2011 IEEE Forum on Integrated and
Sustainable Transportation Systems
Vienna, Austria, June 29 - July 1, 2011

Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT) in Southern California: Design, Implementation, Preliminary Findings, and Integration Plans

Konstadinos G. Goulias[1], Chandra R. Bhat[2], Ram M. Pendyala[3], Yali Chen[1], Rajesh Paleti[2], Karthik C. Konduri[3], Guoxiong Huang[4], and Hsi-Hwa Hu[4]

***Abstract*—**In this paper we describe the recently developed large scale spatio-temporal simulator of activities and travel for Southern California. The simulator includes population synthesis that recreates the entire resident population in this Mega region, provides locations for residences, workplaces, and schools for each person, estimates car ownership and type, and provides other key personal and household characteristics. Then, a synthetic schedule generator recreates for each resident person in the simulated region a schedule of activities and travel that reflects intra-household activity coordination for a day. These synthetic activity and travel daily schedules are then converted to multiple Origin Destination (OD) matrices at different times in a day. These are in turn combined with other OD matrices (representing truck travel, travel from and to ports and airports, and travel generated outside the region) and assigned to the network. The assignment output is then used in the software EMFAC to produce estimates of fuel consumed and pollutants emitted (including CO_2) by different classes of vehicles. The overall model system also includes provision for finer spatial and temporal resolutions and a staged plan to implement them. Numerical examples from each major modeling group are also provided together with next steps.

I. Introduction

The State of California has recently embarked on an aggressive movement towards reducing greenhouse gas emissions that contribute to global climate change, promoting sustainability, and better managing vehicular travel demand. The recent California State Senate Bill 375 explicitly calls for major metropolitan areas in the state to meet ambitious greenhouse gas emission reduction targets within the next several years. Metro areas are considering a range of policies to meet the emission reduction targets including land use strategies, pricing mechanisms, managed lanes, telecommuting and flexible work hours, enhancement of transit and pedestrian/bicycle modes, and use of technology to better utilize existing capacity. The analysis of these policies, and responding to the mandates of legislative actions such as Senate Bill 375 in California, calls for the adoption of model systems that are able to accurately represent activity-travel patterns of humans in a fine-resolution time-space continuum. The Southern California Association of Governments (SCAG), the metropolitan planning agency for the Southern California region, is moving forward with the development of a comprehensive activity-based microsimulation model system of travel demand to enhance its ability to estimate the impacts of a range of policy measures in response to Senate Bill 375 (http://www.scag.ca.gov/sb375/index.htm). SCAG is also required to develop a "Sustainable Community Strategy" through integration of land use and transportation planning and demonstrate its ability to meet the GHG emissions reduction targets by 2020 (8% GHG per capita per day reduction) and 2035 (13% GHG per capita per day tentatively). These are challenging targets for such a vast region, which includes a population of approximately 18.6 million people in 2008 (expected to grow to 23 million by 2035) and offers an extremely complex multimodal and diverse planning context with multiple actors in different jurisdictions. The new activity-based microsimulation model system is developed to address exactly this diversity among persons and contexts and described in this paper.

Here we describe the second phase of the development and application of the Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT), which is tailored to the Southern California region.

Manuscript received May 20, 2011. This work was supported by the Southern California Association of Governments and the UC Lab Fees Program.

[1]GeoTrans & Department of Geography, University of California, Santa Barbara, California, 93106-4060. Email: goulias@geog.ucsb.edu (corresponding author).

[2]Department of Civil, Department of Civil, Architectural and Environmental Engineering, The University of Texas, Austin, Texas 78712.

[3]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, Arizona, 85287-5306.

[4]Southern California Association of Governments, Los Angeles, California, 90017.

II. SimAGENT

The overall model structure is presented in Figure 1 in a schematic cascading form. The set of blocks on the left side represents groups of models that are designed for the first year (baseline) of the simulation that for this application it is 2003 to align with the four-step model of SCAG developed for the 2008 Regional Transportation Plan. Each block of the figure represents a group of techniques and statistical models many of which are developed to address policy actions aiming at replicating the resident population decision making. In essence this first set of models of the left hand side of Figure 1 recreates the resident population and gives to each person a daily schedule and ultimately assigns traffic to the network and computes emissions. The middle four blocks evolve the region economic and demographic landscape over time. This computerized evolution is done using a land use model based on the spatial input-output model called PECAS (Production, Exchange and Consumption Allocation System - http://www.scag.ca.gov/modeling/mtf/index.htm). This is paralleled by a set of algorithms and negotiated with local jurisdictions forecasts of residential development, industrial location, and demographic evolution at aggregate levels (cities and subareas within cities). In the middle of these two blocks is the household evolution module, which microscopically considers every resident household and gives it transitions and changes in time and space. It also converts travel times and the spatial distribution of economic activity and residential locations into accessibility indicators that are used to also drive travel behavior. The right hand side set of blocks is a repetition of the daily activity and travel patterns but at the next and all subsequent years of the simulation.

In this way land use policies of increased density and land use mix can be reflected in location decisions, car ownership and use, and activity participation and destination choices (including decisions to participate in activities and to travel alone or with others). In building policy scenarios, we start with an assembly of data on the entire roadway network and its characteristics (roadway types, intersections and ramps, number of lanes, and speed limits), and the public transportation network (type of service, routes, and schedules). In parallel, we assemble data on the resident population at different levels of geographical aggregation with core data at the zonal level of the pre-existing four-step model to enable use of existing forecasts and comparisons with simpler model systems. We also assemble external information about demographics, social and economic conditions of the study area, as well as any forecasts available. In this paper we focus on the first set of modeling blocks and offer a few schematic and numerical examples of output.

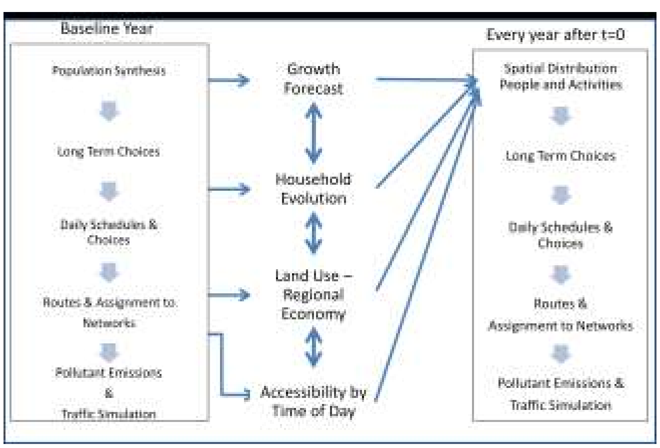

Figure 1 Schematic Representation of SimAGENT Blocks

A. *PopGen (Population Synthesis)*

In SimAGENT, the process starts with PopGen in which the entire resident population is synthetically generated/recreated person-by-person and household-by-household based on the method described in [1] and further enhanced and improved for more recent applications (see http://urbanmodel.asu.edu/popgen.html). The input to this software and block of methods is the spatial organization of the simulated area in the form of zone-specific univariate distributions of person and household characteristics provided by the US Census and SCAG for the baseline year (in this case 2003). As the population is recreated on a person-by-person and household-by-household basis, these distributions are used as the control totals for each spatial unit of analysis (approximately 4,000 Traffic Analysis Zones in this version of the model system) in an iterative algorithm that starts from a multivariate set of relationships (in essence a cross-tabulation) among the person and household variables used as seed. For future years these distributions are SCAG forecasts based on procedures of the growth forecast block in Figure 1 for 2020 and 2035 corresponding to the GHG target years. The multivariate set of relationships can be kept constant (assuming a steady state of demographic relationships) or can be changed to capture the impact of changing population composition and associated relationships including but not limited to age, birth rates, and household size.

B. *Accessibility*

In parallel (and shown in Figure 1 in the middle group of models) we also developed opportunity-based accessibility indicators at the level of the US census block (203,000 blocks cover the entire study area) to represent the ease (or difficulty) of reaching 15 different types of industries (representing the opportunities for activity participation) from each of these blocks within 10, 20, and 50 minutes of roadway travel [2]. The types of industries are: (a) Agriculture, forestry, fishing and hunting and mining; (b) Construction; (c) Manufacturing; (d) Wholesale trade; (e) Retail trade; (f) Transportation and warehousing and utilities; (g) Information; (h) Finance, insurance, real estate and rental and leasing; (i) Professional, scientific, management, administrative, and waste management services; (j) Educational; (k) Health; (i) Arts, entertainment, recreation, accommodation and food services; (m) Armed forces; (o) Public administration; and (p) Other services (except public administration). Different accessibility values are obtained for the morning peak period (6 to 9 AM), midday (9 AM to 3 PM), evening peak period (3 to 7 PM), and at night (7 PM to 6 AM) capturing not only the different roadway conditions but also the patterns of opening and closing of businesses during the day by allowing within each period above to also have different opening and closing hours of each industry type. Figure 2 provides an example of this spatial distribution by time of day. The top left hand quadrant also shows the percent of persons arriving and staying at the workplace (in this example banks and related institutions). As expected after 7:00 pm accessibility to these services is dramatically lower because the striking majority of these businesses are closed. The resident population with its detailed characteristics and a selection of indicators of the accessibility they enjoy are the inputs for the next block. Accessibility indicators are used in many of the behavioral models of the baseline year. They are modified based on the middle blocks of Figure 1 when the spatial distribution of economic activities change and they are also modified based on travel times changing based on network flow.

C. CEMSELTS

This block of models was first developed as part of a larger model system [3] and was modified and tailored to the SCAG region using local data. In CEMSELTS each person and household created in PopGen, and located in each zone of the study region, is given additional characteristics.

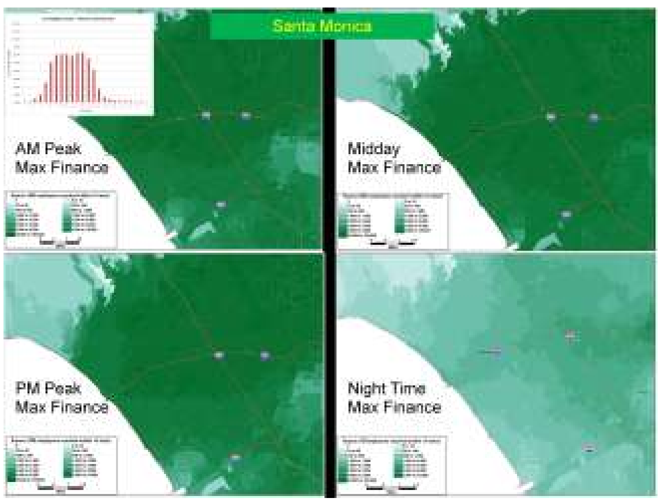

Figure 2 Time of Day Accessibility Map

For persons in college a model is used to assign a college location, which is also a hierarchical function of accessibility. Workers are identified using a labor force participation model that is a function of age, gender, education, and presence of children in the household. Employed persons are then assigned (in a probabilistic way) their type of industry, work location (which is also a function of accessibility), weekly work duration, and work flexibility. Each individual is also assigned a driver's license depending on age, gender, and race. Using these characteristics, household income is computed as a function of race, presence of elderly individuals, education level of members of households, and employment industry of workers in the household. This is followed by a residential tenure model (own or rent) and a housing type model to assign each household to a single-family detached, single-family attached, apartment, and mobile home or trailer type of residence.

An important model in this simulation system is car ownership and type. This type of model in essence determines the predicted non-commercial regional vehicle fleet mix that is used as input to the emission estimation software. This is also particularly important for California because of the expected market penetration of electric cars and the incentive programs created at the state and federal levels in the US to promote this type of technology. A model system like this can be used to assess different incentive structures promoting environmentally friendly technologies in cars.

One of the inhibitors in building car ownership, car type and make models is the existence of many possible alternatives in this choice setting that includes many combinations of available alternatives. The solution here is to use the random utility model called Multiple Discrete-Continues Extreme Value (MDCEV) model developed by Bhat in 2005 [4], which is capable of modeling multiple vehicle holdings, body types, fuel types, age, and use (miles) simultaneously. This model includes 55 alternatives for body type/vintage (9 body types – Subcompact Sedan, Compact Sedan, Mid-size Sedan, Large Sedan, Coupe, Cross-utility car, SUV, Van, Pickup) and 5 vintage categories (New to 1 year, 2-3 years, 4-5 years, 6-9 years, 10-12 years, >12 years), 47 alternatives for vehicle make (Ford, Chevrolet, Toyota, Honda, etc.), and several hundreds of models across the many body type/make combination categories. Both groups of models include as explanatory variables household composition indicators and residential accessibility from section B. of this paper. The model is currently being enhanced to incorporate the additional dimension of fuel type used. It also has the potential of expanding the pool of options to include commercial vehicle fleets. After this step, each of the household vehicles is allocated to a driver in each household based on a probabilistic model using as predictors variables such as gender, education, and employment.

At this point of the simulation cascade on the left side of Figure 1 the model system produces the spatial distribution of all the residents by different social and demographic levels (including race) as well as employment and school locations assigned to each person. In addition, each household is assigned to a housing type. This resembles a complete Census of the resident population and can be done at any level of spatial aggregation. One could also draw samples from this population or proceed to the next step using 100% of the simulated residents. This is particularly convenient and useful in testing different policy scenarios to select a few to study in more detail. It is also possible to pay special attention to a specific subarea (e.g., a city) and perform more detailed analysis and modeling while keeping the rest of the region as an evolving background. The next set of model simulates the life of persons in a day.

D. CEMDAP

For each synthetically generated household and person within each household, daily activity and travel patterns are created in this block of models. To do so, a new modified version of the Comprehensive Econometric Microsimulator of Daily Activity-travel Patterns (CEMDAP http://www.ce.utexas.edu/prof/bhat/CEMDAP.htm) is used as the modeling engine that simulates activity-travel patterns of all individuals in the region for a 24 hour period along the continuous time axis. This model block creates synthetic schedules in two steps: (a) **the generation step** in which work and school activity participation and timing decisions are created, children's travel needs are estimated and an allocation of escort responsibilities to parents takes place, and independent and joint activity participation decisions are modeled; and (b) application of **the scheduling of activities** that produces the sequence of activities, with the departure and arrival times, activity duration(s), mode for each trip, and determination of the location of each activity. The models in this way create a complete description of the movement of each individual over space and time that is congruent with the movements of the rest of the household in which each person belongs. In this way, for each person, we have information about the type of activity, when, where, how long, with whom, in what sequence, and interrelationships with other persons and locations in the engagement pattern.

In the generation step, working and student adults are first passed through models that predict if they will work or go to school in the simulated day. Then, they are given start and end times for their work and school activity. Conditional on this a household level MDCEV model is then used to simulate combination of joint and solo activities for all persons. This provides an intra-household consistent schedule of activities and makes the entire simulation feasible because its formulation decreases the number of alternatives to simulate. Joint and solo activity durations are predicted for shopping, maintenance, social, entertainment, visit, active recreation, eat out, and other. In addition, duration for work-related and other serving passengers are also modeled. Activities are then arranged in tours (complete sequence of stops and trips starting and ending at the same location) and the modes used are predicted accordingly. The end result resembles Figure 3 in which we have two adults that go to work and a child going to school independently. In the evening they all go out for dinner. The trips, stops, activity types, activity start and ends times, modes are all determined by the simulation.

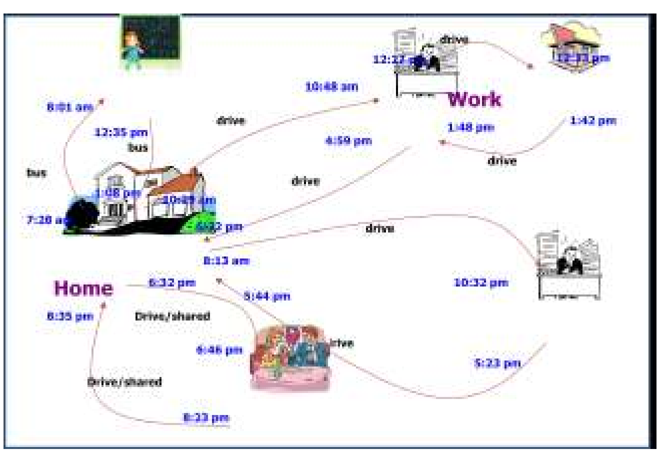

Figure 3 Example of a Daily Household Schedule

To the best of our knowledge, this is one of the most complete activity and travel pattern simulation models systems for a large metropolitan area that includes more than 17 million residents. The output of the model system contains more complete data than activity survey diary database because it recreates the entire population of the SCAG region and does not have any missing data for each person within each household. This is an important consideration for verification and validation because the simulation produces more information than is currently available from other sources to use as the gold standard.

At this point, the output from CEMDAP can be used in many different ways. For example, we have developed policy scenarios and studied their impact in timing decisions of individuals (e.g., advancing or postponing the stating of trips). We also coupled this output with the more traditional four-step model routines to perform traffic assignment and emission estimation. Moreover, we are also advancing along the path of using detailed routing algorithms that can track in the simulation individual vehicles and eventually compute emissions at fine spatial and temporal resolution.

E. Static Network Assignment

The output of CEMDAP in this block is aggregated (i.e., convert person trips to zonal sums of trips) for each of the 4019 zones to create a trip interchange matrix of Origins-Destinations among the traffic analysis zones at different time periods in a day (the four periods presented previously for the accessibility indicators). These are combined with similar ODs for heavy vehicles (trucks), and vehicular OD from and to ports and airports, as well as ODs of traffic generated outside the region. These additional OD matrices are the same as the four-step model used by SCAG in its 2008 Regional Transportation Plan. Using these trip interchange matrices traffic assignment (i.e., vehicular origins and destinations are converted into traffic volumes on the highway network based on behavioral and mathematical principles) produces estimates of traffic volumes on the links of the network allowing comparisons among different methods. Figure 4 is the output of traffic assignment during the AM peak period showing vehicle per hour.

This figure also shows the added output from SimAGENT, which is the number of persons at each location (in this case traffic analysis zone centroids) by each activity type engaged in for that specific hour (7:00 to 8:00 AM in Figure 4).

One way to compare outputs of this type is by examining how close to the observed traffic are traffic volumes estimated by different models. To this end, agencies identify strategic locations forming a ring around major attractors and use them as benchmarks (called screenline and contains 23 locations in this case). Figure 5 shows each of these locations and the relative 'closeness" reproduced by the trip based aggregated four-step model (which is iterated to match these daily traffic counts) and the SimAGENT, which achieves this closeness in one step with no additional adjustments or iterations.

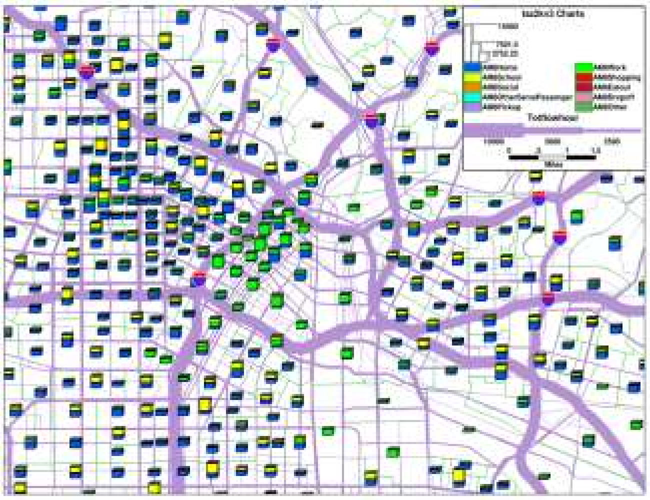

Figure 4 Assigned Traffic Output in SimAGENT and Person Presence with Activity Type Between 7:00 and 8:00 AM

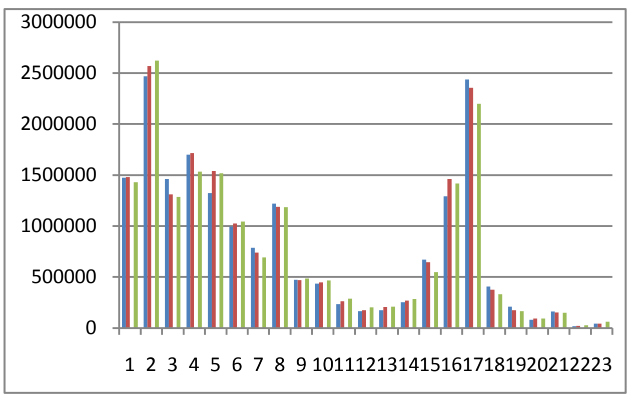

Figure 5 Screenline Daily Traffic Counts Comparisons at 23 different network locations (red is trip based model, green is SimAGENT, and blue is from observed counts)

At the time this summary was written, an alternative to this traffic assignment is not complete. We are experimenting with Dynamic Traffic Assignment (DTA) that is capable of replicating traffic at multiple times in a day while keeping track of the routes vehicles follow and ensure trip continuity (trips can start in one period of assignment and end in another). Our team is also developing a TRANSIMS and a MATSIM application that aim at recreating the detailed stops along routes vehicles experience in the simulation. Testing of emissions estimation was also performed at this stage of model development using the more traditional approach that is also used in conformity studies with the four step model but this time employing SimAGENT assigned traffic.

F. Mobile Source Emissions and Regional Fuel Consumption

The output of the traffic assignment by different time periods in a day is then used as input to the software EMFAC (a California region-tailored emissions calculating software) that produces estimates of fuel consumed and GHG emissions. Table 1 compares the EMFAC output between the two model systems (4-step and simAGENT). Both systems use the same truck traffic, special generators (ports and airports), and external to the study area traffic. The model year for regional fleet composition is 2003. We also experimented with different year by year fleet composition in EMFAC2007 and years of simulation showing a dramatic decrease in emissions but very low sensitivity to CO2 emissions due to the fairly constant fuel efficiency of the technology groups in the software.

Table 1 Comparison Between 4-step and SimAGENT (for the Year 2003)

Emissions and Fuel Consumption	4-Step Model*	SimAGENT Baseline*
Organic Gases (g/mile)	0.943	0.926
CO (g/mile)	9.498	9.348
NOx (g/mile)	1.929	1.955
CO2 (g/mile)	561.340	543.545
Gasoline (gallons/mile)	0.051	0.050
Gasoline (mile/ gallons)	19.377	20.203
Diesel (gallons/mile)	0.102	0.101
Diesel (mile /gallons)	9.833	9.893
Organic Gases (g/person-day)	22.291	21.333
CO (g/person-day)	224.553	215.388
NOx (g/person-day)	45.606	45.050
CO2 (g/person-day)	13271.790	12524.452
Vehicle Miles Travel/person-day	23.643	23.042

*Both models use the same regional fleet distribution for SCAG, which are classified by technology groups for exhaust and other emissions. In the passenger cars fleet these include approximately 70% in technology group 23 (model years 1997 to 2005 that are LEV), 20% in technology group 24 (1997 to 2005 that are ULEV), and 8.4% in technology group 31 that are Partial ZEV), and the rest are a mix of a variety of groups. Additional details can be found at EMFAC2007 version 2.3 User's Guide (http://www.arb.ca.gov/msei/onroad/downloads/docs/user_guide_emfac2007.pdf).

III. NEXT STEPS

Although the SimAGENT version described here appears to work well for the policy analysis purpose that was designed, it is currently being tested and strategically modified using external data for verification and validation. In parallel, a major effort is dedicated to the computation of transit accessibility indicators to improve the models' sensitivity to transit improvements. In addition, a battery of tests using SimAGENT coupling with microsimulation-based traffic assignment and routing techniques commenced.

REFERENCES

[1] Ye X., Konduri K., Pendyala R.M., Sana B., and P. Wadell (2009), "A Methodology to Match Distributions of both Household and Person Attributes in the Generation of Synthetic Populations." Paper presented at the 2009 Transportation Research Board Meeting, Washincton D.C. (also available at http://urbanmodel.asu.edu/popgen/papers/PopulationSynthesizerPaper_TRB.pdf).
[2] Chen, Y., S. Ravulaparthy, K. Deutsch, P. Dalal, S.Y. Yoon, T. Lei, K.G. Goulias, R.M. Pendyala, C.R. Bhat, and H-H. Hu (2010) Development of Opportunity-based Accessibility Indicators, Paper presented at the 90th Annual Transportation Research Board Meeting, Washington D.C., January 23-27, 2011.
[3] Eluru, N., A.R. Pinjari, J.Y. Guo, I.N. Sener, S. Srinivasan, R.B. Copperman, and C.R. Bhat (2008), "Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems," *Transportation Research Record*, Vol. 2076, pp. 171-182..
[4] Bhat, C.R. (2005), "A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions," *Transportation Research Part B*, Vol. 39, No. 8, pp. 679-707.