temporal variables (work and school start and end times and durations, school-home and home-school commute durations, and departure time, activity durations, and travel durations of independent and joint discretionary tours) are in absolute time values. The bounds on work and school start and end times are to allow sufficient time for after-work tours, and before-work tours. The bounds on work and school durations restrict the durations within a reasonable range.

Table 2.6 Available Time Definitions

Available time for…	Definition (in minutes)
Home/work - stay duration for workers	
First after-work tour	1440 – arrival time at home from work
Subsequent after-work tours	1440 – arrival time at home from the previous after-work tour
First work-based tour	Work-based duration
Subsequent work-based tours	1440 – arrival time at home from the previous work-based tour
First before-work tour	Time from 3 a.m. until the departure to work
Subsequent before-work tours	1440 – arrival time at home from the previous before-work tour
Home-stay duration for non-workers	
If non-worker does not undertake pick-up, drop-off, or joint discretionary activity	
First tour	1440
Subsequent tours	1440 – arrival time from the tour before
If non-worker undertakes drop-off activity	
First tour	1440 – arrival time at home from drop-off tour
Subsequent tours	1440 – arrival time from the tour before
If non-worker undertakes pick-up/joint discretionary Activity	
First tour before pick-up/joint discretionary tour	Time from 3 a.m. until departure for pick-up/joint discretionary tour
First tour after pick-up/joint discretionary tour	1440 – arrival time at home after pick-up/joint discretionary tour
Subsequent tours	1440 – arrival time from the tour before
Tour/commute	
After-work, work-based, and before-work tours	Available time for the corresponding work/home-stay duration – work/home-stay duration
Work-home commute	Time from 3 a.m. until the start of work
Home-work commute	1440 – work end time
Non-worker tours	Available time for corresponding home-stay duration – home-stay duration
Activity duration	
First stop in a tour/commute	Available time for the tour/commute
Subsequent stops in a tour/commute	Available time for the previous stop – (activity duration + travel duration for the previous stop)
Travel duration	Available time for activity duration – activity duration

Table 2.7 Temporal Bounds on Worker Home and Work-Stay Duration (as % of available time)

	Lower Bound	Upper Bound
Before-work tours	31.58	86.96
Work-based tours		
One tour, one stop in tour	15.32	64.30
One tour, two or more stops in tour	7.17	56.76
Two or more tours, one stop in tour	11.97	64.11
Two or more tours, two or more stops in tour	7.17	59.87
After-work tours		
One tour, one stop in tour	1.47	38.55
One tour, two or more stops in tour	1.58	28.57
Two or more tours, one stop in tour	1.45	37.24
Two or more tours, two or more stops in tour	1.32	28.17

Table 2.8 Temporal Bounds on Non-Worker Home and Work-Stay Duration (as % of available time)

	Lower Bound	Upper Bound
First tour		
One stop in tour	15.28	63.54
Two stops in tour	15.28	56.25
Three or more stops in tour	13.89	50.00
Second tour		
One stop in tour	2.17	46.19
Two stops in tour	1.41	43.83
Three or more stops in tour	0.84	38.62
Third tour	1.80	37.50
Fourth tour	1.64	29.17

Table 2.9 Temporal Bounds on Worker Activity Duration (as % of available time)

	Lower Bound	Upper Bound
Stops in before-work tours	0.00	61.29
Stops in home-work commute		
One stop in commute	0.00	77.27
Two stops in commute	0.00	70.06
Stops in work-based tours		
One tour, one stop in tour	1.67	30.61
One tour, two stops in tour	0.36	29.51
Two or more tours, one stop in tour	1.67	35.90
Two or more tours, two stops in tour	0.29	31.91
Stops in work-home commute		
One stop in commute	0.17	32.76
Two stops in commute	0.17	27.36
Stops in after-work tours		
One tour, one stop in tour	0.79	41.86
One tour, two stops in tour	0.22	32.14
Two or more tours, one stop in tour	0.49	42.50
Two or more tours, two stops in tour	0.21	32.14

Table 2.10 Temporal Bounds on Non-Worker Activity Duration (as % of available time)

	Lower Bound	Upper Bound
First tour		
One stop in tour	0.09	47.57
Two stops in tour	0.11	42.17
Three stops in tour	0.15	35.36
Four or more stops in tour	0.14	22.22
Second tour		
One stop in tour	0.14	37.74
Two stops in tour	0.29	30.43
Three stops in tour	0.28	32.04
Four or more stops in tour	0.15	19.74
Third tour	0.15	38.05
Fourth tour	0.16	38.63

Table 2.11 Temporal Bounds on Worker Travel Duration (as % of available time)

	Lower Bound	Upper Bound
Stops in before-work tours	1.26	47.37
Stops in home-work commute		
One stop in commute	7.50	83.33
Two stops in commute	3.31	76.19
Stops in work-based tours		
One tour, one stop in tour	0.97	13.33
One tour, two stops in tour	0.59	15.38
Two or more tours, one stop in tour	0.97	14.81
Two or more tours, two stops in tour	0.59	19.69
Stops in work-home commute		
One stop in commute	0.71	8.47
Two stops in commute	0.46	8.93
Stops in after-work tours		
One tour, one stop in tour	0.74	9.30
One tour, two stops in tour	0.62	9.43
Two or more tours, one stop in tour	0.74	9.38
Two or more tours, two stops in tour	0.62	10.03

Table 2.12 Temporal Bounds on Non-Worker Travel Duration (as % of available time)

	Lower Bound	Upper Bound
First tour		
One stop in tour	0.42	10.34
Two stops in tour	0.35	8.57
Three stops in tour	0.39	8.09
Four or more stops in tour	0.28	7.69
Second tour		
One stop in tour	0.44	7.93
Two stops in tour	0.56	11.11
Three stops in tour	0.46	10.64
Four or more stops in tour	0.34	6.42
Third tour	0.37	10.45
Fourth tour	0.67	11.48

Table 2.13 Temporal Bounds on Work and School Start and End Times (absolute time)

	Lower Bound	Upper Bound
School (children)		
Start time (minutes from 3 a.m.)	270.0	390.0
End time (minutes from 3 a.m.)	540.0	900.0
Duration (minutes)	180.0	600.0
Work (adults)		
Start time (minutes from 3 a.m.)	210.0	660.0
End time (minutes from 3 a.m.)	660.0	1020.0
Duration (minutes)	240.0	720.0
School (adults)		
Start time (minutes from 3 a.m.)	240.0	490.0
End time (minutes from 3 a.m.)	498.8	1035.0
Duration (minutes)	120.0	600.0

Table 2.14 Temporal Bounds on Home-to-School and School-to-Home Commute Durations (absolute time in minutes)

	Lower Bound	Upper Bound
School-to-home commute duration		
Auto	5.0	45.0
School bus	10.0	60.0
Walk/bike	3.5	35.0
Home-to-school commute duration		
Auto	3.0	30.0
School bus	10.0	65.0
Walk/bike	4.0	30.0

Table 2.15 Temporal Bounds for Independent Discretionary Tours Undertaken by Children (absolute time)

	Lower Bound	Upper Bound
Departure time (minutes from 3 a.m.)	255.0	990.0
Activity duration (minutes)	10.0	345.0
Travel time (minutes)	1.0	35.0

Table 2.16 Temporal Bounds for Joint Discretionary Tours Undertaken by a Parent and Children (absolute time)

	Lower Bound	Upper Bound
Departure time (minutes from 3 a.m.)		
If parent is worker	Minimum {375.0, work-home commute end time}	Minimum {1080.0, work-home commute end time}
If parent is non-worker	375.0	1080.0
Activity duration (minutes)	15.0	210.0
Travel time (minutes)	2.0	35.0

3. SOFTWARE DEVELOPMENT

The goal of the CEMDAP software development process is to provide a microsimulation platform that can be easily configured for different study areas with different levels of data availability. This chapter describes the software development aspects of CEMDAP. Section 3.1 discusses the use of an object-oriented development paradigm. Section 3.2 describes the CEMDAP system software quality attributes. Section 3.3 describes CEMDAP's software architecture in relation to the system quality attributes. Section 3.4 provides an overview of the strategies adopted to enhance the computational performance of CEMDAP. Finally, Section 3.5 summarizes the improvements in the software architecture, design, and implementation of the recent version of CEMDAP in comparison with that of a previous version (see Bhat et al. [2003] for the previous CEMDAP version).

3.1 The Development Paradigm

Several software development paradigms are currently in use. The two most popular are the procedural and the object-oriented (OO) paradigms. The procedural paradigm focuses on data flow and is based on performing actions on data. The approach entails three stages: (1) start with a *structured* analysis, (2) develop a *modular* design, and (3) write *procedural* programs. Because each stage of the procedural paradigm involves a different technique, the transition from one stage to the next is not direct. Therefore working out what parts of the program code are affected by a change in the requirements is complex. As a result, any requirement changes late in the development process would be difficult to accommodate.

The OO paradigm, in contrast to the procedural paradigm, focuses on *objects* and is based on the data (objects) performing actions on themselves. Thus no conversion is involved in moving from one stage to the next and accommodating late requirement changes is relatively easy. From this perspective, the OO paradigm better serves the goal of continual improvement and enhancements. The OO approach is also more suitable for the CEMDAP design because its fundamental concept of objects parallels the purpose of microsimulation (i.e., modeling the behavior of agents, or objects, in the real world).

The OO approach encompasses two basic techniques: *abstraction* and *encapsulation.* These techniques enable the management of complex simulation systems. Abstraction is a process that involves identifying the crucial behavior of an object and eliminating irrelevant and tedious details. A well-thought-out abstraction provides a greatly simplified representation of the real world from the perspective of the software developer. Abstraction is implemented through encapsulation, which is the mechanism of storing the *abstraction* as one cohesive unit describing the state (or behavior) of an object and the methods that manipulate that object. Encapsulation makes it possible to separate an object's implementation from its behavior, thus restricting access to the object's internal data. This is desirable because, while the fundamental nature of objects in the real world does not change much, the way in which they behave and interact with each other does. This separation of the *what* from the *how* is another reason that requirement changes are easily accommodated within the OO paradigm (Harrington, 1995).

3.2 Software System Quality Attributes

CEMDAP, as a software, has been developed to exhibit several desirable system quality attributes: (1) data integrity, (2) performance, (3) extensibility and modifiability, (4) buildability, and (5) usability. The focus of this section is on describing these qualities.

1. *Data Integrity.* CEMDAP manipulates large amounts of data pertaining to the population, land use, and transportation system of a city or metropolitan area. It is important to properly store, retrieve, and transfer data because data integrity directly affects the other quality attributes such as performance, build ability, and extensibility. Further, CEMDAP demands accuracy; without guaranteeing the accuracy of computation, correct storage, retrieval, and transfer of data are futile efforts. Hence it is important that large amounts of data be handled without compromising the integrity of the data itself.
2. *Performance.* The algorithms being used in an activity-travel system can be computationally intensive. Moreover, the computationally intensive algorithms are applied to a large amount of data, which may lead to very high simulation times. The practical need to limit the computation time to a reasonable period makes performance an attribute of primary importance.

3. *Extensibility and Modifiability.* The vision behind the CEMDAP development process is to develop a comprehensive system in which the activity-travel microsimulator is integrated with other modules such as the sociodemographic, land use, and economic system simulator and the dynamic traffic micro-assignment module. Therefore extensibility of CEMDAP is a desired quality attribute.

   CEMDAP is built with the mindset that it will be a continuing project and that its mathematical models, algorithms, and the simulation sequence (see Chapter 2 for the CEMDAP simulation algorithm) will continue to improve over time. While the extensibility of modeling modules aids evolution, the flexibility to substitute different models or introduce new models and the flexibility to modify the simulation sequence are of vital importance.
4. *Build ability.* Because the primary goal of this project is to demonstrate the applicability of the activity-based travel modeling approach, the software system produced at the end is meant to be a prototype achieved within a reasonable amount of time.
5. *Usability.* Eventual users of CEMDAP are not expected to be programmers or to be software savvy. A friendly and intuitive user interface is considered important for this project to go beyond its creators. In particular, a Microsoft Windows-like user interface is desired because of users' familiarity with the Windows style of applications.

The aforementioned quality attributes must be considered while creating the software architecture, as well as during the design, implementation, and deployment stages. The following section describes CEMDAP's system architecture and presents its salient features in relation to the system quality attributes.

3.3 System Architecture

This section describes the architecture of the CEMDAP software through the *decomposition view* and the *deployment view*. The decomposition view, which shows how CEMDAP's responsibilities are partitioned across modules, is presented in Section 3.3.1. The deployment view, which conveys how the system is set out to run correctly, is presented in Section 3.3.2.

3.3.1 Decomposition View of CEMDAP

This section presents the decomposition structure of CEMDAP and introduces the sub-modules and relations within CEMDAP. As shown in Figure 3-1, the major components of CEMDAP are the Input Database, Data Coordinator, Run-time Data Objects, Modeling Modules, Simulation Coordinator, Application Driver, and Output Files. A brief overview of each of these system components is presented in this section.

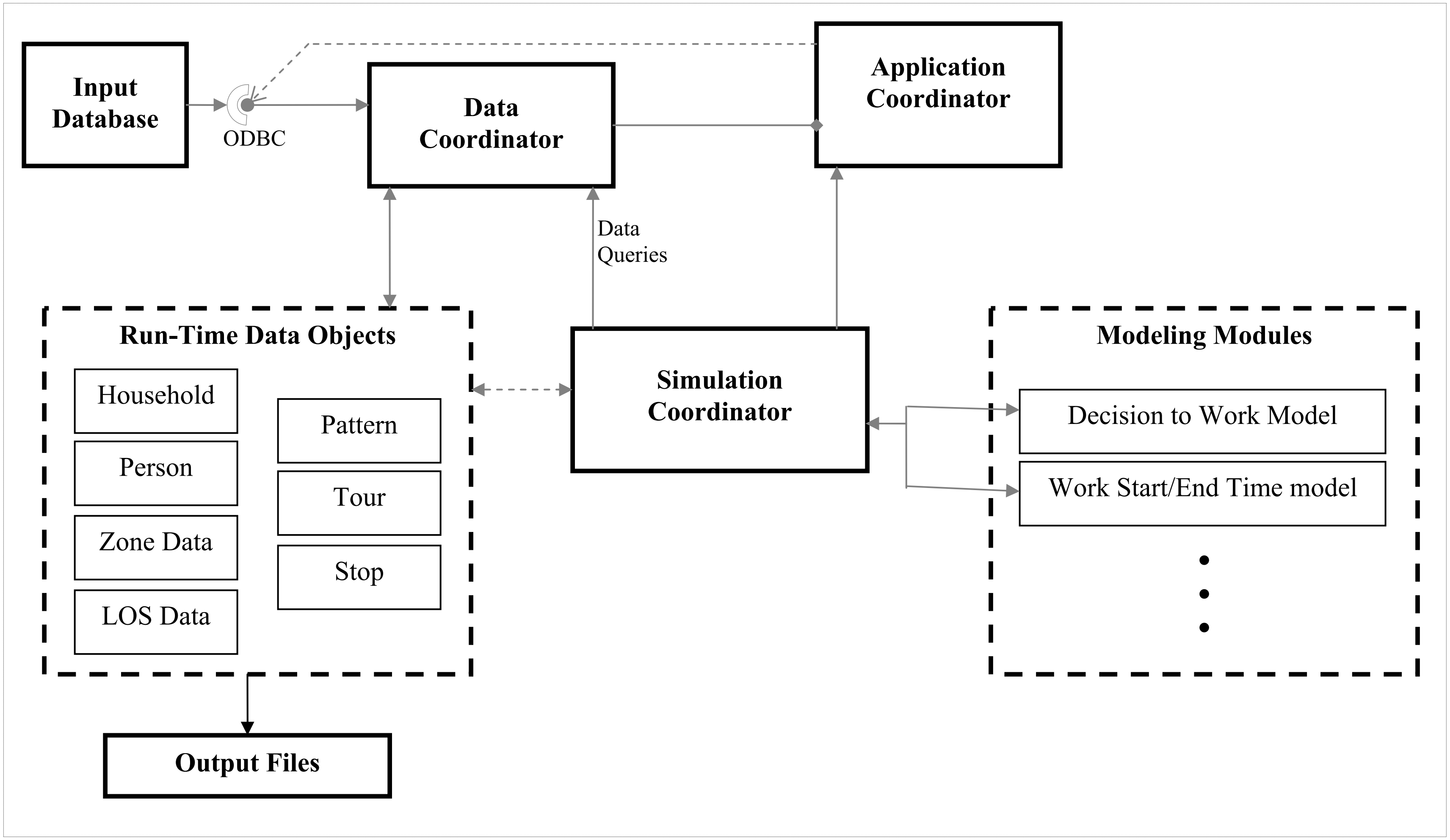

Figure 3-1 Decomposition Structure of CEMDAP Software Architecture

3.3.1.1 *Input Database*

The simulation of activity-travel patterns is a data intensive exercise. Three sets of data are required: (1) disaggregate socioeconomic 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. The details of the input data schema are provided in the CEMDAP user manual (Bhat et al. 2006).

The CEMDAP architecture requires these input data to be stored in a relational database management system (DBMS). Alternatives to using a relational DBMS are storing data as flat-files on the system, storing as xml-files or spreadsheets, or creating a custom data store specific for this application. The reason for choosing a DBMS to store data is to leverage on the last 30 to 40 years of research advances in storage, organization, query, and management of large volumes of data. A DBMS provides the following architectural qualities required of a data store component:

1. *Performance and Security.* Almost every successful commercial DBMS places great importance on the performance of the servers. These systems achieve efficient query execution through internal mechanisms of indexing relations, Relational Query Algebra, and Relational Query Optimization. Most DBMSs also provide higher performance through concurrent executions of multiple queries without corrupting the state of the underlying data or interfering with another query's execution.

2. *Multi-user Access Control.* User management and access control is one of the architectural drivers for using a DBMS rather than its alternatives. Although not being used in the current version of the CEMDAP application, it leaves room to expand the application to provide multi-user access control.

3. *Portability.* Other alternatives restrict data to a particular format, on a particular machine. Using a DBMS alleviates this concern by providing a standard interface to access the data regardless of the internal format the DBMS is storing it in or the machine the DBMS server runs on.

4. *Modifiability.* Restructuring the data schema, as well as adding new data, is handled easily in a DBMS. The server contains internal mechanisms for re-indexing and composing queries to handle the new data.

5. *Flexibility (in query).* Structured Query Language (SQL) is a domain-specific language for generating a program to extract desired pieces of information from a data store. A clear advantage of using a DBMS is that queries can be easily modified and the complexity of generating an efficient program to execute the query rests upon the DBMS. Because CEMDAP requires the extraction of a variety of information, flexibility in querying data is of primary importance.

6. *Reliability.* Many DBMSs provide facility to backup and restore information and guarantee a consistent state of the data. Building a custom solution to this problem is a daunting task.

7. *Scalability.* Depending on the operating system, input data file sizes have an upper limit. Accessing large data files may demonstrate time complexity proportional exponentially to the amount of data. Clearly, these means are not scalable beyond a certain point. Some DBMSs can provide up to few terabytes of storage and, most importantly, efficient query execution that is not proportional to the volume of data.

8. *Robustness.* Because of the commercial nature of DBMS and the volume of user base, they have been thoroughly tested for errors and can generally be considered robust. An alternative to DBMS as data store would need to suffice the robustness criteria.

9. *Affordability.* Several commercial DBMSs provide a much cheaper alternative to developing custom-built data storage solutions. These include MS Access, PostgreSQL, mySQL, and MS SQL Server Desktop Edition. CEMDAP is designed to interact with several DBMSs through an Open Database Connectivity (ODBC) interface, which will be discussed in Section 3.3.2., to provide database portability. However, it should be noted that the portability may be compromised in the use of SQL query statements. Most DBMSs support standard SQL and their proprietary language features. Hence queries may need to be modified after migrating to a new DBMS.

3.3.1.2 Data Coordinator

The Data Coordinator is the component responsible for establishing the ODBC connection and interacting with the external database that contains the input data. It extracts the content and structural information of the data tables and converts data into their corresponding data structures that are used within CEMDAP. It is also responsible for all data queries to the database during the process of simulation. By limiting the database interaction to this one system entity, any changes pertaining to the database can be made more easily. The approach helps achieve the portability objective with respect to database changes.

3.3.1.3 Run-Time Data Objects

These are the main data structures that CEMDAP operates on internally. Instances of household, person, zone, zone-to-zone, and LOS entities are created by the data coordinator from the input database. The remaining entities (i.e., pattern, tour, and stop) are created by the simulation coordinator (discussed in later sections) as required during the simulation process. The run-time data objects also act as a cache for the simulation coordinator that frequently accesses some data. Use of these caches instead of accessing data from the input database addresses the performance quality objective.

3.3.1.4 Modeling Modules

CEMDAP microsimulates the activity-travel patterns by implementing the individual modeling modules in a sequence described in Chapter 2. Each modeling module in the system corresponds to a behavioral model in the framework described in Chapter 2. Each decision variable is associated with an instance of one of these modeling modules. Once a module is configured via the user interface, it possesses knowledge about the econometric structure and all the relevant parameters required to produce the probability distribution for the given variable. When called upon, the module executes a prediction algorithm to determine the corresponding choice.

Although the modeling modules are many, they are derived from a limited number of econometric structures. Currently, six types of econometric models are implemented in CEMDAP as model templates: regression, hazard duration, binary logit, multinomial logit, spatial location choice, and ordered probit models. Additional econometric structures may be added to this library of model templates. By making the modeling modules almost-replaceable units, CEMDAP is addressing the extensibility objective with respect to changing models.

3.3.1.5 Simulation Coordinator

The simulation coordinator is responsible for running the simulation and controlling the flow of the simulation. It coordinates the logic and sequence in which the modeling modules are called. The simulation coordinator holds a reference to the data coordinator and makes use of it during the simulation sequence. The simulation coordinator operates on the run-time data objects; data objects are created and manipulated as the corresponding choice outcomes are predicted with each modeling component. In addition, the simulation coordinator performs any required consistency checks and keeps track of the progress of the simulation as the simulation advances.

It is expected that the simulation sequence will evolve over time as more research is carried out. Changes necessary to the simulation sequence are centralized in one place. Designing the simulation coordinator in this way addresses the modifiability of simulation sequence criteria.

3.3.1.6 Application Driver

The application driver starts and runs the application. On startup, it instantiates the user interface and obtains handles to the simulation coordinator and the data coordinator. It references the ODBC driver for opening and closing the database connection. It also coordinates the functionality offered to the users—such as selecting input data source, choosing the output path, loading and saving the CEMDAP model specification files (see the user manual by Bhat et al., [2006] for details on the specification file), and running the simulation.

3.3.1.7 OutputFiles

The output of CEMDAP is written to flat-files (plain tabbed formatted files) that are selected through the Graphical User Interface (GUI). The reason for choosing flat-files for output data storage rather than a DBMS was mainly to maintain ease and flexibility. Because the output is sequential, it is amenable to being streamed into a flat-file. Also, because the output may need to be processed by other generic applications or imported to various DBMSs, a plain tabbed formatted file is simple and can be read by most other software and DBMSs.

3.3.2 Deployment View of CEMDAP

Figure 3-2 illustrates the deployment structure of CEMDAP. CEMDAP Binary is the core executable component that embodies the functionality of CEMDAP. It is designed to run on a single Microsoft (MS) Windows host machine and makes use of several external components that are also designed for a MS Windows host machine. The "*vc user crt71 rtl* *x86.msm*" library is a generic library on which Microsoft's Visual C++ application relies. The "*vc user mfc71 rtl x86.msm*" library contains Microsoft Foundation Classes (MFC), which contain common Windows-GUI components prewritten by Microsoft. The library "*vc user stl71 rtl x86.msm"* is yet another reusable library supplied by Microsoft. It contains common algorithms as Templated functions. The "*mxxml.dll"* library contains common extended markup language (XML) parsing routines and "*ODBC32.dll"* library contains routines to interact with an ODBC-compliant database. These dependent libraries are provided in the CEMDAP installation package. These reusable libraries aid in reducing the development time and hence help achieve the buildability objective. Moreover, the use of MFC helps achieve the usability objective by providing MS Windows-style user interface.

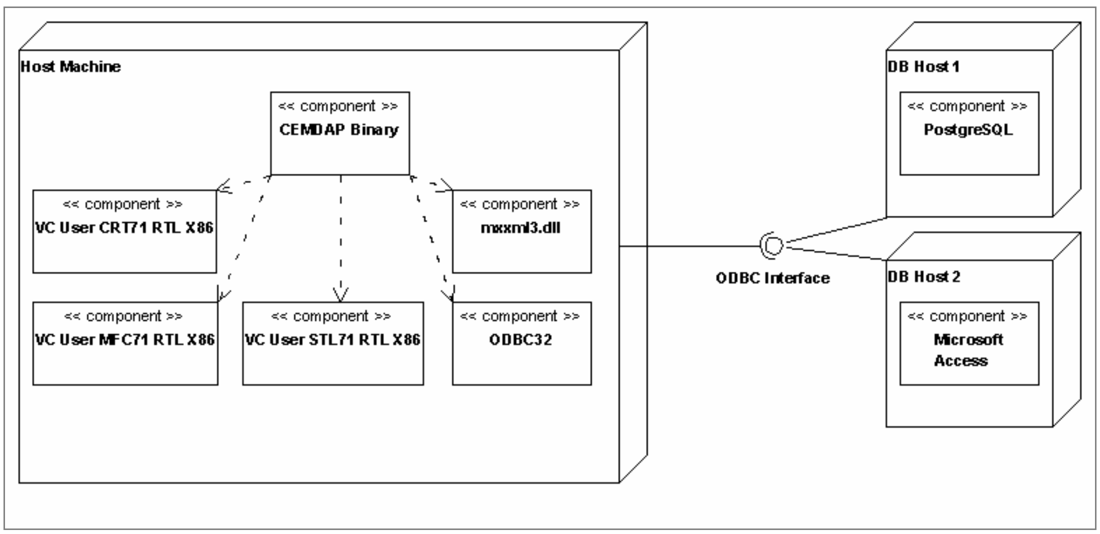

Figure 3-2 Deployment Structure of CEMDAP Software Architecture

As mentioned earlier, CEMDAP interacts with a relational DBMS through an ODBC. ODBC provides a product-independent interface between client applications (CEMDAP, in this case) and database servers, allowing applications to be portable between database servers from different manufacturers. In practice, ODBC has turned out to be a standard mechanism for communicating with a database even if portability is not a key factor. Another advantage of interfacing through an ODBC interface is that the database servers and CEMDAP application can be run on different machines with no additional complexity in interacting with the database over the network. Figure 3-2 illustrates this point by showing that both Microsoft Access and PostgreSQL databases can interact with CEMDAP from different machines. Through the ODBC interface, CEMDAP can access data from DBMS such as Microsoft Access and PostgreSQL and alleviate data management efforts within CEMDAP.

While describing the architecture, it is also necessary to mention the tradeoffs made by selecting an architectural option. The downside to ODBC is the potential performance degradation resulting from the additional processing overhead of accessing the data from the database server. Even though the difference is only on the order of milliseconds, with an application such as CEMDAP that makes frequent accesses to data, such a difference can quickly add up to a significant increase in processing time. Yet the rationale for using ODBC interface is clear. Other forms of database interactions, such as proprietary protocols supported by each database vendor, would compromise the portability of CEMDAP between different databases. In addition, developing routines for custom database interaction would increase the development time.

Admittedly, ODBC interface does result in significant performance degradation. Hence, strategies such as multithreading and data caching are adopted to enhance the computational performance of CEMDAP. These strategies are described in the following section.

3.4 Performance Enhancement Strategies

The computationally intensive algorithms used in CEMDAP are applied to large amounts of data, which further increases simulation time. In addition, data access through the ODBC interface can add to the performance degradation. This section describes two performance enhancement strategies adopted to enhance the performance of CEMDAP-multithreading and data caching.

3.4.1 Multithreading

Multithreading is a way of efficiently utilizing computing resources (for example, the central processing unit or the processor of a computer). In multithreading, the data and information pertaining to multiple tasks (instead of a single task) are loaded into the memory of a processor, which rapidly switches between the various tasks at a fixed time interval called time slice. All the tasks (or parts of the tasks) are handled in a sequence (not simultaneously) by the processor. Although the processor handles one task at a time, loading multiple tasks into its memory enables it to quickly switch between various tasks and improves the performance because the number of data queries and the intensity of data access through the ODBC interface are reduced. Thus, while not compromising on the portability feature enabled by the ODBC interface, we are addressing the performance objective.

In CEMDAP, multithreading is enabled by loading the input data related to several households into the processor. It is to be noted that the time slice has to be small enough to allow a large number of tasks (households in this case) to be handled and, at the same time, it has to be large enough that each task is allocated a sufficient amount of processor time to get useful work done. The number of threads that can be run at a time (or the number of households that can be loaded into the memory of the processor) depends on the processor speed and the Random Access Memory (RAM) of the machine.

3.4.2 Data Caching

CEMDAP manipulates large amounts of data pertaining to the population, land use, and transportation system of a city/metropolitan area. Frequent data access calls to such large databases through the ODBC interface may degrade the overall performance. A strategy adopted to counter such performance degradation is to cache large amounts of data so as to reduce the number of data access calls through the ODBC interface.

The optimal extent of data-caching depends on the machine configuration (RAM and the processor speed) and the size of the input data (especially the LOS files). The input data size varies with the size of the city or metropolitan area to which CEMDAP is being applied and the spatial and temporal resolution at which the LOS files are loaded into CEMDAP.

It may be possible to cache the entire LOS data for achieving greater simulation speeds. However, any move toward finer spatial and temporal resolutions and larger study areas would cause a significant increase in the LOS data size and limit the extent to which the LOS data can be cached. Hence, cleverly designed partial-data caching routines are built into CEMDAP so that only frequently used data is temporarily cached. For example, the LOS data corresponding to an origin zone is cached into CEMDAP until all the households belonging to that particular zone are processed. Thus, the LOS data access calls corresponding to that particular origin zone are avoided until the next household to be processed belongs to a different zone. Similarly, the commute LOS data (the LOS data between residential and employment zones during the commute start and end times) of a worker is cached when he or she is being processed.

The data caching mechanism can be used to cache data in several more possible ways to efficiently handle the data access and usage in CEMDAP. In addition, the data caching and the multithreading mechanisms can be synchronized to further increase the efficiency of data access and usage. Further exploration of the use of data caching and multithreading mechanisms may help increase the data handling efficiency and the simulation speed of CEMDAP.

3.5 An Overview of the Software Enhancements

The recent version of CEMDAP is significantly improved from the previous version in several ways (see Bhat et al. [2003] for details on the previous version of CEMDAP). The following software enhancements are incorporated into the recent version of CEMDAP.

1. CEMDAP now uses PostgreSQL as the DBMS, rather than Microsoft Access. PostgreSQL is an open source database software released under the Berkeley Software Distribution (BSD) license. It is known to be stable at large data loads and accommodates larger data size resulting from a higher resolution in terms of space and time. Thus, CEMDAP, which is not limited by the input data size, can be easily deployed in study areas of varying sizes with varying levels of spatial and temporal configurations of the LOS data. On the other hand, the previous version of CEMDAP was limited to a maximum of 1 gigabyte of input data load capacity.
2. CEMDAP has built-in data caching routines to temporarily store frequently accessed data items in RAM to reduce the number of queries and disk accesses. The data caching routines are written to allow the developers (not the users) to easily customize the extent of data caching depending upon the size of the study area and the spatial and temporal resolution of the LOS data.
3. The system computational efficiency is enhanced by carrying out the simulation over multiple threads. In addition, the system allows the developers to easily customize the extent of multithreading (i.e., the number of threads of computation) for machines of different configurations (processing speed and RAM capacity).
4. In addition to the above three significant enhancements, a new simulation sequence is implemented in CEMDAP to accommodate the newly developed modeling sequence that incorporates activity-travel patterns of children and intrahousehold interactions between adults and children. Also, separate simulation coordinators are implemented to control the simulation sequence for different types of households and a new model module

(similar to a multinomial logit modeling module) is added to the system for jointly simulating work start and end times.

4. SYNTHETIC POPULATION GENERATOR

The preceding chapter summarized the software architecture employed for developing CEMDAP. This chapter discusses the synthetic population generator (SPG), which is the component of CEMDAP that creates the base year initial population, as well as the 2025 forecast year population. The synthetic population generation process creates, as outputs, data records describing the sociodemographic characteristics of individuals and households residing in the study area. The generation process typically involves an aggregate dataset that represents the desired or expected marginal distribution of the variables and a disaggregate dataset that is a collection of records representing a sample of the"real" households and individuals in the population. The aggregate data are typically drawn from aggregate census data, such as the Summary Files (SF) of the U.S. and the Small Area Statistics (SAS) files of the U.K. Examples of the disaggregate dataset, on the other hand, include the Public-Use Microdata Samples (PUMS) of the U.S. and the Sample of Anonymized Records (SAR) of the U.K. Given the aggregate and disaggregate datasets, the population records are produced by selecting sample records from the disaggregate dataset to meet the marginal distribution given by the aggregate dataset.

The rest of the chapter is structured as follows. Section 4.1 discusses the algorithm used to produce the synthetic population records. Section 4.2 discusses the datasets assembled for generating the base year and forecast year populations. Section 4.3 presents the results of a validation exercise that compares the SPG outputs with census data.

4.1 SPG Algorithm

As mentioned earlier, the generation of a synthetic population requires an aggregate dataset and a disaggregate dataset that provide information about the sociodemographic variables considered to significantly impact individuals' activity-travel decisions. Typically, the aggregate dataset comprises a set of cross-tabulations that describe the one-, two-, or multi-way distributions of some (but not all) of the desired sociodemographic attributes at a relatively fine spatial resolution (for example, census block groups that can be as small as two street blocks).

We refer to these attributes with known marginal distributions as the control variables and to the spatial units for which the aggregate distribution information is available as the target areas. The disaggregate dataset, on the other hand, provides information for all the desired sociodemographic variables of interest, but for only a sample of households and individuals. We refer to the spatial units for which the disaggregate distribution information is available as the seed areas. Seed areas are typically larger than the target areas (for instance, the PUMS data are available for the Public-Use Microdata Areas, or PUMA, which are areas of no less than 100,000 in population).

Given a pair of target and seed areas, the SPG creates the synthetic population for the target area based on the algorithm shown in Figure 4-1 (the mathematical details and an example application of the algorithm are provided in Appendix B). In Step 1, the cross-tabulations that provide the marginal distributions of the household-level control variables are used to construct the full multi-way distribution across all the household-level control variables using a procedure known as the Iterative Proportional Fitting Procedure (IPFP). In Step 2, the full multi-way distribution across all the individual-level control variables is also constructed using the IPFP. In Step 3, separate count tables are constructed to keep track of the numbers of households and individuals belonging to each demographic group that have been selected into the target area during the subsequent population generation process. At this point, the cells in the two tables are initialized to zero to reflect the fact that no households and individuals have been created for the target area. During subsequent iterations, these cell values will be updated as households and individuals are selected into the target area. Step 4 entails assigning each PUMS sample household in the corresponding seed area a probability of being selected into the target area. The probability is a function of the multi-way distribution obtained in Step 1 and the distribution of the households already selected into the target area.The selection probability of a sample household decreases as more households from the same demographic group are selected into the target area. In Step 5, a household is randomly drawn from the pool of sample households to be considered for "cloning" and added to the population for the target area based on the probabilities computed in the previous step. Step 6 determines if the randomly selected household should be added to the synthetic population to help meet the multi-way distributions obtained from Steps 1 and 2. If so, the randomly selected household is added to the pool of the synthetic population for the target area in Step 7 and the count tables are updated in Step 8;

otherwise the randomly selected household is removed from the consideration set so that it will never be selected again. Steps 4 through 8 are repeated until either the desired number of households is reached or there are no more households in the consideration set. See Guo *et al.* (2005) and Guo and Bhat (2006) for a more detailed discussion of this algorithm.

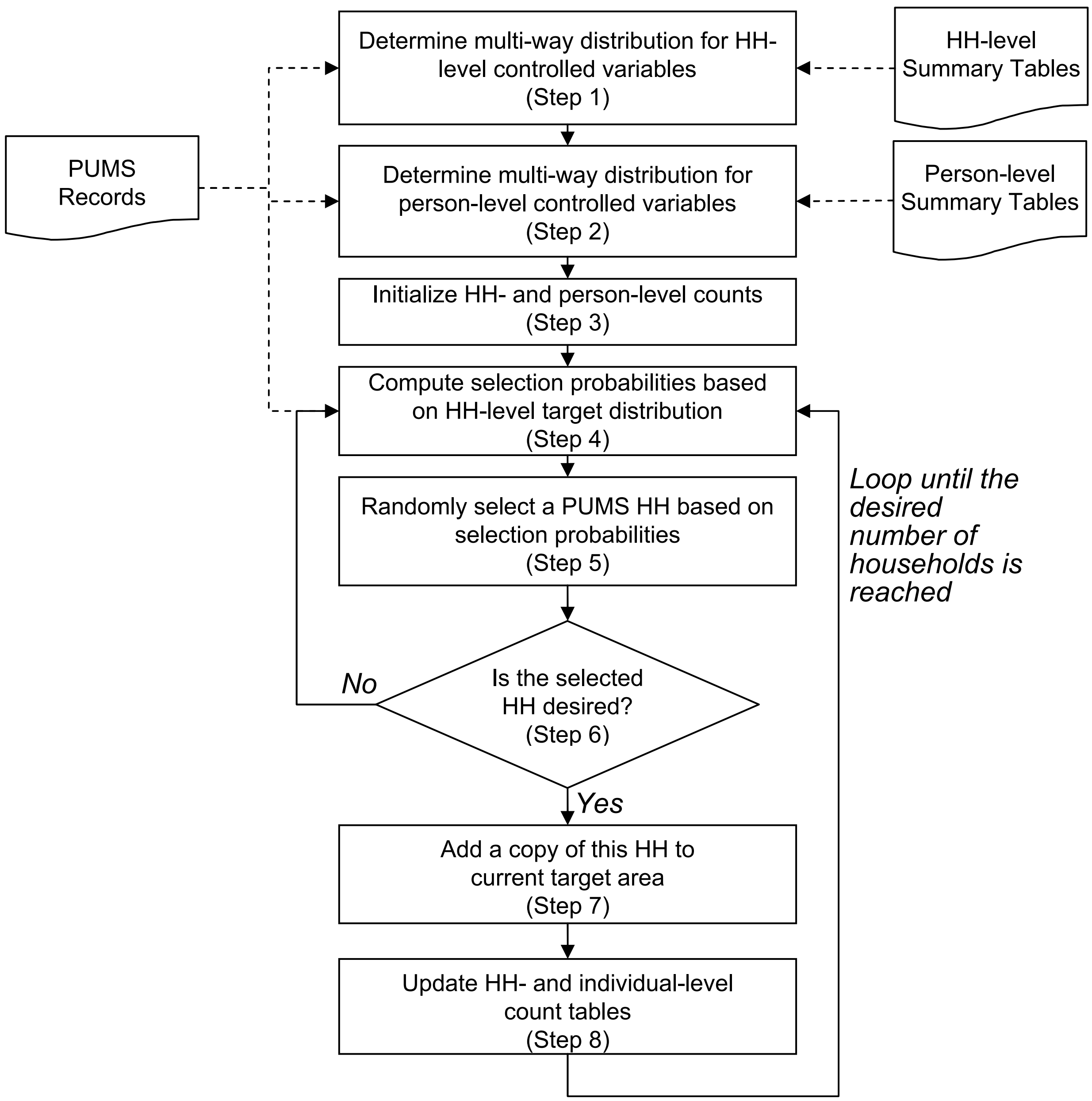

Figure 4-1 Overview of the Population Synthesis Algorithm

4.2 Input Data Sources

The SPG described in the preceding section has been designed to create the synthetic population for any given study area and any given analysis year for which the required aggregate and disaggregate datasets are available. For the purpose of this project, the SPG has been applied to produce the population for the base year (2000) and the forecast year (2025) for the Dallas–Fort Worth area. For this application, five control variables are selected at the household level: family indicator (HH_FAM), household size (HH_SIZE), household type (HH_TYPE), presence of children (HH_CHILDREN), and age of householder (HHR_AGE). Three controlled variables are selected at the individual-level: gender (P_GENDER), race (P_RACE), and age (P_AGE). In the remainder of this section, we describe the definitions of and the data sources for the control variables for the base year (Section 4.2.1) and the forecast year (Section 4.2.2).

4.2.1 Input Data for Base Year

The generation of the DFW base year synthetic population relies on two data sources: (1) Census 2000 summary file SF1 and (2) 2000 five-percent PUMS data. Census SF1 is a collection of summary tables, based on a 100 percent population survey, of household and individual demographic variables for census tracts, block groups, or blocks. Some of the summary tables describe the distribution of a single variable, while other tables are cross-tabulations describing the distribution of multiple variables. In particular, Table P20 of census SF1 describes the joint distribution of four of our household-level control variables (HH_FAM, HH_TYPE, HH_CHILDREN, and HHR_AGE) and Table P26 of Census SF1 describes the joint distribution of two of the household-level control variables (HH_FAM and HH_SIZE). Table 4.1 shows the definition of these control variables. Tables 4.2 and 4.3 provide the mapping between the control variables and the fields in P20 and P26, respectively. Census Table P7 describes the P_RACE individual-level control variable and Table P12 describes the joint distribution of the P_GENDER and P_AGE control variables. The definitions of these individual-level control variables are shown in Table 4.4; the mapping between these variables and the corresponding census tables are presented in Tables 4.5 and 4.6. These four census tables (P20, P26, P7, and P12) provide the desired cross-tabulations for census block groups, which form our target areas for the DFW application. The PUMS data, on the other hand, provide the five-percent sample records of households and individuals in the population. Each record is geographically referenced by a PUMA, which forms our seed areas for the DFW application. A block-group-to-PUMA lookup table, available from the Census Bureau, is used to determine the target-seed area pairings.

Table 4.1 Household-Level Control Variables Defined for the Base Year

Variable Name	Value	Value Description
HH_FAM	0	Family
	1	Non-family
HH_TYPE	1	Family: married couple
	2	Family: male householder, no wife
	3	Family: female householder, no husband
	4	Non-family: householder alone
	5	Non-family: householder not alone
HH_CHILDREN	0	No own children under 18
	1	Own children under 18 years
HHR_AGE	0	15-64
	1	65 and over
HH_SIZE	0	1 person
	1	2 persons
	2	3 persons
	3	4 persons
	4	5 persons
	5	6 persons
	6	7 or more persons

Table 4.2 Mapping between the SF1 Table P20 and the Household-Level Control Variables

Field in P20	HHR_AGE	HH_FAM	HH_TYPE	HH_CHILDR
P020005	0	0	1	1
P020006	0	0	1	0
P020009	0	0	2	1
P020010	0	0	2	0
P020012	0	0	3	1
P020013	0	0	3	0
P020015	0	1	4	0
P020016	0	1	5	0
P020020	1	0	1	1
P020021	1	0	1	0
P020024	1	0	2	1
P020025	1	0	2	0
P020027	1	0	3	1
P020028	1	0	3	0
P020030	1	1	4	0
P020031	1	1	5	0

Table 4.3 Mapping between the SF1 Table P26 and the Household-Level Control Variables

Field in P26	HH_FAM	HH_SIZE
P026003	0	1
P026004	0	2
P026005	0	3
P026006	0	4
P026007	0	5
P026008	0	6
P026010	1	0
P026011	1	1
P026012	1	2
P026013	1	3
P026014	1	4
P026015	1	5
P026016	1	6

Table 4.4 Individual-Level Control Variables Defined for the Base Year

Variable Name	Value	Value Description
P_RACE	0	White alone
	1	African-American alone
	2	American-Indian and Alaska Native alone
	3	Asian alone
	4	Native Hawaiian and other Pacific Islander alone
	5	Some other race alone
	6	Two or more races
P_GENDER	0	Male
	1	Female
P_AGE	0	Under 5 years
	1	5 to 14 years
	2	15 to 24 years
	3	25 to 34 years
	4	35 to 44 years
	5	45 to 54 years
	6	55 to 64 years
	7	65 to 74 years
	8	75 to 84 years
	9	85 and more

Table 4.5 Mapping between the SF1 Table P7 and the Individual-Level Control Variable

Field in P7	P_RACE
P007001	0
P007002	1
P007003	2
P007004	3
P007005	4
P007006	5
P007007	6
P007008	7

Table 4.6 Mapping between the SF1 Table P12 and the Individual-Level Control Variables

Field in P12	P_GENDER	P_AGE
P012003	0	0
P012004+P012005	0	1
P012006+…+P012010	0	2
P012011+P012012	0	3
P012013+P012014	0	4
P012015+P012016	0	5
P012017+…+P012019	0	6
P012020+…+P012022	0	7
P012023+P012024	0	8
P012025	0	9
P012027	1	0
P012028+P012029	1	1
P012030+…+P012034	1	2
P012035+P012036	1	3
P012037+P012038	1	4
P012039+P012040	1	5
P012041+…+P012043	1	6
P012044+…+P012046	1	7
P012047+P012048	1	8
P012049	1	9

4.2.2 Input Data for Forecast Year

As in the case of generating the synthetic population for the base year, the generation of a synthetic population for the forecast year also requires an aggregate and a disaggregate dataset. However, since the future has not taken place yet, the datasets required for SPG are not readily available. Instead, we use the 2000 PUMS as the disaggregate dataset for the forecasting year and the 2000 PUMA as the seed areas, assuming that the 2000 PUMS will be a representative sample of the forecast year population. As for the aggregate dataset that provides the marginal distributions of the various control variables for the forecast year, we create cross-tabulations resembling those used for the base year by synthesizing population data from different sources. As summarized in Table 4.7, these data sources include (1) the 2025 regional and TAZ-level household and population totals predicted by NCTCOG (see NCTCOG, 2003 and NCTCOG, 2006); (2) the 2025 county-level population projections by race, age, and gender provided by Texas State Data Center (TSDC) (see Texas State Data Center, 2006); and (3) the 2000 synthetic population created by SPG. Below, we discuss in more detail the processes by which the household- and individual-level cross-tabulations are created based on these three data sources.

Table 4.7 Forecast Data, Sources, and Application

Forecast Data for 2025	Data Source	Summary File Application
Zonal-level Household Population	NCTCOG	Household
Zonal-level Individual Population		Individual
County-level three-way Race, Age, and Gender Marginal Totals	Texas State Data Center	Individual

4.2.2.1 Creation of Household-Level Cross-Tabulations

For the purpose of creating the forecast year synthetic population, we want to use the same household-level control variables and the same value category definitions used for the base year. These are the HHR_AGE, HH_FAM, HH_TYPE, HH_CHILDREN, and HH_SIZE variables defined in Table 4.1. Moreover, we want to use the same cross-tabulation structures as the ones defined in Tables 4.2 and 4.3 for the base year, except that the new tabulations will reflect the household distribution at the TAZ-level, as opposed to the census block-group-level. This is because TAZ is our choice of target area for generating the synthetic population for the forecast year.

The creation of the two cross-tabulations for year 2025 essentially involves performing the following two steps for each target TAZ - *i*:

Step 1: Populate the HHR_AGE by HH_FAM by HH_TYPE by HH_CHILDREN cross-tabulation, and the HH_FAM by HH_SIZE cross-tabulation, for zone *i* based on the household counts observed in the population synthesized for zone *i* for the base year.

Step 2: Apply an expansion factor ($Total2025_i$ /$Total2000_i$) uniformly to all cells in the two cross-tabulations, where $Total2025_i$ and $Total2000_i$ are the total zonal population for year 2025 (as predicted by the NCTCOG) and year 2000 (as computed for the base year population).

4.2.2.2 *Creation of individual-level cross-tabulations*

For the individual-level cross-tabulations, we again use the same control variables—P_RACE, P_AGE, and P_GENDER—as those defined for the base year (see Table 4.4). However, in order to utilize the projection data from TSDC, we need to use the same category definitions as used by TSDC. As shown in Table 4.8, the category definitions for the forecast year are more aggregated than those for the base year (this can be readily observed by comparing Tables 4.8 and 4.4). Also, instead of producing two separate cross-tabulations as for the base year, we create a single three-way cross-tabulation for the individual-level control variables for the forecast year because we want to make the most out of the county-level three-way tabulation readily provided by TSDC.

Table 4.8 Definition of Individual-Level Variables for Forecast Year

Variable Label	Size	Value	Value Description
P_RACE	4	0	White alone
		1	Black African-American alone
		2	Hispanic alone
		3	Others alone
P_GENDER	2	0	Male
		1	Female
P_AGE	5	0	Under 18 years
		1	18 to 24 years
		2	25 to 44 years
		3	45 to 64 years
		4	65 years and older

The process of creating the individual-level cross-tabulation for a TAZ *i* in county *j* entails the following four steps:

Step 1: Construct a P_RACE by P_GENDER by P_AGE cross-tabulation of percentages (based on the new categorization) for county *j* based on the population counts observed in the population synthesized for county *j* for the base year. We will denote each cell in this tabulation as $Pop2000_{x,y,z,j}$, that is, the percentage of people of race *x*, gender *y*, and age *z* in county *j*.

Step 2: Similarly, construct a P_RACE by P_GENDER by P_AGE cross-tabulation of percentages for zone *i* based on the population synthesized for the base year.

We will denote each cell in this tabulation as $Pop2000_{x,y,z,i}$, that is, the percentage of people of race x, gender y, and age z in zone i.

Step 3: Update the tabulation constructed in Step 2 by adding to each cell value $Pop2025_{x,y,z,i}$ an expansion factor ($Pop2000_{x,y,z,j}$ - $Pop2025_{x,y,z,j}$) for all x, y, and z, where $Pop2025_{x,y,z,j}$ is given by the TSDC data and $Pop2000_{x,y,z,j}$ is given by Step 1. Note that the same expansion factor is applied to all zones in the same county.

Step 4: Multiply all cells in the tabulation resulting from Step 3 by $ExpectedTotal2025_i$, where $ExpectedTotal2025_i$ is the zonal population size as predicted by NCTCOG. This ensures that the zonal population total is consistent with NCTCOG projection.

4.3 Verification

The synthetic populations generated using the process outlined in Section 4.1 and the data described in Section 4.2 are verified in this section. Section 4.3.1 and Section 4.3.2 discuss the verification findings for the base year population and the forecast year population, respectively. The reader should note the purpose of the verification exercise is to ensure that the synthesizing results are consistent, to the greatest extent possible, with the marginal distributions given by the input data. For the validation of the SPG algorithm itself, see Guo and Bhat (2006).

4.3.1 Verification of Base Year Synthetic Population

The base year synthetic population is verified against the marginal distributions given by the census SF1 data. The (observed) marginal totals corresponding to the rows in Tables 4.2, 4.3, 4.5, and 4.6 are computed based on the base year synthetic population and compared against the corresponding (expected) marginal totals given by SF1 tables. This comparison is done for all census block groups. As an example, the observed and expected marginal totals for a set of block groups in Tarrant County are aggregated and shown in Figure 4-2. The codes on the x-axis for Figure 4-2a correspond to combinations of household-level control variables, as mapped in Tables 4.2 and 4.3, while the codes on the x-axis for Figure 4-2b correspond to combinations of person-level control variables, as mapped in Tables 4.5 and 4.6. It can be seen that minor discrepancies exist for a few groups, but, overall, the SPG is able to produce synthetic populations that are consistent with the input aggregate data.

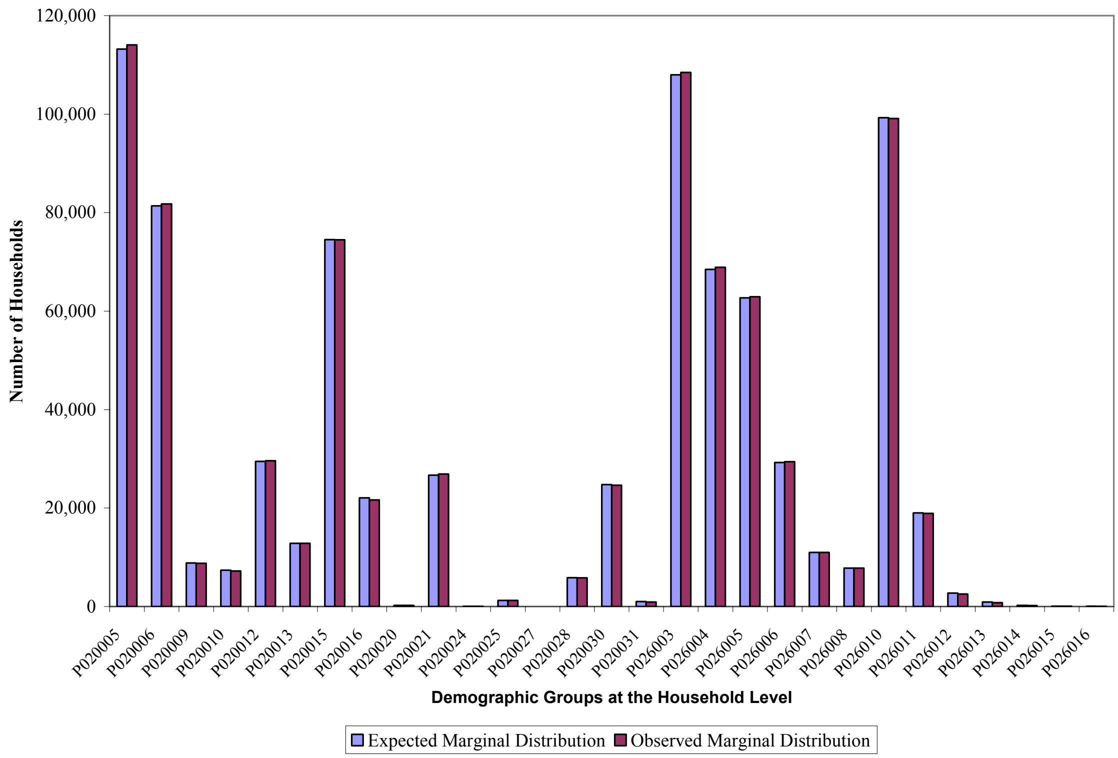

Figure 4-2a Comparisons between Expected and Observed Marginal Distributions for Household-Level Control Variables for the Base Year

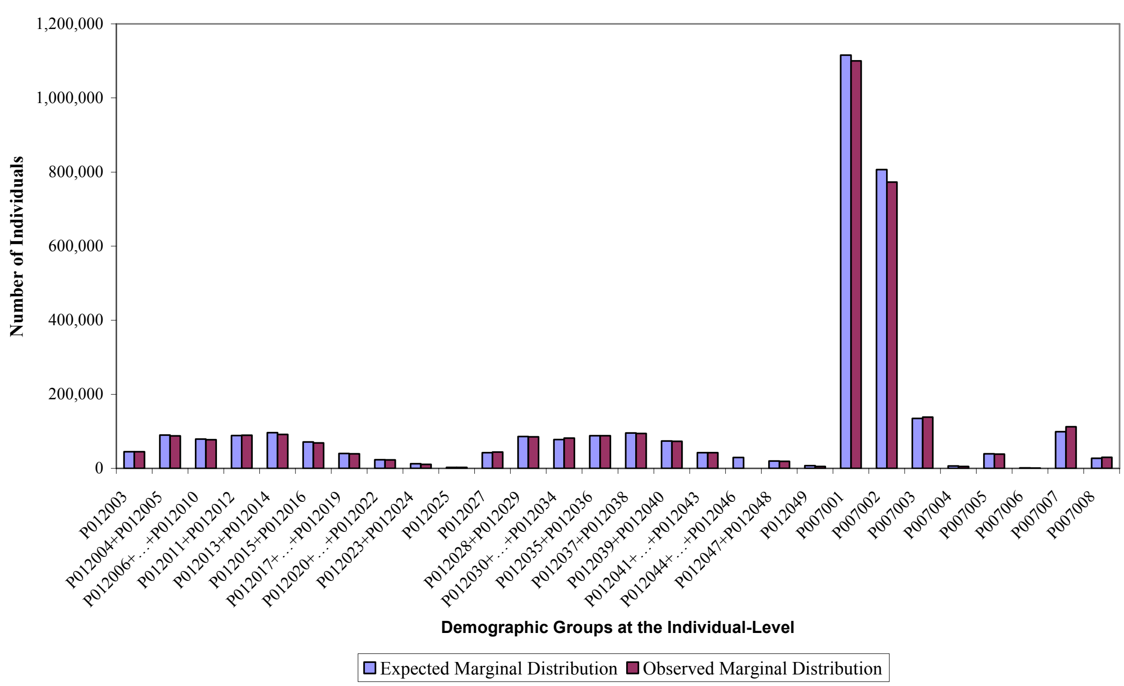

Figure 4-2b Comparisons between Expected and Observed Marginal Distributions for Individual-Level Control Variables for the Base Year

4.3.2 Verification of Forecast Year Synthetic Population

The forecast year synthetic population is verified using the same procedure as described in the preceding section. Since the target area used for the forecast year is the TAZ, the comparison between expected and observed marginal totals is done for all TAZs. The marginal totals for a set of TAZs in Tarrant County are aggregated and reported in Figure 4-3. Again, it can be seen that SPG satisfactorily produced a synthetic population to meet the input aggregate data.

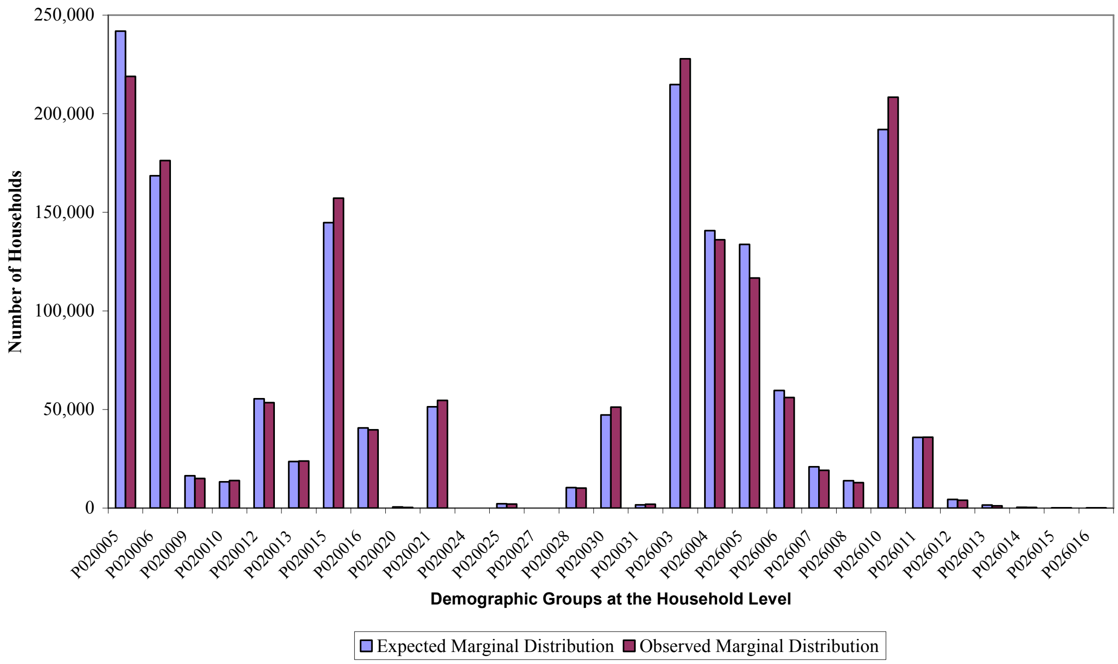

Figure 4-3a Comparisons between Expected and Observed Marginal Distributions for Household-Level Control Variables for the Forecast Year

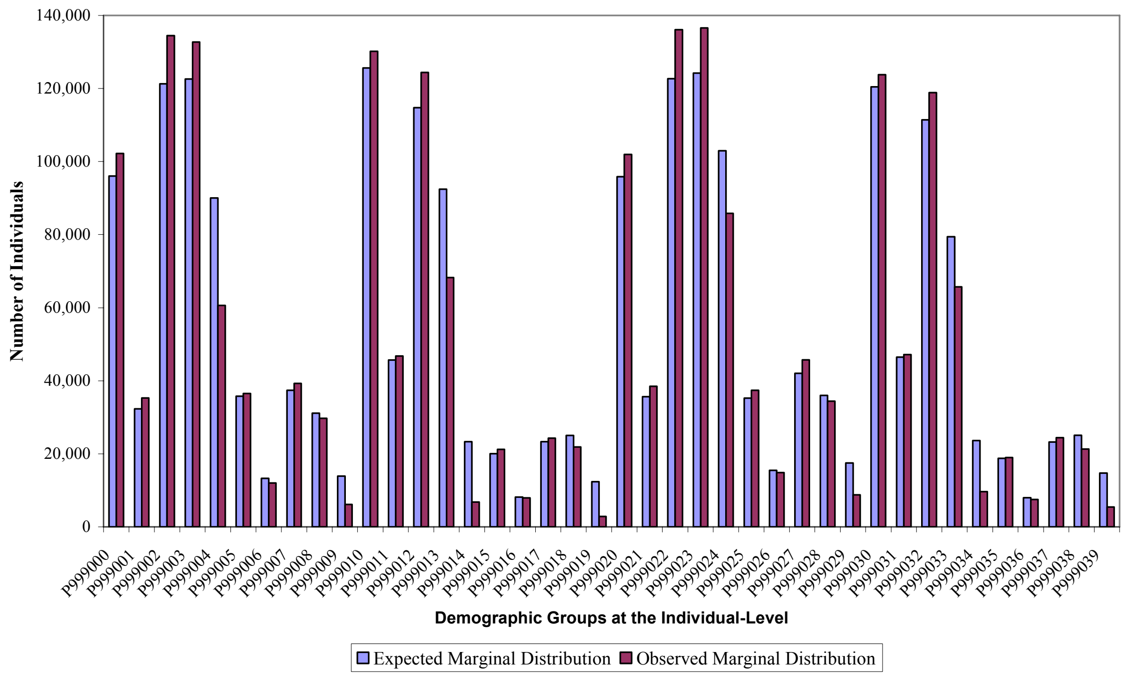

Figure 4-3b Comparisons between Expected and Observed Marginal Distributions for Individual-Level Control Variables for the Forecast Year

5. GENERATION AND VALIDATION OF ANALYSIS YEAR CHARACTERISTICS FOR SYNTHETIC POPULATION

The SPG described in the preceding chapter produces, as output, the values of the control variables for each synthetic household and individual. Although variables other than these control variables are available in the PUMS data, these variables are discarded from the SPG output because their corresponding distributions have not been controlled for. Instead, additional sociodemographic variables about the population that are required as input to CEMDAP are generated using a separate set of Comprehensive Econometric Microsimulator of SocioEconomic Land-use and Transportation System (CEMSELTS) modules. The generation and validation of these variables are the focus of this chapter. Section 5.1 contains a discussion of the structure and the prediction procedure underlying each modules. Section 5.2 describes the implementation of the modules. The validation of the outputs resulting from the application of these modules is provided in Section 5.3.

5.1 CEMSELTS Modules

As discussed in Chapter 4, SPG generates the synthetic population using control variables at the household and person levels. The household-level control variables are: (1) whether the household is a family or not, (2) household type, (3) presence of children, (4) age of household head, and (5) household size. The person-level control variables are: (1) race, (2) gender, and (3) age. Values synthesized for these control variables, together with the residential location of the households, are then taken as input to a suite of CEMSELTS modules to produce additional household- and person-level characteristics required by CEMDAP. The sequence in which these modules are applied is illustrated in Figure 5-1. The details of the individual modules are discussed below.[1]

[1] A detailed description of the entire CEMSELTS modeling system, data sources, and estimation results is available in Guo et al.,[2005]. The estimation results are included as an appendix to this report.

5.1.1 Modules for Generating Person-Level Attributes

5.1.1.1 Education attainment and study status

The data available for modeling education attainment (EA) and study status (SS) are very limited. Thus, we determine these two attributes for each individual based on the following assumptions: (1) all individuals start schooling at the age of 5, (2) all individuals complete primary education at the age of 12, (3) an individual who drops out never returns to school, and (4) an individual completes the final degree without leaving school.

The education attainment and study status are determined as follows. Individuals under 5 years of age have EA=”no schooling” and SS=”not studying.”. Individuals between the age of 5 and 12 have EA=”primary school” and SS=”studying.” For individuals 13 to 18 years of age, progress in secondary school is determined probabilistically. The drop-out rate of an individual at a given year is provided by a probability lookup table (Table C.1) and depends on the individual’s race and gender. If an individual drops out, his or her EA is set to the grade at which the dropout occurred and the SS is updated. For all individuals over 16 years, another lookup table (Table C.2) is employed to determine the highest degree that the individual will attain. The EA and SS are then determined based on the current age of the individual and the assumption that an associate’s degree, a bachelor’s degree, a master’s degree, and a doctoral degree will take 2, 4, 6, and 9 years to earn, respectively.

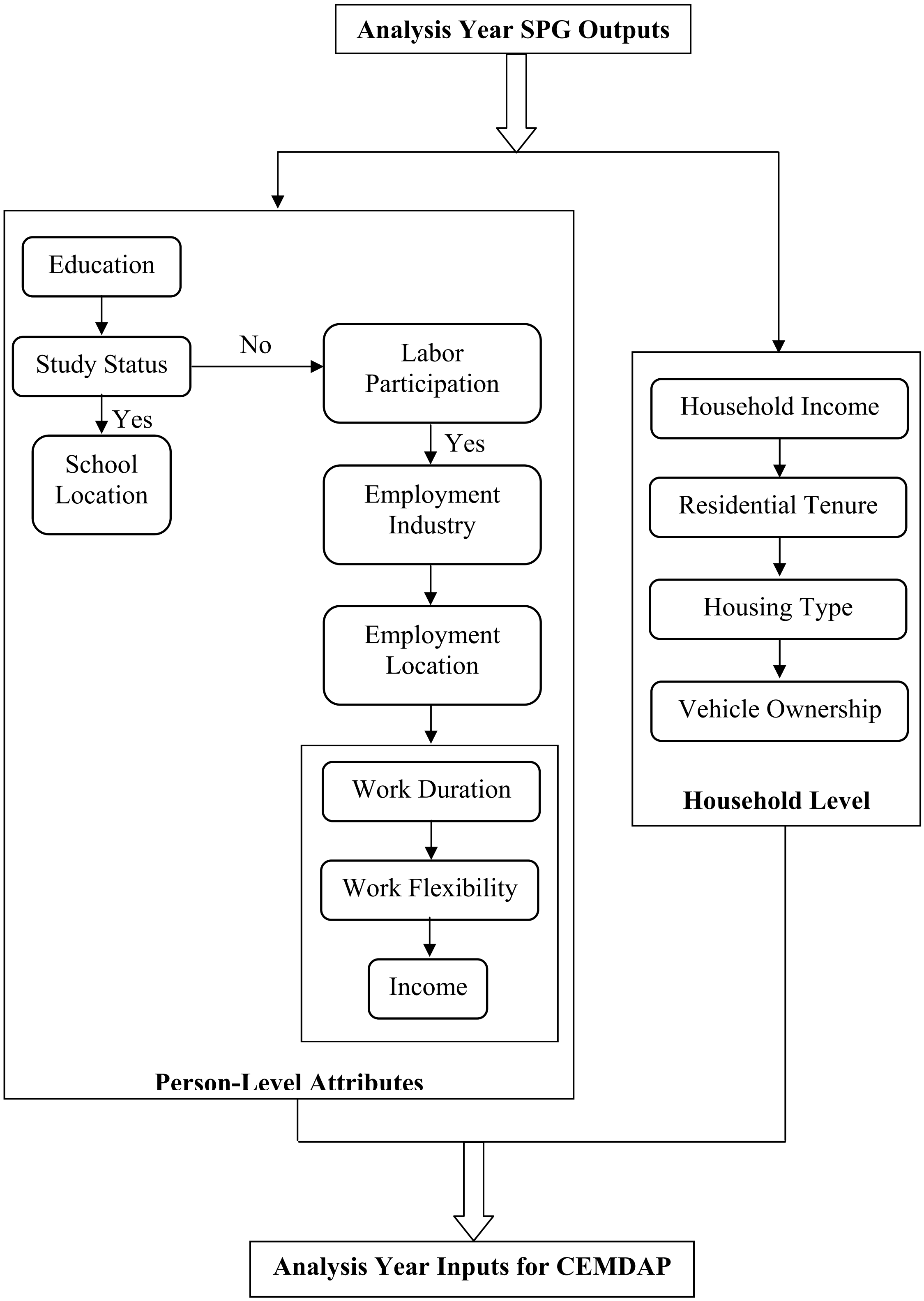

Figure 5-1 Flowchart Detailing the Prediction Framework Employed To Generate Analysis Year Attributes

5.1.1.2 Study location

The study location model is applied to individuals who are attending school or college. For children attending primary, middle, or high school, the closest zone with a school from the child's residence is assigned as the school location. The location of primary and secondary schools in the DFW area is obtained from the DFW school look-up table. An excerpt of the table used is provided in Appendix C (Table C.3). The study location of individuals attending college is determined based on race-specific look-up tables. An excerpt of the table is presented in Appendix C (Table C.4).

5.1.1.3 Labor participation

The labor participation model determines the decision to participate in the labor force for each individual over 12 years of age and currently not studying. The decision-making mechanism assumes a binary logit form and is estimated employing data from the PUMS. The estimates of the binary logit model are presented in Table C.5 of Appendix C. Based on the estimated probability of being employed, a deterministic determination is made in the microsimulation framework

5.1.1.4 Employment industry

For those individuals who enter the labor force, the employment industry model determines the industry in which the individual works. The industry variable was aggregated into six categories: construction and manufacturing, trade and transportation, professional businesses, government, retail and repair, and other. The utility and choice probabilities associated with each alternative industry are computed using the MNL model presented in Table C.6.

5.1.1.5 Employment location

An employment location model is applied to all the individuals entering the work force. The choice alternatives include the 4,874 zones of the DFW region. The choice probability is computed based on an MNL model estimated using the DFW household survey data (Table C.7).

5.1.1.6 Work duration

The work duration model determines the weekly hours of work for individuals who are part of the work force. A grouped response probit structure is employed to model the work duration of the individuals participating in the workforce. These are the model outcomes: less than 35 hours, 35–45 hours, and greater than 45 hours. The parameter estimates of the model based on the DFW household survey data are presented in Table C.8 of Appendix C.

5.1.1.7 Work flexibility

Individuals' work flexibility is characterized as low flexibility, medium flexibility, and high flexibility.[2] The probability associated with each level of flexibility is given by an ordered probit model. The parameter estimates of the ordered probit model are estimated using the DFW household survey data (see Table C.9).

5.1.1.8 Income

The income of each employed individual is modeled at an individual level and is subsequently aggregated up to the household level. The person income model takes the grouped response structure with income grouped into six categories: $0–$9,999; $10,000–$19,999; $20,000–$29,999; $30,000–$39,999; $40,000–$49,999; and $50,000 or more. The parameter estimates of the grouped response model are presented in Table C.10 of Appendix C.

5.1.2 Modules for Generating Household-Level Attributes

5.1.2.1 Residential tenure

The household residential tenure model determines the household's preference to either own or rent a house. A binary logit model is estimated using the 1996 DFW household survey data. The estimates of the model are presented in Table C.11 of Appendix C. Based on the model, the propensity and the probability to own or rent are calculated and a deterministic assignment is implemented.

5.1.2.2 Housing type

The choice of housing type involves a complete market segmentation modeling of owned and rented housing. Within the owned housing segment, the choices are: single-family detached, single-family attached, and mobile home or trailer. The rented housing segment includes: single-family detached, single-family attached, and apartment. Separate multinomial logit models have been estimated for the two market segments. The estimation results obtained using the 1996 DFW household survey data are presented in Table C.12a and C.12b of Appendix C.

---

[2] Work flexibility of employed individuals was categorized based on the individual's response to the question in the survey questionnaire.

5.1.2.3 Vehicle ownership

The number of vehicles owned by a household is modeled using the multinomial logit structure, where the five choice alternatives are defined as having 0, 1, 2, 3, or 4 or more cars. The model estimates, obtained using the DFW household survey, are presented in Table C.13 of Appendix C.

5.2 Module Implementation

The implementation of the aforementioned modules was accomplished employing the software Gauss6.0, a matrix programming platform that is capable of handling large data matrices (see Aptech, Inc. [2006]). The implementation entails writing Gauss code for estimating a limited number of econometric structures, including ordered probit, ordered logit, and multinomial logit models. Generic implementations of these modules enable the reusability of the code. In addition to these modules, additional code was written to obtain the continuous values of attributes. For instance, income category was determined employing a grouped response structure. In order to obtain a continuous income value, a uniform random number was generated and used to obtain a continuous value within the chosen interval. This involves an implicit assumption of uniform distribution of income within the interval. Other more sophisticated approaches, as suggested by Bhat (1994), may also be applied; these will be implemented in the future.

5.3 Validation Statistics

The CEMSELTS modules discussed in the previous sections were implemented for two analysis years: (1) the base year 2000 and (2) a forecast year (2025). The CEMSELTS modules are validated for the 2000 synthetic population by comparing the outputs with the DFW household survey, 2000 PUMS, and census 2000 as appropriate. The validation results for the individual modules models are summarized in Tables 5.1 through 5.10. In each of these tables, the first column lists the alternatives for each choice dimension. The second column shows the predicted share of each alternative outcome. The third column corresponds to the sample share observed in the DFW travel survey data. The fourth column represents the observed share found in either the census summary data or the PUMS data (except for Tables 5.6 and 5.7 where census or PUMS data are not available for the choice dimensions). Overall, the prediction capability of the CEMSELT modules is satisfactory. Among the person-level attributes, the predicted values of educational attainment, labor participation, and employment industry match well with the census and DFW distribution. For the work location dimension, the predicted work locations are aggregated and compared to the PUMS county-county flows. The results (Table 5.4) indicate that the predicted employment locations match reasonably well with the PUMS values. The distribution of work duration and personal income do show some substantial difference from the distribution found in the census and DFW sample. These differences may be attributed to the small sample sizes employed in the estimation of the corresponding prediction modules. The household attributes predicted match the corresponding survey and DFW samples very well (see Tables 5.8 through 5.11). The results corresponding to the 2025 forecast year attributes are presented in Appendix C (Tables C.14 through C.22).

Table 5.1 Education Attainment Module Comparison

Education	Predicted	DFW Sample	Census
No School	8.4	6.4	10.7
Children: Preschool–Grade 4	9.8	8.9	9.6
Children: Grades 5–8	6.4	4.4	5.9
Children: Grades 9–12	6.2	11.1	5.4
Adult: High school or less	47.2	35.2	47.3
Adult: Associate	4.0	20.1	3.6
Adult: Bachelor's	14.3		12.2
Adult: Master's	3.7	13.9[3]	5.3[4]
Adult: PhD	0.2		

Table 5.2 Labor Participation Module Comparison

Labor Participation	Predicted	DFW Sample	Census
Employed	48.1	48.9	49.4
Unemployed	51.9	51.1	50.6

Table 5.3 Employment Industry Module Comparison

Employment Industry	Predicted	DFW Sample	Census
Construction and Manufacturing	18.8	20.1	20.9
Wholesale Trade and Transportation	14.2	13.1	10.8
Professional, Personal, and Financial	33.8	39.6	33.0
Public and Military	5.9	5.2	3.1
Retail and Repair	24.0	22.0	22.8
Other Industry	3.3	0.0	9.4

[3] Value corresponds to the sum of the three categories (1) Adult: Bachelor's, (2) Adult: Master's, (3) Adult: PhD.
[4] Value corresponds to the sum of the two categories (1) Adult: Master's and (2) Adult: PhD.

Table 5.4 Employment Location Module Comparison

County-County Flows: PUMS

		% within County of Work								Total
		Collin	Dallas	Denton	Ellis	Johnson	Kaufman	Other	Tarrant	
County of Residence	Collin	**69.0**	9.7	3.4	0.2	0.5	3.6	14.1	0.7	**11.8**
	Dallas	20.0	**66.8**	6.6	9.0	2.1	17.5	28.3	6.4	**39.0**
	Denton	7.8	8.3	**84.1**	0.8	0.5	0.6	14.9	3.0	**11.7**
	Ellis	0.3	2.0	0.1	**81.0**	1.0	0.4	3.3	0.4	**2.5**
	Johnson	0.1	0.4	0.2	4.1	**84.2**	0.2	5.5	3.9	**2.6**
	Kaufman	1.0	2.8	0.3	1.0	0.1	**76.8**	5.4	0.2	**2.8**
	Parker	0.1	0.2	0.4	0.4	1.0	0.0	7.8	9.6	**2.9**
	Tarrant	1.7	9.8	4.8	3.5	10.7	0.9	20.7	**75.9**	**26.7**
	Other	--	--	--	--	--	--	--	--	--
Total		**8.9**	**51.0**	**6.7**	**1.5**	**1.4**	**1.5**	**1.9**	**27.1**	**100.0**

County-County Flows: CEMSELTS Work Location Module

		% within County of Work									Total
		Collin	Dallas	Denton	Ellis	Johnson	Kaufman	Parker	Rockwall	Tarrant	
County of Residence	Collin	**67.5**	22.6	8.0	0.2	0.0	0.1	0.0	0.6	1.0	**10.5**
	Dallas	9.3	**69.2**	7.3	2.2	0.3	0.7	0.0	1.1	10.0	**45.3**
	Denton	4.6	11.2	**77.1**	0.1	0.1	0.0	0.0	0.1	6.8	**9.2**
	Ellis	0.7	21.8	0.5	**67.1**	3.2	1.0	0.0	0.1	5.7	**2.0**
	Johnson	0.0	1.4	0.2	1.4	**78.9**	0.0	0.2	0.0	17.9	**2.2**
	Kaufman	2.6	51.5	0.8	4.4	0.1	**31.6**	0.0	6.0	3.0	**0.5**
	Parker	0.0	0.9	0.5	0.0	2.5	0.0	**38.9**	0.0	57.3	**0.3**
	Rockwall	15.0	40.8	0.7	0.5	0.1	1.4	0.0	**40.7**	0.8	**0.8**
	Tarrant	0.4	10.3	2.9	0.6	4.1	0.0	0.7	0.0	**80.9**	**29.2**
Total		**10.5**	**45.3**	**9.2**	**2.0**	**2.2**	**0.4**	**0.3**	**0.8**	**29.2**	**100.0**

Table 5.5 Work Duration Module Comparison

Work Duration	Predicted	Sample	Census
Hours 0–20 (Hours 0–14 for Census)	21.0	11.6	3.1
Hours 20–40 (Hours 15–34 for Census)	43.4	53.5	12.9
Hours 40+ (Hours 35+ for Census)	35.5	34.9	84.1

Table 5.6 Work Flexibility Module Comparison

Work Flexibility	Predicted	Sample	Census
Low/No Flexibility	18.4	20.2	N/A
Med Flexibility	14.7	15.5	N/A
High Flexibility	15.0	15.3	N/A
Unemployed	51.8	51.1	N/A

Table 5.7 Personal Income Module Comparison

Personal Income ($)	Predicted	Sample	Census
No Income	22.4	18.3	N/A
0–10,000	18.3	23.3	N/A
10,000–20,000	20.8	17.0	N/A
20,000–30,000	17.6	14.6	N/A
30,000–40,000	11.3	14.4	N/A
40,000–50,000	4.8	10.5	N/A
50,000 +	4.8	1.9	N/A

Table 5.8 Residential Tenure Module Comparison

Residential Tenure	Predicted	Sample	Census
Own	66.7	66.6	60.0
Rent	33.3	33.4	40.0

Table 5.9 Housing Type for Owners Module Comparison

Housing Type for Owners	Predicted	Sample	Census
Single Family Detached	93.1	94.2	89.1
Single Family Attached	3.6	3.5	2.5
Mobile Home/Trailer	3.4	2.3	6.6
Multi-Family/Apartment/Condo	0.0	0.0	1.8

Table 5.10 Housing Type for Renters Module Comparison

Housing Type for Renters	Predicted	Sample	Census
Single Family Detached	26.8	26.5	20.6
Single Family Attached	8.4	9.3	3.9
Multi-Family/Apartment/Condo	64.8	64.3	73.0
Mobile Home/Trailer	0.0	0.0	2.5

Table 5.11 Household Vehicle Ownership for Renters Module Comparison

Vehicle Ownership	Predicted	Sample	Census
No. of Vehicles = 0	6.8	6.6	6.1
No. of Vehicles = 1	40.3	36.7	35.6
No. of Vehicles = 2	37.1	42.5	42.5
No. of Vehicles = 3	12.5	11.2	12.1
No. of Vehicles = 4 or more	3.4	2.9	3.8

6. VALIDATION, SAMPLING, AND SENSITIVITY ANALYSIS

The preceding chapter presented the framework employed for the generation of inputs required by CEMDAP. The objectives of the current chapter are fivefold: (1) describe the validation of CEMDAP outputs against the estimation sample (Section 6.1), (2) discuss the results of sampling tests to reduce overall computing run times (Section 6.2), (3) present aggregate comparisons of CEMDAP outputs with those obtained from NCTCOG's four-step model (Section 6.3), (4) illustrate CEMDAP's applicability as a policy-evaluation tool based on prediction under several different scenarios (Section 6.4) and (5) present the activity-travel forecasts from CEMDAP for a future year (Section 6.5).

6.1 Validation

CEMDAP employs a suite of econometric models to predict the activity-travel patterns of individuals. These models were estimated using the 1996 DFW Household Travel Survey. This section of the chapter describes the results of a validation exercise undertaken to assess the ability of CEMDAP to produce predicted activity-travel patterns that are consistent, reasonable, and close to the observed patterns in the survey.

The following procedure was adopted for validation. First, CEMDAP was used to simulate the activity-travel patterns of the 1910 households from the DFW household travel survey used in the estimation of the models. Next, the predicted patterns were compared with the observed patterns and systematic differences and inconsistencies were noted. Finally, the modeling system and the software were suitably updated (by including additional consistency checks, debugging of code) and used to produce the final predicted patterns.

The rest of this section presents statistics comparing the software outputs with the observed activity-travel patterns. The CEMDAP-predicted activity-travel patterns and the observed DFW survey patterns were compared along several activity-travel attributes. However, for the sake of brevity, only selected measures are reported here. These include comparisons of the following: (1) pattern-level attributes (Section 6.1.1), (2) tour-level attributes (Section 6.1.2), (3) chaining propensity (Section 6.1.3), (4) characteristics of trips and travel by trip type (Section 6.1.4), (5) activity-episode characteristics (Section 6.1.5]) and (6) work start and end time distributions (Section 6.1.6).

6.1.1 Pattern-Level Attributes

The pattern-level measures presented in this section are the average number of worker and non-worker tours. The averages correspond to mean values across all individuals in the sample. The results indicate that the CEMDAP outputs are quite close to the DFW survey patterns for worker and non-worker tours (see Table 6.1). The number of non-school tours for children has the highest magnitude of difference. This variation may be attributed to the fact that the sample from which the number of non-school trips for children was estimated is much smaller than the sample employed to model worker and non-worker attributes.

Table 6.1 DFW Survey vs. CEMDAP - Number of Tours[5]

	DFW Survey	CEMDAP
Avg. no. of before-work tours (workers)	0.04	0.02
Avg. no. of work-based tours (workers)	0.30	0.34
Avg. no. of after-work tours (workers)	0.32	0.39
Avg. no. of tours (non-workers)	1.14	1.19
Avg. no. of non-school tours (children)	0.28	0.18

6.1.2 Tour-level attributes

The tour-level measure presented in this section is the number of stops in a tour. Table 6.2 provides the average values of number of stops within each type of tour (the average values for each tour type correspond to the means across all tours of that type). The CEMDAP outputs are quite similar to the DFW survey results. The greatest difference in the average number of stops is within the commute tours. CEMDAP is overpredicting the number of stops in commute tours.

[5] Averaged over all workers, non-workers, and children.

Table 6.2 DFW Survey vs. CEMDAP - Number of Stops[6]

	DFW Survey	CEMDAP
Avg. no. of stops in before-work tour	1.33	1.36
Avg. no. of stops in work-based tour	1.31	1.27
Avg. no. of stops in after-work tour	1.43	1.41
Avg. no. of stops in home-work commute	0.22	0.15
Avg. no. of stops in work-home commute	0.45	0.39
Avg. no. of stops in non-worker tour	1.71	1.78

6.1.3 Chaining Propensity

Chaining propensity is a measure of the inclination to undertake more than one activity episode (or stop) in a tour. The non-commute chaining propensity for workers is defined as the ratio of the sum of the number of before-work, work-based, and after-work tours to the total number of out-of-home activity episodes undertaken in the before-work, work-based, and after-work tours, respectively. The chaining propensity for non-workers is the ratio of the total number of tours to the total number of out-of-home activity episodes. If each tour comprises only one stop, then the chaining propensity is one. As more stops are included in each tour, the propensity falls below one. Hence, the smaller the value of the chaining propensity measure, the greater the extent of trip chaining.

CEMDAP slightly overpredicts the total number of worker tours and stops (see Table 6.3). However, since it overpredicts both stops and tours, the average chaining propensity is close to the DFW value. For non-workers, CEMDAP outputs match quite well with the DFW survey results for the total number of tours, stops, and chaining propensity.

[6] Averaged within each tour type or commute.

Table 6.3 DFW Survey vs. CEMDAP - Chaining Propensity[7]

	DFW Survey	CEMDAP
Workers		
Avg. no. of non-commute tours	1.27	1.46
Avg. no. of stops in non-commute tours	3.07	3.43
Avg. non-commute chaining propensity	0.43	0.45
Non-workers		
Avg. no. of tours	1.60	1.60
Avg. no. of stops	3.49	3.48
Avg. chaining propensity	0.60	0.61

6.1.4 Characteristics of Trips and Travel by Trip Type

This section compares the characteristics of trips (frequency, travel time, and travel distance) predicted by CEMDAP with the corresponding values observed in the survey. The statistics are provided separately for each of the three commonly used trip types: home-based work, home-based other, and non–home-based.

CEMDAP performs well in predicting the total number of daily trips per person for all trip types (see Table 6.4). CEMDAP, on average, slightly underpredicts the travel times for all trip types. CEMDAP overpredicts the average trip mileage (both per person and per vehicle) for home-based work and home-based other trips. This variation could be attributed, in part, to the fact that CEMDAP employs in-vehicle interzonal travel times, while the data from the survey represents door-to-door travel time.

[7] Averaged over all workers/non-workers.

Table 6.4 DFW Survey vs. CEMDAP – Trip Type[8]

	Avg. no. of daily trips per person	Avg. person min. of travel per trip	Avg. person miles of travel (PMT) per trip	Avg. veh. miles of travel (VMT) per trip
Home-based work				
DFW Survey	1.79	27.67	11.68	12.17
CEMDAP	1.70	26.92	11.96	12.67
Home-based other				
DFW Survey	2.59	18.06	9.38	9.27
CEMDAP	2.65	17.49	10.72	11.05
Non–home-based				
DFW Survey	2.43	17.78	9.78	9.94
CEMDAP	2.57	15.15	8.29	8.86

6.1.5 Activity-Episode Characteristics

This section compares the characteristics of the activity-episodes (frequency, duration, travel time) predicted by CEMDAP with the corresponding values observed in the survey. The statistics are provided separately for each of the eleven activity purposes.

Overall, CEMDAP predicts the average number of activities per person for each activity episode quite well (Table 6.5). However, CEMDAP slightly underpredicts the duration of work-related, household/personal business, and social/recreational activities. It also overpredicts the duration of shopping, other serve-passenger, and joint discretionary activities. The CEMDAP output is similar to the DFW survey results for average travel duration to activity episodes for all activity purposes, except joint discretionary activities. Finally, it is also important to note that, in CEMDAP, the number of drop-off at school and pick-up from school episodes are each fixed to one and the durations of these episodes are fixed to 5 minutes.

[8] Averaged over all individuals who made at least one out-of-home stop.

Table 6.5 DFW Survey vs. CEMDAP - Activity Episodes

	Avg. no. of activity episodes[9]	Avg. min. of activity per pers. per day[10]	Avg. dur. of activity episodes[11]	Avg. travel time to activities (minutes)[10]
Work				
DFW Survey	1.32	490.67	372.40	23.91
CEMDAP	1.34	502.87	375.56	18.30
Work-related				
DFW Survey	1.48	195.52	132.51	21.76
CEMDAP	1.71	103.45	60.63	19.20
Shopping				
DFW Survey	1.22	37.28	30.67	15.49
CEMDAP	1.62	66.28	40.83	14.22
Household/Personal				
DFW Survey	1.55	80.10	51.68	16.04
CEMDAP	1.71	76.38	44.76	15.70
Social/Recreational				
DFW Survey	1.27	131.53	103.43	18.09
CEMDAP	1.69	138.29	81.71	16.74
Eat Out				
DFW Survey	1.13	50.91	45.19	14.50
CEMDAP	1.50	77.45	51.49	14.32
Other Serve-Passenger				
DFW Survey	1.64	10.72	6.55	17.09
CEMDAP	1.88	24.23	12.92	19.23
Drop-off at School				
DFW Survey	1.07	3.22	2.99	9.73
CEMDAP	1.00	5.00	5.00	9.31
Pick-up at School				
DFW Survey	1.07	8.41	7.85	14.19
CEMDAP	1.00	5.00	5.00	12.00
Joint Discret. Activities				
DFW Survey	1.25	79.63	63.84	15.43
CEMDAP	1.00	86.87	86.98	6.29
Indep. Discret. Activities				
DFW Survey	1.00	183.02	183.02	11.93
CEMDAP	1.00	190.01	190.01	13.15

[9] Averaged over individuals who participated in the respective activity at least once.
[10] Averaged over each individual.
[11] Averaged over each activity episode.

6.1.6 Work Start and End Times

Figure 6-1 and Figure 6-2 present the distribution of DFW survey and CEMDAP predicted work start and end times, respectively. The DFW survey has longer tails for both work start and work end times. However, these tails represent a very low percentage of workers. CEMDAP predicts work start times past 7:30 a.m. well but has an increased discrepancy for earlier start times.[12] Apart from the DFW survey peak at 6:30 p.m. that CEMDAP fails to predict, CEMDAP predicts work end time reasonably well. The research team will evaluate and enhance the performance of the work start and end time modules in the future.

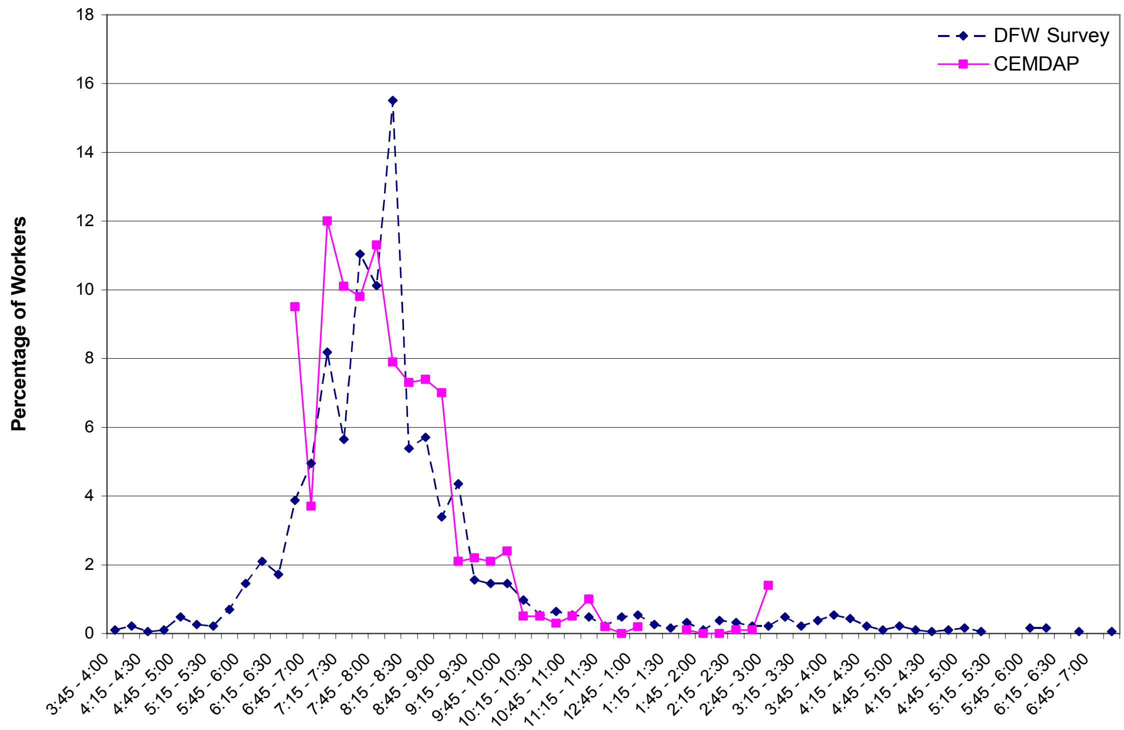

Figure 6.1 DFW Survey vs. CEMDAP - Work Start Time

---

[12] The reader should note here that the spike between 6.45 a.m. and 7.00 a.m. is due to the restriction imposed in CEMDAP. In CEMDAP the earliest work start time is 6.45 a.m. This time is the 95-percentile value of the work start times in the DFW survey.

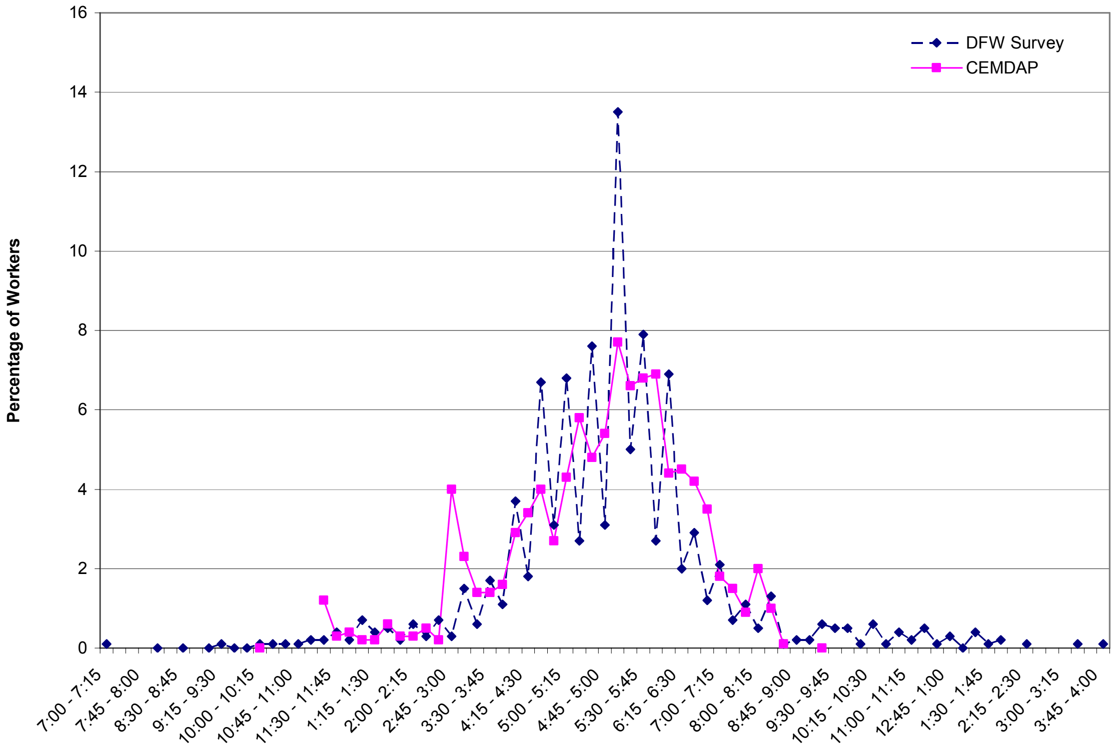

Figure 6-2 DFW Survey vs. CEMDAP - Work End Time

6.2 Sampling

The overall run time of CEMDAP is determined by two factors. The first factor is the size of the population for which the travel patterns are generated. Since CEMDAP simulates the activity-travel patterns of each household in the population through a suite of econometric models, the run time will increase if the population increases. One could reduce computational time by distributing the processing over several computers, but the gain is linear (i.e., two computers will take half as much time as one computer does) and consequently several machines may be required to achieve the desired overall processing time. The second factor that influences run times is the size of the level-of-service files. There are 4,874 TAZs in the DFW area and the corresponding LOS files are large (4,874 X 4,874 rows for each LOS attribute). The large number of rows rules out the possibility of loading the entire set of files into Random Access Memory (RAM). Alternate mechanisms, such as mult-threading and caching (discussed in Chapters 2 and 3), have successfully been employed to reduce runtimes. In spite of these improvements, the computational runtime for CEMDAP on the entire DFW region (for 1.8 million households and 4.7 million individuals) is approximately 25 days on a single machine.

In this context, given our computational resources, it is not practical to generate the activity-travel patterns for all 1.8 million households for each set of validation and sensitivity analysis. An attractive alternative is to run CEMDAP only for a random sample of the population. Prior to the adoption of this strategy, we evaluated different sampling schemes to determine which one, if any, would adequately substitute for running the entire population. In the rest of this section, we compare the results of running CEMDAP on 100%, 50%, 25%, 10%, and 5% of the DFW population.

Our results indicate that the pattern, tour, and stop level attribute predictions are very similar for all sample sizes (see Tables 6.6, 6.7, and 6.8 for a comparison of the 5% and 100% sample results). In order to compare the different sample sizes, we factored each aggregate statistic by the appropriate amount to reach the 100% sample totals (e.g., the 5% sample aggregate results were multiplied by 20). The most significant differences between the 100% and the 5% sample are in person miles of travel (Table 6.8). This is possibly due to differences in the spatial coverage of the 5% sample relative to the 100% sample (i.e., there may be no or inadequate travel predicted between certain zone combinations if only a 5% sample is used). To further investigate this issue, we examined the stop location choice predictions across the different sampling levels at two aggregated spatial levels.[13] The two levels employed are: (1) 500 x 500 (current zones aggregated into 500 units) and (2) 57 x 57 (DFW regional jurisdictions).

Table 6.6 100% vs. 5% Sample - Number of Tours[14]

	100 % Sample	**5% Sample x 20**
Avg. no. of before-work tours (workers)	0.02	0.02
Avg. no. of work-based tours (workers)	0.33	0.33
Avg. no. of after-work tours (workers)	0.40	0.40
Avg. no. of tours (non-workers)	0.23	0.23
Avg. no. of non-school tours (children)	1.45	1.43

[13] It is not appropriate to conduct the tests at the 4,874 zone level because the number of possible locations for the 4,874 x 4,874 system is on the order of 25 million trips. At the same time the possible trips for the entire DFW region is on the order of 15 million trips. Therefore, matching the location choices at any sampling rate is not practical.
[14] Averaged over all workers/non-workers.

Table 6.7 100% vs. 5% Sample - Number of Stops[15]

	100 % Sample	5% Sample x 20
Avg. no. of stops in before-work tour	1.38	1.26
Avg. no. of stops in work-based tour	1.46	1.26
Avg. no. of stops in after-work tour	1.55	1.34
Avg. no. of stops in home-work commute	0.20	0.20
Avg. no. of stops in work-home commute	0.41	0.41
Avg. no. of stops in non-worker tour	1.78	1.79

Table 6.8 100% vs. 5% Sample - Aggregate Number of Trips, PHT, and VMT by Trip Type

	100% Sample	5% Sample x 20
Total Number of Trips (millions)		
Home-based work	2.74	2.74
Home-based non-work	9.44	9.36
Non–home-based	4.94	4.86
Overall	17.12	16.96
Total Person Hours Traveled (millions)		
Home-based work	54.73	54.75
Home-based non-work	148.64	146.13
Non–home-based	74.70	73.31
Overall	278.07	274.19
Total Person Miles Traveled (millions)		
Home-based work	35.32	35.12
Home-based non-work	88.29	82.64
Non–home-based	43.83	40.50
Overall	167.45	158.27

[15] Averaged within each tour.

The Mean Percentage Error (MPE) was employed in the analysis and is defined as:

$$MPE = \frac{1}{N_c^2}\left[\sum_{1}^{N_c}\sum_{1}^{N_c}\left|\left(P_{ij} - P'_{ij}\right)\right|\right]$$

where, N_c represents the number of spatial units at the level of aggregation considered, P_{ij} represents the corresponding ij^{th} cell value of the 100% sample, and P'_{ij} represents the corresponding ij^{th} cell value of the sample under consideration.

The results of the spatial level analysis of the errors due to sampling are presented in Table 6.9. These results indicate that, if the emphasis is on stop location, it is not adequate to employ a 10% sample.

To summarize, the sample size analysis conducted reveals two issues: (1) 5% samples are adequate to represent the pattern, tour, and stop level attributes and (2) a rather large sampling rate is needed if the emphasis is on the spatial location of stops.

Table 6.9 Sampling Analysis of Location Choices

Level of Aggregation	Sampling Rate (%)	MPE
500 x 500	50	39.33
	25	66.21
	10	98.69
57 x 57	50	4.26
	25	7.31
	10	12.28

6.3 CEMDAP Comparison with the Four-Step Model

This section presents a comparative assessment of CEMDAP with the four-step model currently in use for the DFW region.[16].This analysis was performed in the following way. First,

[16] The authors would like to thank the NCTCOG staff for undertaking much of this analysis and providing us with the results.

the four-step model was applied to determine the link volumes and the predicted volumes were compared with observed counts. Second, the first three steps of the trip-based model (i.e.„ trip generation, trip distribution, and modal split) were replaced with outputs obtained from CEMDAP. (Specifically, the CEMDAP activity travel patterns were appropriately repackaged to develop trip origin-destination tables by mode for each of the three time periods: a.m. peak, off peak, and p.m. peak.). Estimates of external trips and truck trips were borrowed from the four-step model and suitably added to the O-D matrices from CEMDAP. The network assignment step was undertaken based on the DFW static assignment procedure. Then, the CEMDAP predicted link volumes were compared with the observed link counts. Finally, the errors between the predicted flows and observed counts were compared for the two cases (DFW model predictions and CEMDAP predictions) using the Root Mean Squared Error (RMSE) measure, defined as follows:

$$RMSE_f = \frac{1}{N_f} \sum_{\forall links} \sqrt{(Actual\ Link\ Count - Predicted\ Link\ Count)^2}$$

$$\%RMSE_f = \frac{1}{N_f} \sum_{\forall links} \sqrt{(Actual\ Link\ Count - Predicted\ Link\ Count)^2}\ x\ 100$$

where N_f represents the number of links of functional class *f*.

The results, presented in Table 6.10, indicate that the CEMDAP model performs close to the DFW model without K factors. The DFW model with K factors performs slightly better than CEMDAP in terms of replicating current link counts. However, it should be noted that CEMDAP results are based on models that do not include any calibration adjustment factors of any kind. Besides, it is important not to use closeness to current link counts as the sole basis for assessing the performance of travel models. Rather, the focus should also be on the level of behavioral fidelity captured in the model. The better the behavioral fidelity of a model, the better it will be in terms of transferability in time (especially if the demographics and travel environment change substantially over time). After all, the value of a travel model is in its ability to forecast well into the future, not replicate current conditions. Finally, it should be noted here that the use of traditional static assignment process does, to an extent, "undo" the benefits of a continuous time activity system. This happens because the patterns are grouped back to three aggregate time periods in the assignment stage and the static assignment process does not consider the dynamics of vehicle delays. However, the results in this section do provide validation that CEMDAP is producing reasonable results.

Table 6.10 Weekday Volume vs. Weekday Counts (% RMSE)

Roadway functional class	DFW model with K factors	DFW model without K factors	CEMDAP 5%	CEMDAP 10%	CEMDAP 100%
Freeways	15.36	21.48	26.00	25.88	25.84
Major Arterials	31.19	36.69	42.36	42.18	42.07
Minor Arterials	40.58	43.02	44.75	44.62	44.61
Collectors	68.43	70.11	70.29	70.19	70.10
Congested Roads	50.83	54.32	68.31	67.37	66.88
Uncongested Roads	71.89	75.76	78.77	79.18	79.88
Overall	36.9	42.6	47.44	47.28	47.23

6.4 Scenarios and Sensitivity Analysis

This section discusses in detail the application of CEMDAP as a tool for policy analysis. Specifically, the activity-travel patterns were simulated, using a 5% sample, for a total of eleven scenarios (in addition to the year 2000 "base case" scenario). These scenarios involve changes to the transportation system and population characteristics. Section 6.4.1 describes each scenario and the corresponding modifications made to CEMDAP. Sections 6.4.2 through 6.4.7 discuss the impact of the transportation system–related scenarios (i.e., changes to IVTT and cost) on the activity-travel patterns. Section 6.4.8 discusses the impact of population changes on the activity-travel patterns. Finally, Section 6.4.9 compares CEMDAP to the DFW model results for one scenario.

6.4.1 Scenario Description and Generation

The sensitivity was tested by constructing eleven scenarios. Table 6.11 provides a description and highlights how the 2000 base year input data were altered for each scenario.

Table 6.11 Scenario Description

Scenario	Description	Changes to Base Year
25% Decrease in IVTT	A 25% decrease in IVTT for drive-alone, shared ride, and transit, across all time periods	LOS tables were altered by multiplying the auto IVTT and transit IVTT by .75 in the a.m., p.m., and off-peak files.
25% Increase in IVTT	A 25% increase in IVTT for drive-alone, shared ride, and transit across all time periods; 2000 base year	LOS tables were altered by multiplying the auto IVTT and transit IVTT by 1.25 in the a.m. and p.m. peak files.
25% Increase in IVTT— Auto Mode	A 25% increase in IVTT for drive-alone and shared ride IVTT for all time periods	LOS tables were altered by multiplying the auto IVTT by 1.25 in the a.m., p.m., and off-peak files.
25% Increase in IVTT— Peak Periods	A 25% increase in IVTT for drive-alone, shared ride, and transit IVTT for the a.m. and p.m. peak time periods	LOS tables were altered by multiplying the auto IVTT and transit IVTT by 1.25 in the a.m. and p.m. peak files.
25% Increase in IVTT— Auto Mode and Peak Periods	A 25% increase in IVTT for the drive-alone and shared ride for the a.m. and p.m. peak time periods	LOS tables were altered by multiplying the auto IVTT by 1.25 in the a.m. and p.m. peak files.
25% Increase in Cost	A 25% increase in cost for drive-alone, shared ride, and transit	LOS tables were altered by multiplying the auto cost and transit cost by 1.25 in the a.m., p.m., and off-peak files.
25% Increase in Cost—Auto Mode	A 25% increase in cost for drive-alone and shared ride for all time periods	LOS tables were altered by multiplying the auto cost by 1.25 in the a.m., p.m., and off-peak files.
25% Increase in Cost—Peak Periods	A 25% increase in cost for drive-alone, shared ride, and transit cost for the a.m. and p.m. peak time periods	LOS tables were altered by multiplying the auto IVTT and transit cost by 1.25 in the a.m. and p.m. peak files.
25% Increase in Cost—Auto Mode and Peak Periods	A 25% increase in cost for drive-alone and shared ride for the a.m. and p.m. peak time periods.	LOS tables were altered by multiplying the auto cost by 1.25 in the a.m. and p.m. peak files.
$2 Increase in CBD Cost—Auto Mode and Peak Periods	A $2 charge is imposed on the auto trips that enter/exit the CBD during a.m. and p.m. peak periods.	LOS tables were altered by adding an additional $2 to the existing cost for trips that originate or end in the CBD in the a.m. and p.m. peak files.
25% Increase in Regional Population	25% increase in the number of people residing in the DFW population.	The households were increased by selecting 25% of the current household population, and adding those records to the existing households.

6.4.2 Pattern-Level Statistics

Several pattern-level attributes for the scenario case are compared with the corresponding attributes for the base case. These are: (1) number of worker tours and stops, (2) trip chaining propensity, (3) average daily duration of activities, and (4) work start and end times. The results are discussed below.

The average number of worker tours and stops does not differ much among scenarios (as shown in Table 6.12). Also, the average chaining propensity does not differ greatly by scenario (Table 6.13). Finally, there is also very little difference among the average (daily) durations of activities for each of the scenarios (Table 6.14). It is important to note here that none of these pattern-level measures are directly impacted by transportation level of service measures in the empirical specifications estimated for the DFW region. For all these scenarios, there is also little change in the work start and end times. That is, work start and end times of the scenarios are quite similar to the work start and end times of the 2000 base year scenario.

Table 6.12 CEMDAP Scenarios—Number of Worker Tours and Stops

	Avg. No. of BW Tours[17]	Avg. No. of Stops in BW Tour[18]	Avg. No. of WB Tours[17]	Avg. No. of Stops in WB Tour[18]	Avg. No. of AW Tours[17]	Avg. No. of Stops in AW Tour[18]	Avg. No. of Stops in HW Commute[17]	Avg. No. of Stops in WH Commute[18]
2000 Base Scenario	0.02	1.26	0.33	1.26	0.40	1.34	0.20	0.41
25% Decrease in IVTT	0.02	1.26	0.32	1.27	0.41	1.35	0.20	0.41
25% Increase in IVTT	0.02	1.24	0.33	1.27	0.39	1.34	0.20	0.40
25% Increase in IVTT—Auto Mode	0.02	1.26	0.33	1.26	0.39	1.34	0.20	0.40
25% Increase in IVTT—Peak Periods	0.02	1.26	0.33	1.26	0.39	1.33	0.21	0.40
25% Increase in IVTT—Auto Mode and Peak Periods	0.02	1.26	0.33	1.26	0.40	1.33	0.21	0.40
25% Increase in Cost	0.02	1.26	0.32	1.27	0.40	1.33	0.21	0.41
25% Increase in Cost—Auto Mode	0.02	1.26	0.32	1.27	0.40	1.33	0.20	0.40
25% Increase in Cost—Peak Periods	0.02	1.25	0.32	1.27	0.40	1.34	0.20	0.41
25% Increase in Cost—Auto Mode and Peak Periods	0.02	1.26	0.32	1.27	0.40	1.34	0.20	0.40
$2 Increase in CBD Cost—Auto Mode and Peak Periods	0.02	1.24	0.32	1.27	0.40	1.34	0.20	0.40

[17] Averaged over all workers
[18] Averaged within each tour

Table 6.13 Trip Chaining Characteristics[19]

	Non-Workers			Workers		
	Avg. No. of Tours	Avg. No. of Stops	Avg. Chaining Prop.	Avg. No. of Non-Commute Tours	Avg. No. of Stops in Non-Commute Tours	Avg. Chaining Propensity
Base Case	1.73	3.89	0.60	1.43	3.30	0.46
25% Decrease in IVTT	1.74	3.90	0.60	1.44	3.33	0.46
25% Increase in IVTT	1.73	3.88	0.60	1.43	3.29	0.46
25% Increase in IVTT—Auto Mode	1.74	3.90	0.60	1.44	3.33	0.46
25% Increase in IVTT—Peak Periods	1.73	3.88	0.60	1.43	3.30	0.46
25% Increase in IVTT—Auto Mode and Peak Periods	1.73	3.89	0.60	1.44	3.30	0.46
25% Increase in Cost	1.73	3.88	0.60	1.43	3.30	0.46
25% Increase in Cost—Auto Mode	1.73	3.88	0.60	1.43	3.30	0.46
25% Increase in Cost—Peak Periods	1.74	3.91	0.59	1.43	3.29	0.46
25% Increase in Cost—Auto Mode and Peak Periods	1.74	3.90	0.60	1.43	3.31	0.46
$2 Increase in CBD Cost—Auto Mode and Peak Periods	1.74	3.90	0.59	1.43	3.31	0.46

[19] Average over all workers and non-workers.

Table 6.14 Average Activity Duration[20]

	Home	Work	School	Work-Related
Base Case	620.78	506.11	381.13	167.59
25% Decrease in IVTT	623.34	507.09	381.21	168.51
25% Increase in IVTT	618.00	506.14	380.74	168.06
25% Increase in IVTT—Auto Mode	617.56	506.02	380.08	170.55
25% Increase in IVTT—Peak Periods	619.99	504.86	380.22	169.51
25% Increase in IVTT—Auto Mode and Peak Periods	618.83	504.49	380.76	171.62
25% Increase in Cost	620.15	507.06	380.63	169.50
25% Increase in Cost—Auto Mode	620.51	505.59	380.18	168.61
25% Increase in Cost—Peak Periods	619.94	506.48	381.04	168.37
25% Increase in Cost—Auto Mode and Peak Periods	620.29	506.67	380.93	169.37
$2 Increase in CBD Cost—Auto Mode and Peak Periods	620.18	506.47	380.78	170.15

[20] Averaged over each activity episode

Table 6.14 (cont.) Average Activity Duration

	Shopping	Household/Personal Business	Social/Recreational	Eat Out
Base Case	77.35	88.03	148.47	86.79
25% Decrease in IVTT	77.05	86.33	149.13	87.24
25% Increase in IVTT	77.52	87.60	148.26	85.21
25% Increase in IVTT—Auto Mode	76.51	87.12	149.48	86.33
25% Increase in IVTT—Peak Periods	77.21	87.05	147.64	86.10
25% Increase in IVTT —Auto Mode and Peak Periods	77.41	87.39	148.94	86.03
25% Increase in Cost	76.94	86.57	147.62	86.50
25% Increase in Cost—Auto Mode	77.74	86.66	148.87	86.45
25% Increase in Cost—Peak Periods	77.58	86.70	150.61	86.76
25% Increase in Cost—Auto Mode and Peak Periods	77.17	87.03	147.89	85.88
$2 Increase in CBD Cost—Auto Mode and Peak Periods	77.89	87.41	149.01	88.10

Table 6.14 (cont.) Average Activity Duration

	Other Serve-Passenger	Drop-off/Pick-up at School	Joint Discretionary Activities	Independent Discretionary Activities
Base Case	24.93	5.00	95.84	181.60
25% Decrease in IVTT	24.89	5.00	97.22	174.33
25% Increase in IVTT	24.51	5.00	96.42	180.18
25% Increase in IVTT—Auto Mode	24.86	5.00	97.04	181.33
25% Increase in IVTT —Peak Periods	24.79	5.00	95.11	178.96
25% Increase in IVTT—Auto Mode and Peak Periods	25.36	5.00	94.76	180.99
25% Increase in Cost	25.45	5.00	96.52	187.06
25% Increase in Cost—Auto Mode	24.41	5.00	96.14	181.93
25% Increase in Cost—Peak Periods	24.56	5.00	96.20	178.85
25% Increase in Cost—Auto Mode and Peak Periods	24.51	5.00	97.62	178.29
$2 Increase in CBD Cost—Auto Mode and Peak Periods	24.24	5.00	97.61	180.88

6.4.3 Aggregate Mode Shares

The aggregate commute mode shares are presented in Table 6.15. The drive-alone trips account for the majority of trips in the DFW area. Vehicular trips (drive-alone and shared-ride together) account for over 90% of trips. In scenarios with an increase in IVTT across auto modes, there is a small shift from drive-alone to shared ride for the commute. There is a greater shift from drive-alone to shared-ride when cost is increased in the peak periods, rather than in the drive-alone mode. The results indicate that with an increase in auto travel times or costs (either for the entire day or only for the peak period), there is a decrease in the share of the drive-alone mode and a corresponding increase in the shared ride and transit modes.

6.4.4 Aggregate Trip Frequency

There is very little difference in aggregate trip frequency for the different scenarios (Table 6.16). The 25% IVTT decrease across all motorized modes and time periods results in a slight increase in the number of home-based non-work trips. Increasing IVTT did not result in substantial changes in the trip frequency. Also, increasing cost did not have significant impacts on trip frequency.

Table 6.15 Commute Mode Shares

	Drive Alone	Shared Ride	Walk/Bike	Transit	School Bus
Base Case	54.58	41.28	2.14	0.84	1.16
25% Decrease in IVTT	54.55	41.25	2.17	0.88	1.16
25% Increase in IVTT	54.50	41.35	2.13	0.84	1.18
25% Increase in IVTT —Auto Mode	54.30	41.42	2.20	0.90	1.19
25% Increase in IVTT—Peak Periods	54.24	41.60	2.16	0.83	1.16
25% Increase in IVTT—Auto Mode and Peak Periods	54.29	41.58	2.11	0.86	1.16
25% Increase in Cost	54.67	41.16	2.15	0.86	1.15
25% Increase in Cost—Auto Mode	54.49	41.37	2.14	0.85	1.16
25% Increase in Cost—Peak Periods	54.23	41.58	2.17	0.86	1.17
25% Increase in Cost —Auto Mode and Peak Periods	54.57	41.26	2.11	0.87	1.19
$2 Increase in CBD Cost—Auto Mode and Peak Periods	54.30	41.45	2.23	0.84	1.18

Table 6.16 Aggregate Trip Frequency by Trip Type (millions)

	Home-Based Work	Home-Based Non-Work	Non–Home-Based	Overall
Base Case	2.74	9.36	4.86	16.96
25% Decrease in IVTT	2.73	9.45	4.86	17.04
25% Increase in IVTT	2.75	9.34	4.86	16.95
25% Increase in IVTT—Auto Mode	2.74	9.35	4.85	16.95
25% Increase in IVTT—Peak Periods	2.74	9.36	4.85	16.95
25% Increase in IVTT—Auto Mode and Peak Periods	2.74	9.39	4.87	17.00
25% Increase in Cost	2.74	9.37	4.82	16.94
25% Increase in Cost—Auto Mode	2.75	9.37	4.84	16.96
25% Increase in Cost—Peak Periods	2.74	9.40	4.85	16.99
25% Increase in Cost—Auto Mode and Peak Periods	2.75	9.42	4.85	17.03
$2 Increase in CBD Cost—Auto Mode and Peak Periods	2.74	9.41	4.85	16.99

6.4.5 Aggregate Person Hours of Travel

A 25% decrease in IVTT for all motorized modes and time periods causes a 7% decrease in total person hours of travel (PHT), while a 25% increase in IVTT for all motorized modes and time periods causes a 7% increase in total PHT (Table 6.17). Increasing IVTT in the auto modes also causes an increase in PHT but increasing IVTT in the peak periods does not alter total PHT. Interestingly, a cost increase in the peak period results in an increase in PHT but other cost increases do not have an affect on PHT.

6.4.6 Aggregate Person Miles of Travel

The total person miles of travel (PMT) increases by 23%, when IVTT is decreased by 25% across the board. The total PMT decreases by 14% when IVTT is increased by 25% across the board (Table 6.18). An IVTT increase in the auto mode has similar affects on PMT as the overall IVTT increase scenario. IVTT increase in the peak periods decreases PMT, but to a lesser extent compared to an IVTT increase in the auto mode. Total PMT for home-based work trips does not change for any scenario. This is due to the fact that work locations for all the scenarios are assumed to remain the same. Therefore, travel distances to work are not affected by changes in network characteristics. There was no significant change in PMT for the cost increase scenarios. For the two scenarios where IVTT is increased in the peak periods, PMT decreases for the peak periods by over 7% but does not change significantly for the off-peak periods.

Table 6.17 Total Person Hours of Travel (PHT) by Trip Type (millions)

	Home-Based Work	Home-Based Non-Work	Non–Home-Based	Overall
Base Case	54.75	146.13	73.31	274.19
25% Decrease in IVTT	41.29	140.66	71.74	253.70
25% Increase in IVTT	68.37	151.04	74.84	294.24
25% Increase in IVTT—Auto Mode	67.45	150.99	74.69	293.13
25% Increase in IVTT—Peak Periods	54.73	146.66	73.25	274.64
25% Increase in IVTT—Auto Mode and Peak Periods	64.24	149.28	74.32	287.84
25% Increase in Cost	55.08	146.52	72.88	274.48
25% Increase in Cost—Auto Mode	54.93	145.84	72.99	273.76
25% Increase in Cost—Peak Periods	64.79	149.25	73.91	287.95
25% Increase in Cost—Auto Mode and Peak Periods	55.17	147.34	73.34	275.85
$2 Increase in CBD Cost—Auto Mode and Peak Periods	54.78	146.92	73.06	274.76

Table 6.18 Total Person Miles of Travel (PMT) by Trip Type (millions)

	Home-Based Work	Home-Based Non-Work	Non–Home-Based	Overall
Base Case	35.12	82.64	40.50	158.27
25% Decrease in IVTT	35.12	105.68	53.83	194.63
25% Increase in IVTT	35.16	68.16	32.05	135.36
25% Increase in IVTT—Auto Mode	35.12	67.87	31.97	134.96
25% Increase in IVTT—Peak Periods	35.09	78.49	37.98	151.56
25% Increase in IVTT—Auto Mode and Peak Periods	35.02	78.54	38.26	151.82
25% Increase in Cost	35.31	83.10	40.27	158.68
25% Increase in Cost—Auto Mode	35.26	82.64	40.28	158.17
25% Increase in Cost—Peak Periods	35.06	83.10	40.37	158.53
25% Increase in Cost—Auto Mode and Peak Periods	35.34	83.55	40.55	159.44
$2 Increase in CBD Cost—Auto Mode and Peak Periods	35.18	83.16	40.27	158.61

6.4.7 Percentage of Stops in the Central Business District by Trip Period

The percentage of stops in the central business district (CBD) for non-commute auto tours was calculated for each time period (Table 6.19). This was done to determine if charging $2 for auto trips involving the CBD in the peak periods would cause a decrease in the number of trips into the CBD. The scenario does not show any pattern in the number of stops into or out of the CBD. We will examine this result carefully in the future.

6.4.8 The 25% Increase in Population Scenario Results

A 25% increase in total population results in a 25% increase in trip frequency, person hours of travel, and person and vehicle miles of travel, for all trips. All other statistics remain the same as the base year.[21]

[21] The statistics are identical to the base year scenario statistics, so the corresponding tables are not provided.

Table 6.19 Percentage of Stops in the CBD for Non-Commute Auto Trips

	A.M. Peak (6:30 a.m. to 8:59 a.m.)	**Mid-Day Off Peak (9:00 a.m. to 2:59 p.m.)**	**P.M. Peak (3:00 p.m. to 6:29 p.m.)**	**Evening Off Peak (6:30 p.m. to 6:29 a.m.)**
Base Case	0.80	1.62	0.72	0.39
25% Decrease in IVTT	0.98	1.74	0.77	0.42
25% Increase in IVTT	1.02	1.72	0.68	0.32
25% Increase in IVTT—Auto Mode	0.74	1.73	0.65	0.30
25% Increase in IVTT—Peak Periods	0.82	1.68	0.63	0.37
25% Increase in IVTT—Auto Mode and Peak Periods	0.97	1.76	0.64	0.37
25% Increase in Cost	0.55	1.69	0.69	0.39
25% Increase in Cost—Auto Mode	0.78	1.72	0.70	0.37
25% Increase in Cost—Peak Periods	0.85	1.73	0.69	0.39
25% Increase in Cost—Auto Mode and Peak Periods	0.77	1.68	0.69	0.37
$2 Increase in CBD Cost—Auto Mode and Peak Periods	0.96	1.73	0.68	0.34

6.5 CEMDAP Forecasting Results: The 2025 Forecast Scenario

The main purpose of CEMDAP is to determine how many people and vehicles will be traveling from one location to another in a future year. This section discusses the results of using CEMDAP for forecasting travel patterns for the year 2025. Specifically, the input data were altered to reflect the DFW population and land-use characteristics in the year 2025. The 2025 forecasting scenario increases the number of persons in DFW by 61%, the number of households in DFW by 66%, and the number of workers in DFW by 76%. Also, we used predicted network characteristics for the year 2025 (as provided by NCTCOG).

Section 6.5.1 and Section 6.5.2 show changes in pattern-level and aggregate statistics between the year 2000 base case and the 2025 scenario, respectively. Section 6.5.3 compares the CEMDAP 2025 scenario outputs with the DFW model predictions for 2025.

6.5.1 2025 Scenario Pattern-Level Statistics

Several pattern-level attributes of the base case are compared with the corresponding attributes in the 2025 scenario. These are: (1) number of worker tours and stops, (2) trip chaining propensity, (3) average daily duration of activities, and (4) work start and end times. The results are discussed below.

The average number of worker tours and stops does not differ much among scenarios (as shown in Table 6.20). Also, the average chaining propensity does not vary greatly between 2000 and 2025 (Table 6.21). There is very little change in the average number of daily trips per person for each trip type (Table 6.22). Average minutes of travel both per person and per trip do increase for each trip type between 2000 and 2025. The greatest increase occurs in home-based work trips. Average person miles of travel and average vehicle miles of travel also increase between 2000 and 2025. Overall, the pattern-level statistics suggest that people are not changing their daily travel patterns but are instead traveling longer and farther to reach their desired destinations.

Table 6.20 2025 Scenario: Number of Worker Tours and Stops

	Avg. No. of BW Tours[22]	Avg. No. of Stops in BW Tour[23]	Avg. No. of WB Tours[7]	Avg. No. of Stops in WB Tour[8]	Avg. No. of AW Tours[7]	Avg. No. of Stops in AW Tour[8]	Avg. No. of Stops in HW Commute[7]	Avg. No. of Stops in WH Commute[8]
Year 2000 Base	0.02	1.26	0.33	1.26	0.40	1.34	0.20	0.41
Year 2025 Forecast	0.03	1.23	0.31	1.28	0.40	1.34	0.21	0.41

Table 6.21 2025 Scenario: Trip Chaining Characteristics[24]

	Non-Workers			Workers		
	Avg. No. of Tours	Avg. No. of Stops	Avg. Chaining Prop.	Avg. No. of Non-Commute Tours	Avg. No. of Stops in Non-Commute Tours	Avg. Chaining Propensity
Year 2000 Base	1.73	3.89	0.60	1.43	3.30	0.46
Year 2025 Forecast	1.70	3.93	0.59	1.44	3.33	0.46

[7] Averaged over all workers.
[23] Averaged within each tour.
[24] Average over all workers and on-workers.

Table 6.22 Trip Type Characteristics[25]

	Avg. no. of daily trips per person	Avg. min. of travel per trip	Avg. person miles of travel (PMT) per trip	Avg. veh. miles of travel (VMT) per trip
Home-Based Work				
Year 2000 Base	1.67	19.92	12.85	13.14
Year 2025 Forecast	1.66	22.21	14.78	15.06
Home-Based Other				
Year 2000 Base	2.80	16.54	9.31	10.08
Year 2025 Forecast	2.76	17.18	9.94	10.73
Non–Home-Based				
Year 2000 Base	2.49	15.64	8.68	8.94
Year 2025 Forecast	2.54	16.18	9.10	9.43

6.5.2 2025 Scenario Aggregate Statistics

The aggregate statistics compared between the 2000 base case and the 2025 forecasting scenario in the current section are these: (1) total trip frequency, (2) person hours of travel, and (3) person miles of travel. The results are discussed below.

The number of home-based work trips increases by 75%, which is exactly the same number of workers added to the DFW region (Table 6.23). However, person hours of travel and person miles of travel for home-based work trips increase by 94% and 101%, respectively. This is consistent with the pattern-level trip-type characteristics, which reveal an increase in per person and per trip travel time and distance for the year 2025 (Table 6.11). Overall, the number of total trips increases by 63%, total PHT increases by 71%, and total PMT increases by 76%. Similar to the home-based work trip increase, the total increase in number of trips is proportional to the increase in population between 2000 and 2025.

[25] Averaged over all individuals who made at least one out-of-home stop.

Table 6.23 2025 Scenario: Aggregate Trip Frequency by Trip Type (millions)

	Home-Based Work	Home-Based Non-Work	Non–Home-Based	Overall
Trip Frequency				
Year 2000 Base	2.74	9.36	4.86	16.96
Year 2025 Forecast	4.79	14.66	8.16	27.61
Percent Difference	74.94	56.61	67.93	62.81
Person Hours of Travel (PHT)				
Year 2000 Base	54.75	146.13	73.31	274.19
Year 2025 Forecast	106.04	237.72	126.42	470.18
Percent Difference	93.69	62.68	72.45	71.48
Person Miles of Travel (PMT)				
Year 2000 Base	35.12	82.64	40.50	158.27
Year 2025 Forecast	70.58	137.95	70.42	278.95
Percent Difference	100.94	66.92	73.88	76.25

6.5.3 CEMDAP versus DFW Model: 2025 Forecasting Scenario

The current section compares CEMDAP forecasted activity travel patterns for 2025 with the DFW model predicted forecasts for the year 2025. The comparison is based on the following statistics: (1) total vehicle miles of travel and (2) total number of vehicle trips. The results of the comparison are reported in Table 6.24 and Table 6.25,[26],respectively. In particular, Table 6.24 indicates that from 1999 to 2025 the DFW model predicts a uniform increase in the trips across all time periods, wheras CEMDAP predicts a higher proportion of increase for the a.m, peak period relative to other periods. Table 6.25 indicates that from 2000 to 2025 CEMDAP predicts a higher percentage rise of auto trips in the a.m. peak period compared to the DFW model changes.

For the year 2025, the DFW model outputs and the CEMDAP outputs differ by only 1% in total VMT and by only 0.23% in total number of trips. For both vehicle miles of travel and number of vehicle trips, the greatest percentage difference between the two models is observed in the morning peak period. Overall, this suggests that the DFW model and CEMDAP are close in their estimations of future travel demand for the year 2025.

[26] Table A5.16 also includes the results of the base year outputs of CEMDAP and DFW models.

Table 6.24 CEMDAP vs. DFW Model: 2025 Scenario Aggregate VMT by Time of Day (millions)

	DFW 1999	CEMDAP 2000	DFW 2025	CEMDAP 2025	Percent Difference between DFW 1999 and DFW 2025	Percent Difference between CEMDAP 2000 and CEMDAP 2025
AM Peak Period	24.09	27.86	43.27	49.93	79.62	79.22
PM Peak Period	33.75	39.72	61.21	65.77	81.36	65.58
Off-Peak Periods	73.72	77.62	133.24	124.41	80.74	60.28
Total	131.57	145.21	237.72	240.10	80.68	65.35

Table 6.25 CEMDAP vs. DFW Model: 2025 Scenario Number of Trips by Mode and Time of Day (millions)

	DFW 1999	CEMDAP 2000	DFW 2025	CEMDAP 2025	Percent Difference between DFW 1999 and DFW 2025	Percent Difference between CEMDAP 2000 and CEMDAP 2025
Drive Alone						
AM Peak Period	1.85	1.88	2.93	3.41	58.38	81.38
PM Peak Period	2.68	2.60	4.29	4.31	60.07	65.77
Off-peak Periods	5.71	4.85	9.15	7.88	60.25	62.47
Total	10.24	9.33	16.36	15.60	59.77	67.20
Shared-Ride No HOV						
AM Peak Period	0.32	0.36	0.52	0.61	62.50	69.44
PM Peak Period	0.61	0.92	1.01	1.40	65.57	52.17
Off-peak Periods	1.48	1.58	2.43	2.49	64.19	57.59
Total	2.41	2.86	3.96	4.50	64.32	57.34
Shared-Ride HOV						
AM Peak Period	0.22	0.25	0.36	0.42	63.64	68.00
PM Peak Period	0.30	0.15	0.50	0.23	66.67	53.33
Off-peak Periods	0.63	0.97	1.03	1.52	63.49	56.70
Total	1.15	1.37	1.88	2.18	63.48	59.12
Trucks						
AM Peak Period	0.07	0.07	0.10	0.10	42.86	42.86
PM Peak Period	0.10	0.10	0.15	0.15	50.00	50.00
Off-peak Periods	0.38	0.38	0.60	0.60	57.89	57.89
Total	0.55	0.55	0.85	0.85	54.55	54.55
All Vehicles						
AM Peak Period	2.45	2.56	3.92	4.55	60.00	77.73
PM Peak Period	3.69	3.76	5.94	6.08	60.98	61.70
Off-peak Periods	8.19	7.78	13.20	12.49	61.17	60.54
Total	14.34	14.10	23.06	23.12	60.81	63.97

7. SUMMARY

This report focused on the development of the latest version of CEMDAP, the activity-travel simulator. Specifically, this report documented (1) the modeling and software enhancements to CEMDAP, (2) the generation of the inputs for CEMDAP using software components SPG and CEMSELTS, and (3) the empirical validation of CEMDAP and the results of sensitivity testing carried out using CEMDAP.

Chapter 2 described the new econometric modeling system and the microsimulation framework embedded within the latest version of CEMDAP for (1) accommodating a finer spatial resolution (4,874 zones instead of 919 zones for the DFW area in Texas), (2) explicitly accounting for children's activity-travel, and (3) explicitly capturing the intrahousehold interactions between the travel patterns of children and their parents (such as escort to and from school and joint participation in discretionary activities). The chapter highlighted the spatial and temporal consistency checks implemented within CEMDAP to ensure that the simulation process does not result in unreasonable or impossible activity travel patterns.

Chapter 3 discussed the software features of CEMDAP, including the object-oriented approach and the software architecture. The choice of object-oriented development paradigm and the benefits it offers were highlighted. The chapter discussed the strategies adopted to enhance the computational performance of CEMDAP. Finally, the improvements in the software architecture, design, and implementation of the recent version of CEMDAP in comparison with that of a previous version were discussed.

Chapter 4 presented details of generating and verifying the synthetic population for the base year (year 2000) and forecast year (year 2025). In particular, the chapter summarized the algorithm employed in the current project. The specific datasets used and compiled for generating the base year and the forecast year populations were described. The chapter also verified the results produced by the SPG for the DFW application.

Chapter 5 described the implementation of CEMSELTS to generate the disaggregate household and person level inputs as required for CEMDAP. In particular, the chapter identified the household- and person-level attributes that need to be generated external to SPG. The chapter discussed the structure and the prediction procedure underlying each of these household- and person-level attributes. The validation of the outputs resulting from the application of these modules was also presented.

Chapter 6 focused on the empirical validation of CEMDAP and the results of sensitivity testing undertaken with CEMDAP. The chapter discussed validation results of CEMDAP against the estimation sample and presented aggregate comparisons between CEMDAP and the NCTCOG's four-step model. The applicability of CEMDAP as a policy tool was illustrated based on the prediction under several scenarios. The forecasting ability of CEMDAP was demonstrated by discussing the results of the activity travel patterns generated for a future year (2025).

REFERENCES

Bhat, C.R. Imputing a Continuous Income Variable from Grouped and Missing Income Observations. *Economics Letters*, Vol. 46, 1994, pp. 311-319.

Bhat, C.R., S. Srinivasan, and J. Guo. Activity Based Travel Demand Analysis for Metropolitan Areas in Texas: Model Components and Mathematical Formulations. Report 4080-2, prepared for the Texas Department of Transportation, 2001.

Bhat, C.R., S. Srinivasan, J. Guo, and A. Sivakumar. Activity Based Travel Demand Analysis for Metropolitan Areas in Texas: A Micro-simulation Framework for Forecasting. Report 4080-4, prepared for the Texas Department of Transportation, 2003.

Bhat, C.R., A. Pinjar, N. Eluru, R. Copperman, I. N. Sener, J. Guo, and S. Srinivasan. CEMDAP User's Manual. Product 4080-P8, prepared for the Texas Department of Transportation, 2006

Guo, J.Y., S. Srinivasan, N. Eluru, A. Pinjari, R. Copperman, and C.R. Bhat., Activity-Based Travel-Demand Analysis for Metropolitan Areas in Texas: CEMSELTS Model Estimations and Prediction Procedures, 4874 Zone System CEMDAP Model Estimations and Procedures, and the SPG Software Details. Report 4080-7, prepared for the Texas Department of Transportation, 2005.

Guo, J.Y., and C.R. Bhat. Population Synthesis for Microsimulating Travel Behavior. Technical paper, Department of Civil and Environmental Engineering, University of Wisconsin–Madison, 2006.

APPENDIX A: MODEL ESTIMATION RESULTS FOR CEMDAP

This appendix presents the complete set of empirical models estimated using travel survey data from the DFW region that constitutes the overall CEMDAP modeling system. This overall modeling system is broadly subdivided into five categories; model estimation results from each of the categories are presented below. The five categories are (1) the generation-allocation model system (Section 1), (2) the worker scheduling model system (Section 2), (3) the non-worker scheduling model system (Section 3), (4) the joint discretionary tour-scheduling model system (Section 4), and (5) the children scheduling model system (Section 5).

A.1 Generation-Allocation Model System

Table A.1 Child's Decision To Go to School (Model GA1)

Explanatory Variables	**Param.**	**t-stat**
Constant	-0.577	-2.18
Highest level of education completed		
No school (base)	--	--
Preschool	0.905	3.32
Kindergarten to grade 4	1.935	7.32
Grade 5 to grade 8	1.863	6.77
Grade 9 or higher	1.620	3.37
Household income (in thousands of dollars)	0.006	2.20

Table A.2 Child's school start time (Model GA2)

Explanatory Variables	Param.	t-stat
Threshold parameters		
THRESH01 (0 to 260.5)	-2.589	-17.19
THRESH02 (260.5 to 270.5)	-1.999	-16.79
THRESH03 (270.5 to 280.5)	-1.454	-14.45
THRESH04 (280.5 to 285.5)	-0.972	-10.54
THRESH05 (285.5 to 290.5)	-0.645	-6.79
THRESH06 (290.5 to 295.5)	-0.415	-4.12
THRESH07 (295.5 to 300.5)	-0.026	-0.22
THRESH08 (300.5 to 310.5)	0.278	1.83
THRESH09 (310.5 to 320.5)	0.552	2.91
THRESH10 (320.5 to 330.5)	0.785	3.39
THRESH11 (330.5 to 350.5)	1.068	3.59
THRESH12 (350.5 to 400.5)	1.330	3.66
Age ≤ 5 years	0.503	3.42
Highest level of education completed		
Kindergarten to grade 4	-0.260	-2.58
Ethnicity		
African-American	-0.239	-2.19
Asian	0.823	2.59
Number of unemployed adults in household	0.131	1.40
Variance of the heterogeneity term	0.215	0.48

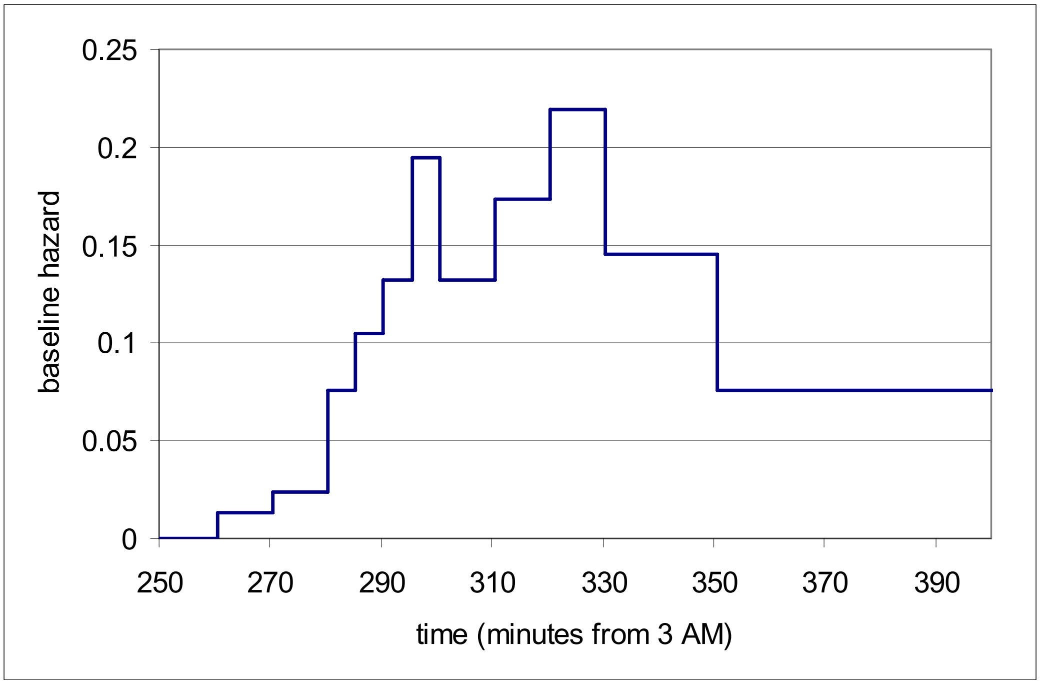

Figure A-1 Baseline hazard function for child's school start time

Table A.3 Child's school end time (Model GA3)

Explanatory variables	Param.	t-stat
Threshold parameters		
THRESH01 (0 to 300.5)	-2.620	-17.11
THRESH02 (300.5 to 400.5)	-2.163	-16.45
THRESH03 (400.5 to 420.5)	-1.587	-14.90
THRESH04 (420.5 to 430.5)	-0.962	-10.78
THRESH05 (430.5 to 440.5)	-0.414	-5.02
THRESH06 (440.5 to 450.5)	-0.099	-1.17
THRESH07 (450.5 to 460.5)	0.109	1.23
THRESH08 (460.5 to 480.5)	0.593	5.22
THRESH09 (480.5 to 550.5)	1.031	6.41
Age $\leq$ 5 years	-2.340	-1.68
Age $\leq$ 5 years * one employed adult	3.015	2.15
Age $\leq$ 5 years * two employed adults	3.521	2.51
Highest level of education completed		
Preschool	-0.467	-3.89
Kindergarten to grade 4	-0.401	-3.91
Variance of the heterogeneity term	0.000	--

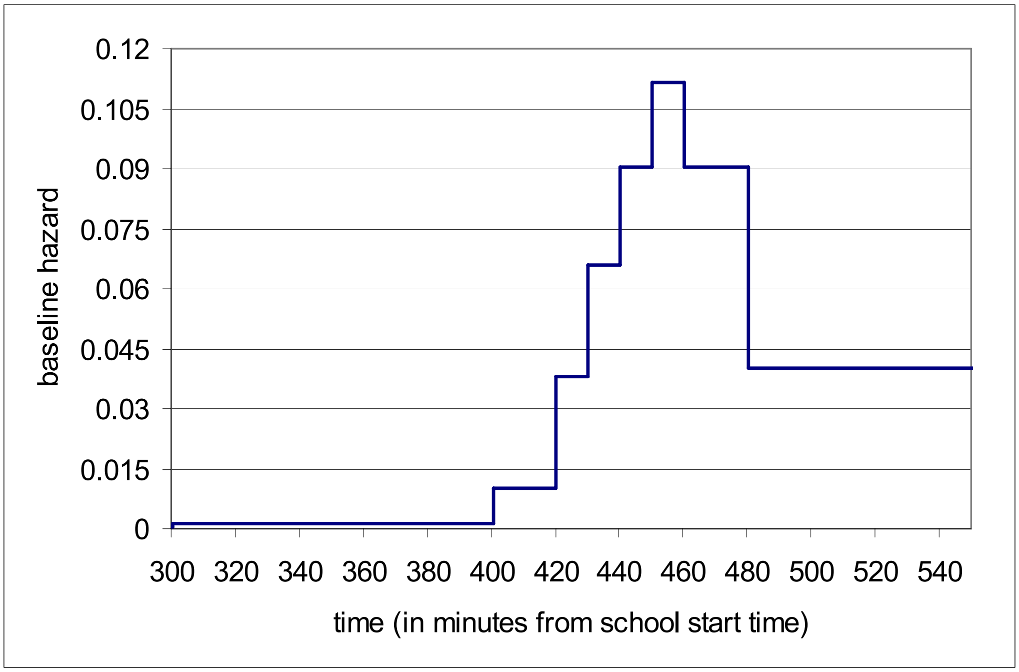

Figure A-2 Baseline hazard function for child's school end time

Table A.4 Decision to go to work (Model GA4)

Explanatory variables	Param.	t-stat
Constant	1.910	9.68
Age	-0.008	-2.07
Ratio of personal income to household income	0.461	3.11
Female	0.316	3.27
Number of non-school going children * Mother	-0.495	-2.85
Weekly work duration		
Between 0 and 20 hours	-1.776	-12.41
Between 20 and 40 hours	-0.450	-4.37
High work flexibility	-1.146	-12.49

Table A.5 Work start and end times (Model GA5)

Explanatory variables	Param.	t-stat
Arrival-time function		
$Sin(2\pi t_a/24)$	-1.896	-1.06
$Sin(4\pi t_a/24)$	2.358	3.99
$Sin(6\pi t_a/24)$	1.066	6.09
$Cos(2\pi t_a/24)$	-7.935	-6.24
$Cos(4\pi t_a/24)$	-4.506	-6.37

Table A.5 (cont.) Work start and end times (Model GA5)

Explanatory Variables	**Param.**	**t-stat**
$Cos(6\pi t_a/24)$	-1.445	-5.22
Departure-time function		
$Sin(2\pi t_d/24)$	6.990	3.57
$Sin(4\pi t_d/24)$	3.609	4.86
$Sin(6\pi t_d/24)$	0.728	3.43
$Cos(2\pi t_d/24)$	-4.339	-3.43
$Cos(4\pi t_d/24)$	-0.768	-1.09
$Cos(6\pi t_d/24)$	0.105	0.45
Duration function		
Duration	3.437	5.44
$Duration^2$	-1.394	-5.58
$Duration^3$	0.313	6.72
$Duration^4$	-0.032	-7.30
$Duration^5$	1.457	7.42
$Duration^6$	-0.025	-7.26
Expected Home-to-Work Travel Time	-0.030	-1.90
Expected Home-to-Work Travel Cost	-0.003	fixed
Size variables		
Num. of 15 min. periods in the arrival time period	0.593	15.17
Num. of 15 min. periods in the departure time period	0.364	5.76
Mother—Departure Time		
$Sin(2\pi t_d/24)$ * Mother	-6.906	-1.56
$Sin(4\pi t_d/24)$ * Mother	-7.837	-1.76
$Sin(6\pi t_d/24)$ * Mother	-3.086	-1.86
$Cos(2\pi t_d/24)$ * Mother	-14.964	-1.87
$Cos(4\pi t_d/24)$ * Mother	-6.997	-1.85
$Cos(6\pi t_d/24)$ * Mother	-1.593	-1.67
High work flexibility—Arrival Time		
$Sin(2\pi t_a/24)$ * High work flexibility	7.381	2.86
$Sin(4\pi t_a/24)$ * High work flexibility	4.636	3.39
$Sin(6\pi t_a/24)$ * High work flexibility	-0.177	-0.56
$Cos(2\pi t_a/24)$ * High work flexibility	-4.250	-3.39
$Cos(4\pi t_a/24)$ * High work flexibility	2.743	2.38
$Cos(6\pi t_a/24)$ * High work flexibility	2.519	4.34

Table A.5 (cont.) Work start and end times (Model GA5)

Explanatory Variables	Param.	t-stat
Work duration > 40 hours/week—Arrival Time		
$Sin(2\pi t_a/24)$ * Work duration > 40 hours/week	2.107	1.10
$Sin(4\pi t_a/24)$ * Work duration > 40 hours/week	-1.886	-2.35
$Sin(6\pi t_a/24)$ * Work duration > 40 hours/week	-0.909	-2.41
$Cos(2\pi t_a/24)$ * Work duration > 40 hours/week	2.241	3.04
$Cos(4\pi t_a/24)$ * Work duration > 40 hours/week	1.805	1.69
$Cos(6\pi t_a/24)$ * Work duration > 40 hours/week	-1.028	-2.85

Table A.6 Decision to undertake work-related activities (Model GA6)

Explanatory Variables	Param.	t-stat
Constant	-0.189	-1.73
Female	-0.703	-6.46
Number of non-schoolgoing children * Mother	-0.669	-2.27
Worker	0.954	3.70
Work-based duration	-0.005	-10.76
High work flexibility	0.319	2.95
Employment type		
Wholesale and Transportation	-0.330	-2.00

Table A.7 Adult's decision to go to school (Model GA7)

Explanatory Variables	Param.	t-stat
Constant	1.011	3.72
Caucasian	0.560	2.11
Highest level of education		
Some college, no degree	-0.861	-2.98
Associate's or bachelor's degree	-1.130	-3.26
Master's or PhD degree	-1.983	-3.71
Household income	0.006	1.49
Presence of non-school going children	-0.810	-1.90

Table A.8 Adult's school start and end times (Models GA8 and GA9)

Explanatory Variables	School start time (Model GA8)		School end time (Model GA9)	
	Param.	t-stat	Param.	t-stat
Constant	5.790	113.69	5.999	71.44
Highest level of education				
Some college, no degree	0.170	3.80	-0.465	-6.81
Associate's or bachelor's degree	0.170	3.80	-0.465	-6.81
Master's or PhD degree	0.276	3.57	-0.728	-6.19
Adult son or daughter in a single-parent or nuclear family household	-0.139	-2.47	--	--
Adult in "other" household type	-0.128	-2.37	--	--
Household income ($1000)	0.001	2.43	-0.002	-2.18
Vehicles per licensed driver	--	--	0.1196	1.63

Table A.9 Child's mode of travel to and from school: Sample shares

		Mode of travel from school				
		Drive by parent	Drive by other	School bus	Walk or bike	Total
Mode of travel to school	Drive by parent	254	66	40	43	403
	Drive by other	17	48	6	8	79
	School bus	6	6	99	6	117
	Walk or bike	11	1	2	103	117
	Total	288	121	147	160	716

Table A.10 Child's travel model to school (Model GA10) and from school (Model GA11)

Explanatory Variables	Mode to school (Model GA10)		Mode from school (Model GA11)	
	Param.	t-stat	Param.	t-stat
Drive by parent				
Age	-0.159	-6.08	-0.236	-7.99
Number of vehicles in household	0.367	2.87	0.751	5.30
Number of workers	--	--	-0.624	-3.97
School-home distance	0.610	5.55	0.641	6.48
Drive by others				
Constant	-2.213	-5.58	-1.762	-3.82
Age			-0.084	-2.47
African-American	-1.300	-2.84	--	--
Number of non–school-going children	0.604	2.87	--	--
Number of non-workers	-0.639	-2.13	--	--
School-home distance	0.527	4.54	0.617	6.17
School Bus				
Constant	-2.509	-6.72	-2.694	-6.27
School-home distance	0.663	5.98	0.677	6.83
Walk or bike				
Constant	-1.166	-3.03	-1.383	-3.16
African-American	--	--	0.695	2.51

Table A.11 Allocation of the drop-off episode (Model GA12)

Explanatory Variables	Father		Mother	
	Param.	t-stat	Param.	t-stat
Constant	-0.799	-3.51	--	--
Work start time	0.004	2.69	0.004	2.69
Work duration	-0.004	-3.96	-0.004	-3.96

Table A.12 Allocation of the pick-up episode (Model GA13)

Explanatory Variables	Father		Mother	
	Param.	t-stat	Param.	t-stat
Constant	-0.735	-1.56	--	--
Age	0.153	1.96	0.153	1.96
Mult. School-going children in hh	-1.889	-2.48	--	--
Work duration	-0.004	-3.86	-0.004	-3.86

Table A.13 Child's decision to undertake joint discretionary activity with parent (Model GA14)

Explanatory variables	Param.	t-stat
Constant	-1.601	-6.95
Personal and household level characteristics		
Household income ($1000)	0.005	1.78
Number of vehicles	0.166	1.66
Household-level activity participation characteristics		
Number of school going children	-0.139	-1.85
Presence of a female worker	-0.569	-3.65
School-related characteristics		
School start time	0.002	2.57
School-based duration	-0.002	-2.88
Mode of travel from school: Driven back by parent	0.324	1.56

Table A.14 Allocation of the joint discretionary episode to one of the parents (Model GA15)

Explanatory variables	Father		Mother	
	Param.	t-stat	Param.	t-stat
Constant	0.089	0.21	--	--
Number of school-going children	-1.266	-1.57	--	--
Work duration	-0.002	-1.93	-0.002	-1.93

Table A.15 Child's decision to undertake independent discretionary activity (Model GA16)

Explanatory variables	Param.	t-stat
Constant	-2.851	-5.88
Individual- and household-level characteristics		
Age	0.088	3.17
Male	0.256	1.29
Caucasian	0.405	1.54
Household income (in thousands of dollars)	0.008	2.25
Household-level activity participation characteristics		
Number of school going children	0.243	2.89
Number of non-school going children	0.317	2.10
Number of workers	-0.458	-2.13
Number of non-workers	-0.842	-2.81
Presence of female workers	-0.518	-1.79
Mode of travel from school to home		
Driven back by parent	-1.091	-3.28
Driven back by others	0.916	3.44

Table A.16 Decision of household to undertake grocery shopping (Model GA17)

Explanatory variables	Param.	t-stat
Constant	-1.019	-7.09
Individual- and household-level characteristics		
Number of vehicles	0.170	3.13
Single-person household	-0.256	-2.23
Household location characteristics		
Distance to nearest major shopping zone	-0.031	-3.56
Household-level activity participation characteristics		
Presence of non-schoolgoing children	-0.180	-1.41
Number of non-workers	0.260	4.68

Table A.17 Decision of an adult to undertake grocery shopping given household undertakes it (Model GA18)

Explanatory variables	Param.	t-stat
Constant	1.303	3.16
Individual- and household-level characteristics		
Age	0.008	1.90
Income (in thousands of dollars)	-0.004	-1.70
Male	-0.727	-3.84
Licensed	1.395	5.73
Household-level activity participation characteristics		
Number of workers	-0.166	-1.38
Number of non-workers	-0.893	-7.48
Number of female workers	-0.384	-2.34
Individual-level activity participation		
Worker	-0.782	-1.97
Worker * female	0.434	1.49
Work-based duration	-0.002	-2.96
Undertakes work-related activities	-0.687	-3.25
Drops off children at school	0.823	2.25

Table A.18 Decision of an adult to undertake household or personal business activities (Model GA19)

Explanatory variables	Param.	t-stat
Constant	-0.823	-4.98
Personal and household level characteristics		
Age	-0.007	-3.34
Licensed	0.484	3.83
Caucasian	0.484	5.34
Household-level activity participation characteristics		
Number of school-going children	-0.120	-2.45
Number of non– school-going children	-0.207	-3.49
Another household adult works	-0.173	-2.14
Individual work characteristics		
Worker	0.740	3.99
Work duration	-0.003	-7.29
Expected no-stop total auto commute time	-0.003	-1.64
Individual non-work participation		
Work related	-0.197	-1.97
Shopping	0.646	8.59

Table A.19 Decision of an adult to undertake social or recreational activities (Model GA20)

Explanatory variables	Param.	t-stat
Constant	-1.396	-7.21
Personal and household level characteristics		
Age	-0.013	-5.15
Income (1000$)	-0.003	-2.07
Household income (1000$)	0.004	3.24
Licensed	0.663	4.42
Caucasian	0.318	3.11
Household-level activity participation characteristics		
Another adult undertakes shopping	0.291	2.01
Number of workers	-0.160	-3.01
Number of non–school-going children	-0.128	-2.01
Individual work characteristics		
Worker	1.535	4.71
Work end time	-0.002	-3.03
Work duration	-0.001	-2.92
Individual non-work participation		
Work related	-0.294	-2.55
Shopping	0.227	1.84
Household/personal business activities	0.597	7.34
Shopping and household/personal business activities	-0.409	-2.51

Table A.20 Decision of an adult to undertake eating out activities (Model GA21)

Explanatory variables	Param.	t-stat
Constant	-2.976	-12.14
Personal and household level characteristics		
Age	-0.007	-2.69
Income (1000$)	0.003	2.45
Household income (1000$)	0.006	5.09
Licensed	0.746	3.80
Caucasian	0.594	5.19
Household-level activity participation characteristics		
Number of workers	-0.149	-2.75
Number of non–school-going children	-0.178	-2.54
Another adult undertakes shopping	0.448	3.01
Individual work characteristics		
Worker	-0.636	-1.97
Work end time	0.001	2.75
Expected no-stop total auto commute time	0.007	4.29
Individual non-work participation		
Work related	0.757	7.28
Shopping	0.327	2.96
Household/personal business	0.841	11.33
Social/recreational	0.517	5.71
Shopping and social recreational	-0.610	-3.33

Table A.21 Decision of an adult to undertake other serve-passenger activities (Model GA22)

Explanatory variables	Param.	t-stat
Constant	-1.692	-6.18
Single person household	-0.384	-2.10
Single parent household	0.664	3.71
Age	-0.010	-2.94
Work duration	-0.002	-6.71
Number of school going children	0.590	10.89
Number of non–school-going children	0.413	5.95
Number of workers in household	0.362	5.02
Number of non-workers in household	-0.310	-3.41
Undertakes household/personal business activity	0.405	4.31
Undertakes social/recreational activity	0.388	3.99
Undertakes eat out activity	0.269	2.75

A.2 Worker Scheduling Model System

Table A.22 Commute mode (Model WSCH1)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike		Transit	
	Param.	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**
Constant	1.307	4.14	-0.248	-0.61	-0.990	-3.01	--	--	0.333	1.61
Person and household-level characteristics										
Age	--	--	-0.029	-3.44	--	--	--	--	--	--
Pers. veh. availability	0.637	3.04	--	--	--	--	--	--	--	--
Employed	--	--	--	--	-0.996	-5.26	-0.996	-5.26		
Mult. adults in hh	--	--	--	--	0.795	2.54	--	--	--	--
Household-level activity participation decisions										
Mult. workers in hh	--	--	0.448	2.88	0.448	2.88	--	--	--	--
Individual activity participation										
Work related	--	--	--	--	-2.245	-2.23	--	--	--	--
Shopping	--	--	--	--	--	--	-0.684	-2.35	-0.684	-2.35
Other serve passenger	--	--	1.023	4.99	--	--	--	--	--	--
Joint discret. activities with children	--	--	1.585	3.28	--	--	--	--	--	--
Level-of-service										
AM peak trav. time (min)	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17	-0.012	-6.17
AM peak travel cost ($)	- 0.001	fixed	-0.001	fixed	-0.001	fixed	-0.001	fixed	-0.001	fixed

Table A.23 Number of stops in the work-to-home (Model WSCH2) and home-to-work (Model WSCH3) commutes

Explanatory Variables	Work-to-home commute (Model WSCH2)		Home-to-work commute (Model WSCH3)	
	Param.	t-stat	Param.	t-stat
Individual- and household-level characteristics				
Female	0.220	3.46	--	--
Student	-0.308	-2.59	--	--
Employed	--	--	0.360	2.74
High work flexibility	-0.185	-2.33	--	--
Person's income ($1000)	0.002	2.09	--	--
Household-level activity participation				
Number of school going children	-0.139	-3.14	0.116	2.36
Number of non-school going children	--	--	0.120	1.79
Individual activity participation				
Work-related activities	0.620	6.66	0.440	4.19
Shopping	0.771	9.05	--	--
Household or personal business	0.611	8.12	0.188	2.07
Social or recreational activities	0.363	4.73	--	--
Other serve-passenger activities	0.773	10.48	1.271	15.60
Shopping and social or recreat. activ.	-0.326	-1.97	--	--
Household or pers. bus. and eating out	0.396	4.15	0.365	3.26
Work and commute				
Work start time	--	--	0.002	8.43
Work end time	-0.002	-6.90	--	--
Commute mode is driver, solo	-0.496	-5.26	-0.167	-1.52
Expected work-to-home commute time *Auto mode	0.007	3.24	--	--
Threshold parameters				
0 and 1 stop	-0.748	-2.96	2.396	12.91
1 and 2 stops	0.354	1.40	3.525	17.74

Table A.24 Number of after-work, work-based, and before-work tours (Models WSCH4, WSCH5, and WSCH6)

Explanatory Variables	Number of after-work tours (Model WSCH4)		Number of work-based tours (Model WSCH5)		Number of before-work tours (Model WSCH6)	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Person and household level characteristics						
Age	-0.011	-3.79	--	--	--	--
Female	-0.228	-3.33	--	--	--	--
Mother	--	--	--	--	0.587	2.16
Father	--	--	--	--	0.867	2.95
Licensed	--	--	0.567	2.34	--	--
Employed	0.472	3.77	-0.347	-2.34	--	--
High work flexibility	--	--	0.259	3.19	--	--
Single-person household	-0.222	-2.46	--	--	--	--
Household level activity participation						
Number of school-going children	--	--	-0.127	-2.49	0.243	2.80
Number of workers in household	--	--	--	--	-0.203	-1.87
Number of non-workers in household	--	--	--	--	-0.739	-2.87
Individual activity participation						
Work related	0.270	2.52	1.310	12.74	0.398	2.00
Drops-off children at school	--	--	0.447	2.31	0.869	3.30
Picks-up children from school	0.535	2.34	0.457	1.73	--	--
Shopping	0.977	11.47	0.288	3.41	--	--
Household/personal business	0.772	10.39	0.554	7.60	--	--
Social/recreation activities	1.423	17.64	--	--	0.293	1.94
Eat-out activities	0.396	5.55	1.279	18.02	-0.293	-1.79
Other serve passenger	0.623	6.74	0.271	2.88	0.682	3.98

Table A.24 (cont.) Number of after-work, work-based, and before-work tours (Models WSCH4, WSCH5, and WSCH6)

Explanatory Variables	Number of after-work tours (Model WSCH4)		Number of work-based tours (Model WSCH5)		Number of before-work tours (Model WSCH6)	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Pattern-level attributes						
Available time in this period	0.006	17.91	0.005	14.78	0.009	13.80
Number of work-to-home commute stops	-0.507	-11.57	-0.289	-6.80	-0.142	-1.62
Number of home-to-work commute stops	-0.436	-6.51	-0.167	-2.48	-0.452	-3.04
Commute mode is driver, solo	--	--	0.286	2.74	--	--
Threshold parameters						
0 and 1 tour	4.212	16.11	4.198	14.84	4.550	13.79
1 and 2 tours	6.467	21.94	6.380	20.60	--	--

Table A.25 After-work tour mode (Model WSCH7)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	1.152	3.99	0.983	3.09	--	--	--	--
Person and household level characteristics								
Female	--	--	-0.823	-4.89	--	--	--	--
Student	--	--	-1.356	-4.63	--	--	--	--
Income	--	--	--	--	--	--	-0.024	-2.05
Number of vehicles	--	--	--	--	--	--	-0.777	-4.32
Number of children	--	--	0.174	2.02	--	--	--	--
Single person household	--	--	-1.841	-5.89	-1.406	-4.19	--	--
Individual activity participation								
Household/personal business	--	--	--	--	--	--	-1.186	-2.69
Eat out	--	--	0.709	3.90	1.143	5.31	--	--
Social/Recreation	-0.588	-2.80	-0.426	-1.89	--	--	--	--
Other serve passenger	--	--	0.290	1.51	--	--	--	--
Num AW tours	--	--	--	--	-0.466	-2.36	--	--

Table A.26 Before-work tour mode (Model WSCH7)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	2.847	5.22	1.309	2.05	0.182	0.30	--	--
Individual activity participation								
Other serve passenger	--	--	2.305	4.30	--	--	--	--

Table A.27 Work-based tour mode (Model WSCH7)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	3.477	13.85	1.038	5.18	1.120	3.35	--	--
Individual activity participation								
Work-related activities	--	--	--	--	--	--	-0.931	-1.74
Shopping activities	--	--	-0.800	-2.44	--	--	--	--
Eat out activities	-1.875	-8.47	--	--	--	--	--	--
Household pers. bus. activities	--	--	-0.507	-2.79	-0.507	-2.79	--	--
Pattern-level characteristics								
Number of home-work stops	--	--	--	--	--	--	-0.675	-1.75
Number of work based tours	--	--	--	--	-0.367	-1.40	--	--

Table A.28 Number of stops in a tour (Model WSCH8)

Explanatory Variables	After-work tours		Before-work tours		Work-based tours	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Individual- and household-level characteristics						
Employed	--	--	--	--	-1.188	-4.85
Single-person household	0.313	2.40	--	--	--	--
Individual activity participation decisions						
Work-related activities	--	--	--	--	1.183	8.11
Shopping	0.967	8.50	0.940	2.73	--	--
Household or personal business	1.105	10.05	--	--	0.993	7.75
Social or recreation activities	0.747	6.99	--	--	0.494	3.78
Eat-out	0.846	8.27	0.532	1.61	0.465	3.63
Other serve-passenger activities	1.240	9.85	--	--	--	--
Drops off child at school	--	--	--	--	0.532	2.05
Pattern-level attributes						
Number of work-based tours	--	--	--	--	-0.198	-1.54
Number of after-work tours	-0.696	-6.06	--	--	--	--
Number of work-to-home commute stops	-0.245	-3.57	-0.988	-2.15	-0.168	-2.22
Number of home-to-work commute stops	-0.687	-5.56	--	--	--	--
Available time	0.003	6.38	0.001	1.59	--	--
Tour-level attributes						
Tour mode is non motorized	-0.526	-1.53	--	--	-0.805	-3.07
Threshold parameters						
1 and 2 stops	3.024	9.61	1.567	3.08	1.935	4.94
2 and 3 stops	4.106	12.45	2.600	4.60	2.741	6.87
3 and 4 stops	4.829	13.91	--	--	3.403	8.24
4 and 5 stops	5.354	14.52	--	--	3.839	8.86

Table A.29 Home or work stay duration before the tour (Model WSCH9)

Explanatory variables	Before-work tours		Work-based Tours		After-work tours	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Household- and individual-level characteristics						
Number of children in the household	-0.162	-2.06	--	--	--	--
Male	--	--	-0.125	-3.24	--	--
Employment type: public and military	--	--	-0.178	-2.06	--	--
Individual activity participation decisions						
Work-related activities	--	--	-0.356	-6.54	--	--
Shopping	0.206	2.62	--	--	--	--
Household or personal business	--	--	--	--	-0.217	-2.84
Social or recreational activities	--	--	0.094	2.01	--	--
Eating out	--	--	0.109	2.58	--	--
Pattern-level attributes						
Number of tours	--	--	-0.231	-4.48	--	--
One tour	--	--	--	--	0.297	3.46
Number of stops in WH commute fewer than two	--	--	--	--	-0.298	-2.15
Tour-level attributes						
Available time for the tour	0.002	10.04	0.002	9.81	0.003	8.17
Tour mode						
Driver, solo	0.166	2.20	--	--	0.143	1.98
Passenger	--	--	-0.146	-2.68	--	--
First tour in this period	--	--	0.284	3.09	--	--
Number of stops in this tour	--	--	-0.104	-4.27	-0.111	-2.50
Constant	4.616	42.51	4.411	34.35	2.709	11.40

Table A.30 Activity type at a stop (Model WSCH10)

Explanatory variables	Work related		Shopping		HH/personal business	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Number of episodes of this type already undertaken	-0.889	-17.35	-0.889	-17.35	-0.889	-17.35
Tour-level attributes						
Number of stops in the tour/commute	--	--	-0.802	-6.24	-0.142	-1.88
Tour mode						
Driver, solo	--	--	0.563	21.12	--	--
Driver, with passenger	--	--	0.791	2.53	--	--
Passenger	--	--	--	--	--	--
Stop-level attributes						
Stop is in						
Home to work commute	--	--	-1.569	-5.12	--	--
Work-based tour 1	--	--	-1.761	-5.71	-0.728	-4.46
Work-based tour 2	--	--	-2.907	-2.69	-0.870	-2.06
After-work tour 1	--	--	1.157	3.10	1.332	3.92
After work tour 2	--	--	--	--	--	--
Position of stop in tour/commute						
Second stop	--	--	1.160	4.84	0.327	1.87
Third stop	--	--	1.476	3.12	0.574	1.52
Constant	--	--	0.608	1.61	0.359	1.79

Table A.30 (cont.) Activity type at a stop (Model WSCH10)

Explanatory variables	Social/recreational		Eat out		Other serve pass.	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Number of episodes of this type already undertaken	-0.469	-15.64	-0.469	-15.64	-0.469	-15.64
Tour-level attributes						
Number of stops in the tour/commute	-0.344	-3.25	--	--	--	--
Tour mode	-0.291	-7.36	-0.246	-6.30	-0.250	-6.42
Driver, solo						
Driver, with passenger	--	--	-0.349	-2.73	--	--
Passenger	-3.515	-2.90	-2.913	-2.47	--	--
Stop-level attributes	--	--	--	--	--	--
Tour/commute of stop						
Home to work commute	--	--	0.665	5.32	-0.510	-3.04
Work-based tour 1	--	--	--	--	-0.469	-2.25
Work-based tour 2	--	--	--	--	--	--
After-work tour 1	--	--	--	--	--	--
After work tour 2	0.355	2.21	-0.265	-1.47	0.747	4.23
Position of stop in tour/commute						
Second stop	0.230	1.08	0.358	1.87	--	--
Third stop	0.813	1.87	0.405	0.99	0.541	1.36
Constant	0.061	0.25	0.529	2.34	-0.012	-0.04

Table A.31 Activity duration at a stop (Model WSCH11)

Explanatory variables	Before-work tour stops		Home-to-work commute stops		Work-based tour stops		Work-to-home commute stops		After-work tour stops	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Pattern-level attributes										
One tour in this period	--	--	--	--	0.445	4.52	--	--	0.384	5.08
Tour-level attributes										
Tour mode										
Driver, solo	0.618	1.83	--	--	--	--	--	--	--	--
Driver with passenger	0.618	1.83	0.331	1.99	0.268	2.58	--	--	--	--
Passenger	--	--	--	--	0.392	3.55	--	--	0.202	2.50
One stop in the tour	--	--	0.487	3.96	0.671	8.25	0.508	6.28	0.335	3.94
Stop-level attributes										
Available time	0.004	4.06	0.009	16.03	0.002	5.04	0.006	18.61	0.002	6.35
First stop in the commute	--	--	-0.373	-2.70	--	--	-0.210	-2.52	0.275	2.10
Second stop in the commute	--	--	--	--	--	--	--	--	0.290	2.26
Activity type at destination										
Work-related activ.	--	--	--	--	1.038	9.60	--	--	--	--
Grocery shopping	--	--	--	--	--	--	-0.375	-3.70	-1.281	-13.05
Eating out	--	--	--	--	0.605	6.64	--	--	-0.674	-6.99
Household or pers. bus.	--	--	-0.533	-4.22	--	--	-0.577	-7.03	-1.136	-13.19
Other serve-pass. activ.	-2.712	-10.69	-1.843	-15.96	--	--	-1.830	-18.9	-2.996	-27.04
Constant	2.087	5.52	2.100	16.16	1.219	7.75	2.738	31.40	2.705	14.83

Table A.32 Travel time to a stop (Model WSCH12)

Explanatory variables	Before-work tour stops		Home-to-work commute stops		Work-based tour stops		Work-to-home commute stops		After-work tour stops	
	Param.	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**
Pattern-level attributes										
One tour in this period	--	--	--	--	0.239	3.39	--	--	0.231	3.17
Tour-level attributes										
Tour mode										
Driver, solo	--	--	--	--	--	--	--	--	0.349	2.50
Driver with passenger	0.518	2.87	--	--	--	--	--	--	0.494	3.49
Passenger	--	--	--	--	--	--	--	--	0.549	3.79
Walk or bike	--	--	--	--	-0.568	-6.22	--	--	--	--
Stop in Tour 1 of this period									-0.201	-2.12
One stop in this commute	--	--	0.170	1.63	--	--	0.100	1.62	--	--
Stop-level attributes										
Available time	--	--	0.004	5.69	--	--	0.002	6.53	0.001	4.12
First stop in the commute	--	--	--	--	--	--	0.424	6.87	--	--
Second stop in the commute	--	--	-0.211	-1.69	-0.318	-4.46	--	--	--	--
Activity type at destination										
Work-related activities	--	--	--	--	0.505	6.54	--	--	--	--
Grocery shopping	--	--	--	--	--	--	--	--	-0.330	-4.65
Eating out	--	--	--	--	-0.188	-3.16	--	--	-0.158	-2.26
Household or pers. bus.	0.252	1.38	--	--	--	--	-0.092	-1.74	-0.220	-3.51
Other serve-pass. activ.	--	--	-0.252	-3.09	--	--	0.112	1.81	-0.137	-1.69
Constant	2.128	16.49	2.228	19.81	2.143	28.00	2.241	44.53	1.777	10.75

Table A.33 Location of a stop (Model WSCH13)

Explanatory variables	**Param.**	**t-stat**
Impedance measures		
Auto IVTT at start of trip	-0.250	-20.20
Auto IVTT at start of trip * Walk mode	-0.685	-6.28
Distance to the ultimate destination	-0.168	-13.22
Distance to the ultimate destination * shopping	-0.163	-4.00
Destination zone adjacent to the origin zone	0.402	4.37
Destination zone same as the origin zone	1.208	10.91
Attraction variables		
Destination zone is the CBD	-1.259	-3.99
LN (service + retail employment) at destination zone	0.254	6.68
LN (service + retail employment) at destination zone * Work-related activities	0.202	1.87
LN (service + retail employment) at destination zone * Household or personal business	0.158	2.58
LN (service + retail employment) at destination zone * Eating out	0.226	3.48
LN (population) at destination zone * Other serve-passenger activities	0.228	4.60

A.3 Non-worker Scheduling Model System

Table A.34 Number of independent tours (Model NWSCH1)

Explanatory variables	**Param.**	**t-stat**
Personal and household characteristics		
Female	-0.146	-2.28
Licensed	0.574	3.76
Student	0.324	2.18
Single-person household	-0.313	-3.85
Single-parent household	-0.296	-1.85
Household-level activity participation decisions		
Number of school going children	0.215	3.99
Individual activity participation decisions		
Work-related activities	0.335	3.89
Shopping	0.832	7.74
Household or personal business	0.822	9.59
Social or recreational activities	1.025	13.04
Eating out	0.634	7.17
Other serve-passenger activities	0.880	10.77
Shopping and household or personal bus. activities	-0.323	-2.44
Shopping and eating out activities	-0.395	-2.97
Thresholds		
1 and 2 tours	2.015	11.48
2 and 3 tours	3.297	17.87
3 and 4 tours	4.103	21.14

Table A.35 Decision to undertake an independent tour before a pick-up or joint discretionary tour (Model NWSCH2)

Explanatory variables	Param.	t-stat
Available time before pick up or joint discretionary tour	0.012	4.60
Individual activity participation decisions		
Drops off children	2.623	2.69
Picks up children	1.810	2.06
Shopping	1.641	2.30
Household or personal business	1.345	2.05
Constant	-9.611	-4.33

Table A.36 Decision to undertake an independent tour after a pick-up or joint discretionary tour (Model NWSCH3)

Explanatory variables	Param.	t-stat
Available time after the pick-up or joint discretionary tour	0.006	3.81
Constant	-4.488	-4.07

Table A.37 Tour mode (Model NWSCH4)

Explanatory variables	Driver, solo		Driver with passenger		Passenger		Walk or Bike	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	1.470	4.99	-0.438	-1.37	-0.066	-0.25	--	--
Personal and household level characteristics								
Female	--	--	1.224	8.77	--	--	--	--
Personal vehicle availability	1.476	4.76	1.476	4.76	0.819	3.01	--	--
Student	-0.616	-2.71	--	--	--	--	--	--
Household-level activity participation decisions								
Number of school-going children	--	--	0.496	6.47	--	--	--	--
Number of non– school-going children	--	--	0.753	8.76	--	--	--	--
Number of workers	--	--	0.211	2.69	--	--	--	--
Number of non workers	--	--	0.378	5.25	--	--	--	--
Individual activity participation decisions								
Household/personal business	0.942	3.55	0.828	2.99	0.556	2.05	--	--
Eat out	0.681	1.81	1.196	3.12	1.527	3.99	--	--
Other serve-passenger	--	--	--	--	-0.545	-3.06	--	--
Pattern-level characteristics								
Two or more tours	--	--	--	--	-0.329	-2.42	--	--

Table A.38 Number of stops in a tour (Model NWSCH5)

Explanatory variables	Param.	t-stat
Individual- and household-level characteristics		
Age	-0.005	-2.64
Father	0.329	2.30
Employed	0.169	2.06
Student	-0.343	-2.28
Household income	0.001	1.85
Household-level activity participation decisions		
Number of workers	-0.142	-3.30
Number of non-workers	-0.138	-2.60
Individual activity participation decisions		
Shopping	0.469	4.63
Household or personal business	0.960	11.09
Social or recreational activities	0.555	10.19
Eat-out	1.182	11.63
Other serve-passenger activities	0.645	9.85
Shopping and household or personal business	0.279	2.47
Shopping and eating out	-0.240	-2.28
Household or personal business and eating out	-0.506	-4.45
Pattern-level attributes		
Available time	0.001	5.44
Total number of tours		
Two	-0.576	-8.31
Three	-0.981	-10.22
Four	-1.508	-11.74
Tour-level attributes		
Second tour	0.427	2.65
Third tour	0.470	2.11
Fourth tour	0.559	1.82
Tour mode is walk or bike	-1.231	-4.68
Thresholds		
1 and 2 stops	2.695	6.79
2 and 3 stops	3.427	8.60
3 and 4 stops	4.045	10.09
4 and 5 stops	4.468	11.09

Table A.39 Number of stops in a tour following a pick-up or drop-off stop (Model NWSCH6)

Explanatory variables	Param.	t-stat
Individual-level characteristics		
Employed	0.600	1.97
Household-level activity participation decisions		
Presence of non–school-going children	-0.753	-2.42
Individual activity participation decisions		
Work-related activities	0.784	1.69
Household or personal business	0.666	2.37
Tour-level characteristics		
Drops-off children in tour	-1.294	-2.38
Tour start time	-0.003	-2.53
Threshold		
0 and 1 stop	-1.539	-1.69

Table A.40 Home-stay duration before a tour (Model NWSCH7)

Explanatory variables	Tour 1		Tour 2		Tour 3		Tour 4	
	Param.	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**	**Param.**	**t-stat**
Constant	5.932	138.99	3.133	13.19	2.102	6.92	2.215	3.89
Individual- and household-level characteristics								
Male	-0.089	-4.37	--	--	--	--	--	--
Employed	0.074	2.91	--	--	--	--	--	--
Student	-0.179	-3.52	--	--	--	--	--	--
Number of children in household	--	--	--	--	0.301	2.26	--	--
Couple household	--	--	-0.194	-2.64	--	--	-0.393	-1.66
Individual activity participation decisions								
Work-related activities	-0.490	-15.23	-0.364	-3.65	--	--	--	--
Shopping	0.066	3.03	--	--	--	--	--	--
Household or personal business	--	--	-0.175	-2.11	--	--	--	--
Eating out	0.069	3.00	-0.165	-2.13	--	--	--	--
Social or recreational activities	--	--	0.134	1.77	--	--	--	--
Other serve-passenger activities	-0.136	-4.80	--	--	--	--	--	--
Pattern-level attributes								
One Tour	0.282	8.92	--	--	--	--	--	--
Two Tours	0.122	3.82	--	--	--	--	--	--
Three or more tours	--	--	-0.825	-9.56	--	--	--	--
Tour-level attributes								
Available time for the tour	--	--	0.002	10.52	0.003	8.65	0.003	3.77
Tour mode	--	--	--	--	--	--	--	--
Driver, solo	-0.041	-1.92	-0.243	-3.20	-0.346	-2.10	--	--
Passenger	--	--	--	--	-0.346	-1.83	-0.447	-1.93
One stop in tour	0.058	2.74	0.225	2.81	--	--	--	--

Table A.41 Activity type at a stop (Model NWSCH8)

Explanatory variables	Work related		Shopping		HH/personal business	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Number of episodes of this type already undertaken	-0.469	-15.64	-0.469	-15.64	-0.469	-15.64
Tour-level attributes						
First tour	--	--	--	--	--	--
Number of stops in tour	--	--	-0.461	-8.55	-0.112	-3.52
Tour mode						
Driver, solo	--	--	0.258	2.05	0.287	2.92
Undertakes pick-up in this tour	--	--	-2.265	-2.30	0.508	1.21
Undertakes drop-off in this tour	--	--	--	--	-1.651	-2.55
Stop-level attributes						
Second stop	--	--	0.622	4.37	--	--
Third stop	--	--	1.075	5.72	--	--
Fourth stop	--	--	1.132	4.47	--	--
Fifth stop	--	--	1.762	5.75	--	--
Constant	--	--	-0.026	-0.13	-0.036	-0.22

Table A.41 (cont.) Activity type at a stop (Model NWSCH8)

Explanatory variables	Social/Recreational		Eat out		Other serve pass.	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Number of episodes of this type already undertaken	-0.469	-15.64	-0.469	-15.64	-0.469	-15.64
Tour-level attributes						
First tour	-0.344	-3.25	--	--	--	--
Number of stops in tour	-0.291	-7.36	-0.246	-6.30	-0.250	-6.42
Tour mode						
Driver, solo	--	--	-0.349	-2.73	--	--
Undertakes pick-up in this tour	-3.515	-2.90	-2.913	-2.47	--	--
Undertakes drop-off in this tour	--	--	--	--	--	--
Stop-level attributes						
Second stop	--	--	0.665	5.32	-0.510	-3.04
Third stop	--	--	--	--	-0.469	-2.25
Fourth stop	--	--	--	--	--	--
Fifth stop	--	--	--	--	--	--
Constant	0.355	2.21	-0.265	-1.47	0.747	4.23

Table A.42 Activity duration at a stop (Model NWSCH9)

Explanatory variables	Stops in Tour 1		Stops in Tour 2		Stops in Tour 3		Stops in Tour 4	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	2.440	22.77	2.626	18.13	2.708	6.55	3.733	18.12
Tour-level attributes								
Number of stops in tour								
One	0.717	12.26	0.713	8.09	0.456	3.11	--	--
Two	0.312	5.03	0.325	3.33	0.288	1.75	--	--
Three	0.308	5.00	0.298	2.82	--	--	--	--
Tour mode								
Driver, solo	-0.424	-6.92	-0.147	-2.17	1.104	2.82	-0.509	-2.29
Driver with passenger	-0.320	-4.36	--	--	1.205	3.00	--	--
Passenger	--	--	--	--	1.540	3.72	--	--
Stop-level attributes								
Available time for activity and travel	0.001	8.68	0.000	1.89	--	--	--	--
Destination activity type								
Work-related activities	1.458	22.53	0.706	5.57	--	--	--	--
Household or personal business	--	--	--	--	-1.155	-7.47	--	--
Eating out	0.309	4.08	0.420	3.90	-0.589	-3.35	--	--
Social or recreational activities	1.053	16.26	1.037	12.13	--	--	0.942	3.57
Other serve-passenger activities	-2.226	-26.90	-1.830	-18.31	-2.968	-18.12	-2.227	-8.42

Table A.43 Travel time to a stop (Model NWSCH10)

Explanatory variables	Stops in Tour 1		Stops in Tour 2		Stops in Tour 3		Stops in Tour 4	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	2.699	59.89	2.284	36.22	1.646	5.21	1.570	5.63
Tour-level attributes								
Two or more stops in tour	--	--	--	--	--	--	0.469	2.21
Tour mode								
Driver, solo	-0.316	-7.00	-0.120	-2.48	0.706	2.22	--	--
Driver with passenger	-0.243	-4.56	--	--	0.706	2.22	--	--
Passenger	--	--	--	--	0.706	2.22	--	--
Stop-level attributes								
First Stop in Tour	0.207	6.36	--	--	--	--	0.484	1.95
Destination activity type								
Work-related activities	0.532	11.61	0.588	5.86	--	--	--	--
Shopping	-0.195	-4.22	--	--	-0.300	-2.16	--	--
Household or personal business	--	--	0.233	3.37	--	--	--	--
Eating out	--	--	0.182	2.03	--	--	--	--
Social or recreational activities	--	--	0.242	3.22	--	--	0.466	2.71
Other serve-passenger activities	--	--	0.270	3.18	--	--	--	--

Table A.44 Location of a stop (Model NWSCH11)

Explanatory variables	**Param.**	**t-stat**
Impedance measures		
Cost	-0.431	-1.84
Auto IVTT at start of trip	-0.229	-12.89
Auto IVTT at start of trip * walk mode	-0.599	-4.62
Auto IVTT at start of trip * household/personal business	0.034	1.82
Distance to the ultimate destination	-0.143	-7.64
Distance to the ultimate destination * work related	0.163	4.43
Distance to the ultimate destination * shopping	-0.162	-4.46
Distance to the ultimate destination * social/recreational	0.061	1.86
Destination zone adjacent to the origin zone	0.442	4.99
Destination zone same as the origin zone	1.320	12.47
Attraction variables		
Destination zone is the CBD	-1.346	-3.23
LN (service + retail employment) at destination zone	0.2885	7.20
LN (service + retail employment) at destination zone * Shopping	0.268	3.79
LN (service + retail employment) at destination zone * HH/personal business	0.249	4.26
LN (service + retail employment) at destination zone * Eat out	0.384	4.43
LN (population) at destination zone * Other serve passenger	0.180	3.20

A.4 Joint Discretionary Tour Scheduling Model System

Table A.45 Departure time for the tour (Model JNTSCH1)

Explanatory variables	Param.	t-stat
Constant	6.510	124.15
Adult's arrival time at home from work(x 10^{-3})	0.260	2.77
Child's arrival time at home from school (x 10^{-3})	0.270	2.70

Table A.46 Activity duration at the stop (Model JNTSCH2)

Explanatory variables	Param.	t-stat
Constant	5.233	12.76
Departure time for the tour	-0.001	-2.69
Adult is a worker	0.707	3.22

Table A.47 Travel time to the stop (Model JNTSCH3)

Explanatory variables	Param.	t-stat
Constant	2.337	18.86
Adult is a worker	0.389	1.91

Table A.48 Travel time to the stop (Model JNTSCH3)

Explanatory variables	Param.	t-stat
Auto in-vehicle travel time at trip start time	-0.267	-4.12
Destination zone same as origin zone	2.420	4.48
Destination zone adjacent to origin zone	1.239	2.60
LN (retail + service employment) at destination zone	0.437	2.98
LN (population) at destination zone	0.244	2.03

A.5 The children scheduling model system

Table A.49 School-to-home (Model CSCH1) and home-to-school (Model CSCH2) commute durations

Explanatory variables	School-to-home duration (Model CSCH1)		Home-to-school duration (Model CSCH2)	
	Param.	t-stat	Param.	t-stat
Constant	2.432	37.62	2.296	38.47
Travel mode from or to school				
School bus	0.635	8.28	0.942	13.07
Walk or bike	0.309	3.90	0.377	5.05
School and home zones are the same	-0.277	-2.90	-0.516	-5.84
School and home zones are adjacent	-0.169	-2.16	-0.380	-5.32
Distance between school and home zone	0.049	6.31	0.038	5.46

Table A.50 Mode for the independent discretionary tour (Model CSCH3)

Explanatory variables	Drive by other		Walk or bike	
	Param.	t-stat	Param.	t-stat
Constant	--	--	0.130	0.37
Male	--	--	0.830	2.44
Goes to school	--	--	-1.140	-3.20

Table A.51 Departure time for the independent discretionary tour (Model CSCH4)

Explanatory variables	Param.	t-stat
Constant	6.179	66.54
Arrival time at home after school (x 10^{-3})	0.100	1.54
Age	0.026	2.71
Male	0.078	1.19

Table A.52 Activity duration at the independent discretionary stop (Model CSCH5)

Explanatory variables	Param.	t-stat
Constant	5.046	19.95
Start time of the tour	-0.001	-2.87

Table A.53 Travel time to the independent discretionary stop (Model CSCH6)

Explanatory variables	Param.	t-stat
Constant	2.441	13.13
Travel mode is walk or bike	-0.270	-1.51
Child goes to school	-0.249	-1.33

Table A.54 Location of the independent discretionary stop (Model CSCH7)

Explanatory variables	Param.	t-stat
Auto in-vehicle travel time at trip start time	-0.159	-3.03
Auto in-vehicle travel time at trip start time * Walk or bike mode	-0.332	-3.32
Destination zone same as the origin	2.952	6.22
Destination zone adjacent to the origin	1.169	2.55
LN (population) of the destination zone	0.347	2.64

Appendix B: Synthetic Population Generator

B.1 Mathematical details of the proposed algorithm

The algorithm includes a number of major steps: (1) determine the household-level multi-way distribution, (2) determine the individual-level multi-way distribution, (3) initialize the household- and individual-level counts, (4) compute selection probabilities, (5) select a sample household, (6) check household desirability, (7) add the selected households to the target area, and (8) update the household- and individual-level counts. We discuss each of these steps is in turn below. An example is also provided in the Appendix to demonstrate the application of our proposed algorithm.

B.1.1 Determine Household-Level Multi-Way Distribution

Given the aggregate (e.g.,. U.S. Census Summary Tables) and disaggregate (*e.g.* U.S. PUMS data) input data, this step creates the full multi-way distribution across all the household-level control variables using the IPFP-based recursive procedure outlined in Figure 1. We denote each cell in the resulting household-level multi-way distribution by HH[v_1, v_2, …, v_k, …], where the index v_k is the value of the k^{th} household-level controlled variable, $v_k = 1, \ldots, M_k$. HH[v_1, v_2, …, v_k, …] gives the expected number of households with attribute values of (v_1, v_2, …, v_k, …) in the target area.

B.1.2 Determine Individual-Level Multi-Way Distribution

This step creates the full multi-way distribution across all the individual-level controlled attributes, also using the procedure presented in Figure 1. We denote each cell in the resulting individual-level multi-way distribution by POP[v_1, v_2, …, v_l, …], where the index v_l denotes the value of the l^{th} individual-level variable, $v_l = 1, \ldots, N_l$. POP[v_1, v_2, …, v_l, …] thus gives the expected number of individuals with attribute values of (v_1, v_2, …, v_l, …) in the target area. It should be noted that the cell values in both HH and POP will be used as they are without being rounded to integer values.

B.1.3 Initialize Household- and Person-Level Counts

Two multi-way tables, *HHI* and, *POPI* are used to keep track of the numbers of households and individuals belonging to each demographic group that have been selected into the target area during the iterative process. At the start of the process, the cell values in the two tables are initialized to zero to reflect the fact that no households and individuals have been created for the target area. During subsequent iterations, these cell values will be updated as households and individuals are selected into the target area.

B.1.4 Compute Household Selection Probabilities

Given the target distribution (HH) and the current distribution (*HHI*) of households already selected into the target area, each PUMS sample household in the corresponding seed area is assigned with a probability of being selected into the target area in the current iteration. The probability of household *i* being selected is computed by

$$P_i = \frac{w_i}{\sum_j w_j \cdot Y^j_{v_1,v_2,\cdots,v_k,\cdots}} \cdot \frac{\mathrm{HH}[v_1,v_2,\cdots,v_k,\cdots] - \mathrm{HHI}[v_1,v_2,\cdots,v_k,\cdots]}{\sum_{u_1,u_2,\cdots,u_k,\cdots}\left(\mathrm{HH}[u_1,u_2,\cdots,u_k,\cdots] - \mathrm{HHI}[u_1,u_2,\cdots,u_k,\cdots]\right)} \qquad (4)$$

In the above equation, w_i is the PUMS weight associated with household *i*. The vector (v_1, v_2, …, v_k, …) reflects the characteristics of household *i*. $Y^j_{v_1,v_2,\cdots,v_k,\cdots}$ takes a value of 1 if the j^{th} household is characterized by (v_1, v_2, …, v_k, …) (*i.e.*, the same as the i^{th} household), and a value of 0 otherwise. The equation implies that the selection probability of a sample household decreases as more households from the same demographic group are selected into the target area.

B.1.5 Randomly Select a Household

Based on the probabilities computed in the previous step, a household is randomly drawn from the pool of sample households to be considered for "cloning" and added to the population for the target area.

B.1.6 Check Household Desirability

Given a randomly selected household characterized by (v_1,v_2,…, v_k,…), we will add a copy of this household into the population for the target area if the following conditions hold:

1. The number of such households already selected into the target area (as given by $\text{HHI}[v_1, v_2, \cdots, v_k, \cdots]$) is lower than a pre-specified maximum threshold. Ideally, this threshold should be set to the target value given by $\text{HH}[v_1, v_2, \cdots, v_k, \cdots]$ so that the number of households characterized by ($v_1, v_2, \ldots, v_k, \ldots$) is never higher than desired. However, such a condition may be undesirable for at least two reasons. First, when incorrect zero cell values are found for certain demographic groups, the target total number of households in the area would never be met unless households of other demographic groups are allowed to be over-selected. Second, since the dual goals of satisfying the household-level target distribution and satisfying the individual-level target distribution may be conflicting in nature, fitting the synthetic population perfectly to the household-level target distribution may prevent the individual-level distribution from being satisfied to any acceptable extent. Therefore, in the proposed algorithm, we allow the threshold values to exceed their respective target values by a user-specified percentage, hereafter referred to as the percentage deviation from target size (PDTS).
2. For each person in the household, the number of such individuals already selected into the target area (as given by $\text{POPI}[v_1, v_2, \cdots, v_l, \cdots]$) is lower than a pre-specified maximum threshold. The threshold values are specified as (1+PDTS) of the corresponding target cell value $\text{POP}[v_1, v_2, \cdots, v_l, \cdots]$.

If any of the above conditions fails, then the household is removed from the consideration set so that it will never be selected again. The selection probabilities of the households remaining in the consideration set are then updated before the next household is randomly selected.

B.1.7 Add Household

If the selected household satisfies the conditions described in Section 0, then the household is added to the pool of the synthetic population for the target area. As part of this step, the household sample weight is decreased by one to implement the 'random draw without replacement' strategy.

B.1.8 Update Household- and Individual-Level Counts

The cell values in the count tables HHI$[v_1, v_2, \cdots, v_k, \cdots]$ and POPI$[v_1, v_2, \cdots, v_l, \cdots]$ that correspond to the selected household and its individuals are incremented accordingly to reflect the reduced desirability of such a household and individuals in subsequent iterations.

B.2 An example application

For the purpose of illustrating the population synthesis algorithm presented, we consider a target area of 20 households and 49 people. Household type (HH_FAM) and household size (HH_SIZE) are selected as household-level control variables, while gender (P_GENDER) and race (P_RACE) are selected as individual-level controlled variables. The PUMS sample records for the corresponding seed area are listed in Figure B-1. Based on the sample records and the marginal distributions of the controlled variables, we first determine the complete household- and individual-level multi-way distribution tables, denoted as HH[HH_FAM, HH_SIZE] and POP[P_GENDER, P_RACE] respectively (this corresponds to the steps described in Section B.1.1 and Section B.1.2). Both tables are shown in Figure B-2. The next step is to set up and initialize the household- and individual-level count tables, denoted as HHI[HH_FAM, HH_SIZE] and POPI[P_GENDER, P_RACE] respectively (this step corresponds to Section B.1.3). As shown in Figure B-3, both tables are filled with values of 0 to reflect the fact that no households have yet been selected into the target area.

A selection probability is then calculated for each sample household based on equation (4) (this step corresponds to Section B.1.4). These probability values and the corresponding cumulative probabilities are shown in Figure B-4. Next, a household is selected based on a random number draw (this step corresponds to Section B.1.5). With a random value of 0.635, the household with SERIALNO = 13687 is selected. Since the household satisfies both the household level selection condition (HHI[1,2]<HH[1,2]) and the individual-level selection condition (POPI[0, 0]<POP[0,0] and POPI[1, 0]<POP[1,0]), the household is now added to the target area (this step corresponds to Section B.1.6 and Section B.1.7). The current iteration completes with updating the count tables (see Figure B-5; this step corresponds to Section B.1.8).

(a) PUMS Housing Unit Record

SERIALNO	HWEIGHT	PERSONS	HHT	Other attributes
2599	6	2	Family: married couple	…
2797	9	3	Family: married couple	…
13687	18	4	Family: married couple	…
21197	18	1	Nonfamily: female living alone	…
15458	6	1	Nonfamily: male living alone	…
24526	6	2	Family: married couple	…
39951	15	2	Family: female householder	…

(b) PUMS Person Record

SERIALNO	PNUM	SEX	RACE	Other attributes
2599	1	male	white alone	…
2599	2	female	white alone	…
2797	1	male	white alone	…
2797	2	female	Some other race alone	…
2797	3	male	Some other race alone	…
13687	1	male	white alone	…
13687	2	female	white alone	…
13687	3	male	white alone	…
13687	4	male	white alone	…
21197	1	female	Black or African American alone	…
15458	1	male	white alone	…
24526	1	male	Asian alone	…
24526	2	female	white alone	…
39951	1	male	Black or African American alone	…
39951	2	male	Black or African American alone	…

Figure B-1 Sample household and person records for the seed area.

(a) HH[H_FAM, H_SIZE]

		H_SIZE (household size)			
		0 (1 person)	1 (2 person)	2 (3 persons or more)	Total
H_FAM (whether household is a family)	0 (No)	3	0	0	3
	1 (Yes)	0	8	9	17
	Total	3	8	9	20

(b) POP[P_GENDER, P_RACE]

		P_RACE			
		0 (white alone)	1 (black alone)	2 (other)	Total
P_GENDER	0 (Male)	16.4	7.6	3	27
	1 (Female)	14.6	7.4	0	22
	Total	31	15	3	49

Figure B-2 Steps 1 and 2: determine household-level and individual-level multi-way distribution tables for the target area.

(a) HHI[H_FAM, H_SIZE]

		H_SIZE (household size)			
		0 (1 person)	1 (2 person)	2 (3 persons or more)	Total
H_FAM (whether household is a family)	0 (No)	0	0	0	0
	1 (Yes)	0	0	0	0
	Total	0	0	0	0

(b) POPI[P_GENDER, P_RACE]

		P_RACE			
		0 (white alone)	1 (black alone)	2 (other)	Total
P_GENDER	0 (Male)	0	0	0	0
	1 (Female)	0	0	0	0
	Total	0	0	0	0

Figure B-3 Step 3: initialize household-level and individual-level count tables.

SERIALNO	Probability	Cumulative Probability
2599	0.089	0.000
2797	0.150	0.239
13687	0.300	0.539
21197	0.113	0.651
15458	0.038	0.689
24526	0.089	0.778
39951	0.222	1.000

Figure B-4 Step 4: compute the household selection probabilities.

(a) HHI[H_FAM, H_SIZE]

		H_SIZE (household size)			
		0 (1 person)	1 (2 person)	2 (3 persons or more)	Total
H_FAM (whether household is a family)	0 (No)	0	0	0	0
	1 (Yes)	0	0	1	1
	Total	0	0	1	1

(b) POPI[P_GENDER, P_RACE]

		P_RACE			
		0 (white alone)	1 (black alone)	2 (other)	Total
P_GENDER	0 (Male)	3	0	0	3
	1 (Female)	1	0	0	1
	Total	4	0	0	4

Figure B-5 Step 8: update the household-level and individual-level count tables.

Appendix C: CEMSELTS

Table C-1 Drop-out rate look-up table

Male					
Age	**Black**	**Asian or Pacific Islander**	**Hispanic**	**Native American**	**White**
13	0.001	0.001	0.003	0.002	0.001
14	0.004	0.002	0.005	0.002	0.001
15	0.018	0.005	0.020	0.01	0.005
16	0.019	0.006	0.021	0.011	0.006
17	0.023	0.007	0.022	0.018	0.008
18	0.021	0.006	0.022	0.017	0.009
Female					
Age	**Black**	**Asian or Pacific Islander**	**Hispanic**	**Native American**	**White**
13	0.002	0.001	0.003	0.004	0.001
14	0.003	0.002	0.006	0.008	0.001
15	0.014	0.004	0.018	0.008	0.005
16	0.014	0.003	0.017	0.014	0.005
17	0.013	0.005	0.018	0.011	0.006
18	0.015	0.005	0.016	0.008	0.007

Table C-2 Educational attainment table

Male						
Education Level	**White**	**Black**	**Hispanic**	**Asian or Pacific Islander**	**Native American**	**Other**
High School	.6667	.7866	.7060	.3140	.8000	.9137
Associate's	.0501	.0418	.0442	.0349	.0667	.0208
Bachelor's	.2146	.1506	.1858	.3721	.1333	.0476
Master's	.0651	.0209	.0615	.2791	.0000	.0149
Doctorate	.0036	.0000	.0025	.0000	.0000	.0030

Female						
Education Level	**White**	**Black**	**Hispanic**	**Asian or Pacific Islander**	**Native American**	**Other**
High School	.6125	.7470	.6598	.5364	.9000	.9041
Associate's	.0576	.0643	.0590	.3000	.1000	.0753
Bachelor's	.2699	.1678	.2258	.1364	.0000	.0103
Master's	.0584	.0209	.0533	.0182	.0000	.0068
Doctorate	.0016	.0000	.0021	.0091	.0000	.0034

Table C-3 DFW school look-up table excerpt

Residential TAZ	Decimal Percent	Elementary School (K–5) TAZ	Middle School (6–8) TAZ	High School (9–12) TAZ
2032	1.00	0	40123	40123
2034	0.64	41183	40123	40123
	0.13	2019	2019	0
	0.23	2067	2078	40055
2039	1.00	2046	2039	40045
2040	1.00	2046	2039	40045
2042	0.98	0	40123	40123
	0.02	2181	2373	2181
2046	1.00	2046	2039	40045
2050	0.99	2046	2039	40045
	0.01	30300	2078	2148
2053	1.00	0	40123	40123
2056	0.78	2067	2078	40055
	0.22	41183	40123	40123
2061	0.97	30300	2078	2148
	0.03	2046	2039	40045
2064	0.60	40055	2134	40055
	0.40	41183	40123	40123
2065	1.00	40055	2134	40055
2067	1.00	2067	2078	40055
2070	1.00	30300	2078	2148
2071	1.00	30300	2078	2148
2074	1.00	40055	2134	40055
2075	1.00	40055	2078	40055
2076	1.00	2067	2078	40055
2077	1.00	2067	2078	40055
2078	1.00	2067	2078	40055
2079	1.00	2080	2078	40055
2080	1.00	2080	2078	40055
2081	1.00	2082	2078	40055
2082	1.00	2082	2078	40055
2084	1.00	2082	2078	2148
2092	0.55	0	40123	40123
TAZ = 0: School lies outside of NCTCOQ area				

Table C-4 College look-up table excerpt

Associate's Degree—Male					
Zone	**White**	**Black**	**Hispanic**	**Asian**	**Other**
6821	0.000	0.000	0.000	0.000	0.000
6354	0.065	0.062	0.106	0.146	0.205
40690	0.022	0.081	0.024	0.009	0.027
3067	0.128	0.023	0.038	0.055	0.028
40497	0.007	0.006	0.004	0.002	0.002
8177	0.020	0.019	0.010	0.005	0.010
6444	0.000	0.000	0.000	0.000	0.000
6390	0.085	0.118	0.111	0.218	0.341
7159	0.073	0.089	0.093	0.070	0.024
7531	0.040	0.139	0.054	0.026	0.018
8078	0.040	0.101	0.109	0.027	0.019
6738	0.076	0.072	0.078	0.137	0.148
8660	0.006	0.004	0.005	0.001	0.002
8482	0.000	0.000	0.000	0.000	0.000
7010	0.034	0.010	0.015	0.021	0.008
16101	0.005	0.003	0.003	0.001	0.002
41072	0.012	0.003	0.006	0.001	0.003
40989	0.130	0.052	0.075	0.093	0.052
41005	0.064	0.019	0.084	0.028	0.021
10540	0.073	0.091	0.087	0.037	0.030
10727	0.075	0.086	0.074	0.106	0.046
10040	0.020	0.015	0.014	0.010	0.009
10327	0.027	0.006	0.008	0.005	0.006
9949	0.000	0.000	0.000	0.000	0.000
10218	0.000	0.000	0.000	0.000	0.000
2100	0.000	0.000	0.000	0.000	0.000
6861	0.000	0.000	0.000	0.000	0.000
2164	0.000	0.000	0.000	0.000	0.000
10262	0.000	0.000	0.000	0.000	0.000
3462	0.000	0.000	0.000	0.000	0.000
7227	0.000	0.000	0.000	0.000	0.000

Table C-5 Labor participation model

Explanatory variables	Param.	t-stat
Constant	-1.774	-12.03
Female	-0.883	-9.74
Age		
16 – 40 years	3.321	26.23
41 – 60 years	2.560	24.72
Education Level		
High School	0.764	5.09
College, associate or bachelors	1.312	9.62
Masters or PhD	1.617	10.06
Presence and age of own children		
Presence of children of age < 16 years	0.351	2.80
Female with own children under 6 years	-1.593	-9.39

Table C-6 Employment industry model

Explanatory variables	Construction and Manufacturing		Trade and Transportation		Professional Businesses		Government		Retail and Repair	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	1.60	16.91	0.64	3.21	2.13	13.16	-1.05	-2.25	2.01	22.42
Female	--	--	--	--	1.42	20.06	0.57	4.28	--	--
Black	--	--	--	--	0.23	2.16	0.50	2.73	--	--
Education Level										
High School	--	--	0.79	4.04	0.66	4.55	1.88	4.07	--	--
Associate's	0.62	3.42	1.27	4.83	0.76	6.89	2.33	4.44	--	--
Bachelor's	0.60	5.75	0.99	4.70	1.55	9.75	2.25	4.74	--	--
Master's+	--	--	0.60	2.45	1.83	10.14	1.97	3.96	-0.97	-5.45
Female w/ assoc.'s degree	--	--	--	--	0.94	3.29	--	--	--	--

Table C-7 Employment location choice model

Explanatory variables	Param.	t-stat
Auto in-vehicle travel time (IVTT)	-0.110	-26.75
LN (total employment)	0.643	25.46
Fraction of retail employment	-0.784	-6.24
Accessibility to population	-0.106	-6.29
Accessibility to retail employment	0.662	8.00
Female x IVTT	-0.012	-2.58
Graduate x IVTT	0.021	4.78
Professional businesses x IVTT	-0.029	-5.85
Home and work location zones in same county	0.197	2.77
Home and work location zones in same or adjacent zones	0.819	5.95
Zone in Dallas county	0.129	1.67
Zone in high employment (>200 jobs) category	0.347	4.94
Retail and repair x fraction of retail employment	1.693	7.86
Professional businesses x fraction of service employment	0.683	5.26
Fort-Worth CBD x IVTT	-0.039	-3.66

Table C-8 Work Duration model

Explanatory variables	Param.	t-stat
Threshold 1	-0.204	-3.21
Threshold 2	1.442	21.73
Male	0.479	13.79
Education Level		
High School degree	0.398	6.10
Associate’s degree	0.462	5.26
Bachelor’s degree	0.599	8.79
Master’s and higher	0.631	8.39
Industry		
Construction and Manufacturing	0.297	6.62
Trade and Transportation	0.211	4.11

Table C-9 Work schedule flexibility model

Explanatory variables	**Param.**	**t-stat**
Threshold 1	-0.541	-10.93
Threshold 2	0.290	5.93
Female	-0.247	-5.91
Race		
Black	-0.345	-4.38
Hispanic	-0.249	-2.24
Industry		
Government	-0.387	-4.01
Retail and Repair	0.104	2.06
Work Duration		
Work duration <20 hours per week	0.254	3.56
Work duration between 20 and 40 hours per week	-0.397	-8.80

Table C-10 Personal income model

Explanatory variables	**Param.**	**t-stat**
Age	0.033	17.15
Male	1.021	19.56
Race		
Black	-0.604	-5.40
White	0.199	2.46
Education (less than high school as base)		
High school	0.542	4.80
Attended college but no degree	1.018	9.56
Associate's degree	1.327	10.00
Bachelor's	2.014	19.34
Master's and higher	2.443	19.84
Professional Degree	1.920	11.11
Employed	0.099	1.49
Retired	-2.730	-8.25
Industry (other industry as base)		
Construction and manufacturing	0.180	1.77
Wholesale trade and transportation	0.182	1.68
Professional, personal, and financial services	-0.546	-5.92
Retail and repair	-0.792	-8.45
Variance	0.973	69.30

Table C-11 Residential tenure choice model*

Explanatory variables	**Param.**	**t-stat**
Constant	-0.672	-3.134
Relocated within a year prior to survey (1996)	-1.758	-15.586
Household annual income ($1000)	0.027	14.199
Household size	0.408	5.417
Number of employed people in the household	-0.202	-2.714
Number of children in the household	-0.417	-4.807
Caucasian household	0.331	2.672
Black household	-0.489	-3.000
Single-adult household	-2.842	-8.740
Age of the adult in the single-adult household	0.048	7.284
Household with unrelated persons	1.168	6.457
Household with elderly persons (age $\geq$ 65)	-0.672	-3.134

*Parameter estimates indicate effect of variables on the propensity to own house.

Table C-12a Housing type choice model for housing owners

Explanatory variables	Single-family detached		Single-family attached		Mobile home or trailer	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	--	--	-2.554	-5.34	-2.471	-5.38
Income	0.040	5.65	0.037	4.65	--	--
Household size	--	--	-0.411	-2.63	--	--
Single-adult household	--	--	1.002	2.97	--	--
Household with elderly persons (age ≥ 65)	--	--	--	--	-1.150	-3.10
White household	--	--	--	--	1.404	3.15
Highest education in household is bachelor’s or higher	--	--	--	--	-1.715	-3.87

Table C-12b Housing type choice model for renters

Explanatory variables	Single-family detached		Single-family attached		Apartment	
	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	-1.734	-7.88	-2.023	-12.05	--	--
Income	--	--	--	--	-0.016	-5.73
Household size	0.410	7.18	--	--	--	--
Single-adult household	--	--	--	--	0.663	4.22
Household with elderly persons (age ≥ 65)	--	--	1.116	3.42	--	--
Household with unrelated persons	-0.591	-2.87	--	--	--	--
Asian household	--	--	--	--	1.923	3.88
Black household	--	--	--	--	0.565	3.26

Table C-13 Vehicle ownership model

Explanatory variables	1		2		3		4 or more	
	Param.	t-stat	Param.	t-stat	Param.	t-stat	Param.	t-stat
Constant	-0.152	-0.60	-2.980	-8.32	-7.967	-17.63	-11.70	-21.32
Income	0.080	9.68	0.104	12.23	0.107	12.44	0.114	12.96
Number of employed adults	--	--	1.059	9.23	2.371	17.29	3.093	19.38
Number of non-working adults	--	--	0.481	4.16	1.470	10.65	2.103	12.73
Single-adult household	--	--	-2.014	-13.19	-2.014	-13.19	-2.014	-13.19
Household with children	--	--	--	--	-0.403	-3.85	-0.403	-3.85
White household	0.782	2.99	1.294	4.63	1.656	5.33	--	--
Black household	-1.274	-5.03	-1.934	-6.60	-2.777	-7.05	--	--
Own housing unit	1.002	4.62	1.868	7.73	2.796	10.28	2.796	10.28
Single-family attached or detached housing unit	0.236	1.77	--	--	--	--	--	--

Table C-14 2025 Education Attainment Module Results

Education Level	Predicted
No School	8.1
Children: Preschool through Grade 4	7.5
Children: Grades 5 through 8	5.8
Children: Grades 9 through 12	6.1
Adult: High school or less	49.3
Adult: Associate	4.0
Adult: Bachelor's	13.9
Adult: Master's	3.7
Adult: PhD	0.2

Table C-15 2025 Labor Participation Module Results

Labor Participation	Predicted
Employed	50.80
Unemployed	49.20

Table C-16 2025 Employment Industry Module Results

Employment Industry	Predicted
Construction and Manufacturing	18.4
Wholesale Trade and Transportation	14.1
Professional, Personal, and Financial	34.3
Public and Military	5.9
Retail and Repair	24.1
Other Industry	3.3

Table C-17 2025 Work Duration Module Results

Work Duration	Predicted
Hours 0-20	12.20
Hours 20-40	56.90
Hours 40+	30.88

Table C-18 2025 Work Flexibility Module Results

Work Flexibility	Predicted
Low/No Flexibility	21.0
Med Flexibility	15.3
High Flexibility	14.5
Unemployed	49.2

Table C-19 2025 Personal Income Module Results

Personal Income ($)	Predicted
No Income	30.3
0–10,000	18.5
10,000–20,000	12.1
20,000–30,000	11.0
30,000–40,000	9.2
40,000–50,000	6.8
50,000 +	12.1

Table C-20 2025 Residential Tenure Module Results

Residential tenure	Predicted
Own	66.0
Rent	34.0

Table C-21a 2025 Housing Type for Owners Module Results

Housing Type for Owners	Predicted
Single-Family Detached	93.0
Single-Family Attached	3.6
Mobile Home/Trailer	3.4
Multi-Family/Apartment/Condo	0.0

Table C-21b 2025 Housing Type for Renters Module Results

Housing Type for Renters	Predicted
Single-Family Detached	26.4
Single-Family Attached	8.5
Multi-Family/Apartment/Condo	65.0
Mobile Home/Trailer	0.0

Table C-22 2025 Household Vehicle Ownership Renters Module Results

Vehicle Ownership	Predicted
Number of vehicles = 0	8.0
Number of vehicles = 1	40.4
Number of vehicles = 2	35.8
Number of vehicles = 3	12.4
Number of vehicles = 4 or more	3.4