Model Design Plan for BMC Activity-Based Model

final

report

prepared for

Baltimore Metropolitan Council

prepared by

Cambridge Systematics, Inc.

September 27, 2013

final draft report

Model Design Plan for BMC Activity-Based Model

prepared for

Baltimore Metropolitan Council

prepared by

Cambridge Systematics, Inc.
Hampden Lane, Suite 800
Bethesda, MD 20814

date

January 2014

Table of Contents

List of Tables

List of Figures

1.0 Overview

This document presents the plan for the new activity-based model that is being developed for the Baltimore Metropolitan Council (BMC). The development of this model is being led by a consultant team consisting of Cambridge Systematics, Inc. (CS), AECOM, Gallop Corporation, and Sabra, Wang & Associates. This report was written by Feng Liu, Arun Kuppam, and Thomas Rossi of CS. This report describes the overall model structure and the components of the model that will be developed by the CS team.

The region faces significant challenges in planning for a transportation system to meet the mobility needs of the region's residents. This planning requires a robust, sophisticated, and practical travel forecasting tool that is capable of analyzing the types of innovative planning and policy alternatives to help address these needs. A variety of policies and strategies have been proposed in the region to help improve and enhance the region's transportation system. Proper consideration of these requires that a number of types of analyses be done. These include, but are not limited to, the following:

- Analysis of road pricing and managed lanes actions;
- Analysis of proposed transit investments;
- Air quality conformity analysis;
- Planning for multimodal freight movements and alternatives to facilitate them;
- Corridor analysis and subarea planning;
- Environmental justice, studying the impacts of transportation policies and investments on different segments of the population; and
- Analysis of walk and bicycle travel.

The development of this model is being led by Cambridge Systematics (CS). This report describes the overall model structure and the components of the model that will be developed by the CS team. The report is organized as follows:

- **Chapter 1.0** provides an overview of the overall model structure, and the sequence of ABM components in estimation and application process.
- **Chapter 2.0** presents an overview of limited dependent variables and discrete choice models.
- **Chapter 3.0** lists the data items necessary for estimation of the models.
- **Chapter 4.0** presents the accessibility measures that will be used in the models.

- **Chapter 5.0** outlines the features of the population synthesizer that will be used with the new activity-based model.
- **Chapter 6.0** presents the long-term choice models while the daily activity pattern models are presented in **Chapter 7.0**.
- **Chapter 8.0** describes the tour-level models, and **Chapter 9.0** presents the trip (stop) level models.
- **Chapter 10.0** discusses the components of the new model system such as the special generator models that are not activity-based.
- **Chapter 11.0** discusses the trip assignment procedures.

1.1 OVERALL MODEL STRUCTURE

The proposed structure for the model is designed to address the requirements of planning for the regional transportation system. These requirements have been discussed with BMC staff through a series of meetings held over Summer and Fall 2013.

Figure 1.1 shows the structure of the proposed new activity-based model. The model components and the sections of this report in which they are discussed are listed in **Table 1.1**. Note that in the interest of keeping the figure readable, the logsum relationships from subsequent to previous model steps are not shown in the figure, nor are the non-activity based components (such as trip assignment).

Figure 1.1. Model Process Flow for Activity-Based Components

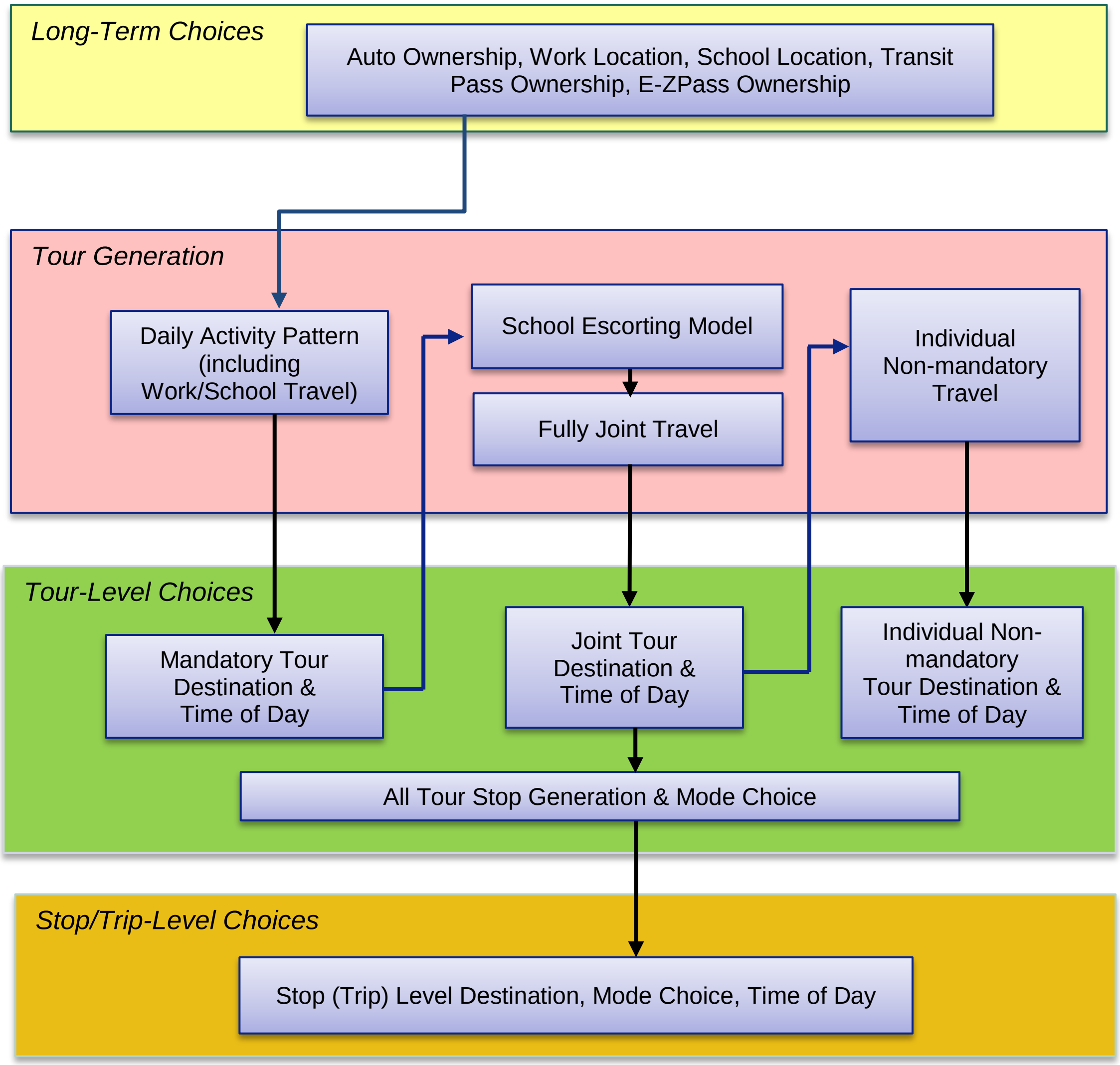


Table 1.1 Components of the Activity-Based Model

Model Name	Level	What is Predicted
Synthetic Population Generator	Households	Household size and composition, household income, person age, gender, employment status, student status
Regular Workplace Location	Workers	Workplace location zone
Regular School Location	Students	School location zone
Auto Ownership	Households	Number of autos owned
Transit Pass Ownership	Households	Whether the household owns a transit pass
E-ZPass Ownership	Households	Whether the household owns an E-ZPass transponder
Daily Activity Pattern	Person Day	0, 1, or 2 tours for each activity purpose 0, 1, or 2 stops for each activity purpose
Joint Travel	Households	Number of fully joint tours with 2 or more household members Which household members participate in each joint tour
School Escorting	Person (Household) Day	On which half tours a student is escorted to/from school Which household member escorts the student Whether escorting is done on a mandatory tour
Work Tour Destination Choice	Work Tours	For work tours – regular workplace or other work location (and its zone)
Work-Based Sub-tour Generation	Work Tours	Number and purpose of any sub-tours made during a work tour
Work Mode Choice	Work Tours	Main tour mode
School Mode and Time-of-Day Choice	School Tours	Main tour mode, the time period arriving at school and the time period leaving school (all school tours are to regular school location)
Work Time-of-Day Choice	Work Tours	The time period arriving at work, and the time period leaving work
Other Tour Time-of-Day Choice	Other Tours	Time period arriving at the primary destination and the time period leaving the primary destination
Other Tour Mode and Destination Choice	Other Tours	Primary destination zone and main tour mode
Intermediate Stop Generation	Half-tour	Number and activity purpose of any intermediate stops made on the half-tour, conditional on day pattern
Intermediate Stop Location	Trip	Destination zone of each intermediate stop, conditional on tour origin and destination, and location of any previous stops
Trip Mode Choice	Trip	Trip mode, conditional on main tour mode
Trip Departure Time	Trip	Departure time, conditional on time windows remaining from previous choices
Special Generators	Zone	Number of trips, trip end location, mode choice
Commercial Vehicle	Zone	Number of trips, trip end location
External Travel	Zone	Number of trips, trip end location
Highway Assignment	Vehicle Trip Table	Link volumes and travel times/speeds
Transit Assignment	Person Trip Table	Transit trips/boardings by route/stop

1.2 TourCast Structure

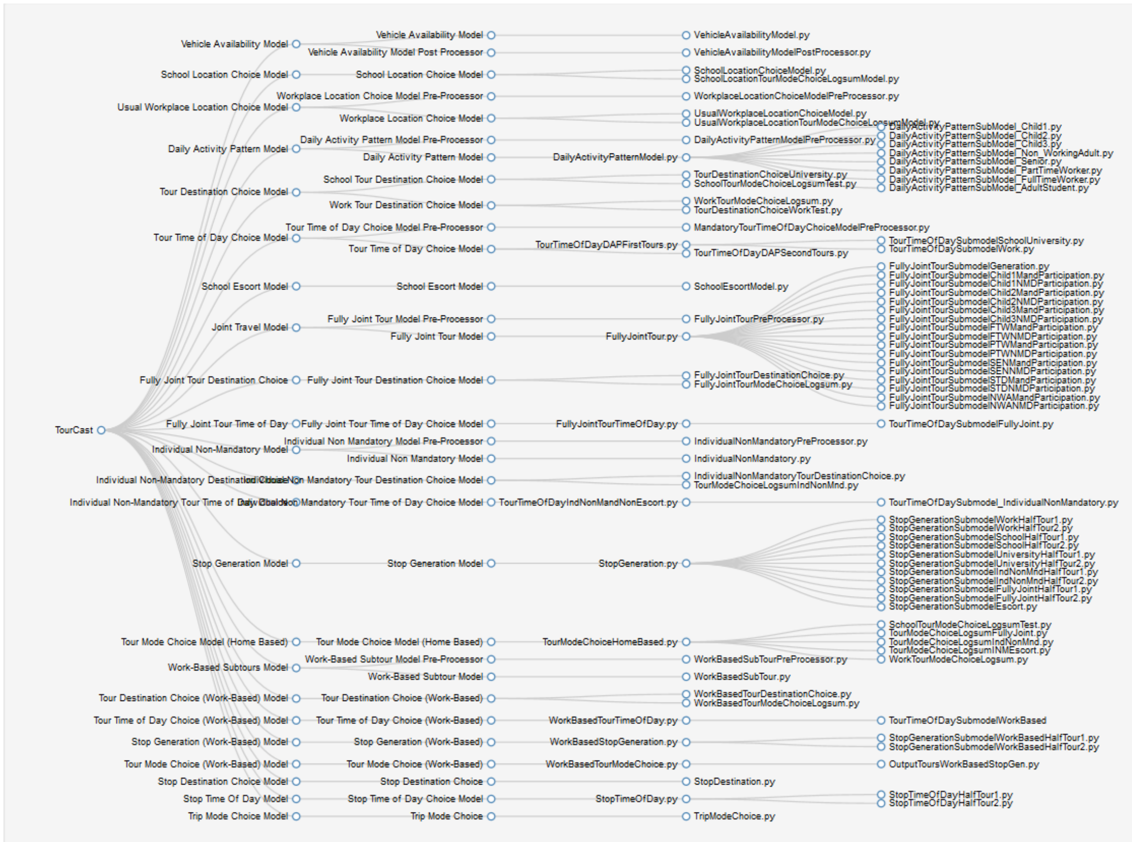
The CS team will implement and deliver the ABM model components using TourCast™, CS' activity-based modeling platform. TourCast, which has been implemented within the context of two ABMs with Cube, combines an extremely powerful computational engine with an established Python-based specification for configuring individual model components and model sequences. TourCast is not a black box, and BMC will have access to all TourCast source code and the flexibility to revise as desired.

TourCast was developed over a period of several years in response to ABM model development efforts for the Denver Regional Council of Governments (DRCOG, the Focus model), SHRP2 Project C10B, and ABM deployments for the Houston-Galveston Area Council (H-GAC, nearing completion) and Met Council (to be completed in 2014). It provides an ideal balance between configurability and model execution speed. TourCast is designed so that it can operate independently, or be integrated into commercial travel demand modeling platforms such as Cube, etc. The deployments at H-GAC and Met Council both rely on Cube integration, similar to the anticipated BMC deployment.

TourCast models critical travel behaviors such as time shifting, telecommuting, transit use, and the interactions of household members. TourCast was designed for practitioners and provides extraordinary control, flexibility, and ease of use. In summary, TourCast offers:

- Extremely fast model execution from a finely tuned model engine;
- Integration with commercial modeling software (including Cube);
- Outputs (intermediate and final) that can be sent to databases, GIS, or report generators;
- A modular, scalable software design with a service-oriented architecture (see Figure 1.2); and
- Desktop, server, and cloud configurations.

Figure 1.2. Overview of TourCast Software Code



1.3 TOURCAST COMPONENTS

As indicated in Figure 1.1 and Table 1.1, there are several model components implemented in TourCast. These components are all interlinked where the outputs of a model component are the direct inputs into a subsequent component. The sequencing of these components is different in application and in estimation as shown in Table 1.2. The estimation process begins with tour level mode choice models that are used to compute logsums by trip purpose to be used as key explanatory variables (accessibility or impedances) in several other model components. These are followed by stop or trip level models as these are conditional upon some of the choices made by the tour-level mode choice models. After this, the tour-level time of day and destination choice models are estimated, followed by person level models such as daily activity pattern, joint travel and school escorting models. The mandatory tour purposes are estimated before the non-mandatory purposes, while all the long term choice models are estimated towards the end of the estimation process.

As shown in Table 1.2, the application of these components starts off with the long-term choice models, after population synthesis, followed by daily activity pattern and tour-level models. After the tour-level models, the trip-level models are applied that determines the location, time of day and mode choice for every half-tour (or trip) in each individual tour. This final step produces a roster of trips for every individual in the population.

Table 1.2 Application Sequence of TourCast ABM Components

Application Sequence	Model Component	Model Applied
0	Population synthesis	Once for entire region
1	Usual Work Location Choice Model (Long-term)	For every worker
2	School Location Choice Model (Long-term)	For every child & (university student)?
3	Vehicle Availability Model (Long-term)	For every household
4	Daily Activity Pattern Model	For every individual
5	Tour destination choice model – mandatory	For every work tour
6	Transit Pass Ownership	For every household
7	E-ZPass Ownership	For every household
8	Tour time of day choice model – mandatory	For every mandatory tour
9	School escorting model	For every child making school tour
10	Joint tour model	For every household (with at least 2 traveling members)
11	Tour destination choice model - joint tours	For every joint tour
12	Tour time of day choice model - joint tours	For every joint tour
13	Joint tour participation	For every traveler in household
14	Tour generation - individual non-mandatory travel	For every individual (not stay-at-home)
15	Tour destination choice model - non-mandatory tours	For every non-mandatory tour
16	Tour time of day choice model - non-mandatory tours	For every non-mandatory tour
17	Stop generation model	For every individual half-tour (where stop was indicated in DAP phase) and all joint half-tours
18	Tour mode choice model – mandatory	For every mandatory tour
19	Tour mode choice model - joint tour	For every joint tour
20	Tour mode choice model - non-mandatory	For every non-mandatory tour
21	Stop destination	For every stop
22	Trip mode choice	For every trip
23	Stop time of day	For every stop

2.0 Discrete Choice Models

As described in Chapter 1.0, the new model will be built upon a series of discrete choice models that are estimated from the household travel and transit on-board surveys. There are several types of discrete choice models that vary by the form of the dependent variable that is predicted in each of the TourCast model components. This chapter describes the various types of discrete choice variables, the appropriate model structure to estimate these dependent variables, and an overview of the mathematical formulation of the major types of discrete choice models used in TourCast.

2.1 Dependent Variables

There are several types of dependent variables that are modeled in the travel demand forecasting processes. These can be classified into two broad groups, quantitative and qualitative.

Quantitative Dependent Variables

The **quantitative** dependent variables are assumed to have normal distributions. These are measured in terms of their quantities expressed in units of measurement. One of the commonly found quantitative variables in transportation models is the number of trips generated from a region, which is often expressed as a function of household characteristics (e.g., household size, number of workers) or employment (e.g., retail). These variables are typically estimated using regression modeling techniques that are usually appropriate in modeling aggregate (zonal) data.

For example, in the current BMC trip-based model, the number of attractions in the trip generation regression model is expressed as a function of the underlying employment[1]:

$$A_{HBW} = 0.838 * E_T$$

where:

A_{HBW} is the total number of home-based work attractions per TAZ for the Baltimore region;

E_T is the total employment for the modeled TAZ; and

---

[1] *Baltimore Region Travel Demand Model Version 4.0: Model Guide*, Prepared for Baltimore Regional Council, August 2011.

0.838 is the estimated coefficient (trip attraction rate) for the number of trip attractions per employed person in the Baltimore region.

Qualitative Dependent Variables

On the other hand, **qualitative** dependent variables are those that cannot be quantified, that is, their values or measurements represent discrete groups. These are also known as discrete choice or limited dependent variables as they are limited to their range because of some underlying stochastic choice mechanism[2]. These can further be classified into categorical and non-categorical variables

Categorical Dependent Variables

The **categorical** variables are those that categorize individuals or individual decision making into different categories. These variables can take on one of a limited, and usually fixed, number of possible values, and are often used to represent categorical data. In the transportation modeling literature, two types of categorical variables are widely modeled:

- Ordered;
- Unordered.

The **ordered** categorical dependent variable has choices that follow a specific order.

In the H-GAC model, the vehicle availability model estimates the number of vehicles available to a household as a long-term discrete choice variable, which is a special case of the ordered categorical dependent variable[3]. This is also referred to as the ***sequential*** categorical dependent variable where the second choice is dependent on the first choice, third choice is dependent on the first two choices, and so on.

The **unordered** categorical variables have discrete choices that are not ordered in any specific way but group individuals or individual decisions into finite discrete choices. A common example include the mode choice model where the choices are, for example, drive alone, shared ride, walk to transit, and auto to transit.

[2] Maddala, G. S. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press, 1983.

[3] Cambridge Systematics, Inc. *Vehicle Availability Model for H-GAC ABM.* Technical Memorandum to Houston-Galveston Area Council, November 2012.

Non-Categorical Dependent Variables

The other broad group of qualitative variables is the **non-categorical** variable which does not classify values into any categories but has a finite set of independent and discrete choices. For example, a destination choice model might predict the probability of an individual's destination choice that can include any of the TAZs in the modeling region. Each of the TAZs form the non-categorical variable choices.

In modeling disaggregate data, the underlying behavior of the individual decision making units is often found not to be continuous, and so the standard regression modeling techniques are inappropriate[4]. This is due to the qualitative discrete choice nature of the outcomes, which are best modeled using probability theory. The ensuing section describes the logit modeling techniques, which are widely known to analyze purely stochastic and non-deterministic systems using probabilistic approaches.

2.2 LOGIT MODELS

This section describes the logit model, the most commonly used discrete choice analysis method in travel forecasting[5]. This background is provided for understanding the parameters of logit models, rather than to provide a detailed discussion of logit model estimation, validation, and application. The principles and the basic mathematical formulation are presented, and the ways it can be used for choice analysis in travel demand modeling, particularly activity based modeling, are discussed. More detailed information about logit models can be found in Ben-Akiva and Lerman (1985)[6] and Koppelman and Bhat (2006)[7].

The basic idea underlying modern approaches to travel demand modeling is that travel is the result of choices made by individuals or collective decision-making units such as households. Individuals choose which activities to do during the day and whether to travel to perform them, and, if so, at which locations to perform the activities, when to perform them, which modes to use, and which routes to take. Many of these choice situations are discrete,

[4] Train, K. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*. Cambridge, Mass.: The MIT Press, 1986.

[5] Cambridge Systematics, Inc. *Travel Demand Forecasting: Parameters and Techniques, NCHRP Report 716*, Transportation Research Board, 2012.

[6] Ben-Akiva, M. and S. Lerman. *Discrete Choice Theory and Analysis*. Cambridge, Mass.: The MIT Press, 1985.

[7] Koppelman, F. and C. Bhat. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. Prepared for USDOT and FTA, June 2006.

meaning the individual has to choose from a set of mutually exclusive and collectively exhaustive alternatives.

The presentation of discrete choice analysis uses the principle of utility maximization. Briefly, a decision maker is modeled as selecting the alternative with the highest utility among those available at the time a choice is made. An operational model consists of parameterized utility functions in terms of observable independent variables and unknown parameters.

The utility represents the individual's value for each option, and its numerical value depends on attributes of the available options and the individual. In practice, it is not unusual for apparently similar individuals (or even the same individual, under different conditions) to make different choices when faced with similar or even identical alternatives. Models in practice are therefore random utility models, which account for unexplained (from the analyst's perspective) variations in utility.

The utility function, U, can be written as the sum of the deterministic (known) utility function specified by the analyst, V, and an error term, e. That is:

$$U = V + e$$

An analyst never knows the true utility function. In effect, the analyst always measures or estimates utility with error, and an error term of unknown size is always present in the analyst's specification of the utility function. This error term accounts for variables that are not included in the data set, or that the analyst chooses to omit from the model (e.g., because he cannot forecast them well), or that are completely unknown to the analyst.

When the true utilities of the alternatives are random variables, it is not possible to state with certainty which alternative has the greatest utility or which alternative is chosen. This inability is because utility and choice depend on the random components of the utilities of the available alternatives, and these components cannot be measured. The most an analyst can do is to predict the probability that an alternative has the maximum utility and, therefore, the probability that the alternative is chosen. Accordingly, the analyst must represent travel behavior as being probabilistic.

In logit formulations used in most travel demand models, the utility function for each alternative is a linear combination of variables affecting the choice. The utility equations have the form:

$$V_n = \beta_{n0} + \sum \beta_{nk} * x_k \tag{2-1}$$

where:

n = Alternative number;

V_n = (Deterministic) utility of alternative *n*;

β_{n0} = The statistically estimated constant associated with alternative n, essentially the effects of variables that influence the choice that cannot be included in the model due to inability to quantify or forecast, lack of data from the surveys used in model estimation, etc.;

β_{nk} = The statistically estimated coefficient indicating the relative importance of variable x_k on choice n; and

x_k = The value of decision variable k.

Variables in utility functions may be alternative specific, meaning that the coefficients must be different in each utility function (i.e., the values of β_{nk} cannot be equal for all values of n), or they may be generic, meaning that β_{nk} is the same for each alternative. In a logit model, the utility of one alternative matters only in terms of its value relative to the utilities of other alternatives.

Multinomial Logit Model

Logit is the most widely used mathematical model for making probabilistic predictions of mode choices. The simplest function used is the multinomial logit (MNL) formulation. In the MNL model, the probability of each alternative is expressed as:

$$Pn = \frac{exp(Vn)}{\sum exp(Vn')} \tag{2-2}$$

where:

P_n = The probability that alternative n is chosen;

exp() = The exponential function; and

V_n = (Deterministic) utility of alternative n (from Equation 2-1)

The MNL structure is the most commonly used formulation in activity based modeling and is used in location choice (or destination choice models), time of day choice (TOD), and daily activity pattern (DAP) models. In the case of location or destination choice models, the utilities are not just a function of level of service (LOS) variables such as time, distance and cost, but also land use and zonal variables that are represented in size functions.

Figure 2.1 shows an example of the alternatives in an MNL-based destination choice model, where all the internal TAZs Zone 1, Zone 2, …., Zone N are the alternatives under the main root.

Figure 2.1. Multinomial Logit Model Structure – Destination Choice Model

Size functions are used to measure the amount of activity that occurs at each destination zone and incorporate this into the utility of alternative variables. This is similar to the way in which trip attractions are used as a variable in conventional trip distribution models. This type of variable is frequently used in destination choice models to account for differences in zone sizes and employment levels. The size variables used in these models are: employment by type (office, government, industrial, retail, medical, education, restaurant, entertainment), college enrollment, and number of households. The size function is included in the utility equation of each destination choice (TAZ) as shown below:

$$U = Coeff_1 * Var1 + Coeff_2 * Var2 + Coeff_3 * Var3 + \ldots\ldots + Size\ function$$

where:

Var1, Var2, Var3 are explanatory variables (e.g., distance, intrazonal, mixed density, etc.);

Coeff1, Coeff2, Coeff3 are coefficients for *Var1, Var2, Var3*;

The size functions used in location choice models may be defined as follows:

$$Size\ function = LSM * \ln\{(Size\ variable1) + \exp(coeff_{22}) * Size\ variable2 + \exp(coeff_{33}) * Size\ variable3 + \ldots\ldots\}$$

where:

Size variable1 is the base variable (e.g., office employment);

Size variables 2 and 3 are other explanatory variables (e.g., retail, education employment);

$Coeff_{22}$ *and* $Coeff_{33}$ are coefficients for size variables 2 and 3; and

LSM is log size multiplier which is a coefficient that is multiplied by the entire size function.

Nested Logit Model

Another logit model form that is often used in activity based modeling is the nested logit model for mode choice models. Under a nested structure, the model pools together alternatives that share similarities, and the choice is represented as a multistep decision.

Consider an example with three alternatives, labeled 1A, 1B, and 2, where 1A and 1B are more similar to each other than either is to alternative 2. In the upper level of the nested model, the probability that an individual would choose alternative 1 (one of alternative 1A or alternative 1B) is given by Equation 2-3.

$$P1 = \frac{exp(V1)}{exp(V1) + exp(V2)} \qquad (2\text{-}3)$$

The probability of choosing alternative 1A conditional on choosing 1 is equal to:

$$P1A/1 = \frac{exp(V1A)}{exp(V1A\ \ )+exp(V2B\ \ )} \tag{2-4}$$

Thus, the probability of choosing alternative 1A is equal to:

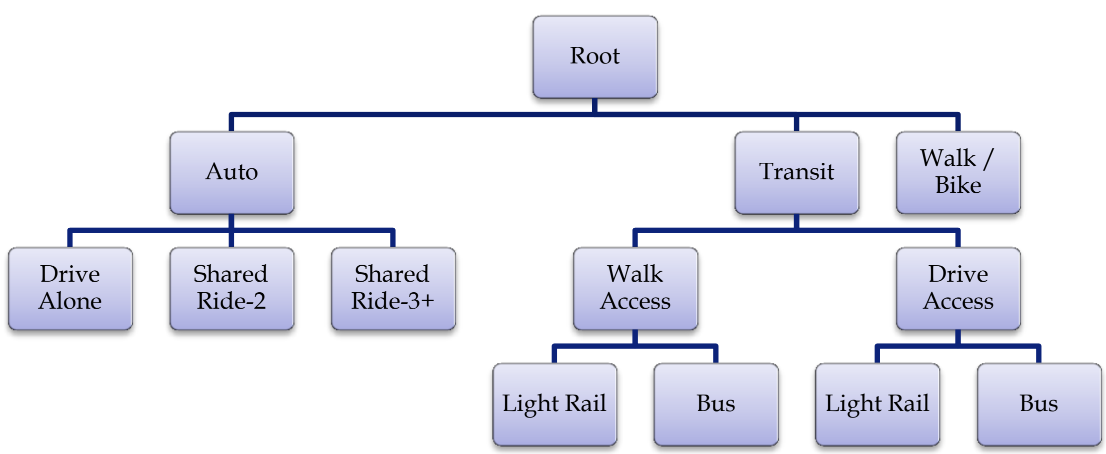
$$P_{1A} = P_{1A/1}\ \ * P_1 \tag{2-5}$$

In a nested model, the utility of an alternative in an upper level is a function of the utilities of its sub-alternatives. The utility for a nest m includes a variable that represents the expected maximum utility of all of the alternatives that compose the nest. This variable is known as the logsum and is given by the formula:

$$Logsum\ nest\ m = ln \sum_{All\_M\_in\_nest\_m}^{n} exp(UM) \tag{2-6}$$

Figure 2.2 shows a nested logit model structure where all the transit modes are pooled together under the 'transit' nest, and the auto modes are pooled under the "auto" nest. The individual transit sub-modes are further nested under the access mode nests – walk and drive access. The coefficients associated with the lower level nests are multiplied by the logsums to compute the expected utility.

Figure 2.2. Nested Logit Model Structure – Mode Choice Model



As an example, consider a model with a simple nest with two alternatives. If the utility of each alternative is the same, say 3.00 (indicating the choice probability of each is 50 percent), then the logsum is equal to ln [exp(3.00) + exp(3.00)] = 3.69, higher than the utility of either alternative. But if the utilities are, say, 5.00 for one alternative and 0.05 for the other (indicating a choice probability for the first alternative of over 99 percent), the logsum is equal to ln [exp(5.00) + exp(0.05)] = 5.01, only slightly higher than the utility of the superior alternative. Thus, the inclusion of a competitive alternative in a nest increases the expected maximum utility of all alternatives while the inclusion of a substantially inferior alternative has little effect on the logsum value.

Note that the logsum is equal to the natural logarithm of the denominator of the logit probability function (Equation 2-2) for the alternatives in nest *m*. A "nesting coefficient" of the logsum term is used in the utility function for nest *m*.

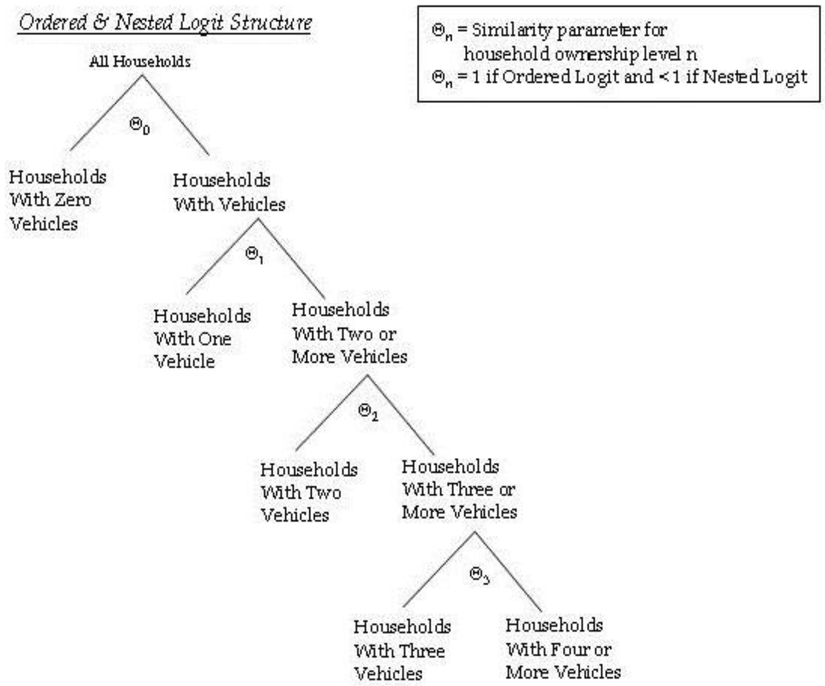
This coefficient must be between zero and one and should be statistically significantly different from zero and one. The primary advantage of nested logit models over (nonnested) multinomial logit models is that nested logit models enable one to reduce the intensity of the "independence of irrelevant alternatives" (IIA) assumption by nesting related choices. The IIA assumption, which is characteristic of all multinomial logit models as well as the lowest level nests in nested logit models, states that the probability of choices does not depend on alternatives that are not relevant. For example, assume in a mode choice model that there are three alternatives—car, red bus, and blue bus—with equal utilities.

Most people would choose between car and any bus, not distinguishing between the bus choices simply due to their color (i.e., they would be perfect substitutes for one another). But, given equal utility for all three of these choices, in a multinomial logit model framework the choice probabilities for each of the three choices would calculate as equal (1/3), leading to a greater probability of choosing any bus than the car alternative simply because the choice is being made among three equal alternatives rather than two (i.e., respecting the IIA assumption means one must not construct such choice sets with irrelevant alternatives).

Ordered Response Logit Model

Another type of logit model that may be used in activity based modeling is the ordered response logit (ORL) model. The ordered categorical or sequential dependent variables such as vehicle availability for a household is often modeled in a ORL structure. Figure 2.3 shows a sequential choice by households, first determining whether to have any vehicles at all, and then how many to have. The ORL structure also assumes that the similarity between the two choices available at each level of the choice structure (as reflected in the *theta* coefficient) is equal. The NL structure also assumes a sequential choice process, but does not assume that the choices at each level of the structure are considered equally. Instead, the *theta* coefficients of this model structure can vary to provide the best model fit to the available data. Experience in vehicle availability modeling shows that the ORL model usually provides a slightly better statistical fit than the MNL model while the NL structure generally shows no advantage.

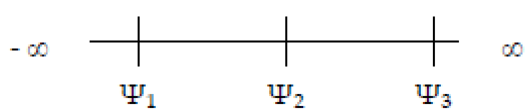
Figure 2.3. Ordered Response Logit Structure – Vehicle Availability Model



The ORL model uses a latent continuous variable y_n^* for modeling ordinal discrete data, where:

$$y_n^* = B'X_n + \varepsilon_n, \qquad \varepsilon_n \sim N(0,1)$$

If there are four independent variables (i.e. vehicle = 0, 1, 2, 3+), then y_n^* is divided into three cutpoints or intercepts as shown below.



If y_n is the observed discrete ordinal variable (i.e. $y_n = 0$ if number of vehicles in household = 0; $y_n = 1$ if number of vehicles in household = 1; etc.), then:

$$\begin{aligned} y_n &= 0 \quad if \quad y_n^* < \Psi_1 \\ &= 1 \quad if \quad \Psi_1 < y_n^* < \Psi_2 \\ &= 2 \quad if \quad \Psi_2 < y_n^* < \Psi_3 \\ &= 3 \quad if \quad \Psi_3 < y_n^* < \infty \end{aligned}$$

Knowing that ε_n is distributed by a standard normal distribution, and substituting in the equation for y_n^*, we can compute the probability of y_n given values of X_n.

$$P(y_n = 0) = P(y_n^* < \Psi_1) = P(B'X_n + \varepsilon_n < \Psi_1) = P(\varepsilon_n < \Psi_1 - B'X_n)$$
$$= \Phi(\Psi_1 - B'X_n)$$

$$P(y_n = 1) = P(\Psi_1 < y_n^* < \Psi_2) = P(\Psi_1 < B'X_n + \varepsilon_n < \Psi_2)$$
$$= P(\Psi_1 - B'X_n < \varepsilon_n < \Psi_2 - B'X_n) = \Phi(\Psi_2 - B'X_n) - \Phi(\Psi_1 - B'X_n)$$

$$P(y_n = 2) = P(\Psi_2 < y_n^* < \Psi_3) = P(\Psi_2 < B'X_n + \varepsilon_n < \Psi_3)$$
$$= P(\Psi_2 - B'X_n < \varepsilon_n < \Psi_3 - B'X_n) = \Phi(\Psi_3 - B'X_n) - \Phi(\Psi_2 - B'X_n)$$

$$P(y_n = 3) = P(\Psi_3 < y_n^* < \infty) = P(\Psi_3 < B'X_n + \varepsilon_n < \infty) = P(\Psi_3 - B'X_n < \varepsilon_n)$$
$$= 1 - \Phi(\Psi_3 - B'X_n)$$

As shown above, the ORL model estimates one equation over all levels of the dependent variable (i.e. the same set of *B* parameters is used for determining the probabilities of each alternative). This is known as the proportional odds assumption.

The Chi-Square Score Test for testing the proportional odds assumption determines whether one should fit a different set of explanatory variable parameters for *B* for each logit function, or whether the sets of explanatory variable parameters are equivalent. Therefore, the null hypothesis is that there is a common parameter vector *B* ($B_k = B$ for all *k*, where *k* is each alternative), and therefore, the assumption of proportional odds holds. If the null hypothesis is rejected, then one rejects the assumption of proportional odds. Distinct B_k parameter vectors should be estimated for each alternative, and therefore, the ordered logit model is not appropriate for the data. The number of degrees of freedom in the chi-square test is $t * (r\text{-}2)$, where *t* is the number of parameters and *r* is the number of dependent variable alternatives.

In the H-GAC ABM, the vehicle availability model was estimated using a ORL structure where the alternatives are[8]:

- 0-vehicles available
- 1-vehicle available
- 2-vehicles available
- 3-vehicles available
- 4-vehicles available
- 5-vehicles available
- 6-vehicles or more available

[8] Lemp, J., A. Kuppam, and T. Rossi. *H-GAC ABM Vehicle Availability Model Estimation Results*. Technical Memorandum to H-GAC, June 4, 2012.

The ORL model is specifically suited for choice contexts where the alternatives follow some natural ordering, as is the case for vehicle availability. The MNL treats each alternative distinctly and estimates the coefficients of linear (latent) utility functions specific for each alternative, while the ORL assumes a single latent function (modeled as a linear function of explanatory variables, similar to MNL) measuring the propensity for a household to own vehicles. The higher the latent variable for a specific household, the more likely it is for that household to own a higher number of vehicles. For H-GAC, both MNL and ORL models were estimated, and it was found that the ORL model had a better statistical fit than the MNL model.

3.0 Data Requirements

This section describes the data currently used within BMC's existing four-step model and the data items that will be needed for the estimation and development of the components that will comprise the BMC activity-based model (ABM).

3.1 NETWORKS/LEVEL OF SERVICE VARIABLES

Time Periods

The current BMC model has four time periods, as follows:

- A.M. peak (6:30 a.m. to 9:30 a.m.);
- Midday (9:30 a.m. to 3:30 p.m.);
- P.M. peak (3:30 p.m. to 6:30 p.m.); and
- Overnight (6:30 p.m. to 6:30 a.m.).

Recent traffic count data and household travel survey indicate a longer PM peak period than the current 3-hour period. An alternative PM peak period of 3:00-7:00 PM is proposed. There is also a desire to have traffic assignment results for peak hour for the planning purpose. Based on the traffic count data and survey data, the peak hours are proposed to be 7:30 – 8:30 AM and 5:00-6:00 PM.

The level of resolution for the time of day choice model should be as fine as possible to minimize aggregation error, but the impact of shorter periods on run times and the sufficiency of survey and count data at finer levels must be considered. It is proposed to use 30 minute periods for the new BMC model. We will also consider aggregation of periods outside peak travel hours although having periods of different sizes can have implications for model estimation. It will be important for model implementation that the level of service can be updated to reflect the final temporal resolution. This will probably require finding a method to adjust level of service skims on the fly, rather than storing up to 96 individual level of service skims, particularly as travel times and costs will most likely be stable outside the peak periods.

Highway Network and Skims

The current BMC model creates highway networks with link information, including but not limited to:

- Functional type;
- Roadway type;
- Area type;
- Managed Lanes;
- Toll code;
- HOV limit;
- Speed limit (Posted speed limit);
- Capacity (from lookup table);
- Free flow speed (from lookup table)
- Number of lanes; and
- Truck restriction.

Other highway-related inputs into the model include turn penalties, auto operating cost, and time-of-day factors. The link information in the highway network for the new model is expected to remain the same.

The highway network is used to produce the following skim data for each of the four time periods:

- Drive alone:
  - time, distance, and toll
- HOV:
  - time, distance, and toll

As discussed in **Section 8.3**, the new model will not have separate alternatives for toll and free auto modes in the mode choice model.

For model estimation, the currently available skim data will be used and expanded to the extent possible into disaggregate time periods. For model application, BMC may desire obtaining disaggregate time period data for tolls and managed lanes. Tour- and stop-level mode choice alternatives will be finalized based on survey data and overall modeling considerations. The skims will be adjusted to reflect the final mode choice alternatives.

Transit Skims

The current BMC model produces transit paths and level of service data for the following transit modes:

- Peak and off-peak drive to local bus;
- Peak and off-peak walk to local bus;

- Peak and off-peak drive to rail;
- Peak and off-peak walk to rail;
- Peak and off-peak drive to commuter rail; and
- Peak and off-peak walk to commuter rail.

The local-bus skims only include paths taken utilizing local bus. The rail skim includes all transit options except for the commuter rail, but with local bus slightly disfavored. The commuter rail skim includes all transit options.

The skims include the following data:

- Drive access time;
- Walk access and egress time;
- Initial and transfer wait time;
- In-vehicle time by mode and total in-vehicle time for all transit modes;
- Number of transfers;
- Transit fare;
- Drive access distance; and
- Mode number of first transit mode accessed on path

Based on the peer review recommendations, a "shallow" mode choice model structure will be implemented, with no transit submodes other than access modes (walk access, park-and-ride, and kiss-and-ride). Therefore, the only mode definitions for the set of transit skims to be developed for the new model to include transit with walk access, transit with park-and-ride access, and transit with kiss-and-ride access (see **Section 8.3**).

Nonmotorized Skims

Nonmotorized modes such as walking and biking currently are not included as mode options within the BMC mode choice model but rather as part of trip generation process. These modes will be included in the new mode choice model. TAZ-to-TAZ distance skims will be prepared for nonmotorized modes, which will include surface streets and bike and walk links in path building. The existing rough walk skim generated in the existing model will be used as a starting point for developing the new skims. An algorithm will be developed to disaggregate TAZ-to-TAZ distance skims to parcel-to-parcel distance skims using the TAZ-to-TAZ distance and parcel-to-parcel orthogonal distance.[9]

[9] Sacramento Area Council of Governments. SACSIM Activity-Based Model, 2008.

3.2 IMPEDANCE MEASURES

Highway Travel Time

Highway measures of impedance include travel time for each time period. In the current BMC model, travel time is measured under peak conditions, represented by the peak-period assignments, and under off-peak conditions, represented by the midday assignments. The parameters used to develop these assignments typically are:

- In-vehicle travel time for each origin-destination zone pair;
- Terminal time for each origin zone; and
- Terminal time for each destination zone.

The in-vehicle travel times are measured in minutes and estimated as a function of free-flow travel time and volume delay curves. Volume delay is determined as a function of the volume to capacity ratio for the time period being estimated. The current volume-delay functions will be used and adjusted if necessary during the assignment validation process.

Terminal times represent the time it takes to travel from one's origin to one's vehicle and from one's vehicle to one's final destination. Terminal times, derived from a look-up table, are higher in denser urban areas, where it is necessary to park further away from the final destination. Terminal times are fixed by traffic analysis zone and the current times will be retained and modified, if necessary.

Network-based models generally calculate the travel time between zones (interzonal time) as a function of the travel time required to traverse from one zone to another. Intrazonal travel times cannot be calculated in this manner, because the modeled trips do not use the roadway network and the time within a zone would be calculated as zero. Intrazonal time is computed as 75% of the average time to the nearest three zones in the Baltimore region and as 50% of the time to the nearest zone in the Washington region.

Travel Cost and Values-of-Time

The highway assignment model uses travel time and toll with a value-of-time in the calculation of generalized costs, which serve as the basis of the skimming and path-building. Since the model system uses generalized costs as inputs to various ABM components (such as destination choice, mode choice, and time-of-day choice), the assumed value-of-time in the highway assignment is related to the rest of the model system in many ways. In the new model, each traveler will have his or her value-of-time simulated from a distribution (see **Section 8.3**) . For highway assignments (see **Section 11.2**), however, average values-of-time will need to be used since aggregate trip tables will be assigned.

Based on previous research and a review of available information on travelers' values-of-time from various sources[10][11][12][13][14][15][16], the specific estimates of the values-of-time will be made.

Distance

Distance between an origin and a destination is calculated total of the link lengths used in the shortest path. If there are multiple paths used, an average for all paths used in the highway network will be used. This can vary by time period since the path from an origin to destination can be affected by congestion in the system. Distance is estimated in miles.

Transit Impedance

Transit impedance is measured in terms of travel time. Transit travel time is estimated for the same peak and off-peak conditions as the highway travel times, using a.m. peak-period and daily assignments. Transit travel impedance can be comprised of the following components:

- In-vehicle travel time;
- Access time;
- Egress time;

---

[10] Parsons Brinckerhoff, Northwestern University, Mark Bradley Research & Consulting, University of California at Irvine, Resource System Group, University of Texas at Austin, Frank Koppelman, and Geostats. *Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand*. SHRP2 Report S2-C04-RW-1, Transportation Research Board, Washington, D.C., 2013.

[11] Sall, E., E. Bent, B. Charlton, J. Koehler, G. Erhardt. Evaluating Regional Pricing Strategies in San Francisco – Application of the SFCTA Activity-Based Regional Pricing Model. Proceedings of the 89th Annual Meeting of the Transportation Research Board (TRB), Washington, D.C., 2010.

[12] Ryan, J. *Travel Forecasting for New Starts, the FTA Perspective,* prepared for the Federal Transit Administration, April 7, 2004.

[13] Cambridge Systematics, Inc. *Assessment of Southeast Florida Road User Costs, Task 1 Technical Memorandum, Travel-Time Values,* prepared for the Florida Department of Transportation for the Southeast Florida Road User Costs Study, 2005.

[14] Cambridge Systematics, Inc. *Results of the Southeast Florida Road User Cost Travel Time Value Survey*, conducted by the Florida Department of Transportation, January 2005.

[15] Brownstone, David and Kenneth Small. *Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations*, University of California at Irvine, June 18, 2003.

[16] Cambridge Systematics, Inc. *Washington State Comprehensive Tolling Study,* prepared for the Washington State Department of Transportation, September 2006.

- Total wait time;
- Transfer time;
- Initial wait time;
- Transfer wait time;
- Number of boardings; and
- Total transit time.

These measures are typically calculated separately for the two primary modes of transit: 1) walk access; and 2) auto access. There are a series of parameters that will affect the development of transit travel times, and these are headways, boarding time, weights for wait time and boarding time. These reflect constraints on travel time (such as the maximum time to wait), as well as factors that account for different perceptions of time (such as the difference in perception between time spent waiting for a bus compared to time spent riding a bus). Travel surveys have shown that time spent waiting for a transit vehicle is more onerous than time spent riding on a transit vehicle.

It is important that estimates of transit impedance be as accurate as possible from the very beginning of the model development process. The effort involved with preparing a model estimation dataset is significant, so all aspects of the transit impedance matrices should be carefully reviewed prior to their use. This review is typically done by developing a transit person trip table from a large scale transit onboard survey, assigning these trips to the transit network, and revising assignment parameters until assigned ridership corresponds to counts to an acceptable degree of accuracy. Key parameters to be answered during this process include:

1. What are the appropriate access choices? The current choice model allows walk access and drive access (Park-and-Ride and Kiss-and-Ride).

2. What are the appropriate transit line-haul choices? The existing model is set up to separately model local bus, express bus, rail, and commuter rail. Recent research[17] indicates there may be alternate ways of identifying transit submodes that may be helpful in reducing the values of alternative specific constants in mode choice models. As discussed above, the new model will follow the recommendation of the peer review panel to have a "shallow" mode choice structure with transit submodes defined only by access mode.

[17] Outwater, M., J. Lobb, B. Sana, N. Ferdous, B. Woodford, D. Schmitt, J. Roux, C. Bhat, R. Sidharthan, R. Pendyala, and S. Hess. *Characteristics of Premium Transit Services that Affect Choice of Mode*. TCRP H-37 Final Report, Transit Cooperative Research Program, Transportation Research Board, 2014 (forthcoming).

3. What are the appropriate path parameters? Test assignments will be used to determine the relative importance of in-vehicle time, waiting time, walking time, transfers and fare. Assignment results will be compared to observed ridership patterns and adjusted as necessary. During this phase of the analysis, it may be determined that a simple assessment of the minimum generalized cost of transit is insufficient to generate appropriate paths. In which case, strategies for favoring some types of transit paths over others will be explored. Options include:

   - Testing the desirability of using different weights for walk time to reflect the fact that walking in some areas (downtown or transit-oriented development locations) is more pleasant than walking in areas involving large blocks, fast moving arterials, or the absence/limited availability of sidewalks.
   - Implementing sub-mode preferences for individual links on a path by discounting the perceived travel time, transfer time, or boarding time for selected "premium transit" modes. (e.g., light rail or express bus).

4. How accurate are the estimates of transit running times? Transit running times must be based on the underlying highway operating speeds and a logical relationship between highway speed and transit speed. Although transit buses often operate less rapidly than cars while in motion, the difference may not amount to more than a 10 percent delay as compared to automobile traffic. The biggest difference in travel times is associated with the need to stop to receive and discharge riders. If this is the case then the appropriate functional form is transit time= (automobile time * 1.1) + (x minutes/passenger) * (anticipated boarding + anticipated alighting passengers). Running time must be calibrated so that the function form is logical and resulting estimates of running time match actual travel times. Alternatively, the following general function can be used to represent bus speeds:

   Bus speed = highway link speed (congested for peak, free flow for off peak) + stop delay time.

   This stop delay time would be dependent on stop spacing and thus some average value of stop delay based on each area type (as a proxy for population and employment density which in most cases drives stop spacing and transit ridership) and facility type would be implemented. These stop delays would need to be calibrated on some level to scheduled running time.

   At the conclusion of this analysis, the modeling team will have an understanding of the key dimensions that affect how transit choices are best represented in the model. Depending on the simplicity or complexity of the resulting trip-based choice structure, decisions can then be made regarding the development of the tour-based mode choice structure.

Non-Motorized Impedance

Distance, rather than time, is proposed as the only impedance measure for non-motorized travel. Walk and bicycle travel times are highly dependent on the speed of the traveler, which in turns depends on the traveler's physical condition and personal preferences, which are not modeled (and for which data are unavailable). Unlike highway and bus speeds, which depend on travel conditions, walk and bicycle speeds are usually determined from the traveler's (mostly) unobserved characteristics. Average speeds could be used, but this is mathematically equivalent to using only distance.

It is possible to obtain non-motorized distance information from the highway network distance skims. There are, however, two major issues with the use of these skims:

1. There are many paths that use facilities other than the roads in the highway network.
2. Non-motorized trips are usually short, and their lengths can vary substantially from the average distances represented by centroid to centroid skims. Furthermore, intrazonal travel is more prevalent for non-motorized travel, and the skims provide no data on intrazonal distances.

Highway skims can be improved if an "all streets" network is used, but non-motorized trips may use facilities that are not roads at all, and issue #2 is not overcome.

There are alternatives to the highway skim distance. For the model estimation data set, point to point distances can be determined from the coordinates of the trip ends, either as straight line or orthogonal (X plus Y) distances. For model application, point locations could be simulated for trip ends within a zone. This type of procedure can be applied for intrazonal as well as interzonal trips.

There has been some experimentation into methods that consider the presence of barriers, such as waterways or railroads, that are not considered when using straight line or orthogonal distances. Such methods will be examined to determine whether they would be practical for use in this project.

It should be noted that these issues also apply to the walk access and egress portions of transit trips. The same distance calculation procedures used for non-motorized travel can be applied for transit walk access and egress, but these distances must be converted to time using average speeds since transit times are required.

3.3 Zone/Land Use Data

The current BMC model uses a 1,809 travel analysis zone (TAZ) system for the 2010 model year. The zonal-level data available as inputs into the model include:

- Regional Planning District (RPD)
- Population
- Institutional Group Quarter Residents
- Non-Institutional Group Quarters Residents
- Number of Households
- Median Household Income
- Labor Force
- Employment by Categories (Retail, Office, Industrial, and Other)
- Total Acreage
- Identification of Truck Activity Zones
- School Enrollment
- Percent Zero-Car Households

The number of households, population, median household income, and labor force data are used to create the distribution of households by five sizes, the distribution of households by four income categories, and the distribution of households by four worker categories. These demographic stratifications are based on the lookup tables that contain the percentage of persons per household for each average household size, the percentage of households within each income category by the ratio of TAZ median household income to the regional median household income, and the percentage of households for average workers per household. Two joint distributions are used to stratify the households in a zone by size and income (5x4) and by workers and income (4x4). These market segmentations can be used within the ABM as control totals for the population synthesizer.

The current BMC model classifies employment into four categories (retail, office, industrial, and other). For the ABM development, a more detailed classification will be considered, including education, government, office, retail, restaurant (food service), entertainment, medical, other services, industrial, and other.

For the estimation and application of the new ABM components, there could be additional zonal variables that will need to be compiled and created with the help of BMC staff. These could include population and/or employment density, area of zones, parking supply data, auto and transit accessibilities, and road/intersection density. BMC also has some of these variables available at the parcel level.

Table 3.1 shows potential variables that could be developed at the parcel/point level, which are useful to the ABM development. Incorporating parcel/point-level data into the BMC ABM will require data preparation for the base year for model estimation, for future year model application, and software development. For the model estimation and base year model development, existing data sources appear to provide the basis for developing most data items, except for parking supply variables. In terms of the quality of existing data, the location and size of a parcel are expected to be pretty accurate. The employment data are also of reasonably good quality as BMC has long processed and cleaned the raw ES202 data. The availability of some other key variables such as the availability and costs associated with parking in different parts of the region and usage of parking facilities, especially those used to access transit routes will be discussed with BMC staff prior to model development.

Table 3.1 Potential Parcel-Buffer Level Variables

Potential Parcel-Level Variables	Details
Employment density	Density within ¼ and ½ mile buffer of a parcel by employment type: Education, Government, Office, Retail, Restaurant (Food Service), Entertainment, Medical, Service, Industrial, Other, Total
Household density	Household density within ¼ and ½ mile buffer of a parcel
Land use mix	Land use mix/diversity within ¼ and ½ mile buffer of a parcel
College student density	University student enrollment density within ¼ and ½ mile buffer of a college
School student density	K-12 student enrollment density within ¼ and ½ mile buffer of a school
Urban design	Street intersection density (street pattern or design variable) such as buffered density or number of intersections with 1-leg, 3-leg and 4+-leg
Accessibility to transit	Distance to nearest transit stop/station

It should be noted that based on the recommendation of the peer review panel, the same level of detail for zones should be used within the internal model area for the new model that is outside the BMC region (i.e., in the MWCOG region) as is used inside the BMC region. This means that the same zone/parcel information must be produced for the area for the new model that is outside the BMC region.

3.4 SURVEY DATA FOR ESTIMATION AND VALIDATION

The household survey data are the first data items required for the model estimation. The surveys were conducted in both Baltimore and Washington regions in 2007/2008, using the same survey designs and generating a combined total of nearly 15,000 completed samples (households). The survey data are organized in three relational databases described as the household file, person file and trip file. The trip files will be reprocessed using several criteria related to activity types, joint travel and intermediate stop-making to develop tour profiles for survey respondents. Each of these files provides key variables necessary to develop tour, trip, and long-term decision-making models. Table 3.2 shows the variables that are available or will be derived from the data and used in model estimation. Other variables may be derived from those listed in the table, but the basic information is fully contained in this table.

Table 3.2 Data Items from Survey Dataset

Description	Details
BASIC PERSON & HH VARIABLES	
Household ID number	Survey ID field
Person ID number	Survey ID field
# people in household	
# vehicles in household	(Dependent variable for auto ownership model)
Total household income level	Categorical household income
Gender	1=male, 2=female
Age	Years
Employment status	1=employed full-time, 2=employed part-time, 0=not employed
Student status	1=enrolled full-time, 2=enrolled part-time, 0=not enrolled
Type of school enrolled in	1=preschool, 2=K-12, 3=post-HS, 0=not enrolled
Relationship to respondent	1=Head, spouse, partner, 2=other HH member, 3=visitor
DERIVED PERSON & HH VARIABLES	**DERIVED FROM BASIC PERSON & HH VARIABLES**
Person type	Derived (e.g., 1=full-time worker, 2=part-time worker, 3=retired 4=other adult, 5=university student, 6=driving age high school student, 7=child age 5-15, 8=child age 0-4)
# employed HH members	Derived by adding across HH members
# student HH members	Derived by adding across HH members
# HH members by person type	Derived by adding across HH members
PERSON/HH LOCATION VARIABLES	
Household residence ID number	Survey ID field
Household residence X coordinate	Geocode
Household residence Y coordinate	Geocode
Household zone	Geocode (Dependent variable for population synthesizer)
Regular work location id	Survey ID field
Regular work X coord.	Geocode
Regular work Y coord.	Geocode
Regular work zone	Geocode (Dependent variable for regular work location model)
DAY PATTERN-LEVEL VARIABLES	**CREATED BY TOUR & PATTERN FORMATION CODE**
# home-based tour records	
# home-based tours by tour type	Dependent variables for day activity pattern models
# work-based subtour records	
# intermediate stops by stop purpose	Dependent variable for day activity pattern models

Table 3.2 Data Items from Survey Dataset (continued)

Description	Details
TOUR-LEVEL VARIABLES	**CREATED BY TOUR & PATTERN FORMATION CODE**
Tour ID number (in priority order)	Created ID field
Subtour parent tour ID (work based subtour only)	Created ID field
Subtour ID within parent tour (work based subtour only)	Created ID field
# of subtours within tour	Dependent variable for subtour frequency/purpose model
Primary destination activity purpose	(1=work, 2=school, 3=serve passenger, 4=personal bus., 5=shopping, 6=meal, 7=social/recreation)
Tour origin outbound departure time	
Primary destination arrival time	Dependent variable for tour times of day model
Primary destination departure time	Dependent variable for tour times of day model
Tour origin return arrival time	
Primary destination location id	Survey ID field
Primary destination X coord.	Geocode
Primary destination Y coord.	Geocode
Primary destination zone	Geocode (Dependent variable for tour destination model)
Tour primary mode	Codes to be decided (Dependent variable for tour mode model)
# trips in outbound tour half	Dependent variable for tour stop frequency/purpose model
# trips in return tour half	Dependent variable for tour stop frequency/purpose model
TRIP-LEVEL VARIABLES	**CREATED BY TOUR AND PATTERN FORMATION CODE**
Trip tour half	1 or 2, Created ID field
Trip ID within tour half	Created ID field
Trip origin activity purpose	Same codes as primary destination activity purpose
Trip destination activity purpose	Same codes as primary destination activity purpose
Trip origin location ID	Survey ID field
Trip origin X coord.	Geocode
Trip origin Y coord.	Geocode
Trip origin zone	Geocode (Tour destination, or destination of previous trip)
Trip destination location ID	Survey ID field
Trip destination X coord.	Geocode
Trip destination Y coord.	Geocode
Trip destination zone	Geocode (Tour origin, or dependent variable for stop location)
Trip mode	Same codes as tours (Dependent variable for trip mode model)
Trip origin departure time	Dependent variable for trip departure time model
Trip destination arrival time	

The data in Table 3.2 are split into six main categories:

1. **Basic person and household variables**. These are the truly exogenous variables. In application, these will be taken from the U.S. Census Public Use Microdata Sample (PUMS) records in the synthetic sample, and so certain variables from the household survey may need to be recoded in a way that is consistent with PUMS coding.

2. **Key-derived person and household variables** . These variables are developed using the definitions of the basic variables. One such important variable is person type, which has been found to be very useful in other activity-based models. While the specific person type categories for this model will emerge from an analysis of the household survey data, typical classifications include full-time worker, part-time worker, driving-age child, child below driving age (and occasionally infant as a separate category), nonworking adult, and senior. Note that additional variables can be derived from these and used in specific models – e.g., a dummy variable for female adults with one or more children aged 0-4.

3. **Person and household location variables**. This is the start of the endogenous variables in the model system. In application, the household location (at the zone level) will be predicted by the population synthesizer, and the regular work zone will be predicted by the choice models.

4. **Day pattern-level variables**. These are created by the code that processes trips into tours. They are person-day counts of the numbers of home-based tours and intermediate stops for each of the seven proposed activity purpose types, plus the count of the number of work-based subtours made. In application, these will be predicted by the day activity pattern model(s).

5. **Tour-level variables**. These are also generated by the tour formation code and contain all the variables needed to model a tour: purpose, timing, destination, mode, the number of intermediate stops on each half-tour, and the correspondence between work tours and subtours. In application, these will all be predicted by the various tour-level models.

6. **Trip-level variables**. Some of these variables are also created by the tour formation code. The variables include trip origin and destination location and purpose, trip departure and arrival time, and trip mode. In application, these will either already be known from the tour-level predictions (e.g., the locations for half-tours with no intermediate stops), or will be predicted by the trip-level models.

To prepare these data items, tour formation code will be used, which sets up the data in these structures. Also note that there are other variables in the survey that might be interesting from a behavioral sense, but there is no means of easily forecasting them, and so it is not proposed to include them in the estimation data or models. These include:

- Residence building type;
- Tenure at residence;
- Auto make and model;
- Auto fuel type;
- Auto own/lease type;
- Bicycle ownership;
- Driving license status;
- Job occupation and industry;
- Job workplace type;
- Job flextime status;
- Travel disability;
- More detailed activity purpose coding than used in models;
- More detailed mode coding/combinations than used in models;
- Activity place type;
- Which vehicle each trip was made in; and
- Self-reported parking cost paid and payment method.

Transit On-Board Survey

BMC also provided data collected from a survey of transit riders in 2007. The survey was conducted on-board transit vehicles including local, MARC, Metro Subway, Light Rail, and Commuter services in the region. Derived from roughly 13,000 usable questionnaires, the survey data contain information on the respondent's current transit trip. This includes trip starting and ending location, trip start time, time spent waiting for the transit vehicle, access and egress modes, as well as several socioeconomic characteristics of the respondent like gender, age, and vehicle availability.

Transit person trip table(s) will be developed from the on-board survey and assigned to the model's transit network to evaluate the modeled path-building procedures, to evaluate access and egress coding for walk, park-and-ride and kiss-and-ride, and compare prediction-success tables between modeled and observed travel patterns.

3.5 OTHER DATA

Regional Air Passenger Survey

The current travel demand model has an air passenger model that was developed and calibrated based on regional air passenger surveys. The Washington-Baltimore Regional Air Passenger Survey has been conducted since 1981 and more recently on a biannual basis. The 2011 Survey collected responses of approximately 21,000 air passengers at Ronald Reagan Washington National, Baltimore/Washington International Thurgood Marshall, and Washington Dulles International Airports. Questionnaires include satisfaction with airport use, trip purpose, trip originations, trip purpose, mode of access, trip destinations, passenger household income, trip pattern by time-of-day and characteristics of air passengers originating from the Baltimore and Washington regions. This survey will be used for re-calibrating the air passenger model.

4.0 Accessibility Measures

There are a wide variety of options and modeling uses for accessibility measures. In some cases, they are simply used as a proxy for logsums where it is impractical to calculate logsums, given the nature of some of the model components. In many cases, the final form of the accessibility measures for each model component will not be known until model estimation has been undertaken; this is due to the fact that accessibility measures tend to be very collinear and often only one will emerge from model estimation. This may not be as severe a problem for models using overall aggregate accessibility, but it may be problematic for mode choice models where the distinction between modes is highly relevant. In these cases, relative accessibilities are sometimes calculated, i.e., calculate the accessibilities where transit access times are within 50 percent of auto access times, which might then be paired with an accessibility measure focusing solely on zones inaccessible by transit (or nonmotorized modes). These are only examples of potential approaches, and do not reflect our official recommendations.

Accessibility measures will be calculated three different ways for different model components:

1. All-day accessibilities used for the long-term models and the day pattern model;
2. Tour-level accessibilities reflecting an intermediate level of temporal resolution used for the tour-level models; and
3. Time-period specific accessibilities used for the time-of-day models and the stop models.

The primary difference between the approaches will be the composite nature of the all-day and tour-level accessibilities. Typically, the accessibility measures will be calculated at a refined level, as detailed as the skims will permit, but then will be "rolled up" depending on the precise nature of the measure being constructed. For instance, a measure used in the mandatory tour submodel might reflect a 40 percent AM, 10 percent MD, 30 percent PM and 20 percent NT weighting to reflect the proportions found in the household survey (taking both legs into account). Indeed, one of the key questions to be answered is whether the accessibility measures should reflect the relative weights of both legs or only those of the outbound leg. Traditionally, accessibility measures are assumed to capture the accessibility of the originating zone; this would tend to argue for an outbound leg approach only. At the same time, if one is considering long-term decisions like residential location or long-term job location, an inbound leg that was particularly burdensome (and recurring) ought to be reflected in the model however it is to be captured. Up to this point, in practice, traffic conditions on inbound legs have rarely been found to be more of a travel disincentive than the outbound AM peak conditions, so in most cases, AM peak conditions are substituted for PM peak conditions in the calculations. CS will check to make sure this is the case in the BMC modeling region.

These accessibility measures may be considered approximate logsums, which are calculated with time periods assigned by an external process, such as a Monte Carlo draw (since the modeled time period is not known at this stage of the model process), or they might involve a more radical break from the logsum approach and measure accessibility in a more direct fashion. The first approach was employed in the Denver Regional Council of Governments (DRCOG) model and will be explored in further detail below. The second approach has been used in New York and Columbus, Ohio.

In the DRCOG activity-based model, the approximate, or aggregate, logsum is calculated in the same basic way as a true logsum, by calculating the utility of multiple alternatives, and then taking expectation across the alternatives by calculating the log of the sum of the exponentiated utilities. However, the amount of computation is greatly reduced, either by ignoring some differences among decision-makers, or by calculating utility for a carefully chosen subset or aggregation of the available alternatives. The approximate logsum is precalculated and used by several of the model components and can be reused for many persons. The categories of decision-makers and the aggregation of alternatives are chosen so that in all choice cases an approximate logsum is available that closely approximates the true logsum.

Table 4.1 gives an indication of how the accessibilities will be generated and applied at different stages in the model. Certain aspects may be updated if certain variables, such as grid connectivity[18], are not available in the base year (or for future year scenarios). Currently, TourCast operates only at the TAZ level and not at the parcel or point location level. In the DRCOG model, the point locations were simulated (although not from parcel data) for activity locations within the chosen zones. These point locations were used in computing distance and time measures between locations for short trips/tours and walk access/egress to transit. The recommended approach for the BMC model is to do something similar to Denver but to use the parcel database to simulate the point (parcel) locations within zones.

[18] Grid connectivity typically is a measure of how "connected" the street network is and whether the local streets in a particular zone form more of a grid-like pattern or are more suburban in nature with multiple cul-de-sacs. In the latter case, driving straight through the zone (or accessing it from a bus route network) is more circuitous.

Table 4.1 Measurement of Accessibility in the Model Hierarchy

Model	Direct Measures of Travel Impedance	Direct Measures of Spatial Attributes	Tour Mode Choice Logsum	Simulated Conditional Outcomes
Regular Workplace Location (Section 6.1)	Distance Distance from school	Employment, enrollment, households. Parking and employment mix. Grid connectivity.	To work	Time of day
Regular School Location (Section 6.2)	Distance	Employment, enrollment, households.	To school	Time of day
Auto Ownership (Section 6.3)	(Simulated) distance to transit stop	Parking price near home.	To work To school	Time of day
Daily Activity Pattern/Joint Travel (Section 7.0)		Mixed use density near home. Intersection density near home.	To work To school	Time of day
Work-Based Subtour Generation (Section 8.2)		Commercial employment near work. School enrollment near work.		
Work Mode Choice (Section 8.3)	All LOS variables	Parking costs, transit accessibility, mixed use density, grid connectivity.		
Work/School Time-of-Day Choice (Section 8.4)	All LOS variables (as generalized cost)			

5.0 Population Synthesizer

The new activity based model will be applied at the disaggregate level where socio-economic and demographics characteristic of every person in the region is an input into the model. In order to develop these data for every person in the region, a population synthesizer is necessary as disaggregate household and person data for the entire population is impossible to obtain. A method to synthesize the population from existing/available data is necessary while controlling the regional totals (distributions) for key household and person characteristics. The population synthesis process generates a synthetic population by expanding disaggregate sample data to reflect known marginal distributions of these household and person characteristics.

The typical characteristics of <u>persons</u> in a synthetic population include:

- Identifier of the household in which the person resides;
- Relationship to others in the household (spouse, child, etc.);
- Age;
- Gender;
- Worker status (full-time, part-time, non-worker, retired);
- Student status; and
- Race.

The typical characteristics of <u>households</u> in a synthetic population include:

- Which (synthetic) people reside in the household;
- Income level; and
- Location (zone).

Other household characteristics such as number of persons or workers that may be used as model variables can be derived from this information while others such as auto ownership will be modeled as long term choices (described in **Chapter 6)**.

5.1 METHODOLOGY

A population synthesizer for the BMC region has already been developed and implemented as part of another project. This model, called "PopGen," was conceived by Dr. Ram Pendyala and developed at the Arizona State

University[19][20]. PopGen has been implemented in other regions, for example, as part of the new ABM for the Southern California Association of Governments (SCAG) and the TourCast implementation for Twin Cities region. It has also been used in several other research efforts. PopGen also includes a population evolution component that has been tested for a subset of the Baltimore region.

PopGen is currently implemented using open source software that can be downloaded from Arizona State University. It has an on-line user's manual (http://simtravel.wikispaces.asu.edu/PopGen+User's+Manual, accessed November 8, 2013).

Iterative Proportional Fitting Procedures

The basic concept of any population synthesizer is based on Iterative proportional fitting (IPF) which involves a "seed" matrix that is repeatedly revised, with each iteration producing a matrix that has closer marginal (row/column) totals to a desired set of totals than the previous iteration. The "seed" matrix is usually multi-dimensional matrix where the dimensions are pre-determined by the user.

A common two-dimensional example of IPF in travel demand models is the Fratar process, where a zone-to-zone trip table is iteratively adjusted until both origin and destination (or production and attraction) totals for each zone are satisfied, with alternate iterations attempting to match origin and destination totals respectively.

Multidimensional IPF can be used to develop a synthetic population. Control totals for variables of interest (for example, household income level, number of workers, and number of persons) are generated at the appropriate geographic (e.g., zone) level, and a seed matrix with all relevant characteristics is generated from a sample of households (e.g., PUMS).

It should be noted that the persons and households in the synthetic population have a full set of all variables (characteristics) included in the sample population used for the seed matrix, or at least those the analyst chooses to retain. This is the case even though the number of variables for which control totals are used may be limited.

---

[19] Pendyala, R.M. and K.C. Konduri. "Population Synthesis for Travel Demand Modeling: Data Needs and Application Case Studies." Presentation to the Using Census Data for Transportation Applications Conference, Irvine, CA, Oct 25-27, 2011. http://onlinepubs.trb.org/onlinepubs/Conferences/2011/Census/Presentations/10-26_830-10amKillough/3Pendyala.pdf. Accessed March 29, 2012.

[20] Arizona State University. "PopGen User's Manual." http://simtravel.wikispaces.asu.edu/PopGen+User's+Manual. Accessed March 29, 2012.

IPF population synthesis procedures are applied by drawing random households using selection probabilities based on the sample population used for the seed matrix. When cells in the final matrix are "filled up," no further households are drawn for them. After all cells have been filled, totals are checked for goodness of fit (perhaps including comparisons for variables not used in the IPF process), and if necessary households can be redrawn.

IPF procedures have been used in several population synthesis methods, including the TRANSIMS model (although it is not used in most current TRANSIMS applications) and the San Francisco, Denver, and Atlanta activity based models. These have been applied using simple custom programs.

PopGen also uses elements of IPF, where it extends the process to estimate sample household weights such that both household and person distributions are matched. Weights are reallocated among sample households of a type to account for differences in household composition.

Evolution of Population

IPF methods can be performed to synthesize populations for future years if forecasts of the marginal totals can be generated. For example, if household income level, number of workers, and number of persons are the control variables, forecasts of the number of households in each category for these variables will be needed. To some extent, this information is equivalent to what is needed to apply conventional aggregate models for future years, for example, for cross-classification trip production models. One issue with developing forecast year synthetic populations using IPF is that sample populations for developing seed matrices can only be available for past years.

As an alternative, we could draw on the decades of research into the evolution of synthetic populations from a base year to future years. Population evolution involves starting from the synthesized base year population and simulating changes in individuals and households over time. Besides the obvious aging of synthesized individuals, these changes could include:

- Births and deaths;
- Changes in employment status (entering the work force, retirement, and job changes) and education status (students entering school and graduating);
- Migration in and out of the region;
- Moving within the region; and
- Household formulation (adult children moving away from parents, or marriage) and household dissolution (divorce, or moving into nursing homes).

Although a simplification, population evolution procedures have the potential to develop better estimates of future year synthetic populations because the process is designed to reflect most of the changes that occur in populations over time. There is also greater consistency between the populations for different analysis years than in an IPF-based process.

However, there are a number of challenges associated with population evolution. The main challenges involve the development of the probabilities of changes in status for the types of changes listed above—for example, mortality and fertility rates, migration rates, marriage and divorce rates, rates of entering and leaving school and college, and rates of entering and leaving the workforce. It is impossible, of course, to know these rates for the future, and so assumptions are often made based on past trends, with varying accuracy in terms of future projections. Most of these rates can vary by region, and so direct transferability of a model from another area is questionable.

Other challenges include:

- The lack of data on some types of population changes such as non-family household formation;
- The difficulty in considering the effects of separate changes in a person's characteristics on one another such as the correlation between being a student and being a worker or even between being a student and marrying); and
- The difficulty in considering the effects of changes in one person's characteristics on another such as the increased likelihood of a person being a worker if no one else in the household is a worker.

5.2 Updates to PopGen

For this project, PopGen will be reviewed thoroughly with BMC and any necessary updates will be undertaken. Based on CS' experience with implementing PopGen for the Twin Cities region as well as BMC's experience with implementing it to small areas within the region, a plan for PopGen updates will be prepared. Some of the considerations for the updates include the following:

- The age distribution of the population is not currently a control total for the BMC PopGen implementation, in part because BMC does not have population forecasts stratified by age group that could be used to generate synthetic populations for forecast years. As a result, current synthetic population results do not match the existing age distributions for past years. To address this concern, BMC has applied PopGen using a "small area" model[21] developed to attempt to address the age distributions of the synthetic populations at the zonal level. This analysis, which used the 2000 data to simulate 2010, met with only limited success[22], and further adjustments are felt to be needed. It is felt that further work must be done to obtain more reasonable age distributions.

- CS analyzed the PopGen outputs as part of the Met Council project. One of the key findings was that depending upon the maximum limit on the household size variable, the synthesized population could vary significantly from the right control total. For example, in the Met Council project, the household size was capped at 7, but the population seed matrix file had a few records where the household size was more than 7. This led to some substantial differences in the synthesized population because PopGen randomly draws records based on geographic ID from households with more than 7 members (i.e., 7, 8, 9, 10, 11+) until the household size 7 category targets are met. However, since it does not distinguish among the "over 7" household categories, the final results are dependent on the households drawn in each run and will definitely vary from one run to the next. Once the maximum was increased to 8, the differences became small enough. Similar tests will be done for BMC.
- A review of the algorithms that generate synthetic populations and of the household and person-level attributes is needed. This will be a collaborative effort of the CS team and BMC.

[21] Mishra, S., X. Zhu, T. Welch, and F. Ducca. "A Framework for Travel Demand Model Related Household and Person Level Control." Prepared by National Center for Smart Growth Research and Education, University of Maryland College Park, 2012.

[22] Pandey, B. "Small Area Model Performance to Estimate Population by Age." Memorandum to C. Baber, July 10, 2013.

6.0 Long Term Choice Models

6.1 Regular Work Location Model

The regular workplace model determines whether each employed person in the synthetic population has a regular work location to which they usually (or regularly or customarily) go, and for those who do, simulates the location. For cases in which a person usually works at home or has no regular work location, the regular work location model will include "work-at-home" and "no regular location" as special outcomes. (The household survey data set will be checked to determine the sufficiency of the sample of work at home observations to inform the way in which they will be treated.)

The workplace location choice models will take into account several measures including:

- Accessibility measures (i.e., logsums) from home to work;
- Location measures such as home and work in same district or home in suburb and work in downtown; and
- Interactions of the labor force and employment characteristics such as worker's occupation and employment by sector specified in size functions.

The work-at-home model can be based on worker and household attributes such as worker's age and education, whether it is part-time work, or whether there are children in the household as well as accessibility measures such as jobs within a certain distance. This model could also be done as a simpler estimate of work-at-home percentages and is a potential tradeoff with other advanced features.

Note that the destinations of workers who do not go to their regular workplaces or do not have regular workplaces are modeled in the tour destination choice model for work tours (see **Section 8.1**). Based on analysis of the household survey data, this is a relatively small fraction of total work tours.

6.2 Regular School Location Model

The school location choice model assigns a school location to every student in the synthetic population, predicting the zone to which they go for their school tours. The household survey data show that students almost never have school purpose destinations other than their regular school locations, and so the design excludes "no regular location" from the choice set.

This model follows a similar structure to that of the regular work location models and takes into account:

- Accessibility measures from home to school such as logsums or distance;
- Location measures such as school districts, to the extent data are available; and
- Interaction measures such as income that may influence the choice of public versus private school).

Typically, school location choice models have fewer variables since distance to school or school district boundaries may direct most students to their school location.

If there are enough response options indicating that a person usually is educated at home, then the school location model can also include "at-home" as a special outcome. These options can be represented in a manner similar to the work-at-home model but may not warrant an additional model for representation.

For university and younger students who are also employed, the school location outcome (along with the home location) conditions the work location choice; for other workers who are also students, the work location conditions the school location choice. Both of these models will need to use disaggregate work tour logsums to capture the effect of level of service on regular work and school location choice.

6.3 Vehicle Availability Model

The number of motor vehicles available to a household has a major impact on the travel behavior of the members of the household. As a result, many MPOs have incorporated models of household vehicle "availability" or vehicle "ownership" into their travel forecasting model systems.

For the BMC ABM, the CS team has chosen to model vehicle availability rather than ownership for two reasons.

- First, household vehicle availability, a measure of the total number of motor vehicles available for use by household members (including both passenger cars and trucks owned, leased, and/or provided by employers), is more closely related with the level of household mobility than the more limited household car ownership measure.
- In addition, data on vehicle availability are collected in the ACS and therefore are available in the Census Transportation Planning Products (CTPP) and PUMS datasets.

The vehicle availability model relates the number of vehicles available to a household to explanatory household person, zonal, and transportation variables. These variables could include household income, number of adults per household, number of workers per household, density measures at the residential zone, employment accessible by auto and transit, transit accessibility from home, interaction variables between household size and number of vehicles, household size and income levels, and number of adults and vehicles.

The primary data source for this model will be the household travel survey. The distributions of the household records by different vehicle availability levels will be examined to determine the number of choices or alternatives that can be specified to estimate these models. The objective here will be to ensure a sufficient number of observations for each alternative to be modeled. Previous modeling experience has shown that there should be a minimum of 50 observations for each alternative to be modeled. If this criterion is not met within some of the vehicle availability levels, then two or more alternatives can be combined to form larger groups with fewer alternatives.

There are three discrete choice model formulations or structures that could be considered for the vehicle availability model. As discussed in Section 2.2, these formulations are the multinomial logit (MNL), ordered response logit (ORL), and nested logit (NL) models. The ORL model is specifically suited for choice contexts where the alternatives follow some natural ordering, as is the case for vehicle availability, whereas MNL treats each alternative distinctly and estimates the coefficients of linear (latent) utility functions specific for each alternative. ORL assumes a single latent function measuring the propensity for a household to own vehicles. The higher the latent variable for a specific household, the more likely it is for that household to own a higher number of vehicles. Based on experience with vehicle availability models elsewhere, it is recommended to use an ORL formulation.

6.4 TRANSIT PASS OWNERSHIP MODEL

A binary logit model will be developed using data from the household survey to estimate the probability of an individual from a household having a transit pass. This model will include several explanatory variables including:

- Characteristics of the individual (age, gender, worker/student status, etc.) and his/her household (e.g. income);
- Locational and accessibility measures that reflect the person's household location relative to the transit system; and
- Outputs from other long term choice model, such as workplace location and vehicle availability.

These models can be validated against current data if a list of transit pass holders and their socioeconomic information is made available to the CS team.

6.5 E-ZPASS TRANSPONDER OWNERSHIP MODEL

A binary logit model will be developed using data from the household survey to estimate the probability of a household owning an electronic transponder for the E-ZPass system. This model will only be applicable to households that own at least one automobile and will include several explanatory variables including:

- Socio-demographic data such as household income; and
- Accessibility measures that map the household's current location relative to the HOT/HOV lane system.

This model can be validated against current data if a list of E-ZPass transponder holders and their household locations and other information is made available to the CS team.

7.0 Daily Activity Pattern Models

One aspect of advanced models that sets them apart from conventional models is the concept of a daily activity pattern (DAP) that can be established at the individual level. Connected with this concept is the understanding that each individual has a limited amount of time per day that can be engaged in activities (including the associated travel time).

Separate models will be estimated for different person types. For example, the following person classification scheme was used in Houston:

- Child less than 5 years old
- Child 5-15 years old
- Child 16 years old or greater
- Non-working adults
- Seniors (all adults 65 years old or greater)
- Part-time workers
- Full-time workers
- College students

The DAPs will be simulated through a series of models, including the following:

- **Daily Activity Pattern**, in which the number of mandatory (work, university, and school) tours is estimated, and whether the tour(s) will have other stops besides the primary (mandatory) activity, and if there are no mandatory tours, whether there are non-mandatory tours;
- **School escorting**, which simulates whether children with a school tour are escorted by another family member;
- **Joint tour participation**, which simulates the number of joint non-mandatory tours undertaken by members of the same household, if needed; and
- **Non-mandatory non-joint** tour generation.

Some DAP models attempt to account for all stop making as part of this first modeling stage since tours are frequently exchanged for stops as travel times lengthen and less "free time" remains.

For the new BMC ABM, the recommendation is to model the presence of intermediate stops (i.e., that the tour is not a simple home-to-destination-to home tour) as part of the day pattern and to model the number of stops on tours where stops are present as a separate tour level model, described in **Section 9.1**.

This series of models will produce the set of tours by purpose made by each synthetic person in the region. The mode, destination, and time of day of each tour and the number of stops are modeled later, as described in **Chapter 8.0**.

The analysis of tour detail conducted by CS using the household survey and the resulting recommendations on the level of complexity will be discussed again with BMC staff.

7.1 Daily Activity Pattern

The main daily activity pattern model determines the number of mandatory tours by type (work, university, and school). Preliminary examination of the household survey data indicates that the number of tours will be limited to two. These may include two tours of any of the three mandatory tour purposes, or one tour of one purpose and one of another. Depending on person type, some tour types may not be available. Children may not have university tours, young children may not have work tours, and adults may not have school tours.

For any patterns with work tours, the alternatives will be separated by whether each tour has any stops. For example, patterns with two work tours will have three associated alternatives: no stops on either tour, stops on one tour, and stops on both tours. There will be four alternatives associated with patterns with no mandatory tours. These include non-mandatory travel only, stay at home, external travel only, and out of area (meaning the person is absent from the Baltimore region on the travel day). It is also worth noting that individuals making mandatory tours may also make non-mandatory tours.

Figure 7.1 provides an example, from the Houston model, of the alternatives for each person type. The final set of alternatives for the BMC model may differ depending on the household survey data analysis and the model estimation results.

Figure 7.1. DAP Model Alternative Availability by Person Type

DAP	DAP Availability							
	Child <5yrs	Child 5-15yrs	Child 16+ yrs	Non-Worker	Senior	Part-Time Worker	Full-Time Worker	College Student
1 Work Tour - No Stops	0	0	1	1	1	1	1	1
1 Work Tour - 1+ Stops	0	0	1	1	1	1	1	1
2 Work Tours - No Stops	0	0	0	0	1	1	1	1
2 Work Tours - 1+ Stops on one, no stops on other	0	0	0	0	1	1	1	1
2 Work Tours - 1+ Stops on both	0	0	0	0	1	1	1	1
1 University Tour & 1 Work Tour - No Stops	0	0	0	0	0	0	1	1
1 University Tour & 1 Work Tour - 1+ Stops	0	0	0	0	0	0	1	1
1 School Tour & 1 Work Tour - No Stops	0	0	1	0	0	0	0	0
1 School Tour & 1 Work Tour - 1+ Stops	0	0	1	0	0	0	0	0
1 University Tour	0	0	0	1	1	1	1	1
2 University Tours	0	0	0	0	0	0	0	1
1 School Tour	1	1	1	0	0	0	0	0
2 School Tours	0	1	1	0	0	0	0	0
Non-Mandatory Travel Only	1	1	1	1	1	1	1	1
Stay-at-Home	1	1	1	1	1	1	1	1
Out-of-Area	1	1	1	1	1	1	1	1
External Travel Only	1	1	1	1	1	1	1	1
Total number of alternatives	**5**	**6**	**10**	**7**	**10**	**10**	**12**	**13**

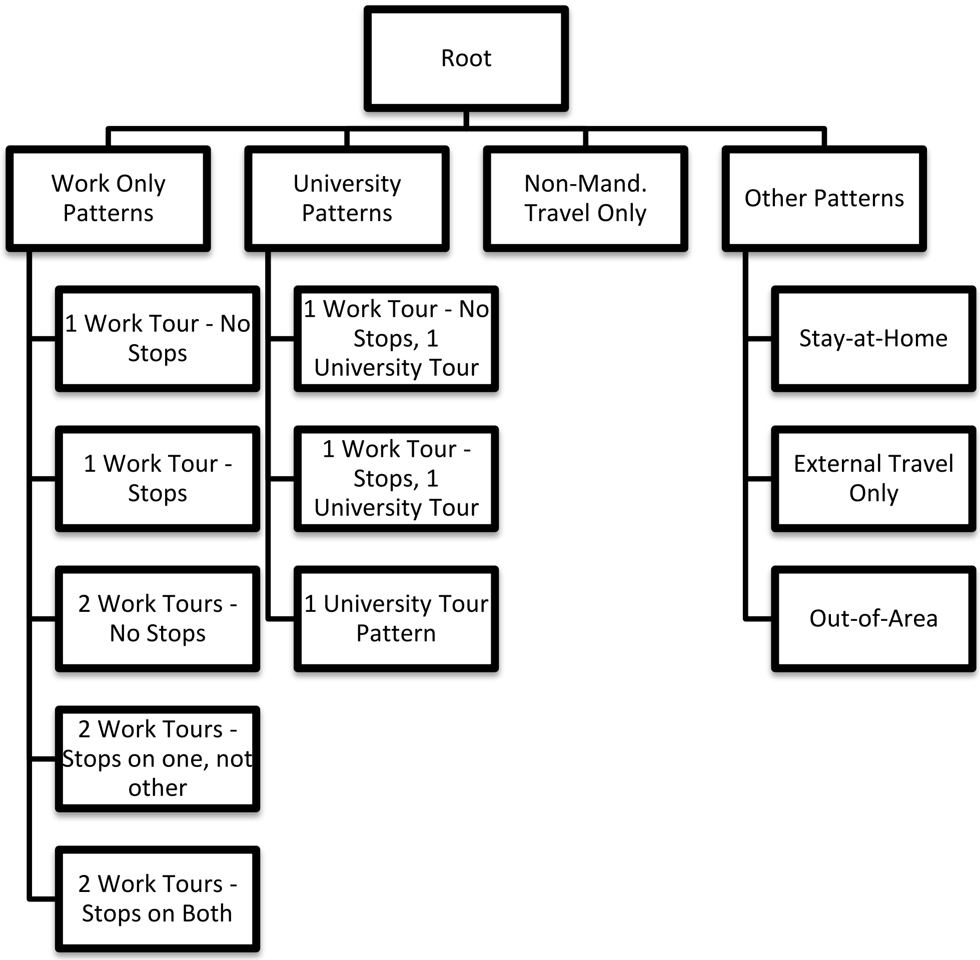
The main daily activity pattern model is expected to use a nested logit form. An example of the nesting structures for full time workers is shown in **Figure 7.2**.

The DAP choices of individuals in the same household are interrelated. To account for these intra-household interactions, the models are sequenced strategically in application, and each model is applied conditionally on the DAP outcomes of previously simulated household members. The sequence is expected to be as follows: Children will be simulated first, since they are considered to be the most travel dependent individuals in a household (i.e., they rely on adults in the household for their travel needs). Younger children will be simulated before older children. The order in which adults will be simulated will reflect the most likely household members responsible for care-taking of children. So non-working adults will be simulated first, followed by seniors, part-time workers, full-time workers, and college students.

The DAP outcomes of previously simulated household members are critical to DAP choice, particularly for the household members that are simulated toward the end of the sequence (e.g., workers and college students). For the youngest children, these variables will not appear since these children are first in the household to be simulated. The DAP choice outcomes of young children in the household, however, appear as explanatory variables for each other person type. On the other end of the spectrum, college students are simulated last, and so their DAP choices depend on the DAP choices of household members of all other person types, but the DAP choices of college students do not affect the choices of other household members.

Of particular importance are the DAP choice outcomes of young children since they are most dependent on adults to meet their travel needs. In addition, these children, particularly the youngest ones, must be under constant supervision. Thus, a simulated stay-at-home pattern for young children is very important to the choice of stay-at-home pattern for adults in the household.

Figure 7.2. Sample Nesting Structures for the Daily Activity Pattern Model for Full Time Workers



Inputs to the Model

- Person and household characteristics, including income;
- Household vehicle availability (from vehicle availability model);
- Household location attributes such as mixed use density, presence of nearby transit, accessibility, and proximity to the regional boundary; and
- Transit pass ownership (if data available).

Outputs from the Model

- Number of mandatory tours by purpose, and presence of stops on each tour.

7.2 School Escorting Model

The school escorting model captures the choice of whether each child traveling to school is escorted by a household member in each direction, and if so, by whom. For each child having a school activity, the escorting model simulates escorting for both the outbound half-tour (travel to school) and the return half-tour (travel home from school).

Each school tour is divided into outbound and return half-tours. For each half-tour, three mutually exclusive and collectively exhaustive options are considered:

1. Traveling with another household member who also implements a mandatory tour for work, university, or school purpose.
2. Pure escorting by a household member who does not have any mandatory activity on this tour. Maintenance or discretionary activity may be carried out on this tour as well, but it would be considered secondary to the escorting function and will be covered by the stop generation model.
3. No escort when the child travels to school and back home on his/her own (by school bus, transit, driving, or walking) or is escorted by non-household members.

It is important to note that the destinations and time periods are simulated for mandatory tours prior to the application of the school escorting model, as shown in **Figure 1.1**. This allows for the time for the mandatory activities to be blocked out in determining whether an adult is available to escort a child to school, as described below.

The first two options listed above will have an alternative in the model for each "available" adult. Adults are split into three groups. The first are adults that could perform the escorting task as a stop on a previously generated mandatory tour (for example, dropping a child off at school on the way to work). The second are adults that could perform the escorting task as a stand-alone tour. The third are adults who are unavailable to escort the child because the child is traveling to school during a previously scheduled mandatory activity. Each adult can appear only in one of these three categories.

There are a number of availability conditions that must be processed in order to generate the choice set for each school tour. For adults with mandatory tours, the timing of the mandatory tour (which is simulated prior to school escorting) must meet two conditions for the adult to be considered available for escorting as part of the mandatory tour. First, the adult's mandatory activity must begin after the school activity (for escorting on the outbound half-tour) or end before the school activity ends (for escorting on the return half-tour). If this condition is not met, the adult is not considered a viable escort candidate and is considered to be in the third category described above. Second, the adult's mandatory activity must begin or end within three half-hour periods of the start or end of the school tour. If this condition is not met, the adult can still be considered a viable escort candidate, but only in the stand-alone school escorting category, not as part of the adult's mandatory tour. Only if both conditions are met is the adult considered a candidate for school escorting as part of the mandatory tour.

Since each school tour has two travel components (outbound and return), there are two choice dimensions, handled simultaneously in the model. In other words, each alternative represents the joint choice for outbound half-tour escort choice and return half-tour escort choice.

One issue in modeling school escorting choices is how to handle multiple children in the same household traveling together to or from school. To accommodate this behavior, such children were grouped for the escorting model. The groups are defined by two conditions: (1) the school tours share a common school location, and (2) the school tours arrive in the same period or return in the same period. When these conditions are met, we expect that the rates at which children make identical escorting choices in the household survey data set will be very high, as they have been in other models.

By treating groups as units in the school escorting model, it is possible to control for effects these child groups may have on the propensity of school escorting. It also solves some issues dealing with availability rules of adults. Suppose these school tours were treated separately. If the first was simulated to be escorted by adult A, that would essentially block out a period in the adult's schedule for escorting the child. Then when the second child is simulated, it is not easy to account for the possibility that the same adult A could also escort the second child's school tour, when that tour occurs in the same period as the first. By treating the school tours as a single unit, the availability conditions become simpler.

Children will be divided into three person types:

1. Driving age school children (16 years and older);
2. Pre-driving age school children (6-15 years old); and
3. Preschool children (under 6 years old).

Potential household escorts are divided into six mutually exclusive and collectively exhaustive person types:

1. Full-time workers;
2. Part-time workers;
3. University students;

4. Non-workers under 65 years old (homemakers, unemployed);
5. Senior (65 years or older); and
6. Driving-age children (16 years or older).

Driving-age school children may be escorted by the other household members; they could also play the escort role for the younger household children.

Children within the household will be ordered and modeled by age from youngest to oldest. The behavioral assumption behind this rule is that, all else being equal, a younger child has limited mobility compared to an older child, and escorting younger children would be considered first in the household decision-making process. Older children will be escorted only if escorts are available after escorting younger children.

The maximum number of adult household members considered as potential escorts will be limited to three, based on practical experience with other models. For the infrequent cases with households with both four or more adults and at least one child under 18 years old, the set of alternatives for model estimation will be created by including the actual escort plus two other adults selected randomly. Adults with mandatory travel that do not have compatible schedules with school children become ineligible to be considered as escorts. It is also worth noting that nonworking adults who are chosen to make pure escorting tours as a result of this model will need their time windows updated before the non-mandatory joint tour submodel is applied.

Inputs to the Model

- Level of service attributes for the trip between the home and school and "detour" attributes for adding the school stop to mandatory tours for potential escorts;
- Characteristics of the potential escorting adults;
- Characteristics of the student or student group;
- Household characteristics including vehicle availability (from the vehicle availability model) and income; and
- Household location attributes.

Outputs from the Model

- For each student, whether he or she is escorted to school and from school; and
- If escorted, which household member performs the escorting and whether it occurs as part of a mandatory tour.

7.3 JOINT TRAVEL

Discussion/Concepts

For the purposes of modeling households, joint travel refers only to travel between members of the same household; carpooling between work colleagues or neighbors will not be classified as joint travel and will not be explicitly modeled but instead will be captured through the mode choice alternatives shared ride 2 and shared ride 3+ for individual-level tours. Joint travel can be categorized as **fully joint travel** or **partially joint travel**.

Fully joint travel means that all important aspects of tour-making are shared by two or more household members, including origin, destination, mode, time of departure from the origin and destination (as well as arrival times), and purpose. While it is possible that two household members would make a fully joint tour with different purposes at the destination (for example, accompanying a household member to a medical appointment), the number of individuals reporting these split purposes is very few, and it will not distort the overall picture greatly to recode one traveler's purpose in order to make the purposes match. Work tours are rarely fully joint.

Partially joint travel means that at least some aspects of the tour are not shared by the household members traveling together. This might mean that the outbound or inbound leg is fully joint but the household members split up after the activity at the destination, or that they begin separately but meet at a destination to share the inbound leg. Partially joint travel may also mean that the outbound or inbound leg is not fully shared and the two members simply depart at the same time (this is particularly the case for drop-off/pick-up activities). The most important category of partially joint travel is escorting children to school, in terms of both the prevalence in surveys and policy impacts because of the concentration of parents driving children to school in a very narrow time window.

Modeling Approach

While the analysis of the household survey data with regard to joint tours is not yet complete, it is expected that the incidence of partially joint tours that do not involve escorting someone to school and of fully joint work and university tours is relatively low. It is therefore proposed to model fully joint non-mandatory travel directly and to explicitly model only school escorting among partially joint travel possibilities. Other partially joint travel will primarily be captured through trip mode choice. The benefits for getting all partially joint travel correct do not fully offset the increased complexity of the model structure.

The fully joint tour model consists of two submodels: tour generation at the household level (which simulates the number and purpose of joint tours for a household) and tour participation at the individual level (which simulates which household members will participate in each tour). The joint tour generation model will be applied for all households that meet the condition of two or more household members with travel in their daily activity pattern choice (i.e., a mandatory or non-mandatory only travel pattern). The participation model is estimated and applied for all members of households where the household consists of three or more household members with active daily activity pattern choices.

The proposed approach for the BMC ABM is shown in the middle sections of **Figure 1.1**. Modeling individual mandatory travel will be given top priority, but variables will be considered that affect the scheduling of the mandatory tours to increase the likelihood of joint tours in households where joint travel is expected.

The joint tour generation model will be estimated as a multinomial logit model with the following set of 15 alternatives considered for each household having at least two household members with travel-making daily activity patterns:

- No tours
- One tour by travel purpose:
  - Shopping
  - Meal
  - Personal business
  - Social-recreation
- Two tours by travel purposes:
  - Shopping/shopping
  - Shopping/meal
  - Shopping/personal business
  - Shopping/social-recreation
  - Meal/meal
  - Meal/personal business
  - Meal/social-recreation
  - Personal business/personal business
  - Personal business/social-recreation
  - Social-recreation/social-recreation

It is possible that combinations that occur rarely in the household survey data set will not be included as alternatives.

After joint tour generation is modeled at the household level, the participation model is run for each eligible household member. This model is a binary logit model with two alternatives available – to participate or not to participate. Household members are sequenced strategically for estimation/application based on person type (note that this sequence need not be the same as for the daily activity pattern model). No strict requirements will be enforced on the estimation to ensure that model application will result in a valid fully joint tour. In other words, application of the model could result in zero or one household members participating. In such cases, the model will be rerun until a valid tour is constructed.

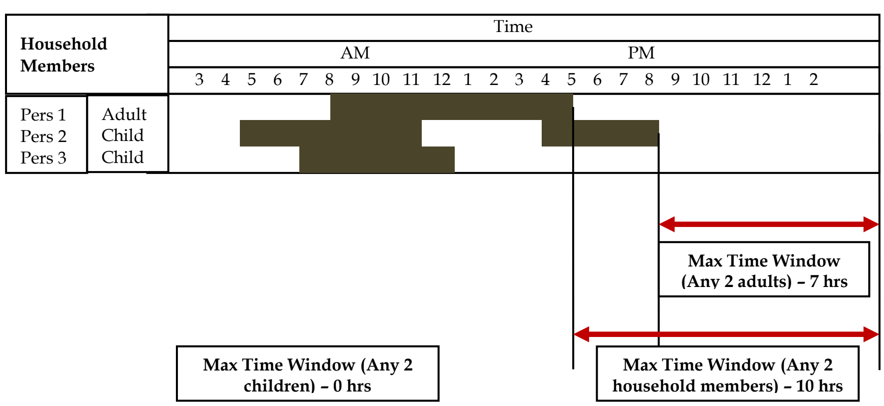
Inputs to the Joint Tour Generation Model

- Household characteristics;
- Daily activity patterns of the potential joint tour participants;
- Time window overlaps among household members; and
- Household location/accessibility attributes.

The term "time windows" refers to the portions of the day for which an individual has not already had a tour simulated and are therefore available for other activities to take place. For the joint tour model, the tours that have already been simulated, including their start and end times, include the mandatory and school escorting tours. The time windows available for joint tours are therefore the times when neither mandatory nor school escorting tours are made.

Time window variables are critical in defining the amount of time available for two or more household members to actually engage in joint activities. For instance, in a household with two workers that have different work shifts, the overlapping available time for engaging in joint activities would be rather small. On the other hand, a household with one part-time worker and one non-worker may have a rather large overlapping window in which joint activities could be scheduled. Figure 7.3 illustrates how these time window overlaps are computed in a simple case with two adults and one child, each having a mandatory tour and one of the adults having a second mandatory tour.

Figure 7.3. Time Window Illustration



Outputs from the Joint Tour Generation Model (for each household)

- Number of joint tours by purpose; and
- For each tour, which household members participate.

Inputs to the Participation Model

- Household and person characteristics; and
- Joint tour composition attributes. These include:
  - Number and type of household members already simulated to participate in joint tour
  - Indicators for whether the joint tour members include only household members having mandatory patterns or only household members having non-mandatory only patterns
  - Maximum time window remaining if the person chooses to participate in the joint tour
  - A joint tour size ratio variable:
    - Equals: max(0, [2-joint tour size before simulated person]/ [Number of remaining candidates in household])
    - Purpose is to encourage participation the closer the simulation gets to last individual in household, if the tour does not have at least 2 members set to participate already.

7.4 Non-Mandatory Tour Generation

The individual non-mandatory tour generation model predicts the number and type of individual (as opposed to joint tours among household members, which are modeled separately) non-mandatory tours for each individual in the synthetic population. This model is a function of the number of available time windows, number of mandatory and joint tours already simulated, whether there was only one partially coordinated tour simulated implying the need to pick-up or drop-off through an escort tour, accessibility to various types of employment, and other household and person attributes.

There are a total of 56 possible alternatives for this model. These include all possible combinations of up to three tours for meal, shop, personal business, escort and social recreation purposes. A nested logit structure is proposed, with the upper nest representing the total number of tours (0, 1, 2, or 3).

Inputs to the Model

- Household and person characteristics;
- Household location/accessibility attributes; and
- Tour attributes. These may include:
    - Number of children and adults with mandatory daily activity patterns;
    - Number of workers and non-workers with non-mandatory only daily activity patterns;
    - Number of fully joint tours;
    - Number of work tour with stops pattern;
    - Number of school escorting tours;
    - Number of available (half-hour) time periods left in the day after previously simulated tours-including all mandatory, school escorting, and joint non-mandatory tours and previously simulated individual non-mandatory tours-are simulated; and
    - Number of available (half-hour) time periods left in the afternoon.

Outputs from the model

- Number of individual non-mandatory tours by purpose

8.0 Tour Level Models

The tour-based choice models incorporate interrelationships among trips that are components of a "tour" that departs from home, visits one or more activity locations, and then returns home. Additionally, there are also subtours that depart from the workplace and return back there. These tours will be generated from the trip files included in the household survey database. Hierarchical rules will be established to identify the appropriate nature of the tour. For instance, tours that include a mandatory destination such as work will be defined as a work tour irrespective of other destinations serviced as part of this tour.

These tour-level models provide an improved framework over trip-level models to model daily travel decisions since they retain information regarding previous or subsequent trips within a tour. Overall, tour-based models account for information on modes, time-of-day, group travel, and other characteristics of travel that are clearly interrelated across trips within a tour. In the process, they involve origin and destination zones, as well as intermediate zones to provide a refined assessment of travel among all zones.

8.1 Work Tour Destination Choice

After the number of work tours has been established via the day pattern model, the details of the work tours must be filled in. For all full-time and part-time workers, a regular work location model (**Section 6.1**) has already been run to determine the location of their regular workplace (if any). While in the majority of cases, the destination of a work tour is the regular workplace, some workers work at other locations, and there are some workers who do not report any regular workplace. For these workers, the regular workplace zone choice is the null set and the entire set of zones will be found in the other nest, collapsing this model to a standard destination choice model.

Estimation Cases

There is an estimation case for every work tour.

Model Structure

The model form will be nested logit, with the upper level representing the binary choice between the regular work location and other locations, and the lower level representing the set of destination zones.

Explanatory Variables

- Person type characteristics (see **Section 3.4**)

- Vehicle availability level
- Accessibility measures and/or approximate logsums, round trip or one-way (see **Chapter 4.0**)
- Other socioeconomic variables that can be tested, such as gender, presence of children, and income
- Land use variables
- Dummy variables representing the home zone and the regular workplace location zone, if any
- Size function based on employment at the destination zone

The output from this model is the destination (work location) zone for each work tour. One way to check for the validity of the model is to evaluate the number of cases where the workplace zone identified by this model matches the workplace zone identified by the long term work location model. We expect that in a majority of the cases, the two zones will be identical.

8.2 WORK-BASED SUBTOUR GENERATION

For every work tour, the work-based subtour model will be run. This model determines the number of "at-work" subtours, which begin and end at the workplace location. Examples of these subtours include traveling to a different location for lunch or traveling to a business meeting and returning to the workplace.

Estimation Cases

Up to two work based subtours will be simulated for each work tour, with the possible subtour purposes being work, meal, shopping, escort, social-recreation, and personal business. Rarely chosen combinations of subtour purposes will be eliminated from the choice set.

Model Structure

The model structure will be nested logit. Different nesting structures will be tested, with the upper level based on subtour purpose (for example, alternatives with work subtours).

Outputs from the model will include the list of subtours for each work tour.

Explanatory Variables

Inputs to the model include:

- Person and household characteristics;
- Accessibility measures and/or logsums (see **Chapter 4.0**);

- Household location attributes; and
- Characteristics of the work tour, the daily activity pattern of the individual, and the time of day for the individual's tours.

8.3 Work Tour Mode Choice

The goal of the mode choice models is to determine the main mode for tours, as well as the modes for all trips made as part of tours. The models will both reflect the unique mode choice behavior and tradeoffs of area residents while being consistent with FTA guidance for ridership forecasting.

These models will provide the basis for the logsum measures used in the tour destination choice models. The mode choice models differ from traditional "trip-based" mode choice models in that there are two distinct sets of mode choice models. The tour mode choice model determines the primary mode for the tour while the trip mode choice models determine the mode for each individual trip made on that tour, based on the mode chosen for the tour. There is one of each model (tour and trip) for each tour purpose (work, school, shopping, personal business, meal, social/recreation, and work-based sub-tours) although models for separate purposes may be estimated in a combined manner.

The tour mode choice models are applied after the stop generation models. This means that the number of stops on each half tour is known and can be used to help inform the mode choice decision.

Alternatives

The final list of modal alternatives to be included in the tour mode choice model will be determined after the survey data has been fully examined. Based on the peer review recommendations, a "shallow" mode choice model structure will be implemented, with no transit submodes other than access modes (walk access, park-and-ride, and kiss-and-ride). Therefore, the following modes will be used in the tour and trip mode choice models:

- Drive alone;
- Shared ride (2 occupants);
- Shared ride (3+ occupants);
- Transit with walk access;
- Transit with park-and-ride access;
- Transit with kiss-and-ride (dropoff/pickup) access;
- Bike; and
- Walk.

Some of these modes may be aggregated for specific tour purposes if there are not enough samples to support model estimation or if the data does not allow an extensive mode split within the transit mode nest.

There are a few additional notes on the selection of these modes:

- Some mode choice models, including the existing BMC model, make a distinction between respondents who will or who are at least willing to take toll roads called the drive-toll alternative and those who do or will not (drive-free). The recommendation is to not model toll road choice at the mode choice level, but to leave it as a route choice issue, with market segmentation to account for different path choices involving priced and free roadways. By using different values of time for different vehicle classes, individuals with higher values-of-time will be more likely to use the toll facilities. This is discussed in more detail below.
- An analysis of the household and transit onboard surveys will help determine if kiss-and-ride and park-and-ride to transit should be combined into a single drive access mode. This decision will be based on the following considerations:
  - Adequacy of the number of kiss-and-ride trips for each transit mode in the survey to estimate mode choice models for this mode; and
  - Whether the underlying drive skims for each mode are the same, unless there are significantly high costs at the park-and-ride lot. We will develop rules to separate kiss-and-ride and park-and-ride with a submode model after mode choice.
- If park-and-ride lots are reaching capacity, the model can address this using a time or cost penalty at the park-and-ride lot in an iterative process to limit the number of vehicles parked at the lot at any given time. This is not an issue for model estimation and does not affect the skims, but will be considered during model application and validation.
- The nature of work subtours argues against using any drive to transit modes as the auto is presumably at a transit system parking lot and unavailable at the work end.

The trip-based mode choice model is conditional on the tour mode reflecting an implicit hierarchy. The availability of alternatives for the trip-based models is different than for the tour-based models, as shown in **Table 8.1**. This table, which will be refined as the specific mode definitions for the new model are finalized, indicates not only which modes are available for trips comprising a tour given the tour mode, but also how tour modes are defined. For example, if a tour includes the walk to local bus mode but no other transit modes, it is a walk to local bus tour even if there are segments that use auto or nonmotorized modes. If a tour includes school bus for any segment, the tour mode is school bus; conversely, the tour mode is walk only if every segment is made by walking.

Table 8.1 Mode Alternatives at Trip Level

Available TRIP Modes	School Bus	Transit Park-and-Ride	Transit Kiss-and-Ride	Transit Walk Access	Shared Ride 3+	Shared Ride 2	Drive Alone	Bike	Walk
School bus	●								
Transit Park-and-Ride	●*	●							
Transit Kiss-and-Ride	●*	●*	●						
Transit Walk Access	●*	●	●	●					
Shared ride 3+	●	●	●	●	●				
Shared ride 2	●	●	●	●	●	●			
Drive alone	●	●	●	●*	●	●	●		
Bike	●*	●*	●*	●*	●*	●*	●	●	
Walk	●	●	●	●	●	●	●	●	●

Note: Choices that will likely be rare and may be excluded are shown with an asterisk.

Estimation Cases

The cases to be used for estimation will be all tours with:

- A valid person-day activity pattern (the day and all tours start and end at home;
- Valid modal data for all activities in the person day; and
- Valid tour origin and destination location geo-codes for the tour.

Model Structure

Both the tour and trip mode choice models will be estimated as nested logit models. A number of different nesting structures will be considered.

Model Segmentation

Tour mode choice will vary by purpose, using separate models.

Explanatory Variables

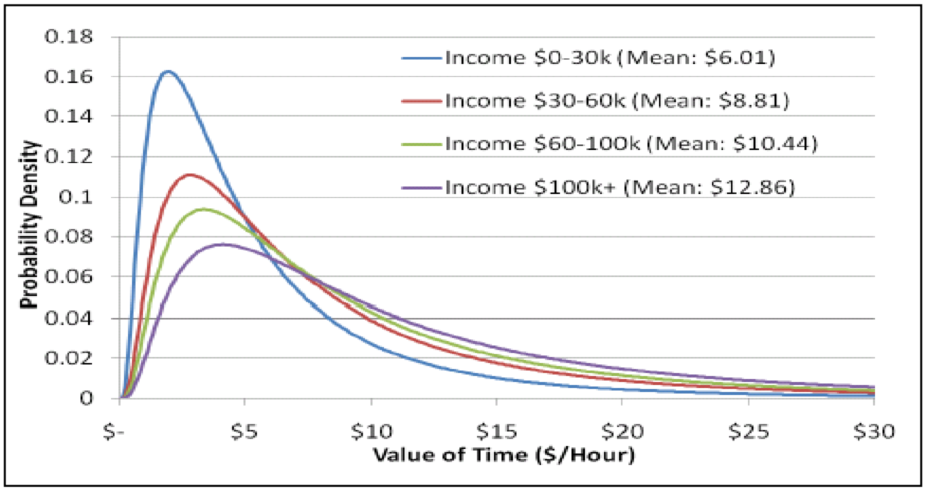
There are four types of variables that will be tested for the tour- and trip-based mode choice models:

- **Level of Service Variables** – These will include in-vehicle travel time, out-of-vehicle travel time, number of transfers, and cost. Out-of-vehicle travel time will include terminal time for auto trips and wait, walk, and transfer time for transit trips. Cost will include auto operating cost, parking cost, tolls, and transit fares. Combinations of these levels of service variables may also be tested, such as the ratio of drive access to transit time to in-vehicle time or the proximity of a rail stop to home, which would increase the likelihood of a transit trip.

- **Land Use Variables.** The land use variables most likely to be significant for mode choice are those that affect pedestrian movements such as the connectivity of streets and walkways, the ease of crossing streets, the provisions of sidewalks, the mix of land uses, and terrain. These variables may offer significant explanatory power for walk, bike, and transit modes. In addition, there are a variety of density variables such as population or employment density by type and urban form variables such as the Central Business District that may be introduced.
- **Number of Stops**. The number of stops by purpose on each half tour affects mode choice.
- **Number of Subtours.** Assuming that subtours are modeled before work tour mode choice, the number of subtours made by the worker will increase the likelihood that the work tour (and trip-level decisions) is made by automobile reflecting the flexibility it offers.
- **Other Variables.** These include demographic variables such as vehicle availability, household income, age, and household size. Other variables may also be geographic-specific areas such as a university. There may also be time period variables in the trip-based models to account for the fact that some modes (like bike and walk) are significantly less likely at night.
- **Distributed Values of Time**. Another potential feature of the mode choice model is distributed values of time. However, before a final determination for whether it is feasible and/or reasonable to pursue, more exploration of the data will be needed. One way this feature could be implemented is by imposing a distribution on the coefficient of travel time. In this way, each individual would be assumed to have his or her own distinct valuation of travel time, rather than all individuals sharing the same one. The parameters of the travel time coefficient's distribution would be determined via model estimation. Often for work travel, a log-normal distribution is imposed, since it ensures the coefficient has the correct sign, but also because there is evidence suggesting the shape of a log-normal may be more suitable than a symmetric distribution such as the normal distribution. **Figure 8.1** shows the value of time distributions that emerged in the San Francisco County Transportation Authority (SFCTA) model using

stated preference data.[23] In this case, income segmentation was pursued in addition to variable value of time.

Figure 8.1. SFCTA Work Value of Time Distributions



If the survey data do not include sufficient information to estimate the parameters of a value of time distribution, it may be considered to use the San Francisco value of time distributions in the new mode choice models (and correspondingly throughout the remainder of the modeling process). Experience shows that local survey data is likely to be insufficient to estimate the parameters of the distribution. Transferring the San Francisco parameters will allow for the value of time differences among individuals to be considered in model application. While the highway assignment process will be aggregate, the values of time can be considered in route choice by creating segments for the highway trip tables associated with value of time ranges. This is discussed further in **Section 11.2**.

- **Modeling Priced Roadways**. As discussed above, the mode choice model will not include separate alternatives for priced and free roadways. A new approach to introducing segmentation in the mode choice model has been proposed to minimize the costs and challenges associated with

---

[23] Sall, E., E. Bent, B. Charlton, J. Koehler, and G. Erhardt (2010). Evaluating Regional Pricing Strategies in San Francisco – Application of the SFCTA Activity-Based Regional Pricing Model. Proceedings of the 89th Annual Meeting of the Transportation Research Board (TRB), Washington, D.C., 2010.

modeling priced roadways[24]. The proposed approach takes advantage of two recent enhancements to travel modeling:

- The use of simulated values of time from a distribution in a disaggregate model application (see, e.g., Sall et al., 2010); and
- Segmentation of trip tables used in aggregate highway assignment by value of time level (which is being implemented in the Houston activity based model).

The proposed approach will define a segmentation scheme based on value of time levels (ranges) to be used for both highway assignment and mode choice. These levels would be defined based on the value of time distributions which are assumed to be used in the activity based model. A set of highway skims would be developed using the implied average value of time for each level. In mode choice application, the skims used for a particular traveler would be those for the value of time range in which the traveler's simulated value of time falls. Highway assignment would be performed using separate trip tables for each value of time range segment, and skims for the next iteration of the model would be developed for each segment.

The segmentation is not used to create separate alternatives in the mode choice model. Rather, mode choice is applied separately for the travelers in each segment, and the segments are retained for the highway assignment. Value of time segmentation is not expected to be as limited as toll/non-toll segmentation for mode choice estimation; more than two segments may be created. While there is no guarantee that a "free" path will be used in developing travel time skims, the likelihood of a free path would be high for the segment with the lowest value of time.

8.4 WORK AND SCHOOL TOUR TIME-OF-DAY CHOICE

This model predicts simultaneously the time periods that the person arrives and leaves the tour primary activity location – either the workplace or the (regular) school location.

[24] Rossi, T., B. Pandey, J. Lemp, and M. Milkovits (2013). Improving the Treatment of Priced Roadways in Mode Choice. Paper submitted to the 5th Innovations in Travel Demand Forecasting Conference, Transportation Research Board, Baltimore, Maryland, 2013.

Alternatives

It is proposed to use 30-minute periods for time of day choice. There will be 48 periods:

1. 3:00 to 3:29 a.m.
2. 3:30 to 3:59 a.m.
3. 4:00 to 4:29 a.m.

etc…..

46. 1:30 to 1:59 a.m.
47. 2:00 to 2:29 a.m.
48. 2:30 to 2:59 a.m.

Since arrival and departure time at the primary activity will be modeled simultaneously, if all combinations were used, there will be 1,176 alternatives.

Availability of Alternatives

Availability of alternatives is determined after adjusting for the time periods used by all previously simulated (higher-priority) tours. Because the tour to the regular workplace or regular school location is nearly always the highest-priority tour of the day, all of the alternatives would be available for the large majority of tours, at least for the first work or school tour simulated.

Estimation Cases

The cases used for estimation will be all tours with:

- A valid person-day activity pattern such that all tours start and end at home;
- Valid start and end time data for all activities in the person day;
- Valid origin and primary destination location geocodes for the tour; and
- A valid primary mode for the tour.

Model Structure

The time-of-day choice models will be multinomial logit across the available alternatives.

Model Segmentation

Separate models will be estimated for work tours and school tours.

Model Variables

Within the models, further segmentation variables will be tested related to:

- Person attributes (age, gender, employment status);
- Household attributes (income, presence of children, number of workers); and
- Day-pattern types (e.g., persons with multiple work tours in a day).

Each of these variables can be specified in the model in three main ways:

- **Period-Specific Effects**. Certain activities may tend to occur in or avoid specific periods; for example, school tours tend to occur during normal school and after-school activity hours.
- **Time-of-Day Shift Effects**. Certain activities may usually be carried out earlier or later than others.
- **Duration Shift Effects**. Certain activities may tend to be carried out for shorter or longer durations than others.

Two kinds of "Shift" variables will be computed, namely "Shift Early" and "Shift Later," which measure the difference between the time period indicator (on a scale from 1 through 24 with 0.5 increments) and the pivot-point. "Shift Early" will be used when the time period indicator is less than the pivot-point, and "Shift Later" will be used when it is greater. The shift variables are defined as follows:

"Shift Early" for AM = max(P – T, 0)

"Shift Later" for AM = max(T – P, 0)

where:

T = Hour – 1, 2, 3, …, 24

P = Pivot-point

Duration "shift" variables will be defined at least for the work tour model. The pivot-point will be set at the peak work tour duration in computing duration "Shift Early" and "Shift Late" variables. During model estimation, these "Shift" variables will be multiplied by household and person attributes to see the effects of individual attributes on time-of-day choice.

Other possible variables in the models include the following:

- **Alternative-specific constants** for specific departure and arrival periods and combinations (but there will not be a constant for every single alternative).
- **Variables for Partially "Used" Periods**. For example, if a previously simulated home-based tour leaves home at 8:15 a.m., then the probability of any other home-based tour beginning or ending in the

period 8:00-8:59 is much smaller than if that period were completely "unused."

- **Time Window Effects**. These variables are derived from the amount of time available after the person's daily activity pattern is simulated. This gives an indication of time available to indulge in other activities and the propensity to either arrive/depart late or early from/to primary destinations. People will tend to stay at the activity location for a shorter duration if they have other activities and tours to carry out during the day. If it is known how much time is already "used up" by previously simulated tours and which periods are fully and partially left, this information may be used in these models. However, typically the work and school tours will be scheduled first and will have a "blank slate," at least for the first work or school tour. In cases where two or more work or school tours are made, the first tour scheduled will impact the second one. However, it will be known how many tours still need to be simulated for the person day after the current tour, and so other "time pressure" variables can be specified. If households are expected to make joint travel (see **Section 7.3**), then there is an incentive to make shorter mandatory tours. The following will be considered.
    - Total available time remaining
    - Available time remaining if other mandatory tours have yet to be scheduled (applies to mandatory tours only)
    - Available time remaining if other non-mandatory tours have yet to be scheduled (applies to individual non-mandatory tours only)

- **Level of Service**. The travel time and cost between the home and primary activity location will be represented using a measure estimated from the network skims. In addition, the delay resulting from congestion will be computed as a function of the time traveled in the chosen period and the free flow travel time. This variable will be used to estimate the propensity for peak spreading since the peak periods are the most congested. The greater the difference between the congested and free flow time travel times, the greater the congestion effect.

8.5 SCHOOL MODE CHOICE

Since school location is modeled as a long term choice (see **Chapter 6.0**), there is no separate school destination choice model.

The school mode choice model will be similar to the work tour model with some key differences:

- School subtours will not be modeled. So, the number and nature of subtours will not influence school mode choice.

- Drive alone will be restricted to students aged 16 or higher.
- The school bus mode will be included in the choice set for the school tour mode choice model.

8.6 Non-Mandatory Tour Destination, Mode and Time-Of-Day Choice

This section discusses the destination, mode and time-of-day choice models for non-work, non-school tours. These include shopping, meal, personal business, social-recreational, non-school escorting tours, as well as joint tours for these purposes (see **Section 7.3**). Separate models will be estimated for joint non-mandatory tours, individual non-mandatory tours, escort tours, and work based subtours.

Destination Choice Models

The maintenance and discretionary tour destination choice models will be similar to the work tour destination choice model when the worker did not have a regular workplace identified. Accessibility measures and size functions will be the primary factors in these models.

The size function serves as a measure of attractiveness of a zone for a given trip and provides a measure for its capacity to accommodate the stop's activity purpose, similar to the number of attractions in a conventional trip distribution model.

- The size function consists of several utility-like terms that are combined in the utility function in a form that corresponds with utility theory for aggregate alternatives.
- A size function is used instead of a single size variable because the defined activity purposes and size attributes do not have a simple one-to-one correspondence.
- Rather, several attributes can indicate capacity for accommodating a given purpose. For example, personal business could be conducted at many types of places, such as restaurants, stores, or office buildings.
- The estimated coefficients give different weights to different size variables for a given purpose, and a scale parameter captures correlation among elemental activity opportunities within zones.
- Other qualitative variables can be used to better describe the attractiveness of a set of zones over and above the size effects. For example, variables that characterize the concentration of leisure, entertainment, shopping, academic, or religious activities in certain groups of zones can be used.

Mode Choice Models

The mode choice models will be of the same form as the mode choice models for the mandatory purposes, as described in **Section 8.3**. The main differences are in the incidence of various modes—as in conventional models, transit is less frequently chosen for discretionary purposes—and in the choice set of available modes. The availability of alternatives will vary for each tour type. All modes will be available for these non-mandatory purposes (except school bus) with a few exceptions.

- For escort tours, the drive alone mode will be defined as unavailable, and transit modes are unlikely to be available.
- For joint tours, the drive alone mode will be defined as unavailable, and the drive to transit mode is unlikely to be available.
- For work based subtours, drive to transit and bike modes are likely to be unavailable.

Time-of-Day Choice Models

The tour time-of-day models for maintenance and discretionary will be similar to the work tour time-of-day models although the explanatory power will largely be driven by the available time remaining to each individual after the mandatory purposes have been scheduled. Joint tours will be scheduled before individual non-mandatory tours. The same model structure and time period alternatives used for mandatory tour time of day choice will be used for non-mandatory tours.

9.0 Trip-Level Models

The model components described this far result in a roster of tours for each person in the synthetic population, with the tour type (mandatory, school escorting, joint non-mandatory, or individual non-mandatory) purpose, destination, and time of day identified. The final tour level model in the application sequence is tour mode choice, which will add the tour mode identifier to each tour. Prior to the application of tour mode choice, stop generation is modeled for each tour.

In the stop generation model, for each tour the purpose and sequence of intermediate stops on each half-tour (outbound and return) are modeled. The stop and trip characteristics are simulated in sequence, first for stops on the outbound half-tour, and then for stops for the return half-tour.

- Stops before the tour primary destination are simulated in reverse temporal sequence. First the stop's location, then its trip mode, and finally the time period of the arrival at the tour destination are simulated. These results also determine the time period in which the trip from the stop location begins, since the trip mode and travel level of service are known.
- This continues for additional stops, constructing the trip chain from the tour primary destination to the tour origin in reverse chronological sequence. The reason for simulating in reverse chronological sequence for the first half-tour is the hypothesis that people aim to arrive at the primary destination at a particular time and adjust their tour departure time so as to enable completion of the desired intermediate stops.
- The same process is followed for the stops on the return half-tour, except that stops are modeled in actual, not reverse, chronological order.

This section describes the modeling of stop generation, stop location, trip mode choice, and trip departure time.

9.1 Intermediate Stop Generation

Model Structure

For each half-tour (including subtours), this model predicts how many, if any, intermediate stops are made on that half-tour for each stop purpose. Note that for mandatory tours, the presence of stops is indicated by the daily activity pattern model (see **Section 7.1**), and this model will be applied only for tours that have been identified as having stops. On these tours, there may be stops on the outbound half-tour, the return half-tour, or on both.

A nested model structure will be tested, with the number of stops at the top level and the combinations of stop purposes at the lower level. To keep the number of alternatives to a manageable number for application and to eliminate alternatives that rarely or never occur, the household survey data set will be examined to identify the maximum number of stops that are likely to occur on tours and to eliminate combinations of stop purposes that are unlikely. For example, if the maximum number of stops is determined to be 3, then there would potentially be 120 alternatives given the seven activity purposes:

- The zero-stop alternative;
- Seven one-stop alternatives;
- Twenty-eight two-stop alternatives; and
- Sixty-four three-stop alternatives.

Many of these potential alternatives will be eliminated due to very low incidence in the household survey dataset.

Model Variables

The variables for model estimation may include the following, which can be interacted with particular numbers or combinations of stops by purpose:

- Person and household attributes (age, gender, income, number of workers, auto availability);
- Day-pattern characteristics (presence of other tours in the day, time window lengths, etc.);
- Tour characteristics (main purpose, mode, time of day, primary versus secondary versus work-based); and
- Characteristics, such as employment density, of the home location and the primary activity location.

9.2 Intermediate Stop Location

At the time that a stop's location is modeled, information about the tour, such as origin, destination, time period arriving and departing the primary destination, and tour mode, are known and can be used to explain the location choice. The number of stops in each half-tour and their purposes are also known. Additionally, details about any stops nearer to the primary destination are also known, including the location, trip mode, and the time of departure toward the tour destination (or arrival from the tour destination on the second half-tour).

However, at the time a stop's destination is modeled, several things are not known. These include the trip mode for the trip between the stop and the previous or next stop, and the departure and arrival times of stops that have not yet been simulated. The arrival time from the stop nearer to the tour origin (or departure time to that stop on return half tour) is also unknown because it will be modeled along with stop location and trip mode for the next stop further from the tour origin.

As a result of this modeling approach, two known locations serve as anchor points for calculating travel impedance. These are the stop location immediately toward the tour destination (the tour's primary activity itself for the first stop simulated on a half tour), which is called the stop origin, and the tour origin (the home, or the workplace for a work based subtour).

Model Structure

The choice of zone for an intermediate stop will be estimated using a multinomial logit model.

Model Variables

The variables for model estimation may include the following (the definitions are similar to those used in tour level destination choice, as described in **Sections 8.1** and **8.6**):

- Size function measuring attractiveness of a zone for a given trip, based on the trip purpose;
- Trip characteristic variables, including stop purpose, stop position on tour, tour mode, tour purpose, multiple stops on half-tour;
- Person characteristics, including household income, presence of children, person type and age, car availability level;
- Time window available;
- Impedance variable ("detour" impedance", calculated based on the notion that the perceived impedance of an intermediate stop is a function of the time and cost along the path from the last prior known stop location to the intermediate stop location, and on to the first subsequent known stop location.

9.3 TRIP MODE CHOICE

The trip mode choice model determines the mode for each individual trip made on that tour, based on the mode chosen for the tour. The trip mode choice model will be applied after stop generation and stop location, but before the trip departure time model. Therefore, the location of the trip origin and destination, destination stop purpose, and trip position within the tour will be known, as well as time-of-day information based on the tour time-of-day.

The trip-level mode choice model is conditional on the tour mode, and so the availability of alternatives for the trip-level models is different from the tour level models and in fact is dependent on the simulated tour mode (see **Table 8.1**).

Position in the tour is a key input variable into the trip mode choice model that is not found in the tour level model. The mode choice may vary a great deal depending on whether the trip is on the outbound or return half of the tour, and whether the trip is from the tour origin, the tour destination, or an intermediate stop. For example, drive to transit mode combinations occur almost exclusively for trips that are either leaving or returning to the home location.

Model Structure

The trip mode choice models will be estimated as nested logit models. Consistency with FTA guidelines will be examined here to guide the development and application of models that are consistent with New Starts evaluation rules.

Model Variables

The explanatory variables used to estimate the trip mode choice models will be similar to those for the tour-based mode choice model, including level of service, land use, and person and household variables. The level of service and land use data will be acquired based on the origin and destination of the trip, and not on the level of service characteristics of the entire tour.

9.4 TRIP DEPARTURE TIME-OF-DAY CHOICE

This is the "last" model in the activity simulator, predicting the timing of each trip, at the same level of resolution (30 minutes) as for the tour level time of day choice model. A key concept for this model is that the simulation order is always from the tour primary destination back towards the tour origin. That means that this model predicts either the departure or arrival time of each trip depending on which half-tour is being simulated:

- For the outbound half-tour, the trips are simulated in **reverse chronological order**. The first trip simulated in the half-tour is the trip that arrives at the primary destination, and the model predicts the **period of arrival**. Then, if there are any intermediate stops on the half-tour, the model predicts the **period of arrival** at each of those stops, each time getting **earlier** and closer to the tour origin.
- For the return half-tour, the trips are simulated in **chronological order**. The first trip simulated is the trip that leaves the primary destination, and the model predicts the **period of departure**. Then, if there are any intermediate stops on the half-tour, the model predicts the **period of**

departure from each of those stops, each time getting **later** and closer to the tour origin.

For the first trip simulated in each half-tour (the trip with one end at the primary destination), the time has already been predicted, and so this model does not have to be applied. For subsequent trips on each half-tour, this model can essentially be thought of as a duration model for intermediate stops. The periods that are available are all periods from the arrival time at the intermediate stop (departure time for the first half-tour), up until any other tour begins (ends) at the tour origin.

Model Structure

The time-of-day choice models will be simple multinomial logit across the available alternatives. Separate models may be estimated depending on the activity at the stop, but such segmentation may not be needed.

Model Variables

The variables may include the following:

- Alternative-specific constants for specific periods or groups of periods (e.g., the a.m. peak period);
- Stop purpose (serve passenger, shop, etc.);
- Tour characteristics, including tour half (outbound, return), and tour mode;
- Person and household attributes (age, gender, income, number of workers, auto, availability, etc.);
- Day-pattern types (e.g., persons with multiple work tours in a day);
- Tour origin and destination characteristics;
- How much time is already "used up" by previously simulated tours and trips, and which periods are fully and partially left; and
- The travel time by the trip mode from the current location to the next location at the time periods defined by the network skims is also known.

10.0 Other Models

In addition to the activity-based model components described in previous chapters, there are other components outside of this framework that will need to be integrated with the overall model system. These models are discussed in this section.

10.1 Air Passenger Models

The current air passenger model is a stand-alone model with the conventional four-step process - trip generation, trip distribution, mode choice, and trip assignment. The model is segmented by residency (resident vs. non-resident) and purpose (business and pleasure) and was calibrated based on regional air passenger surveys. Estimated annual enplanements at the BWI airport are converted to resident and non-resident trips traveling for business and pleasure. The observed and forecasted enplanements were based on the 2007 Washington-Baltimore Regional Air Passenger Survey. The latest 2011 survey will be used to update the annual enplanements, transfer rates, and shares of trips by four market segmentations.

10.2 External Travel

External travel refers to any travel with at least one end outside the model region. It is estimated in terms of external shares of motorized person trips as a decay function of the travel distance to the nearest external station. This function varies by trip purposes and by regions (Baltimore vs. Washington). These functions will be re-calibrated for the new base year 2012, using the latest data available.

10.3 Truck/Commercial Vehicles

The existing BMC modeling system has truck and commercial vehicle trip models that estimate vehicle trip tables by periods for medium trucks, heavy trucks, and light-duty commercial trips. The truck trip generation model uses employment-based trip rates while the distribution model is a gravity model. Truck trips are adjusted by area type factors. As part of the model calibration using the observed truck counts, origin-destination-specific adjustment factors were developed to capture the unexplained variations and are used for model applications. These adjustment factors were developed for the base year 2008. For a new base year 2012, the truck/commercial trip models will be re-calibrated and new adjustment factors will be re-established using the new truck counts.

10.4 TOLL MODEL

The existing BMC model reflects the effect of toll on travel through composite impedance skim in trip distribution, tolls in mode choice functions, and toll diversion model in highway assignment. Composite time skims are used in the gravity model to distribute trips where travel times and highway tolls are combined. The highway tolls are converted into time equivalents using values of time by trip purpose and income level. In the new activity-based model, destination choices will be modeled at the tour and trip levels (as well as in the longer term choice models), and tour purposes will supersede trip purposes for modeling attributes such as values of time. In addition, the incorporation of distributed values of time will be considered.

Based on experience in developing toll methodologies for other regions, it is recommended that the following steps be undertaken:

- Average values of time from reliable local sources should be derived for a variety of trip purposes and market segments, so that the existing composite skims can be updated and calibrated to observed data. These updated composite skims will be used only for the purposes of model estimation of the various BMC ABM components.
- In the new model, values of time will be simulated using a variable function that distributes values of time within and across various market segments. Since the highway assignments will be aggregate equilibrium assignments, individual values of time from the ABM cannot be used directly, but it will be possible to create market segments based on value of time levels, which can be translated into separate vehicle trip tables to be used in multiclass highway assignment.
- The average values of times used in the multiclass traffic assignments generally are different from (higher than) those used in the core ABM microsimulation model.
- The average values of time for trucks should be higher than those for autos. Heavy trucks should have higher values of time than the medium trucks which will have higher values than the commercial vehicles.

The estimates of hourly value of time for drivers, passengers, and trucks should be collected or compiled from other studies in order to cross-check the existing values of time. These values of time estimates typically include wages, average vehicle occupancy, and cargo inventory value for commercial vehicles. This information could be derived from local studies on traveler and freight characteristics and be segmented by income group, trip purpose, and vehicle type.

CS evaluated the advantages and disadvantages of segmenting mode choice auto alternatives into "willing to pay (a toll)" and "non-willing to pay (a toll)" and recommends that the pay vs. no pay choice be incorporated directly in highway assignment. The toll modeling approach options will be evaluated in the peer review process to reach a decision on the preferred approach.

11.0 Assignment

The trip assignment model is the last step of the modeling process. Trip assignment estimates the volume on each link in the transportation system for both highway and transit modes. In addition, the trip assignment model generates specific performance measures, such as the congested speed or travel time on a highway link or the boardings and alightings on a transit route. Trip assignment is performed separately for each mode (auto and transit) and time period (for example, a.m. peak, mid-day off-peak, p.m. peak, and overnight).

There are two primary objectives for the trip assignment model. The first objective is to assign trip tables and produce measures of impedance for most of the ABM components. The second objective is to assign the trip tables and produce volumes for auto and transit networks. These are described separately in the following sections.

11.1 Time Periods

The trip assignment model currently is set up for four time periods:

- A.M. peak (6:30 a.m. to 9:30 a.m.);
- Midday (9:30 a.m. to 3:30 p.m.);
- P.M. peak (3:30 p.m. to 6:30 p.m.); and
- Overnight (6:30 p.m. to 6:30 a.m.).

As discussed in **Section 3.1**, BMC has indicated a desire to separately assign the a.m. peak hour within the a.m. peak period, and the p.m. peak hour within the p.m. peak period. Based on the traffic count data and survey data. The peak hours are proposed to be 7:30 – 8:30 AM for the AM peak hour and 5-6 PM for the PM peak hour. Furthermore, recent traffic count data and household travel survey indicate a longer p.m. peak period than the current three-hour period. An alternative p.m. peak period of 3:00-7:00 is proposed for consideration.

11.2 Highway Assignment

The current BMC model uses fixed numbers of iterations for the four time periods, with all trips assigned on each iteration and the final volumes representing a weighted average of all iterations. The convergence criteria include the following:

- Relative gap of 0.000001
- Relative average absolute volume difference of 0.005

- Root mean squared error of the differences of 0.1
- Average absolute volume difference of 0.5

The iteration and convergence process will be evaluated and modified as necessary to achieve a stable, reasonable convergence of the highway assignment results.

One recommendation for the highway assignment is a static, aggregate equilibrium procedure to assign trips to the roadway network for different time periods. This is a user optimal procedure that is based on the assumption that each traveler chooses a route that has the lowest generalized cost path.

Classes

The number of classes of trips to be assigned in the multiclass assignment will be decided in conjunction with BMC and its planning partners. These classes will be based on certain stratifications such as mode, income group, value of time range, trip purpose, and vehicle type. These could include but are not limited to the following:

- **By Mode –** This is critical to account for the occupancy of vehicles, especially useful for evaluating HOV and managed lanes in the transportation system.
  - Single-occupant vehicle (SOV); and
  - High-occupant vehicle (HOV) which can further be classified by auto occupancy (2, 3+);
- **Value of time/income ranges.** To account for differences in travelers' values of time, especially their effects on route choices between priced and free roadway facilities, it is proposed to simulate each individual's value of time (see **Section 8.3**). To preserve this disaggregate information for use in highway assignment, it is proposed that the auto trip tables generated from the demand models are segmented based on the simulated values of time for travelers, consistent with the way in which these segments will be defined for the mode choice model. An average value of time for each segment would be used in developing impedance measures combining time and cost. The result will be that the trip tables representing higher values of time will be more likely to be assigned to priced roadways than those representing lower values of time.
- **By Vehicle Type –** This is critical to account for passenger car equivalents (PCE) for commercial vehicles. The following three truck types are currently in the BMC model:
  - Commercial vehicles (light-duty);
  - Medium trucks; and
  - Heavy trucks.

Volume-Delay Functions

The highway assignment procedure is applied in an iterative fashion, where travel times are updated after each iteration to reflect congestion occurring on the network. These updates to travel time are based on a volume-delay function for each link. The existing volume-delay functions are the modified versions of the standard BPR functions with parameters varying by facility types, which will be reviewed and modified if necessary during validation. The free-flow time is based initially on the network data provided for each link and then updated in each iteration to represent the travel time resulting from the assigned traffic volumes from the last iteration.

Turn Penalties

Turn penalties are included in the trip assignment model to either prohibit certain turn movements or to penalize certain turn movements. These are included in the model by identifying specific turn movements by their node numbers, and then coding the penalty function that will apply to these turn movements. It is assumed that the current model turn penalties will be retained.

11.3 TRANSIT ASSIGNMENT

The transit assignment uses a best path algorithm, which is one of the two assignment algorithm methods in Cube Public Transport. The best path is the single path for each traveler that minimizes the perceived travel time, including time spent walking, waiting, and riding. Time spent waiting for a transit vehicle is calculated based on the fact that there may be many transit vehicles traveling from a specific origin to a specific destination and the traveler will choose to take the first vehicle that arrives.

Modes

The current BMC model classifies the transit mode into four submodes: local bus, express bus, rail, and commuter rail. Two modes of access are used: walk and drive. As a result, eight combinations of access and mode are used for peak and off-peak assignments in the transit assignment process. Since the transit mode alternatives from mode choice are different for the new model (see **Section 8.3**), the mode definitions will be different for transit assignment as well, to be consistent with the new mode choice model.

Transit Time Functions

Transit time functions are used to account for the fact that transit vehicles have to stop and pick up passengers along the route, and typically travel at slightly slower speeds than passenger cars due to their size and weight. The transit time functions are used to estimate transit travel time as a function of highway travel time. In the current BMC model, transit lines use the congested time on the links during the peak periods, off-peak period transit times are estimated from a look-up table of fixed speeds by area types that is not sensitive to roadway congestion. The current transit time functions will be reviewed and modified if necessary during transit assignment validation.