DEVELOPMENT AND TESTING OF A PROTOTYPE POPULATION EVOLUTION MODEL SYSTEM: A BALTIMORE METROPOLITAN COUNCIL (BMC) CASE STUDY

Ram M. Pendyala, School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287-5306. Ph: 1-480-727-9164; Email: ram.pendyala@asu.edu

Karthik C. Konduri, School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85287-5306. Ph: 1-480-727-9164; Email: karthik.konduri@asu.edu

Xin Ye, Department of Civil Engineering, California State Polytechnic University, 3801 W Temple Ave, Pomona, CA 91768. Ph: 1- 909-869-2488; Email: ye@csupomona.edu

Maren Outwater, Resource Systems Group, Inc., 917 E. Spooner Road, Milwaukee, WI 53217. Ph: 1-414-446-5402; Email: maren.outwater@rsginc.com

Birat Pandey, Baltimore Metropolitan Council, Offices @ McHenry Row, 1500 Whetstone Way, Suite 300, Baltimore, MD 21230. Ph: 1-410-732-0500 x1050; Email: bpandey@baltometro.org

Charles Baber, Baltimore Metropolitan Council, Offices @ McHenry Row, 1500 Whetstone Way, Suite 300, Baltimore, MD 21230. Ph: 1-410-732-0500; Email: cbaber@baltometro.org

---

INTRODUCTION

The implementation of advanced microsimulation-based activity-travel demand model systems generally requires that a representative synthetic population of the model region be generated so that activity-travel choices can be simulated at the level of the individual traveler. There have been many advances in the development of synthetic population generators with recent population synthesizers capable of controlling for both household and person level attributes in the synthesis process. A synthetic population generator utilizes known information about the marginal distributions of various household and person attributes in the population to repeatedly draw households from a sample such that the resulting synthetic population is representative of the true population in the region.

The application of microsimulation models for forecasting purposes calls for the evolution of a base year synthetic population into the future. While there have been some attempts at building population evolution models for forecasting purposes, the state-of-the-practice (not to mention the state-of-the-art) of population evolution remains in a state of infancy. In virtually all applications of advanced microsimulation-based activity-based travel model systems, a new synthetic population is being generated for each horizon year. This requires that an agency develop land use and socio-economic forecasts at the zonal level (or desired level of geography) for a number of control variables, and then use these forecasts to synthesize a new population for each forecast year. While this procedure is practical from an operational perspective, it would be preferable to have the ability to evolve a base year population over time – perhaps on an annual basis. A population evolution model would allow one to generate a synthetic population for any forecast year, without being restricted to those forecast years for which the agency has explicitly generated exogenous land use and socio-economic forecasts. In addition, a well-specified population evolution model provides the ability to reflect socio-economic dynamics that are inevitably prevalent in the real world, including such key phenomena as migration patterns, changing household composition, aging of the population, and economic cycles that invariably affect employment and occupation distributions. A population evolution model system provides the ability to reflect the effects of alternative socio-economic and demographic shifts (real or hypothetical) on the population composition, and consequently on the activity-travel patterns of a region. A population evolution model takes a base year synthetic population through a series of lifecycle events and cycles over time, and updates the population in each time step (say, a year) so that a new population is generated subject to pre-specified future targets or expectations regarding the population in a future year.

The Baltimore Metropolitan Council (BMC) has embarked on a major effort to develop the next generation of advanced travel model systems in the region. As a first step towards the development of an activity-based travel demand model system, BMC has implemented a synthetic population generator for the region. As the agency recognized the importance of being able to generate a synthetic population for any forecast year of potential interest, the agency also undertook the development of a prototype population evolution model that is capable of evolving a base year synthetic population through a series of lifecycle events and processes.

This paper briefly describes the population evolution model prototype and the open-source software system that has been developed to implement the evolution model. The population evolution model system will be implemented at the agency as part of the future activity-based travel model system and is therefore seen as a critical first step in the transition to advanced microsimulation-based models.

DESIGN OF A POPULATION EVOLUTION MODEL SYSTEM FOR BMC

As the first step in the development of a demographic modeling process for activity-based travel demand forecasting, the Baltimore Metropolitan Council (BMC) has implemented PopGen, a synthetic population generator capable of synthesizing a population while simultaneously controlling for both household and person level attributes (http://urbanmodel.asu.edu/popgen.html). The synthetic population generator has been integrated in the four-step travel modeling process as a means of obtaining robust cross-tabulations of household and person demographic attributes that may be used for cross-classification based trip production analysis. The synthetic population generated by PopGen will be used in the future in conjunction with an activity-based travel demand model system to simulate activity-travel patterns of all persons and households in the region.

The development of an initial population evolution model system was undertaken with the view that forecasting activity-travel demand in the future is best accomplished by evolving a base year population over time through the series of lifecycle events that define population dynamics. This section includes a brief description of the various components and the methods adopted in the population evolution model system for BMC. Some of the key criteria in the design of the model system included the desire to utilize data local to the BMC region or the State of Maryland, take advantage of methods programmed in PopGen, and cover the most important and critical lifecycle events and lifestyle choices of a population over time.

The population evolution model system includes an ***Emigration*** and an ***Immigration*** model. The migration of households and individuals into and out of a region is critical to accurately reflecting the dynamics of population growth and composition over time. The Maryland Department of Planning (MDP) provides data on emigration and immigration patterns for every county in the BMC model region (http://planning.maryland.gov/msdc/census/cen2000/Migration/Census2k_migration.shtml). This data includes information on the attributes of households and persons that have emigrated from and immigrated to each county in the region. The model system uses joint distributions from the census PUMS data and applies the iterative proportional fitting (IPF) method to generate expanded joint distributions of key household and person attributes that match the controls provided by the Maryland Department of Planning. The iterative proportional updating (IPU) algorithm is applied to all households in the base year synthetic population to determine the probability that a household should be eliminated (emigrated out) from the evolved population. Next, the same iterative proportional updating (IPU) algorithm is applied to all sample households in the PUMS data to determine the probability that a household in the PUMS data should be drawn into the synthetic population to reflect immigration. The IPU algorithm ensures that households are eliminated from or drawn into the synthetic population in such a way that joint distributions of key attributes of emigrant and immigrant households are matched. Immigrant households are located into traffic analysis zones according to simple multinomial logit based residential location choice models.

The emigration and immigration models are followed by models of ***Aging***, ***Mortality (Death)***, and ***Birth***. The model of aging is a simple deterministic model with each individual aged by one year in an annual time step. Data from the Centers for Disease Control and Prevention (CDC) can be used to determine mortality rates for individuals in different socio-economic groups. For the BMC prototype, mortality rates were computed by gender, race, and age using the CDC online tool (http://wonder.cdc.gov/mcd-icd10.html), and applied as a probabilistic rate-based model to simulate mortality (death) of individuals. In order to reflect birth of individuals, the BMC prototype includes a binary choice model that reflects a women's probability of giving birth to a child in any given year. While it is recognized that the birth of a child may entail the choice of two individuals – a male and a female – the fact is that only a woman actually experiences the event. As the BMC population evolution prototype is modeling the event (rather than the choice per se), the binary choice model is estimated and applied exclusively for women in the population. The binary choice model is estimated using data from the 2006-2010 National Survey of Family Growth (NSFG), which includes information about the history of child birth (among many other demographic phenomena of interest) for any woman in the sample (which includes 7356 females and 6139 males). Household attributes are updated upon the birth of a child (household size is increased by unity, for example). The gender of the new individual is randomly assigned as male/female; the race of the new individual is based on that of both parents (if applicable) or only that of the mother (if the race of the father is unknown).

The next model in the BMC Population Evolution Model System is the ***Education*** model. This model is intended to reflect the educational status of an individual over time and the highest level of education at which an individual would cease further educational pursuits. Using data from the census PUMS for the State of Maryland, a hazard-based survival model is developed to reflect the probability that an individual will drop out of school at any particular stage. Children are entered into school at the age of five years and then progressively advanced through school; however, the hazard-based survival model is applied in each year to simulate the individuals who choose to drop out of school. Once an individual has dropped out of school, the initial prototype does not accommodate a re-entry into the educational system; future refinements of the prototype will include the ability to simulate individuals going back to school after a hiatus. The hazard-based survival model considers gender, race, and age as covariates in determining the likelihood that an individual will drop out of school after attaining a certain educational status. A few simplifying assumptions are incorporated at higher degree levels to reflect the duration of time that it might take for an individual to complete a Masters degree or a Doctoral degree.

The BMC population evolution model prototype includes the ability to simulate the choice of residential arrangement for college-going individuals. The system includes a ***college student residential type*** model to reflect the possibility that children may leave the household and make alternative residential location arrangements at any point in the college education years. A child reaching and transitioning through college may make at least three possible living arrangements – continue living with parents at home, live off-campus as an independent individual, or live on-campus in group quarters (dormitories). A multinomial logit choice model of college student residential arrangement is estimated using census PUMS data for the State of Maryland. The application of this model provides for the updating and creation of new households where appropriate. If a college student chooses to live on-campus, then he or she should be moved from the original household to the group-quarter population.

If a college student chooses to live off-campus, then a new household should be created and located in a traffic analysis zone according to a probabilistic residential location choice model. When a child moves out, the household attributes need to be updated (household size decreases by unity, for example) accordingly. At this time, the prototype does not include a roommate matching model, thus placing each individual living off-campus into a separate household.

The next series of models reflect the economic status of the individuals and households in the population. These include models of ***labor force participation, occupation,*** and ***employee income.*** The ***labor force participation*** model is a binary choice model reflecting whether an individual 16 years or age or older is employed in a wage-earning occupation. The model is able to identify home-based employees in addition to traditional workplace-based employees. The binary choice model is estimated using the census PUMS data for the sample of individuals 16 years or older from the State of Maryland. At this time, the prototype considers labor force participation as a non history-dependent process, i.e., each individual makes a choice of whether to participate in the labor force on an annual basis. However, in reality, labor force participation is likely to be a multi-year process with considerable history dependency. A future development will involve estimating a hazard-based survival model to determine the duration for which an employee will remain in the workforce prior to exiting the labor force (regardless of reason). At this time, the absence of historical employment data prevents the estimation of such a history-dependent labor force participation model. The ***occupation*** choice model is a multinomial logit model that determines the occupation industry for each individual in the labor force and the model is estimated on a sample of employees drawn from the census PUMS for the State of Maryland. Once again, history dependency is ignored in the initial BMC population evolution prototype, with the model allowing employed individuals to exercise a choice of occupation on an annual time step. Finally, the employee ***income*** model is an ordered logit model that determines the ordered income category into which an individual falls. The model once again ignores history dependence and determines income for each individual on an annual basis.

The evolution model system includes a ***marriage decision*** and ***spouse matching*** model system to reflect household formation. Using data from the 2006-2010 National Survey of Family Growth (NSFG), which includes detailed marital history of a sample of females and males, binary choice models of the decision to get married are estimated separately for females and males in the synthetic population and applied to the unmarried individuals in the population on an annual basis. It is found that age, race, and employment status are significant explanatory factors of marriage decision. Once a pool of males and females who decide to get married in a year is formed using the marriage decision model, a ***spouse matching*** model is applied to form couples in a manner consistent with population characteristics. Using data from the NSFG, a joint distribution of the ages of males and females at the time of marriage can be constructed. Using data from the census PUMS, joint distributions of the race and education levels of couples may be constructed. Using marginal distributions on age, race, and education from the marriage decision model output, the joint distributions in the spouse matching model are updated using the iterative proportional fitting (IPF) algorithm. The iterative proportional updating (IPU) algorithm of PopGen is then applied to determine the likely matches between males and females in the pool of those who decided to get married in any year. It is possible that some males and females will remain without a match at the end of this process – an outcome that is potentially reflective of real-world situations where individuals are not able to find a partner even when they intend to get married. These individuals remain in the pool of potential partners for the next annual cycle.

Finally, the model system includes a ***household dissolution/separation*** model. Using data from the 2006-2010 National Survey of Family Growth (NSFG), a binary logit choice model is estimated to reflect the event of a divorce. Unlike marriage, divorce is an event that can take place if either the male or the female decides to dissolve the household. In order to simplify the process in the initial prototype, the binary logit model of household dissolution is estimated on and applied to females. When the model is applied, a new household is created with one of the spouses (chosen at random) and located in a traffic analysis zone according to a residential location choice model. Children are randomly assigned to one of the separating spouses in this initial prototype, although refinements to this approach are being implemented based on data regarding child custodial arrangements. Household attributes are updated to reflect the changes in household composition.

There are a variety of other choice processes and phenomena of interest in a population evolution context. However, the initial BMC prototype includes those described above with placeholders to accommodate future enhancements including the development of models to reflect such events as the acquisition of a driver's license or transit pass, the acquisition of information and communication technologies, the acquisition of bicycle(s), and the acquisition and evolution of the household vehicle fleet. In addition, the model system may be enhanced to constrain the evolutionary processes to match established exogenous forecasts developed by planning agencies. These model enhancement efforts are ongoing at this time.

EXPECTED RESULTS AND IMPLICATIONS FOR MODELING PRACTICE

The initial prototype of the BMC population evolution model system has been programmed within an open-source software enterprise and has been modularized such that it can be interfaced with any synthetic population generation model or activity-based travel demand model. The model system is being extensively tested at this time using the BMC model region as a case study application. For the base year of 2010, the BMC model area includes a population of 5,441,141 in a 10-county region encompassing 1421 zones. In the case study application, PopGen was applied to synthesize a base year population for 2010. Then, the evolution model prototype is applied to the base year synthetic population to simulate the population in each subsequent year. The simulated population in select future years will be compared against existing BMC forecasts to examine the extent to which the population evolution model is able to replicate established and agreed-upon future year demographic forecasts. The model system will also be tested with respect to its ability to evolve the population under alternative assumptions regarding various rates and demographic phenomena. For example, emigration and immigration rates, mortality and birth, and education survival rates will be altered to see how the evolution model is able to respond to changes in evolutionary patterns that might emerge in the future. Preliminary results obtained thus far from an application of the initial prototype are promising and the presentation at the conference will include complete results including comparisons against established agency forecasts, and results of sensitivity tests with altered input parameters.

Agencies around the country are very interested in deploying population evolution models that can be used in conjunction with both existing four-step travel demand models and emerging activity-based microsimulation models of travel demand. The BMC population evolution model prototype is capable of addressing this need and results from this study will help inform future developments in population evolution models. The open source nature of the software makes it possible for agencies and practitioners to experiment with the evolution model and determine how best to adapt it to their contexts. The presentation at the conference will include information and insights on how agencies may accomplish this.