CAMBRIDGE SYSTEMATICS

Think Forward

Memorandum

TO: Metropolitan Washington Council of Governments

FROM: Cambridge Systematics, Inc.

DATE: September 14, 2016, Revised September 22, 2016

RE: Task Order 16.2: Task #13, Develop Census and Household Travel Survey Database

---

This documentation summarizes activities that Cambridge Systematics, Inc. (CS) performed under Task Order 16.2: Advice and Testing, Task 13 Develop Census and Household Travel Survey Database. Specifically, the task activities include a review of existing Census and survey data and their potential use, such as for model estimation and calibration, in the upcoming model updates in trip-based and activity-based models. This document contains two distinct parts. The first identifies the data enhancements needed for the short-term trip-based model improvements. The second part contains a description and specification of the survey data to be used in estimating an activity-based model for the region.

Survey Data Enhancement for Trip-Based Model Improvements

Existing practice

COG/TPB staff has conducted model estimation and calibration in the past. The current version of the regional (trip-based) travel demand forecasting model is known as Version 2.3.57a, but the mode choice model was calibrated in 2011 (this was an earlier version of the 2.3 model). The calibration included the following two steps:

1. Statistical estimation: Statistical estimation was used for some coefficients (e.g., in-vehicle time and cost for income group 1), but the values of other coefficients were set by fiat, typically based on rules of thumb (e.g., out-of-vehicle time coefficients are some multiple of in-vehicle time coefficients).[1]

2. Calibration: Calibration typically involves adjusting one or more constants, such as alternative-specific constants, to ensure that the model will fit observed data. In the case of the COG/TPB mode choice model, staff made adjustments to the nesting constants, which are based on a series of geographic market segments.[2] More specifically, staff used

[1] Ronald Milone et al., *Calibration Report for the TPB Travel Forecasting Model, Version 2.3, on the 3,722-Zone Area System*, Final Report (Washington, D.C.: National Capital Region Transportation Planning Board, January 20, 2012), 6–14, http://www.mwcog.org/transportation/activities/models/documentation.asp.

[2] Ibid., 6–18.

a series of "calibration targets," defined as the number of person trips, for an average weekday, for each trip purpose (5), travel mode (15), and geographic market segment (20). Calibration targets were divided into two groups: transit person trips and auto person trips. Transit person trip calibration targets came from on-board transit surveys, but the auto person trip calibration targets came from the 2007/2008 Household Travel Survey. As staff noted, "although it would be more consistent to get both transit person trip targets and auto person trip targets from one source, such as the HTS, the reality is that by using on-board transit surveys for the transit person trip targets, we were able to take advantage of more data (compare ca. 5,000 unweighted transit person trips in the HTS versus ca. 50,000 unweighted transit person trips in the on-board transit surveys."[3] The following transit on-board surveys were used to develop transit person trip targets:

a. 2008 Metrorail Passenger Survey;

b. 2008 Regional Bus Survey (supplemented by the Fairfax Connector Bus Survey);

c. 2007-2008 On-Board Survey of Maryland Transit Administration (MTA) Riders, which would include survey information from riders of the Maryland Area Regional Commuter (MARC) train service; and

d. 2005 Virginia Railway Express (VRE) Passenger Survey.

Household travel survey

Household survey data will be instrumental to estimating the mode choice models. Household 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 a relational database containing three tables/files: the household file, person file, and trip file. The household file contains household-level information for each person surveyed, including household size, number of workers, household income, and number of vehicles, which will be important variables to the mode choice models. The person file likely will not be needed for estimation of the model choice models. The trip file contains all of the critical trip information, including origin location, destination location, trip purpose, trip mode, and time of day.

In addition to the 2007/2008 Household Travel Survey (HTS), the 2011/12 TPB Geographically-Focused Household Travel Survey (GFHTS) surveys may be useful, particularly to the non-motorized model enhancements.

The two surveys will need to be enhanced with additional attributes such as the Census block numbers associated with household and workplace locations and trip ends, in order to facilitate development of more refined variables related to urban form and built environment, as outlined in the technical memorandum for Task Order 16.4: Non-Motorized Model Enhancements.

Skimmed highway and transit attributes will be attached to the data for model estimation. Since skimmed attributes are coded at the TAZ-to-TAZ level and survey data is coded at the Census

[3] Ibid.

block level, a crosswalk between TAZ and Census block will be used in attaching the skim attributes. To enhance the accuracy of measurements for short-distance trips, the orthogonal distance (ΔX plus ΔY) can be used to replace the TAZ-TAZ skim distance when the skim distance is less than a certain threshold (e.g., 3.75 miles for the Baltimore model). For the model estimation data set, the orthogonal distances can be determined from the coordinates of the trip ends. Alternatively, if privacy rules do not allow for the use of the trip end coordinates, one could also use the coordinates of the Census block centroids associated with those trip ends.

Transit on-board survey

The transit on-board survey was conducted on many transit vehicles including local, regional, and commuter bus providers, Metrorail subway, MTA light rail, and MARC commuter rail[4]. Derived from roughly 13,000 usable questionnaires[5], the survey data contain information on the respondent's current transit trip. This includes trip starting and ending stations, home 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.

The transit on-board survey can be used in two ways. First, it can be added to the household survey dataset to create a more robust dataset for model estimation. Since all records in the on-board survey dataset use the transit mode, careful attention needs to be paid to the weights assigned to on-board survey records to ensure that the overall share of transit suggested by the survey weights is reflective of the region's transit share. The larger, more robust combined dataset should allow us to obtain more precise estimates of the different coefficient estimates we plan on testing in the mode choice model (e.g., different in-vehicle time weights by transit technology, and non-traditional transit variables).

The second way the data can be used is to test the path-building weights used to generate transit paths. It is currently planned to use Cube Public Transport (PT) for transit path-building. Cube-PT can be used to generate either a single best path between each origin and destination or multiple paths between each O & D. When the single path option is selected, Cube-PT finds the single best path (based upon the path building parameter set) and reports the travel attributes of that single path in the skim variables. When the multiple path option is selected, Cube-PT finds all paths that would be reasonably chosen by a traveler for the OD pair. Cube-PT will ultimately generate "averaged" paths, meaning the transit skim attributes will represent averages of the attributes across all used paths, but it can also be used to generate best paths and potentially second-best paths. These paths generated by Cube-PT can then be compared against the on-board survey data. While the on-board survey data does not contain all path-related information,[6] it contains information about the type of transit the survey was conducted on, which can be compared to the skimmed paths. In other words, the comparisons will not be perfect, but the on-board survey data can be used to tweak the path-building parameters.

[4] Cambridge Systematics, Inc., Model Design Plan for BMC Activity-Based Model. Prepared for the Baltimore Metropolitan Council. 2014.

[5] Baltimore Metropolitan Council, 2007 On-Board Transit Survey – BMC Analysis, Task Report 10-1, 2010. Accessed via http://www.baltometro.org/reports/On-Board-Transit-Survey-2007.pdf

[6] For example, the 2007 Metrorail Passenger Survey did not include the production-end mode of access to the first transit vehicle in a path, only the production-end mode of access to Metrorail. This is one of the reasons COG/TPB staff decided to use the 2008 Metrorail Passenger Survey instead of the 2007 survey.

Similar to the HTS, the transit on-board survey database can be enhanced with more refined geocoding such as Census block numbers. To enhance the accuracy of measurements for transit walk access and egress, the orthogonal distances for transit walk access and egress can be used to replace the TAZ-TAZ skim distance, using the coordinates of the trip ends and the nearest transit stop/station.

Activity-Based Model Estimation Data

Household Survey Data

The household and person variables for model estimation for an activity-based model are shown in Table 1[7], which includes both original survey response variables and some of the variables that can be derived from the survey responses.

Table 1: Person and Household Data

Description	Details	Type
Person and Household Demographic Variables		
Household ID number	Survey ID field	Integer
Person ID number	Survey ID field	Integer
Household size	Count of people in household	Integer
Household vehicles	Count of vehicles in household	Integer
Total household income level	Categorical household income	Categorical
Gender	male, female	Categorical
Age	In years	Integer
Employment status	employed full-time, employed part-time, not employed	Categorical
Student status	enrolled full-time, enrolled part-time, not enrolled	Categorical
Type of school enrolled in	preschool, K-12, post-HS, not enrolled	Categorical
Relationship to respondent	Head, spouse, partner, other HH member, visitor	Categorical
Person and Household Derived Variables		
Person type	Derived (e.g., full-time worker, part-time worker, retired other adult, university student, driving age high school student, child age 5-15, child age 0-4)	Categorical
# of employed HH members	Derived by adding across HH members	Integer
# of student HH members	Derived by adding across HH members	Integer
# of HH members by person type	Derived by adding across HH members	Integer
Person and Household Location Variables		
Household residence ID number	Survey ID field	Integer

[7] Cambridge Systematics, Inc., Model Design Plan for BMC Activity-Based Model. Prepared for the Baltimore Metropolitan Council. 2014.

Description	Details	Type
Household Census block	Geocode (Dependent variable for population synthesizer)	Integer
Regular work location id	Survey ID field	Integer
Regular work Census block	Geocode (Dependent variable for regular work location model)	Integer

The trip file is processed to create an additional file/table of tours. The tours consist of a series of at least two trips that start and end at the same location – home or work. One stop in the tour is designated as the location of the primary activity. This is the tour's destination and its purpose is determined by the primary activity. All trips before the primary activity are in the first half of the tour. Trips that follow it are in the second half. Table 2 shows the estimation variables necessary from the trip and tour databases.

Table 2: Trip and Tour Data

Description	Details	Type
	Trip-Level Variables	
Trip tour half	1 or 2, Created ID field	Integer
Trip ID within tour half	Created ID field	Integer
Trip ID	Survey ID field	Integer
Trip origin activity purpose	Same codes as primary destination activity purpose	Categorical
Trip destination activity purpose	Same codes as primary destination activity purpose	Categorical
Trip origin location ID	Survey ID field	Integer
Trip origin zone	Zone Number (Tour destination, or destination of previous trip)	Integer
Trip destination location ID	Survey ID field	Integer
Trip destination zone	Zone Number (Tour origin, or dependent variable for stop location)	Integer
Trip mode	Same codes as tours (Dependent variable for trip mode model)	Categorical
Trip origin departure time	Dependent variable for trip departure time model	Continuous
Trip destination arrival time		Continuous
	Day Pattern-Level Variables	
# home-based tour records		Integer
# home-based tours by tour type	Dependent variables for day activity pattern models	Integer
# work-based subtour records		Integer
# intermediate stops by stop purpose	Dependent variable for day activity pattern models	Integer
	Tour-Level Variables	
Tour ID number (in priority order)	Created ID field	Integer

Description	Details	Type
Subtour parent tour ID	Created ID field (work based subtour only)	Integer
Subtour ID within parent tour	Created ID field (work based subtour only)	Integer
# of subtours within tour	Dependent variable for subtour frequency/purpose model	Integer
Primary destination activity purpose	(work, school, shopping, meal, social/recreation, etc.)	Categorical
Tour origin outbound departure time		Integer
Primary destination arrival time	Dependent variable for tour times of day model	Integer
Primary destination departure time	Dependent variable for tour times of day model	Integer
Tour origin return arrival time		
Primary destination location id	Survey ID field	
Primary destination zone	Zone Number (Dependent variable for tour destination model)	Integer
Tour primary mode	(Dependent variable for tour mode model)	Categorical
# 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	

The tables above contain six main categories of data:

1. **Basic person and household variables**. These are the truly exogenous variables. In model 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. **Trip-level variables**. These variables include trip origin and destination location and purpose, trip departure and arrival time, and trip mode. They come directly from the

survey, with the exception of the tour half ID. In model application, the order is reversed with tours preceding trips – trip characteristics 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.

5. **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).

6. **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.

Other Data for Model Estimation

In addition to the information from the household travel survey, other data items needed for model estimation include:

- Land use data such as aggregate household and employment and other data at the zonal level for the model year

- Parcel/parcel point-based data as outlined in the technical memorandum for Task Order 16.2: Subtask 12 Develop Parcel-Level Database

- Highway network and skims such as auto travel distance, time and toll

- Transit network and skims such as transit travel time, distance and cost, as well as transfers and first and last transit stops/stations.