

**POPULATION UPDATING SYSTEM STRUCTURES AND MODELS EMBEDDED
WITHIN THE COMPREHENSIVE ECONOMETRIC MICROSIMULATOR FOR
URBAN SYSTEMS (CEMUS)**

Naveen Eluru

The University of Texas at Austin, Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712
Phone: (512) 471-4535; Fax: (512) 475-8744; Email: naveeneluru@mail.utexas.edu

Abdul Rawoof Pinjari

The University of Texas at Austin, Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712
Phone: (512) 471-4535; Fax: (512) 475-8744; Email: abdul.pinjari@mail.utexas.edu

Jessica Y. Guo

University of Wisconsin-Madison, Department of Civil and Environmental Engineering
1206 Engineering Hall, 1415 Engineering Dr, Madison, WI 53706
Phone: (608) 890-1064; Fax: (608) 262-5199; Email: jyguo@wisc.edu

Ipek N. Sener

The University of Texas at Austin, Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712
Phone: (512) 471-4535; Fax: (512) 475-8744; Email: ipek@mail.utexas.edu

Sivaramakrishnan Srinivasan

University of Florida, Department of Civil and Coastal Engineering
365 Weil Hall, PO Box 116580, Gainesville, FL 32608
Phone (352) 392-9537 Extn. 1456; Fax: (352) 392-3394; Email: siva@ce.ufl.edu

Rachel B. Copperman

The University of Texas at Austin, Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712
Phone: (512) 471-4535; Fax: (512) 475-8744; Email: RCopperman@mail.utexas.edu

Chandra R. Bhat*

The University of Texas at Austin, Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712
Phone: (512) 471-4535; Fax: (512) 475-8744; Email: bhat@mail.utexas.edu

*corresponding author. This paper was written when the corresponding author was a Visiting Professor at the Institute of Transport and Logistics Studies, Faculty of Economics and Business, University of Sydney.

ABSTRACT

This paper describes the development of a population update modeling system as part of the development of the Comprehensive Econometric Microsimulator for SocioEconomics, Land-use, and Transportation Systems (CEMSELTS). CEMSELTS itself is part of the Comprehensive Econometric Microsimulator for Urban Systems (CEMUS) under development at The University of Texas at Austin. The research in the paper recognizes that modeling the linkages among demographics, land use, and transportation is important for realistic travel demand forecasting. The population update modeling system focuses on the modeling of events and actions of individuals and households in the urban region. An analysis framework is proposed to predict the future-year population characteristics by modeling the changes to all relevant attributes of the households and individuals. The models identified in the analysis framework are estimated for the Dallas-Fort Worth region. The econometric structures used include deterministic models, rate-based probability models, binary logit models, multinomial logit models, and ordered-response probit models. To verify the outputs from these models, the predicted results for the year 2000 are compared against observed 2000 Census data.

1. INTRODUCTION

Conventional wisdom has long indicated that socioeconomics, land use, and transportation are intricately linked (for example, see 1-5). While socioeconomics 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 among geographically dispersed activity sites. The recognition of the linkages among socioeconomics, land use, and transportation is important for realistic forecasts of travel demand. Conventional methods, however, use aggregate forecasts of socioeconomics and land use to feed into travel models and, consequently, cannot capture the multitude of interactions that arise over space and time among the different decision makers.

The shortcomings of the conventional approach have led researchers to develop disaggregate behavioral approaches that capture land-use and travel behavior processes in an integrated manner, while accommodating the moderating role of socioeconomic characteristics (for example, see 6-8). Such behavioral approaches emphasize the interactions among population socioeconomic processes, the households' long-term choice behaviors, and the economic markets within which households act (9). These integrated land-use transportation modeling systems need to consider three important issues. First, over a long-range multi-year forecasting time frame, individuals go through different life-cycle stages and household compositions. Such socioeconomic processes need to be modeled endogenously (*i.e.*, within the integrated land-use transportation model system) to ensure that the distribution of population attributes (personal and household) are representative at each point of time and are sufficiently detailed to support the travel-related behavioral decision models being used. Second, as the socioeconomic process unfolds, individuals may begin/finish schooling, move onto different life-cycle stages, enter/exit the labor market, and change jobs. Similarly, households may decide to own a house as opposed to rent, move to another location, and acquire/dispose off a vehicle. If these longer-term behavior choices concerning the housing and labor market are treated merely as exogenous inputs to the activity-based travel models, then the possibility that households can adjust with combinations of short- and long-term behavioral responses to land-use and transportation policies is systematically ignored (10). A significant increase in transport costs, for example, could result in a household adapting with any combination of daily activity and travel pattern attribute changes, job location changes, and residential location changes. Thus, a framework accounting for this interdependency between short- and long-term behaviors is required to evaluate the impacts of land-use and transport policies. Third, interactions between households and other decision makers (such as businesses, institutions, and real estate developers) within the housing, labor, and transportation markets ultimately shape land-use patterns. If the behavior of households is to be captured properly, the behavior of these other actors in the markets also needs to be explicitly considered.

The broad objective of the current paper is to discuss our current efforts at designing and developing a Comprehensive Econometric Microsimulator for Urban Systems (CEMUS) that is behaviorally oriented and places the focus on the underlying decisions of households and individuals, and businesses and developers, which manifest themselves in the form of aggregate passenger travel patterns. As shown in Figure 1, CEMUS takes as input information on the aggregate socioeconomics and the activity-travel environment characteristics in the urban study region for the base year, as well as policy actions being considered for future years (the activity-travel environment includes the land-use, urban form, and transportation level-of-service characteristics). The aggregate-level base year socioeconomic data are first fed into the synthetic

population generator (SPG) to produce a disaggregate-level synthetic dataset describing a subset of the socioeconomic characteristics of the households and individuals residing in the study area [see Guo and Bhat (11) for information on the SPG module]. Additional base-year socioeconomic attributes related to mobility, schooling, and employment at the individual level, and residential/vehicle ownership choices at the household level, that are difficult to synthesize (or cannot be) synthesized directly from the aggregate socioeconomic data for the base year are simulated by the Comprehensive Econometric Microsimulator for SocioEconomics, Land-use, and Transportation System (CEMSELTS).¹ The base year socioeconomic data, along with the activity-travel environment attributes, are then run through the Comprehensive Econometric Microsimulator for Daily Activity-travel Patterns (CEMDAP) to obtain individual-level activity-travel patterns [see Bhat *et al.* (12) and Pinjari *et al.* (13) for details on the CEMDAP module]. The activity-travel patterns are subsequently passed through a dynamic traffic micro-assignment scheme to determine path flows, link flows, and transportation system level-of-service by time of day [see Lin *et al.* (14) for a discussion of recent efforts on interfacing between CEMDAP and the Visual Interactive System for Transportation Algorithms or VISTA]. The resulting transportation system level-of-service characteristics are fed back to CEMSELTS to generate a revised set of activity-travel environment attributes, which is passed through CEMDAP along with the socioeconomic data to generate revised individual activity-travel patterns. This “within-year” iteration is continued until consistency and base-year equilibrium is achieved. This completes the simulation for the base year.

The next phase, which takes the population one step forward in time (*i.e.* one year), starts with CEMSELTS updating the population, urban-form, and the land-use markets (note that SPG is used only to generate the disaggregate-level synthetic population for the base-year and is not used beyond the base year). An initial set of transportation system attributes is generated by CEMSELTS for this next time step based on (a) the population, urban form, and land-use markets for the next time step, (b) the transportation system attributes from the previous year in the simulation, and (c) the future year policy scenarios provided as input to CEMUS. The CEMSELTS outputs are then input into CEMDAP, which interfaces with a dynamic micro-assignment scheme in a series of consistency/equilibrium iterations for the next time step (just as for the base year) to obtain the “one time step” outputs. The loop continues for several time steps forward until the socioeconomics, land-use, and transportation system path/link flows and transportation system level of service are obtained for the forecast year specified by the analyst. During this iterative process, the effects of the prescribed policy actions can be evaluated based on the simulated network flows and speeds for any intermediate year between the base year and the forecast year.

The focus of the current paper is the CEMSELTS module of CEMUS. Further, and while we have developed a comprehensive conceptualization and structure of CEMSELTS (9), the specific emphasis in this paper will be the component of CEMSELTS that deals with updating the population’s socioeconomic characteristics. To date, a number of demographic and socioeconomic updating modules have been developed in the field of sociology, including

¹ The base year synthetic disaggregate-level sociodemographic data generated by SPG and the base-year activity-travel environment attributes are used by CEMSELTS to generate additional disaggregate-level base-year socioeconomic data (see Figure 1). The reader will note that an advantage of using stochastic models in CEMSELTS to generate some of the base year socioeconomic characteristics is that the synthetic population has more variation than would be obtained by simply expanding the disaggregate-level sample (usually the Public-Use Microdata Samples or PUMS data) employed in the SPG module.

DYNAMOD (15), DYNACAN (16), NEDYMAS (17), and LIFEPATHS (18). These modules explicitly model demographic processes at a high level of detail. However, they are not well suited for application in the context of an activity-based travel microsimulation system because generating the necessary land-use and transportation system characteristics required for an activity-based travel microsimulator with these models is not straightforward. At the same time, within the travel demand forecasting community, the experience with demographic and socioeconomic updating methods for use in microsimulation systems is relatively limited (19). A recent research effort focusing specifically on simulating demographic evolution for the purposes of travel forecasting is the DEMOgraphic (Micro) Simulation (DEMOS) system (20). Other population updating systems that have been developed in the travel demand forecasting community (and with varying levels of detail and sophistication) include the Micro-analytic Integrated Demographic Accounting System (MIDAS) (21), and the Micro-Analytical Simulation of Transport Employment and Residences (MASTER) (22). Earlier land-use transportation modeling systems that focus on modeling certain aspects of the population evolution processes, such as residential relocations and automobile ownership, include TRANUS (23), MEPLAN (24), URBANSIM (25), and ILUTE (26). The research presented in this paper adds to the existing body of work on population updating for travel demand forecasting.

The remainder of the paper is organized as follows. Section 2 describes the scope and challenges of our CEMSELTS development and presents the analysis framework developed for updating population socioeconomic characteristics within CEMSELTS. Section 3 discusses the estimation and verification of the constituent models. Section 4 concludes the paper.

2. POPULATION UPDATE MODELING FRAMEWORK

The structure of the modeling system we have developed for population updating within the CEMSELTS module of the CEMUS framework is presented in Figure 2. This structure comprises two major subsystems: (1) the migration model system, and (2) the socioeconomic evolution model system. The **migration model system** comprises models that determine the movement of existing households out of the study region (*i.e.*, emigration) and the movement of new households and individuals into the study region (*i.e.*, immigration). The migration model system is strategically placed at the top of the overall modeling structure to avoid any in- or out-flow of households and individuals during the rest of the simulation cycle. Once the population is determined, the **socioeconomic evolution (SE) model system** focuses on simulating the changes in the population. This model system in turn comprises three major components: (1) individual-level evolution and choice models (modeling births, deaths, schooling, and employment) (2) household formation models (modeling living arrangement, divorce, move-ins, and move-outs from a family), and (3) household-level long-term choice models (modeling residential moves, housing characteristics, automobile ownership, information and communication technology adoption, and bicycle ownership). Together, the migration and the SE model systems determine the changes in population characteristics, residential pattern, and employment patterns over the course of one simulation year. The structures of the two model systems are discussed in turn in the next two sections. But prior to this discussion, there are two major issues that we would like to bring attention to in the modeling of the above mentioned systems. First, the current state of knowledge regarding the complex nature of, and the interdependency among, socioeconomic processes is arguably quite limited (this is not to underplay the substantial theory and literature that exists in several fields on migration and socioeconomic evolution systems that we cannot review in this paper due to space

considerations). The several elements of the socioeconomic processes are potentially simultaneous or sequential or perhaps a mixture of both. Arriving at the nature of this interdependency poses a substantial challenge, especially because the nature of the interdependency itself is likely to vary across individuals and households. Second, the data for modeling these processes are typically not at the level of micro-detail that would be most desirable for microsimulation-based modeling approaches.

2.1 Migration Model System

As mentioned earlier, the migration model system includes models for both emigration and immigration. The household emigration model determines the likelihood that a household in the study area will move out of the study region during the simulation year. Note that this model is focused on modeling the move of the entire household. In addition, it is also possible that one or more individuals of the household will move out and others in the household will remain in the study region. The household formation models capture such transitions (see Section 2.2.2 for a discussion on household formation models).

Unlike emigration, modeling immigration comprises several models to determine the characteristics of the population moving into the study region. In this regard, it is helpful to distinguish between the immigration of entire non-single individual households (with their constituent household members) into the study region and the immigration of individuals not belonging to immigrating non-single households. *For the immigration of entire non-single households*, the immigrant models determine the different aggregate characteristics (such as age, gender, ethnicity, household composition, education level, and automobile ownership) of the incoming households and individuals. During implementation, these models are used to synthesize “new” households and the constituent individuals to be added to the study area. Thus, the characteristics of the population for each simulation year corresponding to the immigration of entire non-single households are completely synthesized, and this population is not taken through the SE model system. *For the immigration of individuals not belonging to immigrating non-single households*, the immigrant models determine the aggregate characteristics of this population, which are then used to synthesize individual immigrants for each simulation year. However, these immigrating individuals can serve as candidates for new household formations via marriage and move-ins (see Section 2.2.2). The arrow from “New households and individuals in the study region” to the “household formation models” in Figure 2 is shown to accommodate such potential household formations.

2.2 Socioeconomic Evolution (SE) Model System

The SE model system determines the individual- and household-level changes to the population that continue to stay in the study area, and determine household formations associated with the immigration of individuals who are not part of immigrating non-single adult households. As shown in Figure 2, this model system is composed of three sub-components, each of which is discussed in turn in the subsequent three sections. The assumed sequencing of the modeling sub-components represents a systematic approach to enable the practical modeling of a large number of potentially interrelated processes.

2.2.1 Individual-Level Evolution and Choice Models

This is a suite of models for simulating individual-level evolution processes, including (a) demographics related to aging, deaths, and births, (b) personal mobility-related choice of obtaining a driver's license, and (c) schooling and employment choices (see Figure 3).

The first demographic evolution process modeled is aging. Aging, unlike the other evolution processes in Figure 3, is a deterministic process. Hence, a simple counter (rather than a probabilistic choice model) is adequate to implement the aging process.

Mortality is the next individual-level evolution process modeled within our analysis framework. This model determines the likelihood of the death of an individual. In addition, the model also prescribes an upper-limit cutoff point on the age beyond which individuals are assumed not to live. During implementation, an individual predicted to die based on the mortality model is removed from the system and is subjected to no further processing. It is possible that the death of one or more individuals in a household may result in a household composed only of children (individuals 15 years and younger are classified as children and the rest as adults). A secondary model is developed to transfer the children to other households with one or more adults.

Birth is the final demographic evolution process modeled, and is applied to women between the ages of 10 and 49. This birth model determines the number of children born by gender. During implementation for forecasting, this model determines the number of new individuals to be synthesized and added to the household.

The next three model components in Figure 3 focus on the mobility, schooling, and employment choices of individuals. The individual-level mobility choice relates to obtaining and maintaining a valid driver's license. Correspondingly, there are two dimensions of individual mobility decisions: (1) The decision to obtain a driver's license (this dimension is particularly relevant to children turning sixteen), and (2) The decision to maintain a valid driver's license. Due to the specific characteristics of the Dallas-Fort Worth (DFW) area for which we applied our framework, it is assumed that all individuals obtain a driver's license at the age of sixteen and maintain it subsequently.

The schooling-related choices of individuals are modeled by first segmenting the population into two groups: (1) children and adults who have not graduated from high school, and (2) adults who already have graduated from high school. The schooling decisions of the first set of individuals are determined using two models. The **join-school** model determines the likelihood of a non-student starting to attend school. For a student choosing to attend school, the model also determines the school type (*i.e.*, primary school, middle school, or high school) and the school location. The choice of school type is largely dictated by the age of the individual and, in turn, determines the candidate school locations. The **continue-school** model is applicable for individuals who currently are students. This model determines whether the student continues at the same school or not. Termination of attendance at a school may be a result of graduation, dropout, or move to another school. Hence, the continue-school model determines the likelihood of the occurrence of one of the four possible outcomes identified above.

Analogous to the models for the first set of individuals, the schooling choices for the second set of individuals (adults who are already high school graduates) also comprise two main components: the **join-college** model for non-students and the **continue-college** model for current students. The first model for non-students determines whether or not an adult chooses to pursue an education beyond high school. For those choosing to undertake higher-level studies, the model determines the degree level (bachelors, masters, or doctoral) and the school location. The

first choice determines the length of stay in school and candidate school locations, and subsequently also influences the employment-related choices. In modeling the choice of school location, adults can choose to attend schools outside the study area (unlike in the case of school location choice of children). The second model (*i.e.*, the continue-college model) is analogous to the continue-school model for children. This model determines one of four possible outcomes for persons already in college: continue attendance, graduate, dropout, or move to another school.

The final individual-level set of choice models focuses on the employment-related choices of children over the age of 12 years and adults (children 12 years or younger are not allowed to be employed in our system). As in the case of schooling models, modeling employment-related choices also requires that we distinguish the choices of unemployed persons from the choices of those who are employed. Correspondingly, we identify the following two major model systems: the **start-employment** model and the **continue-employment** model. The start-employment model determines whether or not a currently unemployed individual starts working during the simulation year. For those choosing to work, the model determines the following employment characteristics: (1) employment industry, (2) employment location, (3) weekly work duration, (4) work flexibility, and (5) annual personal income. The second major employment-related model system (*i.e.*, the continue-employment model) determines the continuation or termination of work of the individual at his or her current employment location. The termination of employment at the current location may be a consequence of one of the following: retirement, quitting job, or switching jobs. Thus, the continue-employment model determines one of four possible outcomes (*i.e.*, continue, retire, quit, or switch). Subsequent to modeling these choices, secondary models determine the changes to work characteristics (especially personal income changes) for individuals continuing to work and the new employment characteristics for persons choosing to switch jobs. An important issue that needs to be addressed by the models focusing on employment-related choices is the interdependencies in these decisions across household members. At this time, we do not consider such interdependencies and leave this for subsequent research.

2.2.2 Household Formation Models

The next major component in the overall analysis framework comprises models for household formation (Figure 4). We identify four processes that lead to formation of new households: (1) marriage/cohabitation choice, (2) divorce, (3) move-ins, and (4) move-outs. Each of these is discussed in turn next.

The marriage/cohabitation models describe the decision of single adults to marry or enter a cohabiting arrangement (*i.e.*, common-law union) with another single adult and consequently form a new household. These decisions are modeled using a set of three models. The first model, the **living arrangement** model, determines the choice of single individuals to stay single, or marry, or enter into a cohabiting arrangement. The second model, the **spouse/partner characteristics** model, determines the characteristics of the spouse/partner that an individual wishes to marry/cohabit with. The estimation and application of this model is undertaken from a female perspective. That is, it is assumed that it is the woman who decides her spouse/partner from the pool of males that wants to marry or cohabit (as determined in the living arrangement model). Finally, the third model, the **new family residence** model, determines whether the individual continues to live at the same residence after marriage/cohabitation, moves to the spouse/partner's residence, or moves to a completely new residential unit. Currently, we do not have data available for estimation of the new family residence model, and hence assume equal

probabilities of each residential choice. During implementation of these models for forecasting, the living arrangement model is run for all single adults in the population to identify the individuals who are “in the market” to marry or cohabit. Subsequently, the spousal characteristics models are used to match these individuals (*i.e.*, those who choose to marry or cohabit) with their spouses and partners using a female-dominant perspective, and the new family residence models are used to locate newly formed families appropriately.

The divorce models describe the dissolution of existing households as a consequence of a divorce between married couples (or because of the termination of cohabiting arrangements) and the subsequent formation of new households. These decisions are modeled using four models. First, the **marriage/cohabitation dissolution** model determines the likelihood that a marriage will end in divorce or that a cohabiting arrangement will terminate (undertaken again from a female perspective). The next three models describe the subsequent formation of the new households as a consequence of a divorce (or termination of cohabitation). The **child custody** model determines which of the parents gets custody of the children in the original household. The **resource allocation** model determines how the collective resources of the family (house, wealth, automobiles, *etc.*) are distributed between the couple and, finally, the **individual residence** model determines the new residential choices of the two separating adults. Due to lack of data, we currently have not estimated formal models for the child custody, resource allocation, and individual residence components. Instead, we adopt simple non-data based rules for these models (such as providing custody to the mother in the case of a divorce), but hope to improve upon these rules as data becomes available.

Move-in and move-out of adults represent the two other major processes that can result in the formation of new households (see Figure 4). The move-ins occur predominantly in two types of settings: move in of a new roommate and the return home of a young adult. In each case, the move-in model determines the characteristics of the adult who is to move into the household conditional on an adult being predicted to move into the household. As in the case of move-ins, the move-outs also occur predominantly in two types of settings: move out of a roommate and move out of an adult from the parental home. The move-out model determines the likelihood that an adult will move out. A person moving out can form a new independent household within the study region, join an existing household in the study region, or move out of the study area. For those individuals who move out to form an independent household, their residential choice characteristics are obtained by “passing” them through the residential moves model of the household long-term choice model system (see next section). At this time, we have not estimated models for move-in and move-out because of lack of good data to estimate such models. Thus, we use simple non-data based rules to implement these models in forecasting mode.

2.2.3 Household-Level Long-Term Choice Models

The final set of models in our analysis framework corresponds to the long-term decisions of households. Specifically, we identify four major decisions (Figure 5): residential moves, automobile ownership, ICT (information and communication technologies) adoption, and bicycle ownership. These four processes are modeled sequentially and separately by a suite of models.

The residential moves models focus on the decision of a household to remain in its current residential unit as opposed to moving to a new house. Conditional on choosing to relocate, households face three major decisions: housing tenure (rent or own), housing (dwelling unit) type (single family detached, single family attached, apartment, or mobile home/trailer), and the new household location. The automobile ownership models are concerned with the

number of vehicles to own (in terms of the household's transaction decision to not change vehicle fleet composition, buy a vehicle, sell a vehicle, or trade a vehicle) and the vehicle type (passenger car, pick-up truck, SUV, *etc.*). The ICT adoption models determine households' choices relating to the ownership of (1) phones/mobile phones/pagers, (2) computers, and (3) Internet connectivity. Lastly, the bicycle ownership model determines the number of bicycles owned by each household. The ICT and the bicycle ownership models have not been estimated in the current stage of implementation because of inadequate data from the DFW area on these dimensions of choice.

3. ESTIMATION AND VALIDATION

The first step in implementing the proposed population update modeling system is to empirically develop, estimate, and verify all the constituent model components. Section 3.1 outlines the econometric structures of these models and the data sources employed in the estimation procedure. Section 3.2 presents model verification results obtained from comparing the CEMSELTS population updating model predictions against observed data.

3.1 Model Estimation Results

Data collection for estimating the various model components outlined in Section 2 was a significant challenge. Ideally, since the current implementation of CEMDAP is customized for the DFW area, detailed data describing the various evolution processes of the population in the DFW area should be used for model development. However, such data were found to be unavailable for several demographic processes such as deaths, births, and marriages. Therefore, alternative cross-sectional data sources at the state and national levels were used. After data collection, a suite of models were estimated. The modeling frameworks employed in the empirical analysis include: rule-based models, rate-based probability models, binary logit models, multinomial logit models and ordered-response probit models.

A summary of the models described in the analysis framework of Section 2 is presented in Tables 1a and Table 1b.² Table 1a presents the models relevant to individuals. Table 1b summarizes models related to households. The tables present, for each choice process, information regarding the econometric structure employed, the independent variables used, the data source, the choice alternatives in the model and any remarks corresponding to the model. The actual estimation results for the various models components are not presented here to adhere to paper length limitations [for a detailed presentation of probability rate look-up tables, model estimation results, and their interpretations, the reader is referred to Guo *et al.* (9) and Pinjari *et al.* (13)].

The models identified in Table 1a and 1b are used to determine the probabilities of choosing various alternatives at each stage of the population socioeconomic updating process. A microsimulation framework is then used to translate these probabilistic predictions to deterministic predictions. In the case of discrete choices, the chosen alternative is determined through a random draw from a pseudo sample containing all the alternatives in proportion to their probabilities as predicted by the corresponding model. Thus, it is ensured that the chosen discrete outcome is not the same for all observationally similar decision makers. For the continuous choice instances, the choice is determined by a random draw from the probabilistic

² Only model components for which a deterministic rule was used, or for which a model was estimated based on actual data, are listed in Tables 1a and 1b. Thus, some models in Figure 4 (household formation) and Figure 5 (household-level long-term choice models) do not appear in Tables 1a and 1b.

distribution of the choice variable defined by the associated econometric model. For more details on the simulation procedure and the implementation of the various modules to obtain the socioeconomic characteristics for any given year, the reader is referred to Bhat *et al.* (12) and Pinjari *et al.* (13).

3.2 Model Verification

The model system discussed in Sections 2 and 3.1 has been applied to predict the base year population for 2000 as well as the forecast year of 2025 in the DFW area of Texas. In this section, we confine our prediction results to the base year because actual data to verify the 2025 predictions are not available. Future efforts will undertake an evaluation of the prediction results in the context of performance over time (for example, comparing the 2005 year predictions with the observed 2005 population characteristics).

For the base year, and as discussed in Section 1 and Figure 1, the synthetic population generator (SPG) is used to generate a synthetic population describing the households and individuals residing in the study area. The total numbers of households and individuals, as well as several socioeconomic attributes of individuals and households, are controlled for in this generation process. The control attributes include household family type, household size, and number of children at the household level, as well as age, race, and gender at the person level (11). Thus, for the base year, there is no need to implement the CEMSELTS components corresponding to the (1) migration model system, (2) individual-level models of aging, death, and birth, (3) household formation models, and (4) residential location and mobility models within the household-level long-term choice models. However, CEMSELTS is used to generate the base-year socioeconomic attributes related to (1) schooling and employment at the individual-level, (2) household non-location residential choices (tenure, housing type), and (3) vehicle ownership. It is these attributes that we focus on to examine the results from CEMSELTS.

The CEMSELTS models corresponding to the population update modeling system have been implemented using Gauss, a matrix programming platform that is capable of handling large data matrices (27). The implementation entails writing Gauss code for estimating a 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 discrete choice modules, additional code was written to obtain the continuous values of attributes. For instance, income category was determined employing a grouped response structure. 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 (28), may also be applied, and this will be implemented in the future. The research team is also pursuing the development of a visual C++ platform for implementing the many components of CEMSELTS and integrating with the other existing elements of the CEMUS visual C++ platform.

The results for schooling and employment at the individual-level are summarized in Tables 2 and 3. The first column in Table 2 lists the alternatives for each choice dimension. The second column shows the predicted share of each alternative in the base year population obtained after applying SPG and CEMSELTS. The third column corresponds to the sample share observed in the DFW travel survey data. The fourth column represents the observed share found

in the 2000 Census summary data for the DFW area (except for work flexibility and personal income models where Census data are not available for the choice dimensions).

In general, the prediction capability of the system is quite satisfactory for the individual-level attributes. In particular, the predicted values of educational attainment, labor participation, employment industry, and work flexibility match up quite well with the DFW survey and Census distributions. The distributions of weekly work duration and personal income do show some substantial differences from the distributions found in the Census data. These differences may be attributed to the small sample sizes employed in the estimation of the corresponding prediction modules (note that we have not introduced any “calibration constants” to adjust the predicted results). The predicted county-to-county work flows are compared with data obtained from the 2000 Public Use Microdata Sample (PUMS) to verify the employment location module (work flows, as used here, represent the number of individuals residing in each county that work at each of the other counties; this should not be confused with trip flows). Table 3 provides the results for the five-county region in Dallas-Fort Worth, where the five counties are wholly located in the NCTCOG (North Central Texas Council of Governments) planning area. The table indicates the percentage of overall work flow originating from each county in the five-county region (see last column), work flow estimated for each county (see last row) as well as the intercounty flows. The “observed” (*i.e.*, PUMS-based) and predicted flows match up reasonably well (exceptions include Dallas - Ellis, Ellis - Ellis flows).

Finally, the household non-location residential choices (tenure and housing type) and vehicle ownership results are presented in the same format as Table 2, and indicate that the predicted household attributes match the corresponding survey and DFW samples well (see Table 4).

4. CONCLUSION

This paper describes the development of a population update modeling system as part of the development of the Comprehensive Econometric Microsimulator for SocioEconomics, Land-use, and Transportation Systems (CEMSELTS). CEMSELTS itself is part of the Comprehensive Econometric Microsimulator for Urban Systems (CEMUS) under development at The University of Texas at Austin. The population update modeling system outlined in the current paper includes a migration model system and a socioeconomic evolution model system. These model systems determine the future-year population characteristics by modeling the changes to all relevant attributes of households and individuals.

The models identified in the proposed framework have been estimated for the DFW region based on disaggregate level data specific to DFW, and regional and national data sources. The econometric structures employed in the analysis include deterministic rules, rate-based probability models, binary logit models, multinomial logit models, ordered-response probit models, and grouped-response probit models. To verify the outputs from the population updating system, the model system was employed to generate disaggregate information required for CEMDAP for the base year of 2000. The results from the 2000 year micro-simulation are compared against Census and DFW sample data (as appropriate). These results are satisfactory and encouraging for the ongoing development of CEMUS. At the same time, it is also important to note that the CEMSELTS model element is not without its limitations. A few components of the CEMSELTS framework are yet to be estimated, due to lack of appropriate data. Other components can be improved in their modeling if more appropriate data became available. Further, the sequence employed within the CEMSELTS framework might not be an accurate

representation of reality because some of the decisions for individuals may be bundled together (instead of separate models). However, the system represents an implementable framework with currently available data.

ACKNOWLEDGEMENTS

The research in this paper was funded by a Texas Department of Transportation (TxDOT) project entitled “Second Generation Activity-Based Travel Modeling System for Metropolitan Areas in Texas Accommodating Demographic, Land Use, and Traffic Microsimulation Components”. The authors would like to thank Janie Temple and William Knowles of the Transportation Planning and Programming Division of TxDOT for their input and suggestions during the course of the TxDOT project. The authors appreciate Lisa Macias’s help in typesetting and formatting this document. The corresponding author also acknowledges the support of an International Visiting Research Fellowship and Faculty grant from the University of Sydney. Four reviewers provided helpful comments on an earlier version of the paper.

REFERENCES

1. Mitchell, R., and C. Rapkin. *Urban Traffic – A Function of Land Use*. Columbia University Press, New York, 1954.
2. Jones, P., M. Clarke, and M. Dix. *Understanding Travel Behavior*. Gower, Aldershot, 1983.
3. Jones, P. ed. *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*. Avebury, Aldershot, 1990.
4. Banister, D. *Transport Planning*. Chapman & Hall, London, 1994.
5. Hanson, S. *The Geography of Urban Transportation*. Macmillan Publishing Company, New York, 1996.
6. Bhat, C. R., and J. Y. Guo. A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels. *Transportation Research Part B*, Vol. 41, No. 5, 2007, pp. 506-526.
7. Pinjari, A. R., R. M. Pendyala, C. R. Bhat, and P. A. Waddell. Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice. *Transportation*, Vol. 34, No. 5, 2007, pp. 557-573.
8. Cao, X. and P. L. Mokhtarian. How Do Individuals Adapt Their Personal Travel? Objective and subjective influences on the consideration of travel-related strategies for San Francisco Bay Area commuters. *Transport Policy*, Vol. 12, No. 4, 2005, pp. 291-302.
9. 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, Center for Transportation Research, The University of Texas at Austin, 2005.
10. Waddell, P., C. R. Bhat, E. Ruiters, S. Bekhor, M. Outwater, and E. L. Schroer. *Review of the Literature and Operational Models: Final Report to the Puget Sound Regional Council on Land Use and Travel Demand Forecasting Models*, 2001.
11. Guo, J. Y., and C. R. Bhat. Population Synthesis for Microsimulating Travel Behavior. *Transportation Research Record*, Vol. 2014, 2007, pp. 92-101.
12. Bhat, C. R., S. Srinivasan, J. Guo, and A. Sivakumar. Activity-Based Travel Demand Modeling for Metropolitan Areas in Texas: A Microsimulation Framework for Forecasting. Report 4080-4, Center for Transportation Research, The University of Texas at Austin, 2003.
13. Pinjari, A. R., N. Eluru, R. Copperman, I. N. Sener, J. Guo, S. Srinivasan, and C. R. Bhat. Activity-based travel-demand analysis for metropolitan areas in Texas: CEMDAP Models, Framework, Software Architecture and Application Results. Report 4080-8, Center for Transportation Research, The University of Texas at Austin, 2006.
14. Lin, D-Y., N. Eluru, T. Waller and C.R. Bhat. Integrating an Activity-Based Model (CEMDAP) with a Dynamic Traffic Assignment tool (VISTA). Forthcoming, *Transportation Research Record*, 2008.

15. King, A., H. Baekgaard, and M. Robinson. DYNAMOD-2: An Overview. Technical Paper no. 19, National Centre for Social and Economic Modelling, University of Canberra, Australia, 1999.
16. Morrison, R. (1998, November), slightly adapted by B. Dussault (2000) Overview of DYNACAN - a full-fledged Canadian actuarial stochastic model designed for the fiscal and policy analysis of social security schemes.
www.actuaries.org/CTTEES_SOCSEC/Documents/dynacan.pdf
17. Nelissen, J. H. M. Demographic Projections by Means of Microsimulation. The NEDYMAS model, part A+B, Tilburg University Press, Tilburg, 1995.
18. Gribble, S. LifePaths: A Longitudinal Microsimulation Model Using a Synthetic Approach. In *Microsimulation in Government Policy and Forecasting*, Eds. Gupta, A., and V. Kapur, Elsevier, Amsterdam & New York, Ch. 19, 2000.
19. Miller, E. J. Microsimulation. In *Transportation Systems Planning: Methods and Applications*, Eds. K. G. Goulias, CRC Press, Boca Raton, Ch. 12, 2003.
20. Sundararajan, A., and K. G. Goulias. Demographic Microsimulation with DEMOS 2000: Design, Validation, and Forecasting. In *Transportation Systems Planning: Methods and Applications*, Eds. K.G. Goulias, CRC Press, Boca Raton, Ch. 14, 2003.
21. Goulias, K. G., and R. Kitamura. A Dynamic Model System for Regional Travel Demand Forecasting. In *Panels for Transportation Planning: Methods and Applications*, Eds. Golob, T., R. Kitamura, and L. Long, Kluwer Academic Publishers, Boston, Ch. 13, 1996, pp. 321-348.
22. Mackett, R. L. *MASTER Mode*. Report SR 237, Transport and Road Research Laboratory, Crowthorne, England, 1990.
23. Barra, T. de la. *Integrated Land Use and Transport Modelling*. Cambridge University Press, Cambridge, 1989.
24. Hunt, J. D. A Description of the MEPLAN Framework for Land Use and Transport Interaction Modeling. Presented at 73rd Annual Meeting of the Transportation Research Board, Washington, D.C., 1993.
25. Waddell, P. UrbanSim, Modeling Urban Development for Land Use, Transportation, and Environmental Planning. *Journal of the American Planning Association*, Vol. 68, 2002, pp. 297-314.
26. Miller, E. J., J. D. Hunt, J. E. Abraham, and P. A. Salvini. Microsimulating Urban Systems. *Computers, Environment and Urban Systems*, Vol. 28, 2004, pp. 9-44.
27. Aptech. Gauss 8.0 Quick Start Guide. Aptech Systems, Inc., 2006.
28. Bhat, C. R. Imputing a Continuous Income Variable from Grouped and Missing Income Observations. *Economics Letters*, Vol. 46, 1994, pp. 311-319.

List of Figures:

Figure 1 The structure of CEMUS

Figure 2 Overview of the analysis framework

Figure 3 Individual-level evolution and choices models

Figure 4 Household formation models

Figure 5 Household-level long term choice models

List of Tables:

Table 1a Migration and Individual Level Components of the Population Socioeconomic Updating System in CEMSELTS

Table 1b Household Formation and Household-level Long Term Choice in SE Module of CEMSELTS

Table 2 CEMSELTS Output Verification Results: Individual-level models

Table 3 CEMSELTS Employment Location Module Comparison

Table 4 CEMSELTS Output Verification: Household-level models

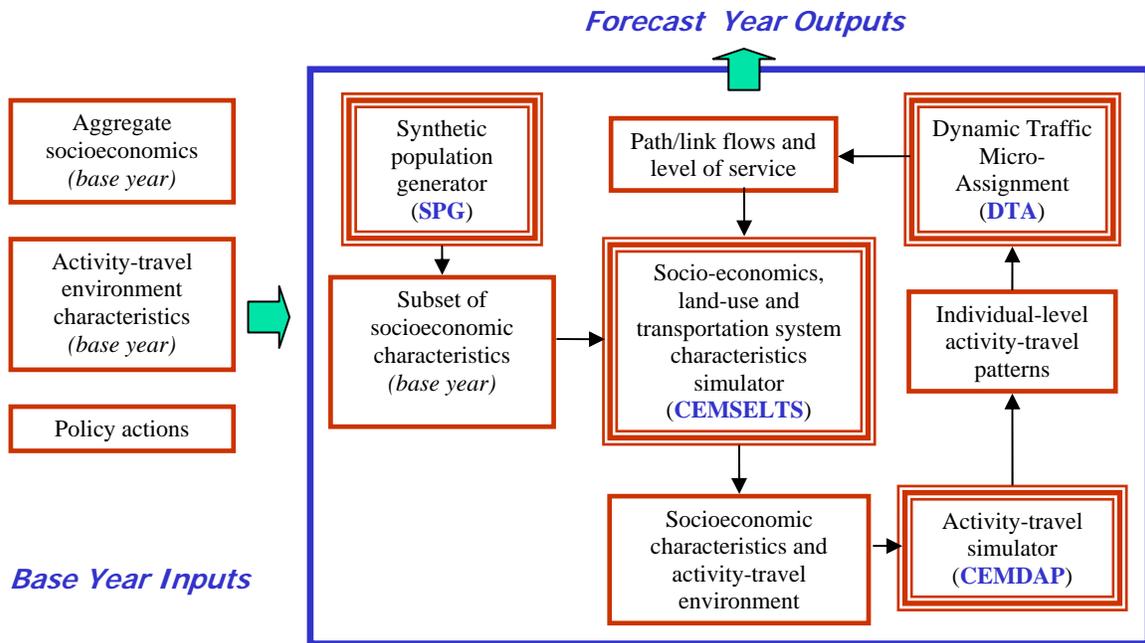


FIGURE 1 The structure of CEMUS.

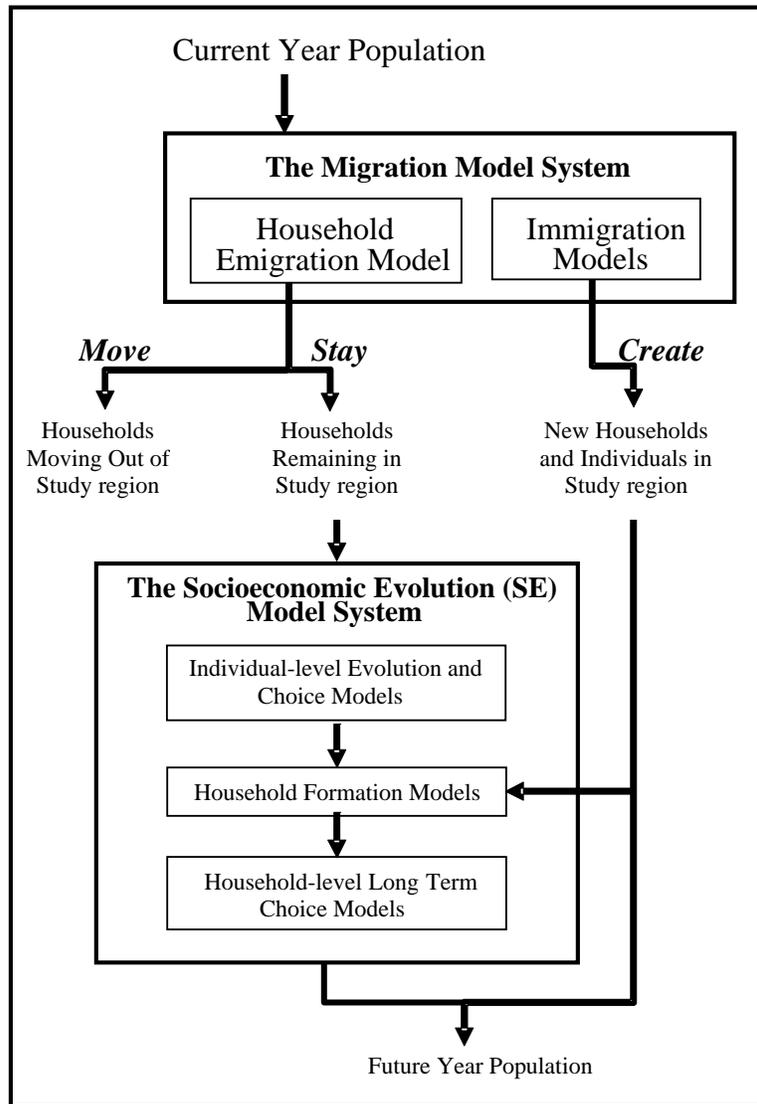


FIGURE 2 Overview of the analysis framework.

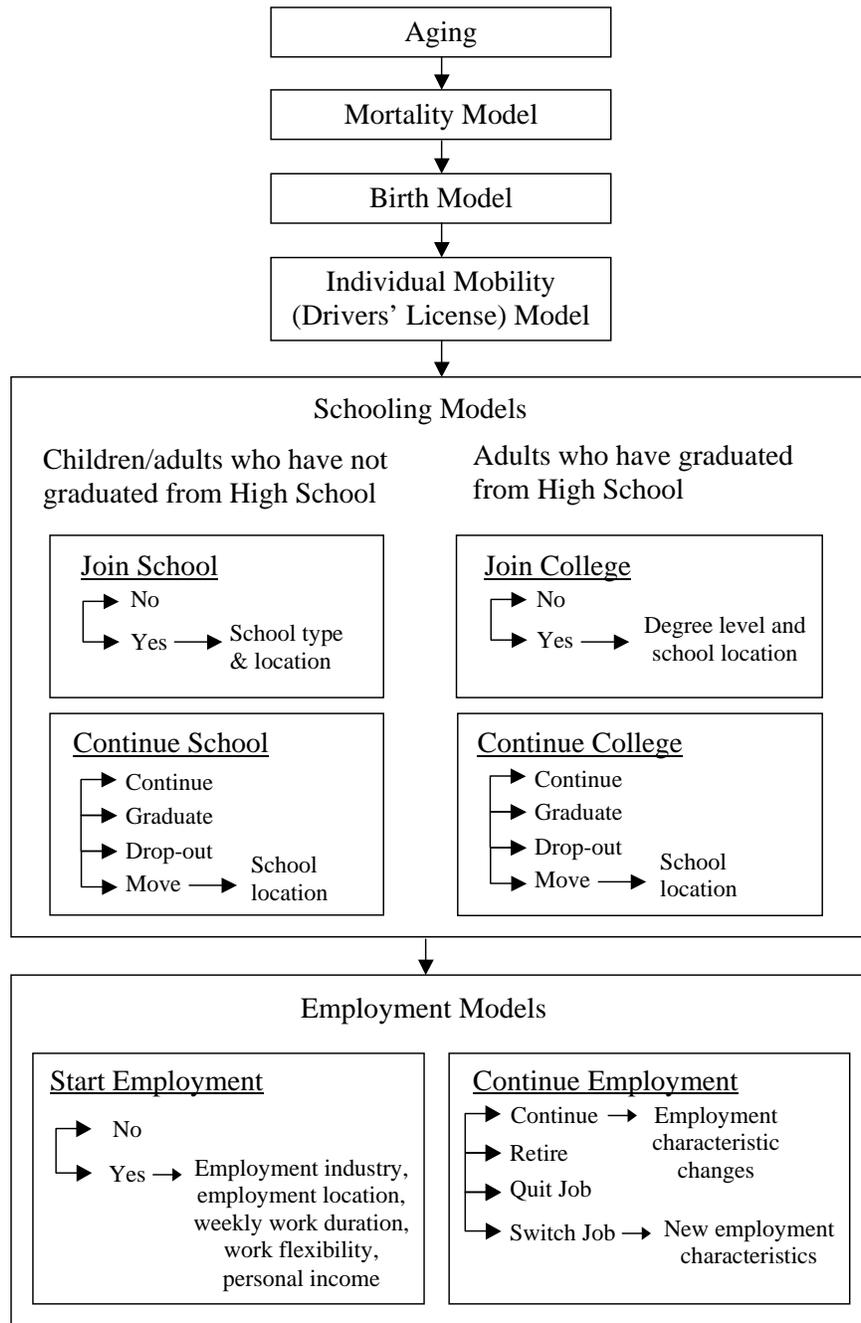


FIGURE 3 Individual-level evolution and choices models.

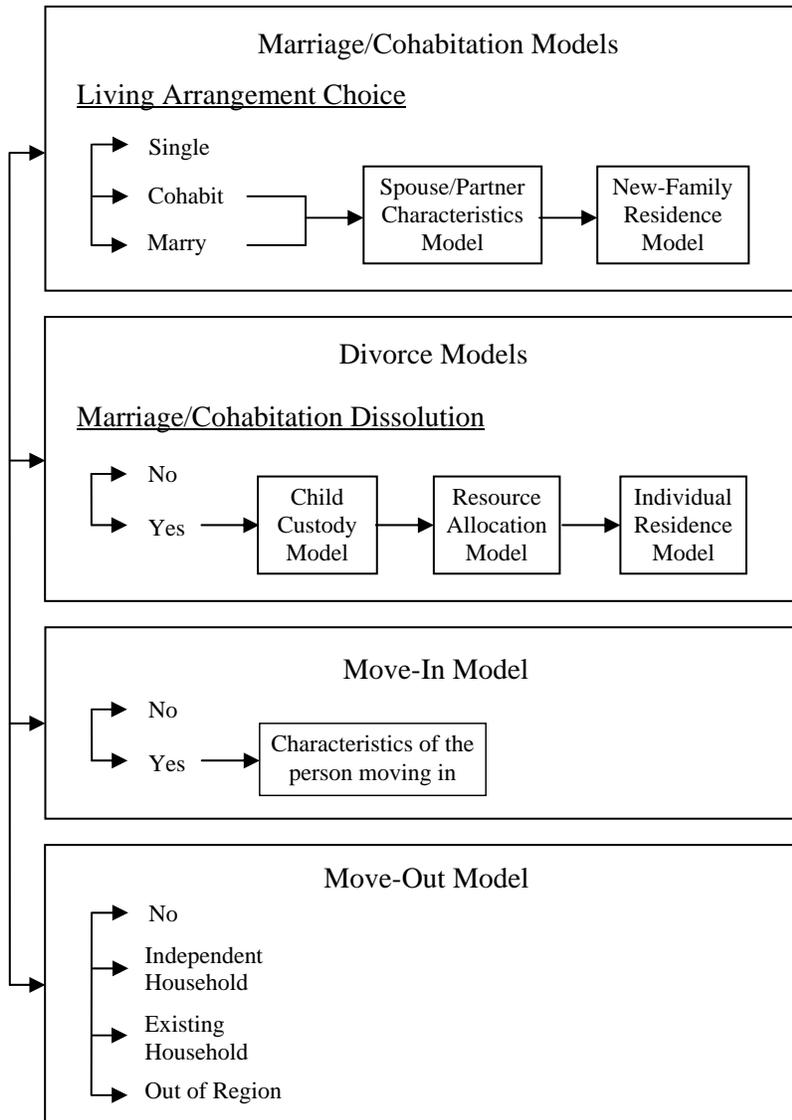


FIGURE 4 Household formation models.

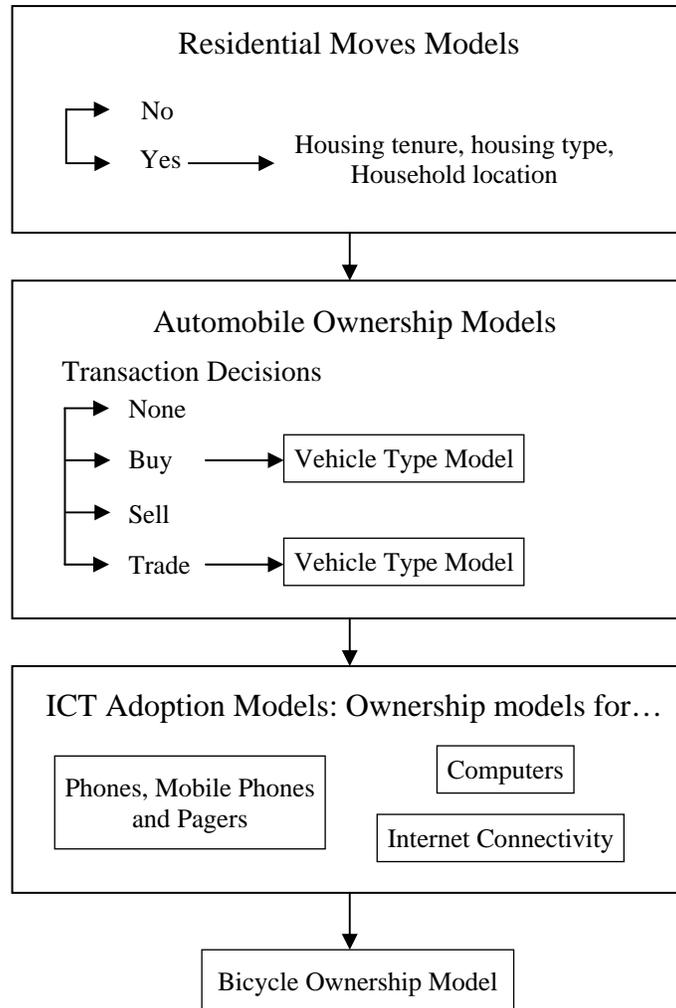


FIGURE 5 Household-level long term choice models.

TABLE 1a Migration and Individual Level Components of the Population Socioeconomic Updating System in CEMSELTS

Model name	Econometric Structure / Rule Based and Independent Variables	Data Source	Choice Alternatives/Comments
Migration Model System Immigration and emigration	Rate-based probability model based on age, gender, and race	Data from Texas State Data Center for 1990-2000 net migration rates for Texas	If net migration rate into the urban region is positive, individuals and households are generated. If net migration rate is negative, households are removed
Socioeconomic Evolution Model System			
Individual-level evolution and choice models:			
<i>Aging</i>	Deterministically updated yearly	---	Increase by one annually
<i>Mortality</i>	Rate-based probability model based on age, gender and race	National Vital Statistics, 2002	Individual dies or individual lives
<i>Birth</i>	Rate-based probability model for females aged 10-14 and 45-49 based on age and race. A binary logit model for females aged 15-45; independent variables include age, race, employment status, education level, and number of children	National Survey of Family Growth Data, 2002 and National Vital Statistics, 2003 and 2004	Female gives birth or female does not give birth
<i>Individual Mobility</i> Driver license (obtain and maintain a license)	Rule based model for individuals (individuals 16 years or older are licensed)	----	Individual is licensed or not
<i>Schooling</i> For children aged < 5 years	Rule based model – all children under the age of 5 are considered as not going to school	---	Schooling or no schooling
For children between 5 to 12 years	Rule based model – all children between 5-12 years are assumed to attend school, and their grade is based on age	---	Grades K through 7
If age between 13 and 18 years	Rate-based probability model depending on age, race, and gender	Texas Education Agency	Continue school, drop-out, or complete schooling. If drops out, grade is set to the grade at which drop-out occurs
If age > 18 years	Rate-based probability model for education level based on race.	2000 PUMS data for DFW area	Associate degree, bachelors, Masters, Ph.D.
School location of children attending primary, middle, or high school	Rule based assignment to closest zone (from residence) with a school in the independent school district of the household residence.	DFW school lookup table	Traffic analysis zones in DFW belonging to the independent school district of the household residence
School location of individuals attending college	Rate-based probability model by race and degree pursued	2000 PUMS data for DFW area	Traffic analysis zones in DFW with colleges or outside DFW

<i>Employment</i> Labor participation model	Binary Logit model; independent variables include age, gender, years of education, marital status and presence of children	PUMS data	Employed, Not employed (Applied for individuals over 12 years of age and not studying)
Employment industry model	Multinomial logit model; ; independent variables include age, race, gender and education level	DFW survey data	Construction and Manufacturing, Trade and Transportation, Professional businesses, Government, Retail and Repair, Other (relevant for employed individuals)
Employment location model	Spatial location choice model; independent variables include employment density, transportation level of service, accessibility to population and employment, and zones in central business district	DFW survey data	4874 TAZs of DFW area
Weekly work duration model	Grouped response model; independent variables include gender, education level and industry	DFW survey data	< 35 hours, 35-45 hours, and > 45 hours (the results are post processed to estimate a continuous “work hours” variable for each employed individual)
Work flexibility model	Ordered probit model; independent variables include age, gender, race, parent, education level, employment industry, and hours worked.	DFW survey data	Low flexibility, Medium flexibility, High flexibility (flexibility level definition is based on individual response in the survey)
Personal income model	Grouped response model; independent variables include age, gender, race, education level, employment status, and employment industry	DFW survey data	\$0 - \$9,999, \$10,000 - \$19,999, \$20,000 - \$29,999, \$30,000 - \$39,999, \$40,000 - \$49,999, and \$50,000 or above (the results are post processed to estimate a continuous “income” variable for each employed individual)
Employment mobility model	Rate-based probability model based on industry	Bureau of Labor statistics	Individual continues in his/her job or terminates job

TABLE 1b Household Formation and Household-level Long Term Choice in SE Module of CEMSELTS

Model Name	Econometric Structure and Independent Variables	Data source	Choice Alternatives
Household formation models			
Marital Status			
Marital status change model	Multinomial logit model; independent variables include age, race, employment status, and female gave birth to a child	National Survey of Family Growth Data, 2002	Single, married/cohabiting, or divorced (At any instance an individual has only two of the three choices available). Note that this model combines the living arrangement choice model for single adults (under marriage/cohabitation models) and the marriage/cohabitation dissolution model (under divorce models) into a single unified model. This unified model does not distinguish between marriages and cohabitation at this point, due to lack of data.
Spouse characteristics (Age, ethnicity and education)	Multinomial logit models for age of spouse (independent variable is age difference between adults), spouse race (independent variables includes female race and education), and spouse education (independent variable includes female age, race and education)	National Survey of Family Growth Data, 2002	For Age: <22, 23-28, 29-35, and >35 For Race: White, Black, Hispanic, and other For Education: Associate, Bachelor's, Master's and higher, and other
Household-level long term choice models			
Household Income model	Sum of incomes of individuals in the household	---	Continuous household income value
Residential Choices			
Residential mobility model	Binary logit model; independent variables include household income, number of adults, race, presence of elderly people and presence of unrelated persons	DFW survey data	Household decides to move or stay
Residential location model	Multinomial logit model; independent variables include level of service variables, accessibility variables interacted with presence of children, household income, residential tenure, and residential type	DFW survey data	4874 TAZs of DFW area
Residential tenure model	Binary logit model; independent variables include household income, household size, number of employed people, number of children, race, presence of elderly people, single-adult household and presence of unrelated people	DFW survey data	Own or rent house
Housing type model	Multinomial logit model; independent variables include household income, race, presence of elderly people, single-adult household, presence of unrelated people and highest education level in the household	DFW survey data	Single-family detached, Single-family attached, Apartment, and Mobile home or trailer (Separate MNL models estimated for households who own their house and households who rent)
Vehicle ownership model	Multinomial logit model; independent variables are household income, number of employed and unemployed adults, presence of children, own house, single-adult household	DFW survey data	0, 1, 2, 3, and 4 or more

TABLE 2 CEMSELTS Output Verification Results: Individual-level models

	Predicted	DFW Sample	Census
<u>Schooling Choices</u>			
<i>Education Attainment</i>			
No School	8.4	6.4	10.7
Children: Preschool thru Grade 4	9.8	8.9	9.6
Children: Grade 5 thru 8	6.4	4.4	5.9
Children: Grade 9 thru 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: Bachelors, Masters, and PhD	18.2	13.9	17.5
<u>Employment Choices</u>			
<i>Labor Participation</i>			
Employed	48.1	48.9	49.4
Unemployed	51.9	51.1	50.6
<i>Employment Industry</i>			
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
<i>Weekly Work Duration</i>			
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
<i>Work Flexibility</i>			
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
<i>Personal Income</i>			
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 3 CEMSELTS Employment Location Module Comparison

County-County Flows: PUMS

		% Within County of Work					Total % of workers by county of residence
		Collin	Dallas	Denton	Ellis	Johnson	
County of Residence	Collin	71.0	11.1	3.6	0.2	0.6	17.5
	Dallas	20.6	76.6	7.0	9.5	2.4	57.7
	Denton	8.0	9.5	89.1	0.8	0.6	17.3
	Ellis	0.3	2.3	0.1	85.2	1.1	3.7
	Johnson	0.1	0.5	0.2	4.3	95.3	3.8
Total % of workers by county of employment		12.8	73.4	9.6	2.2	2.0	100.0

County-County Flows: CEMSELTS Work Location Module

		% Within County of Work					Total % of workers by county of residence
		Collin	Dallas	Denton	Ellis	Johnson	
County of Residence	Collin	68.7	10.5	4.9	0.8	0.0	15.2
	Dallas	23.0	78.4	12.1	23.4	1.8	65.4
	Denton	8.1	8.3	82.8	0.5	0.2	13.3
	Ellis	0.2	2.5	0.1	71.9	1.7	2.9
	Johnson	0.0	0.3	0.1	3.4	96.3	3.2
Total % of workers by county of employment		9.8	80.9	6.6	1.1	1.5	100.0

TABLE 4 CEMSELTS Output Verification: Household-level models

	Predicted	DFW Sample	Census
<u>Residential Choices</u>			
<i>Residential Tenure</i>			
Own	66.7	66.6	60.0
Rent	33.3	33.4	40.0
<i>Housing Type for Owners</i>			
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
<i>Housing Type for Renters</i>			
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
<u>Household Vehicle Ownership</u>			
<i>Vehicle Ownership</i>			
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