MAG CT-RAMP Activity-Based Model
Phase I

Model Estimation Results

Prepared for the Maricopa Association of Governments

Prepared by Parsons Brinckerhoff
In association with
Arizona State University

August 2010

Table of Contents

List of Tables

List of Figures

CHAPTER 1.

OVERVIEW OF ESTIMATED MODELS AND DATA SOURCES

1.1. MODELS ESTIMATED AT PHASE 1

The general structure of the MAG Activity-Based Model (ABM) is shown in Figure 1 below with the models implemented at Phase 1 shadowed in dark gray. This advanced model structure has many innovative components that are explained in detail in the Model Design & Development Plan document. In this report, model estimation results are presented for all model components that were estimated at Phase 1. All choice models have been estimated based on the Phoenix and Tucson data with the exception of the model for participation in special events, which is currently a placeholder with default coefficients. This report contains a description of the data sources used for model estimation and corresponding data-processing steps (including the calculation of accessibility measures that integrate upper-level choices with lower-level model components), and a description of the following model components:

- The system of models for university students, which includes residential type choice (dorm, off-campus rent apartment, off-campus household) and residential TAZ choice for off-campus choices.
- Usual workplace type (from home, usual outside home, variable) for workers by occupation.
- Usual schooling type (from home, outside home) for students by school type (elementary, mid, high, college/university).
- Usual workplace location TAZ. For choice of out-of-home workplace, advanced non-linear forms of impedance functions are presented. These functions include composite mode and time-of-day choice logsums, and various distance-decay terms.
- Household and person mobility attributes including car ownership, transit pass, free parking eligibility, and toll transponder. The innovative component of this sub-model relates to simultaneous treatment of household-level and person-level choices in one extended nested logit model.
- Coordinated daily activity pattern type model for all household members. This model includes a combinatorial nested structure (previous models of this type were all simple multinomial logit), and integration of the choice of the daily activity pattern for each household member with the choice of joint activity episodes among two or more household members (these two components were separated in previously developed model systems). The latter feature (integration with the choice of joint activity episode) has been also included and tested in the San Diego CT-RAMP ABM.

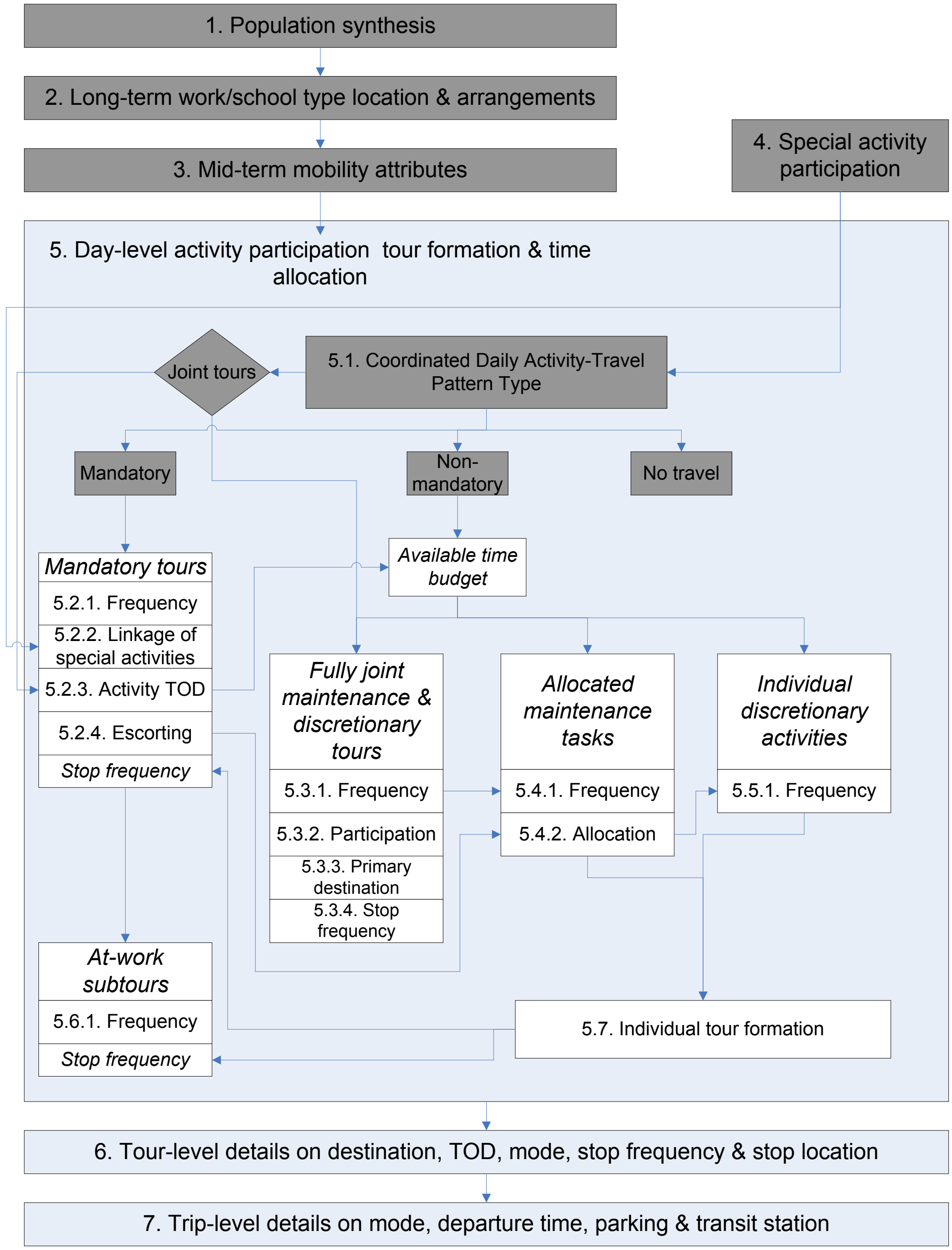

FIGURE 1: MODELS ESTIMATED AT PHASE 1 IN THE MAG ABM SYSTEM

1.2. BUILDING DATABASE FOR MODEL ESTIMATION FROM NHTS, 2008

1.2.1. Data Processing Steps

The National Household Travel Survey (NHTS) was the primary data source for estimating and calibrating the MAG ABM. The NHTS was implemented as an add-on for the Phoenix and Tucson Metropolitan Areas in fall 2008. A total of 7,000 households were collected with a full record of activities and trips for 24 hours. Only households who were assigned a weekday travel day (5,000) were used in the model estimation.

The structure of an ABM requires certain transformations of the original data files in order to create the necessary travel and activity dimensions utilized in data analysis and model estimation. The complexity of data processing for ABMs requires a flexible programming approach where different summaries and outputs can be created from a core set of data file structures. A suite of Visual Foxpro software programs developed over time in various ABM development projects was used to process the NHTS data accordingly.

The current memo outlines the database structure in its current form, as well as further steps to create a fully-functional data processing system for MAG that would serve the needs of further ABM development and Phases 2 and 3.

Data processing for ABM development can be broken into two stages: 1) Core data transformation; and 2) preparation of model-specific data files for estimation – see Figure 2 below. Core data transformation requires transforming travel diary data into trips and tours; as a consequence, it is typically fairly standard across different ABMs. . The second stage is dependent on the adopted structure of the travel model and components, which may be specific to regional conditions, such as mode choices. For this stage, only some general data-processing guidelines can be formulated in advance, while the details are finalized in the process of model development and estimation.

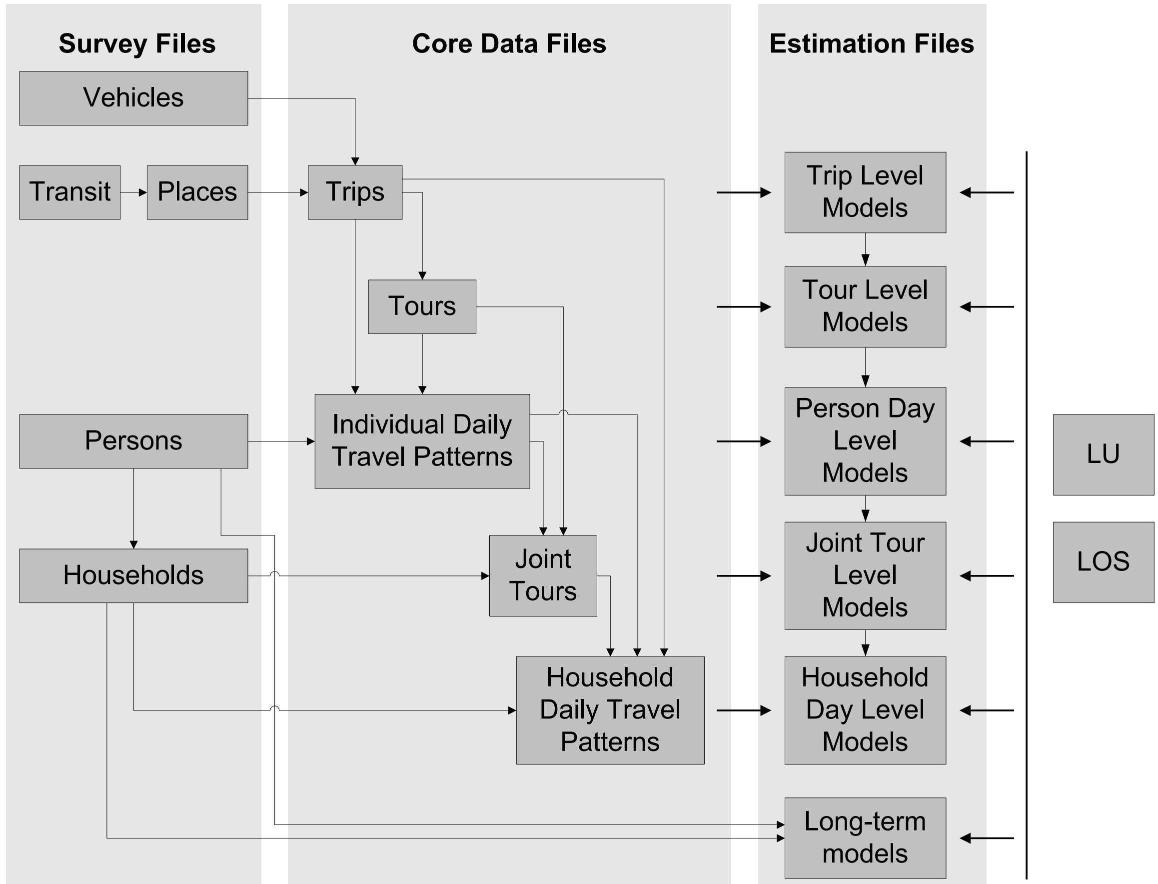

FIGURE 2: MAIN DATA-PROCESSING STEPS WITH NHTS

The following four original files were provided in the NHTS set:

- *Households:* This file contains information on household characteristics.
- *Persons:* This file contains information on person characteristics.
- *Places/Trips:* This file contains information on all places visited by the person and associated activities; this file provides information on all trips since each trip is associated with changing locations. (In some household travel surveys, this file is supplemented by a separate file for transit trips providing an exact itinerary, location of all boarding and alighting stops, etc)
- *Vehicles:* This file contains specific information on the vehicles used by the survey households, and is typically used only for advanced car ownership and car allocation models that include car-type choice.

The first stage of core data transformation includes the following five file-construction steps:

- *Trips:* This file contains trip records with additional information relevant for modeling purposes; this is a major data file for estimation and calibration of a 4-step model.
- *Tours:* This part of database contains two files:
    - Extended trip file in a trip-record format but with a tour ID assigned to each record: This file is derived from the Trips file (one record per trip) but contains additional fields related to the tour structure.
    - Tour file in a tour-record format: This file contains one consolidated record per tour.
- *Individual Daily Activity-Travel Patterns (DAP):* This file contains main travel (trips, tours) and activity-related parameters summarized for each person.
- *Joint tours:* This file contains fully and partially joint tours (when two or more household members travel together); this file is specifically needed for an advanced ABM of the CT-RAMP type
- *Joint Household Daily Activity-Travel Patterns (Joint DAP):* This file contains main travel (trips, tours) and activity-related parameters summarized for each household and including individual and joint travel.

The second stage of file preparation for estimation is dependent upon the adopted model system structure that defines the set of models and their structural characteristics. However, all models can be broken into the following six groups based on the decision-making unit associated with each underlying choice:

- Trip-level models,
- Tour-level models,
- Person-day-level models,
- Joint-tour-level models,
- Household-day level models,
- Household / person long-term models.

The CT-RAMP family of ABMs such as those developed for Columbus, Atlanta, San-Francisco Bay Area, and San-Diego include all six types of models. Certain ABMs do not explicitly model joint tours, and therefore require only five types of models. For a 4-step model system, only trip-level models and person or household day-level models are normally applied depending on the trip generation model structure (either person-based or household-based). Also, the long-term models for a 4-step system include normally a household car ownership model only, while for an ABM system, they also include usual workplace and school location for each household member as well as a more extended set of individual mobility attributes in addition to car ownership.

1.2.2. Key Fields and Linkages across Data Files

The content of each file includes several key fields and a set of associated data items. The key fields are crucial for a proper organization of the whole database including the original survey files, core data files, and files built for estimation of particular models. The data items are always indexed by one or several key fields and can be divided into two groups – static variables and situational variables.

Static variables include data items that are independent of the model structure, for example, age and gender of the modeled person. In the model estimation and application, static items are predetermined in the input files. Situational variables include data items that describe person and household activities / travel, for example, number of activities of a certain type undertaken in the course of the day. In the model estimation, situation variables are also available in the input files though they normally require a certain data transformation. However, in the model application, situational variables are available only if they are modeled at the higher levels of modeling hierarchy relative to the given model. Stated otherwise, situational variables relate to the outcomes of the previously modeled decision-making steps (choices). Thus, calculation and storage of the situational variables depends on the adopted model structure. However, based on the extensive experience with ABMs of different structures (developed for San Francisco County, New York, Columbus, Atlanta, Bay Area, and San-Diego) and vision of the prototype ABM structure for MAG described in the Model Design Document, we can formulate almost all possible situational variables and incorporate them in the core data files in advance.

The most important key fields (by which the files are normally indexed in data processingindices) and their use in the data files are presented in the Table 1 below. They are divided into two groups – indices originally present in the survey files and indices added during the construction of the core data files.

TABLE 1: KEY FIELDS (INDICES) IN THE DATABASE

Index	Survey files					Core Files					Estimation Files					
	Households	**Persons**	**Places**	**Transit roster**	**Vehicles**	**Trips**	**Tours**	**Individual DAP**	**Joint tours**	**Household DAP**	**Trip level**	**Tour level**	**Person day level**	**Joint tour level**	**HH day level**	**Long-term**
Household ID (sampn)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Person ID (perno)		X	X	X		X	X	X	XX	XX	X	X	X	XX	XX	X
Day ID (dayno)			X	X			X	X	X	X	X	X	X	X	X	
Place ID (plano)			X	X		X					X					
Transit segment ID (trnno)				X												
Location ID (locno)			X			X					X					
Trip ID (tripno)			X	X		X					X	XX				
Tour_ID						X*			XX		X	X				
Primary tour destination						X*			X					X		
Sub-tour ID (subt_id)						X*					X	XX				
Primary subt destination						X*										
Joint tour ID (j_tour_ID)						X*			X	XX	X	X	XX	X	XX	
Outbound stop no						X*					X	XX				
Inbound stop no						X*					X	XX				
Vehicle ID (vehno)			X		X	X	X		X		X	X				
Residential TAZ	X	X						X	X	X		X	X	X	X	X
Usual workplace TAZ		X						X		XX			X		XX	XX
Usual school TAZ		X						X		XX			X		XX	XX
Trip origin TAZ						X					X					
Trip destination TAZ						X					X					
Tour origin TAZ							X		X			X		X		
Tour destination TAZ							X		X			X		X		
Time bin occupied 0-48			X			X	X	X	X	X	X	X	X	X	X	

X – single index.

XX – multiple index combined of a predetermined number of single indices.

X* - added to the trip file in the tour-construction procedure.

1.2.3. Quality Checks for the NHTS 2008 Add-On for Phoenix and Tucson

In general, the following quality / consistency checks are implemented as part of the tour construction procedure:

- Validity of the *entire-tour mode* and each half-tour mode (outbound and inbound); the entire-tour mode is defined by a predetermined hierarchy of all trip modes used on the tour; the half-tour mode are defined in the same way but only for trips on the half-tour; the following cases are considered invalid, though with different level of severity:
  - Unknown mode on one or more trips on a half-tour, while mode is known for all trips on the opposite half-tour;
  - Unknown mode for one or more trips on both half-tours, which prevents identification of half-tour modes;
  - Unknown mode for all trips on the tour ;
  - Known half-tour mode that contradicts mode availability rules,
- *Completeness (or closeness) of tours* in terms of starting and ending at home; the following cases are distinguished:
  - Start from a location other than home – valid reason (trip from airport, etc),
  - Start from a location other than home – invalid reason (frequently, missing trip),
  - End at location other than home – valid reason (trip to airport, etc),
  - End at location other than home – invalid reason (frequently, missing trip),
- *Consistency of time-related tour attributes*; the following cases are considered invalid:
  - Missing departure / arrival trip time,
  - Conflicting trip/activity time chain (moving backward in daily schedule),
  - Unrealistic trip time versus skims; in general, it is defined by setting a threshold for maximum and minimum speed for each mode using either highway or areal distance; this flag however, has to be considered as an indication on a possible range of problems including miscoded trip origin and/or destination, miscoded mode, etc, and not always it is possible to separate one particular problem
- Completeness of *trip destination coding*; the following cases are distinguished:
  - Missing / unknown destination zones,
  - Destination outside the modeled region (intercity trips),
- Mode symmetry between outbound and inbound directions; normally most of the tours have the same outbound and inbound modes; there can be valid cases of asymmetric modes, for example being a car passenger on the outbound half-tour while taking transit on the inbound half-tour; however, percentage of asymmetric tours should be low.

The quality control procedure is first implemented for elemental trips and then, in a more sophisticated way, for tours. Trips and tours are identified by corresponding validity flags. The results of the most important validity checks of the MAG/PAG NHTS 2008 at the trip level are presented in Table 2. For evaluation of the survey quality, the same tabulation is presented for the Household Survey implemented for the Chicago Metropolitan Agency for Planning (CMAP) in 2007 in Table 3 below and

San-Francisco Bay Area Travel Survey implemented in 2000 (BATS2000) in Table 4 below (with a slightly different numbering convention for trip purposes). Both surveys are comparable to MAG/PAG NHTS2008 in terms of the survey technology, sample size, and complexity of the region, though the survey questionnaire for BATS2000 was somewhat simpler compared to the CMAP2007 and NHTS2008 questionnaire.

TABLE 2: QUALITY CONTROL, TRIP LEVEL, MAG/PAG, NHTS 2008, REGULAR WORKDAY

Destination purpose	Number of unlinked trips				
	Total	Valid mode	Valid timing	Valid destination	All valid
Absolute number of trips in the survey (not expanded):					
0=Home	14,844	14,740	14,829	14,806	14,693
1=Work	4,729	4,666	4,724	4,592	4,535
2=University	331	331	331	329	329
3=School	1,446	1,435	1,445	1,381	1,370
4=Escort	2,867	2,857	2,867	2,749	2,739
5=Shopping	5,100	5,068	5,097	4,856	4,826
6=Other maintenance	4,168	4,137	4,165	3,980	3,955
7=Eating out	2,451	2,424	2,449	2,308	2,283
8=Visiting	1,059	1,044	1,057	976	959
9=Other discretionary	3,719	3,688	3,717	3,434	3,409
10=Buy Gas	518	513	516	471	465
11=Unknown	212	185	210	178	158
Total	41,444	41,088	41,407	40,060	39,721
Row percent:					
0=Home	100.0%	99.3%	99.9%	99.7%	99.0%
1=Work	100.0%	98.7%	99.9%	97.1%	95.9%
2=University	100.0%	100.0%	100.0%	99.4%	99.4%
3=School	100.0%	99.2%	99.9%	95.5%	94.7%
4=Escort	100.0%	99.7%	100.0%	95.9%	95.5%
5=Shopping	100.0%	99.4%	99.9%	95.2%	94.6%
6=Other maintenance	100.0%	99.3%	99.9%	95.5%	94.9%
7=Eating out	100.0%	98.9%	99.9%	94.2%	93.1%
8=Visiting	100.0%	98.6%	99.8%	92.2%	90.6%
9=Other discretionary	100.0%	99.2%	99.9%	92.3%	91.7%
10=Buy Gas	100.0%	99.0%	99.6%	90.9%	89.8%
11=Unknown	100.0%	87.3%	99.1%	84.0%	74.5%
Total	100.0%	99.1%	99.9%	96.7%	95.8%

TABLE 3: QUALITY CONTROL, TRIP LEVEL, CMAP, HTS 2007, REGULAR WORKDAY

Destination purpose	Number of trips				
	Total	Valid mode	Valid timing	Valid destination*	All valid
Absolute number of trips in the survey (not expanded):					
0=Home	60,361	60,209	60,343	60,361	60,191
1=Work	22,369	22,239	22,365	21,346	21,268
2=University	614	610	614	588	585
3=School	4,845	4,834	4,845	4,780	4,769
4=Escort	11,367	11,354	11,364	10,951	10,936
5=Shopping	20,914	20,880	20,913	20,296	20,261
6=Other maintenance	18,622	18,533	18,620	17,818	17,738
7=Eating out	10,515	10,501	10,512	10,027	10,010
8=Visiting	7,533	7,509	7,530	6,688	6,679
9=Other discretionary	12,642	12,559	12,641	12,106	12,039
10=Changing mode	1,940	1,845	1,938	1,864	1,776
11=Loop	117	114	117	103	100
12=Unknown	6	6	6	4	4
Total	171,845	171,193	171,808	166,932	166,356
Row percent:					
0=Home	100.0%	99.7%	100.0%	100.0%	99.7%
1=Work	100.0%	99.4%	100.0%	95.4%	95.1%
2=University	100.0%	99.3%	100.0%	95.8%	95.3%
3=School	100.0%	99.8%	100.0%	98.7%	98.4%
4=Escort	100.0%	99.9%	100.0%	96.3%	96.2%
5=Shopping	100.0%	99.8%	100.0%	97.0%	96.9%
6=Other maintenance	100.0%	99.5%	100.0%	95.7%	95.3%
7=Eating out	100.0%	99.9%	100.0%	95.4%	95.2%
8=Visiting	100.0%	99.7%	100.0%	88.8%	88.7%
9=Other discretionary	100.0%	99.3%	100.0%	95.8%	95.2%
10=Changing mode	100.0%	95.1%	99.9%	96.1%	91.5%
11=Loop	100.0%	97.4%	100.0%	88.0%	85.5%
12=Unknown	100.0%	100.0%	100.0%	66.7%	66.7%
Total	100.0%	99.6%	100.0%	97.1%	96.8%

* - included valid TAZ for internal destinations and valid purpose/mode code for external destinations

TABLE 4: QUALITY CONTROL, TRIP LEVEL, BATS 2000, REGULAR WORKDAY

Destination purpose	Number of trips				
	Total	Valid mode	Valid timing*	Valid destination**	All valid
Absolute number of trips in the survey (not expanded):					
1=Work	36,754	36,456	36,518	33,976	33,511
2/3=School/University	11,930	11,834	11,884	11,375	11,242
4=Escort	18,004	17,978	17,884	15,926	15,820
5=Shopping	21,376	21,296	21,265	19,378	19,211
6=Other maintenance	14,890	14,794	14,737	13,169	12,995
7=Eating out	15,738	15,662	15,622	13,884	13,733
8=Visiting	5,122	5,088	5,082	4,158	4,103
9=Other discretionary	14,463	14,319	14,333	12,352	12,146
10=Home	73,488	73,061	73,369	73,488	72,945
11=Changing mode	16,799	16,126	16,754	14,450	13,888
12=Unknown	2,752	2,690	2,750	1,748	1,728
Total	231,316	229,304	230,198	213,904	211,322
Row percent:					
1=Work	100.0%	99.2%	99.4%	92.4%	91.2%
2/3=School/University	100.0%	99.2%	99.6%	95.3%	94.2%
4=Escort	100.0%	99.9%	99.3%	88.5%	87.9%
5=Shopping	100.0%	99.6%	99.5%	90.7%	89.9%
6=Other maintenance	100.0%	99.4%	99.0%	88.4%	87.3%
7=Eating out	100.0%	99.5%	99.3%	88.2%	87.3%
8=Visiting	100.0%	99.3%	99.2%	81.2%	80.1%
9=Other discretionary	100.0%	99.0%	99.1%	85.4%	84.0%
10=Home	100.0%	99.4%	99.8%	100.0%	99.3%
11=Changing mode	100.0%	96.0%	99.7%	86.0%	82.7%
12=Unknown	100.0%	97.7%	99.9%	63.5%	62.8%
Total	100.0%	99.1%	99.5%	92.5%	91.4%

* - after imputing values for missing departure and arrival time.

** - included valid TAZ for internal destinations and valid code for external destinations.

In general the MAG/PAG NHTS2008 is characterized by a very good quality of reported mode, timing, and destination trip data with almost 96% of trip records usable for tour-construction purposes across all dimensions (mode, time-of-day, and destination). In relative terms, it is close to the quality of the CMAP HTS2007 data (97%) and significantly better than the quality of the BATS2000 survey (91%) as well as all previous regional surveys use for ABM development (in New York, Columbus, and Atlanta), where the percentage of usable trip records was in the range of 88-91%.

In the tour-construction procedure, trip records with some invalid characteristics are not automatically discarded. Instead, an attempt is made to utilize other reported information to identify the missing information. A tour was only flagged as invalid in cases where it was not possible to identify a crucial attribute. For example, if the primary destination of the tour did not have a valid zone, the whole tour is flagged as invalid-destination record. However, invalid stop zones would not result in an invalid-destination record, since tours with invalid stop zones would still be usable for almost all model estimation task except for stop-location choice.

The most important quality controls at the tour level are summarized in Table 5 below and compared to the same tabulation for CMAP HTS2007 in Table 6 and for the BATS2000 survey in Table 7.

TABLE 5: QUALITY CONTROL TOUR-CONSTRUCTION PROCEDURE, MAG/PAG NHTS 2008, REGULAR WORKDAY

Primary destination purpose	Number of tours						
	Total	Valid mode	Closed	Valid timing	Valid primary destination	All valid	Symmetric mode
Absolute number of tours in the survey (not expanded):							
1=Work	3,532	3,509	3,292	3,527	3,367	3,215	2,925
2=University	258	258	248	258	257	247	222
3=School	1,340	1,335	1,319	1,339	1,281	1,255	1,073
4=Escort	1,354	1,351	1,327	1,354	1,289	1,262	581
5=Shopping	2,558	2,546	2,480	2,554	2,407	2,334	2,425
6=Other maintenance	2,027	2,015	1,963	2,025	1,932	1,880	1,846
7=Eating out	776	769	749	775	728	701	745
8=Visiting	581	578	512	580	540	490	488
9=Other discretionary	2,681	2,669	2,522	2,678	2,512	2,392	2,357
10=Unknown	437	425	398	429	0	0	12
Total	15,544	15,455	14,810	15,519	14,313	13,776	12,674
Row percent:							
1=Work	100.0%	99.3%	93.2%	99.9%	95.3%	91.0%	82.8%
2=University	100.0%	100.0%	96.1%	100.0%	99.6%	95.7%	86.0%
3=School	100.0%	99.6%	98.4%	99.9%	95.6%	93.7%	80.1%
4=Escort	100.0%	99.8%	98.0%	100.0%	95.2%	93.2%	42.9%
5=Shopping	100.0%	99.5%	97.0%	99.8%	94.1%	91.2%	94.8%
6=Other maintenance	100.0%	99.4%	96.8%	99.9%	95.3%	92.7%	91.1%
7=Eating out	100.0%	99.1%	96.5%	99.9%	93.8%	90.3%	96.0%
8=Visiting	100.0%	99.5%	88.1%	99.8%	92.9%	84.3%	84.0%
9=Other discretionary	100.0%	99.6%	94.1%	99.9%	93.7%	89.2%	87.9%
10=Unknown	100.0%	97.3%	91.1%	98.2%	0.0%	0.0%	2.7%
Total	100.0%	99.4%	95.3%	99.8%	92.1%	88.6%	81.5%

TABLE 6: QUALITY CONTROL, TOUR-CONSTRUCTION PROCEDURE, CMAP 2007, REGULAR WORKDAY

Primary destination purpose	Number of tours						
	Total	Valid mode	Closed	Valid timing	Valid primary destination	All valid	Symmetric mode
Absolute number of tours in the survey (not expanded):							
1=Work	17,238	16,455	16,369	17,235	16,502	15,714	15,579
2=University	520	516	510	520	500	491	484
3=School	4,788	4,783	4,731	4,788	4,723	4,661	3,966
4=Escort	4,876	4,836	4,844	4,876	4,735	4,674	2,584
5=Shopping	9,362	9,334	9,289	9,362	9,093	8,999	9,176
6=Other maintenance	7,054	6,872	6,870	7,054	6,752	6,575	6,622
7=Eating out	3,137	3,125	3,090	3,137	3,017	2,981	3,024
8=Visiting	4,191	3,677	3,615	4,183	3,617	3,243	3,509
9=Other discretionary	9,265	9,032	9,014	9,264	8,884	8,717	8,693
10=Changing mode	97	63	57	97	89	53	25
11=Loop	971	911	971	970	967	907	13
12=Unknown	71	2	55	71	55	2	2
Total	61,570	59,606	59,415	61,557	58,934	57,017	53,677
Row percent:							
1=Work	100.0%	95.5%	95.0%	100.0%	95.7%	91.2%	90.4%
2=University	100.0%	99.2%	98.1%	100.0%	96.2%	94.4%	93.1%
3=School	100.0%	99.9%	98.8%	100.0%	98.6%	97.3%	82.8%
4=Escort	100.0%	99.2%	99.3%	100.0%	97.1%	95.9%	53.0%
5=Shopping	100.0%	99.7%	99.2%	100.0%	97.1%	96.1%	98.0%
6=Other maintenance	100.0%	97.4%	97.4%	100.0%	95.7%	93.2%	93.9%
7=Eating out	100.0%	99.6%	98.5%	100.0%	96.2%	95.0%	96.4%
8=Visiting	100.0%	87.7%	86.3%	99.8%	86.3%	77.4%	83.7%
9=Other discretionary	100.0%	97.5%	97.3%	100.0%	95.9%	94.1%	93.8%
10=Changing mode	100.0%	64.9%	58.8%	100.0%	91.8%	54.6%	25.8%
11=Loop	100.0%	93.8%	100.0%	99.9%	99.6%	93.4%	1.3%
12=Unknown	100.0%	2.8%	77.5%	100.0%	77.5%	2.8%	2.8%
Total	100.0%	96.8%	96.5%	100.0%	95.7%	92.6%	87.2%

TABLE 7: QUALITY CONTROL, TOUR-CONSTRUCTION PROCEDURE, BATS 2000, REGULAR WORKDAY

Primary destination purpose	Number of tours						
	Total	Valid mode	Closed	Valid timing	Valid primary destination	All valid	Symmetric mode
Absolute number of tours in the survey (not expanded):							
1=Work	28,527	27,846	27,690	28,075	26,345	25,043	22,507
2/3=School/University	10,563	10,058	10,375	10,494	10,098	9,463	8,448
4=Escort	6,223	6,170	6,078	6,142	5,655	5,485	2,252
5=Shopping	8,564	8,463	8,346	8,484	7,820	7,561	6,329
6=Other maintenance	5,153	5,089	4,985	5,006	4,532	4,314	3,877
7=Eating out	4,101	4,039	3,893	4,017	3,685	3,503	3,261
8=Visiting	2,548	2,510	2,365	2,425	2,078	1,928	2,051
9=Other discretionary	9,126	8,963	8,725	8,889	7,918	7,574	7,727
0=Unknown	2,218	2,061	573	1,027	1,305	246	392
Total	77,023	75,199	73,030	74,559	69,436	65,117	56,844
Row percent:							
1=Work	100.0%	97.6%	97.1%	98.4%	92.4%	87.8%	78.9%
2/3=School/University	100.0%	95.2%	98.2%	99.3%	95.6%	89.6%	80.0%
4=Escort	100.0%	99.1%	97.7%	98.7%	90.9%	88.1%	36.2%
5=Shopping	100.0%	98.8%	97.5%	99.1%	91.3%	88.3%	73.9%
6=Other maintenance	100.0%	98.8%	96.7%	97.1%	87.9%	83.7%	75.2%
7=Eating out	100.0%	98.5%	94.9%	98.0%	89.9%	85.4%	79.5%
8=Visiting	100.0%	98.5%	92.8%	95.2%	81.6%	75.7%	80.5%
9=Other discretionary	100.0%	98.2%	95.6%	97.4%	86.8%	83.0%	84.7%
0=Unknown	100.0%	92.9%	25.8%	46.3%	58.8%	11.1%	17.7%
Total	100.0%	97.6%	94.8%	96.8%	90.1%	84.5%	73.8%

In the tabulation above, at-work sub-tours are not explicitly identified for simplicity of analysis. However, this is a separate purpose in the CT-RAMP model system. At-work sub-tours are generated, distributed, and their mode is modeled with separate sub-models.

n general, the quality of the resulted tour records in the database appears to be relatively good. It is significantly better than the corresponding statistics from BATS2000 and other earlier surveys. It is, however, somewhat lower than the final statistics for the CMAP HTS 2007. Overall, more than 88% of tours satisfied all main validity criteria compared to 85% for BATS2000 and 93% for CMAP HTS2007. In certain cases, relative drops in quality (for example, closeness of tours and destination zones for visiting) are associated with a large number of short non-motorized trips and tours that are frequently not reported in full.

Another important detail is that “not-closed” tours (either origin or destination end not at home) are not necessarily discarded. Many of these tours correspond to valid cases of long trips outside the region (sometimes by air or train). These cases are normally modeled as one-directional half-tours explicitly in intercity / statewide models since their share becomes significant for intercity travel. For urban models (like the ABMs developed for NYMTC, MORPC, ARC, MTC, and SANDAG) “not-closed” intercity tours are typically excluded and an auxiliary (aggregate) model for external traffic is included in the model system.

CHAPTER 2.

DEVELOPMENT OF ACCESSIBILITY MEASURES

2.1. GENERAL FORMS OF ACCESSIBILITY MEASURES

There are multiple accessibility measures used in the MAG ABM that are conceptually similar to the set of accessibility measures applied in the Sacramento and San-Diego ABMs but with some additional refinements. Most of the applied accessibility measures represent simplified destination choice logsums, which is the composite utility of travel across all modes to all potential destinations from an origin zone to all destination zones in different time-of-day periods. These accessibility measures are zonal characteristics that can be stored as a vector indexed by TAZ. Another type of accessibility measure describes of the amount of impedance between zones. Accessibilities of this type are stored as TAZ-to TAZ matrices.

These accessibility measures are primarily needed to ensure that the upper-level models in the ABM hierarchy such as car ownership, daily activity pattern (DAP), and (non-mandatory) tour frequency are sensitive to improvements of transportation level-of-service across all modes, as well as changes in land use. Accessibility measures are similar in nature to density measures, but take into account the accessibility between zones as well as the opportunities to engage in various types of activities in those zones. Accessibility measures are needed since it is infeasible to link all choices by full logsums due to the number of potential alternatives across all dimensions (activities, modes, time periods, tour patterns, and daily activity patterns). Accessibility measures reflect the opportunities to implement a travel tour for a certain purpose from a certain origin (residential or workplace). They are used as explanatory variables in the upper level models (daily activity pattern type and tour frequency) and the corresponding coefficients are estimated along with the coefficients for person and household variables.

In the MAG ABM as well as in the San-Diego ABM we tried to completely avoid area-type dummies (such as CBD, urban, suburban, and rural dummies frequently used in models to explain such choice as car ownership, tour/trip frequency, and mode choice).

The applied *zonal accessibility measures* have the following general form:

$$A_i = \ln\left[\sum_{j=1}^{I} S_j \times \exp\left(TMLS_{ij}\right)\right] \qquad \textbf{Equation 1}$$

where:

$i, j \in I$	=	origin and destination zones,
A_i	=	accessibility measure calculated for each origin zone,
S_j	=	attraction size variable for each potential destination zone,
$TMLS_{ij}$	=	time-of-day and mode choice logsum as the measure of impedance.

The composite travel impedance between zones (the *origin-destination (OD) accessibility measure*) is calculated as a two-level logsum taken over time-of-day periods and modes:

$$TMLS_{ij} = \mu \ln\left[\sum_{t=1}^{2} \exp\left(MLS_{tij} + \alpha_t\right)\right] \qquad \textbf{Equation 2}$$

where:

$t = 1,2$	=	time-of-day periods (currently peak and off-peak are used),
MLS_{ij}	=	mode choice logsum for a particular time-of-day period,
α_t	=	time-of-day-specific constant,
μ	=	nesting coefficient for mode choice under time-of-day choice.

In this form, the destination choice accessibility measure is essentially a sum of all attractions in the region discounted by the travel impedance. Note that this measure is sensitive to travel improvements in both peak and off-peak periods. The relative impact of each period is regulated by the time-of-day-specific constant that is estimated for each travel segment (or activity type).

Accessibility measures are linearly included in a utility function of an upper-level model. To preserve consistency with random-utility choice theory, the coefficient for any accessibility measure should be between 0 and 1; though it is not as restrictive as in a case of a proper nested logit model.

The general logic of inclusion of accessibility measures in travel models is as follows. For models that generate activity patterns, tours, and trips where specific destinations are not known yet, zonal accessibility measures are applied that describe the accessibility of all potential activity locations from the household or tour origin. For models where the destination is known, OD accessibility measures should be used.

2.2. SIZE VARIABLES BY ACTIVITY TYPE

Size variables measure the quantity of potential activities (or activity participation) in each TAZ; they are equivalent to trip attraction equations used in gravity models. They are prepared for each TAZ and segmented by activity type (trip purpose). The zonal size variables are calculated as linear combinations of the relevant land-use variables. The corresponding coefficients were estimated by means of regressions of the observed tour ends (expanded from the NHTS, 2008) on the available land-use variable, primarily employment types by NAICS codes. At this stage we decided to pre-calculate the size terms in advance and separately of the impedance functions. A more theoretically consistent procedure would involve a simultaneous estimation of the size terms and impedance functions in the destination choice context by Equation 1. This would require re-structuring model estimation steps in a manner not consistent with the sequence of activities prescribed for Phase I development. Therefore, simultaneous estimation of size term parameters is reserved for Phase 2, when destination choice models for non-work purposes will be estimated. The estimation results for all activity types are presented in Table 8 for non-work purposes (numbered from 4 through 10 in the CT-RAMP model structure) along with a combined non-work attraction measure for all home-based non-work (non-mandatory) purposes (11). The explanatory variables in the rows are referred to by their tokens used in the model application where “nxx” implies employment for NAICS code “xx”. The resulted size variables in the columns are referred by the purpose number “px” and short token indicated the purpose.

TABLE 8: ZONAL SIZE VARIABLES FOR ACCESSIBILITY MEASURES BY ACTIVITY TYPE

Explanatory variables		Size variables by activity type							
Variable	**Description**	**p4_esco**	**p5_shop**	**p6_main**	**p7_eati**	**p8_visi**	**p9_disc**	**p10_atwo**	**p11_alln**
total_HH	Total number of households	1.0000				0.1421	0.3595		0.5016
retail	Retail employment (n44+n45)		4.2810	1.4185	1.2908		0.4387	0.5403	7.4291
n51	Information			0.7091					0.7091
n52	Finance & Insurance							0.1265	
n53	Real Estate Rental Leasing			2.4753					2.4753
n55	Management of Companies & Enterprises							1.3759	
n56	Administrative & Support							0.2357	
n62	Health Care, Social Assistance			1.0618		0.2349			1.2968
n71	Arts, Entertainment, Recreation				0.3224		0.9049		1.2273
n72	Accommodation, Food Services		1.1224		1.0458		0.4422	0.2809	2.6104
n92	Public Administration			0.5356				0.2265	0.5356
total_emp	Total employment							0.1578	

For escorting purpose (purpose=4) the size variable is set to the total population. This is a special purpose where the accessibility to potential destinations does not directly relate to the household decision to escort one of the household members, because escort tours most frequently involves picking up/dropping off a child at school or daycare.. The density of schools around the household does not intuitively lead to more frequent escort tours. On the contrary, if a child can walk to the nearby school escorting will not be needed. Population density is the only accessibility measure for which both negative and positive signs can be accepted in the tour/activity frequency model. All other accessibility measures are accepted only if they have a logical positive sign. This accessibility measure may eventually be dropped, when an explicit model for escorting children to school is applied in Phase 2.

For shopping tours (purpose=5), the main attractions are logically associated with retail employment and food services. Food services are frequently intertwined with shopping and it is difficult to completely separate these two land-use types. It is equally true for both major shopping malls and small street shops or restaurants. It is recommended in the future to enrich shopping size variables with such explanatory variables as floor area to better distinguish between large shopping malls and small street shops. The (household) maintenance purpose (purpose=6) that includes a wide range of activities such as personal business, banking, visiting post office, visiting doctor or dentist or lawyer, etc, is scattered over a wide range of related employment types including retail, information, real estate, rental, leasing, health care, social assistance, and public administration.

Eating out (purpose=7) and discretionary (purpose=9) activities are closely intertwined and frequently combined in the same tour. They share the same attraction variables including retail employment, recreation & entertainment, and food services, although the coefficients are logically different. In addition, the discretionary purpose includes population as an additional attraction variable, which serves as a proxy for such factors as sport facilities and playing grounds. It is recommended to add non-employment variables like land or floor areas for public parks and sport facilities in the future, to enrich the attraction model for discretionary activities.

Visiting relatives and friends (purpose=8) is a special purpose where the major attraction factor is population (number of households). Visiting also frequently occurs at a hospital (measured by health employment).

Attraction factors for tours originating at the workplace (purpose=10) includes many variables, reflecting the multitude of potential activity purposes for at-work travel. Often at-work tours involve eating out during the lunch break- reflected by size variables such as retail employment and food services. At-work sub-tours may also include business trips for meetings - reflected by size variables such as management of companies and administration (most probable places for business meetings) and some proportion of total employment. Finally, workers often use their lunch break for personal business and shopping, reflected by retail, finance, insurance, and public administration size variables.

Finally, a size variable that is based on total attractions for all non-mandatory home-based purposes (4-9) includes a mix of all corresponding employment types and population. Logically, retail employment plays a major role in this mix.

In addition to the size variables for non-mandatory activities, the MAG ABM requires several size variables for zonal accessibility measures to mandatory activities. They are primarily used in the choice models for work from home and schooling from home. They are currently set with a size term equal to the total employment or enrollment for the corresponding segment. We plan to revise this at Phase 2 and estimate a more elaborate size term for workers by occupation which would relate the occupation of the worker more consistently with the types of jobs the workers are likely to engage in. For work from home, employment for the relevant worker occupation category is used as the size variable. The model uses five worker occupation categories defined by aggregating the occupation code from the NHTS data; these are related to employment by NAICS code, which is available at the zonal level. For schooling from home, enrollment in the school type corresponding to the student type is used as the size term. There are three categories of student type: K-8 (elementary or mid school), 9-12 (high school), university or college. The corresponding size variables are summarized in Table 9. The zonal size variables for mandatory activities are numbered p12-p19 to continue the numbering introduced for non-mandatory activities before (p14-p11). "Whom" and "Shom" labels stand for work and schooling from home that are the sub-models where these size variables are applied.

TABLE 9: ZONAL SIZE VARIABLES FOR MANDATORY ACTIVITIES

Explanatory variables		Size variables	
Variable	**Description**	**Variable**	**Description**
n42	Wholesale Trade	p12_whom1	Sales or marketing
n52	Finance and Insurance		
n44	Retail Trade	p13_whom2	Clerical administrative or retail
n45	Retail Trade		
n53	Real Estate and Rental and Leasing		
n71	Arts, Entertainment, and Recreation		
n72	Accommodation and Food Services		
n92	Public Administration		
n11	Agriculture, Forestry, Fishing, Hunting	p14_whom3	Production, construction, manufacturing, or transport
n21	Mining, Quarrying, Oil & Gas Extraction		
n22	Utilities		
n23	Construction		
n31	Manufacturing		
n32	Manufacturing		
n33	Manufacturing		
n48	Transportation and Warehousing		
n49	Transportation and Warehousing		
n51	Information	p15_whom4	Professional, managerial, or technical
n54	Professional, Scientific, and Technical Services		
n55	Management of Companies and Enterprises		
n56	Administrative and Support and Waste Management and Remediation Services		
n61	Educational Services		
n62	Health Care and Social Assistance		
n81	Other Services (except Public Administration)	p16_whom5	Person care and services
Enroll1	Enrollment K-8	p17_shom1	Enrollment primary & mid
Enroll2	Enrollment 9-12	p18_shom2	Enrollment high school
Enroll3	Enrollment university & college	p19_shom3	Enrollment university & college

2.3. IMPEDANCE FUNCTIONS BY PERSON, HOUSEHOLD, AND ACTIVITY TYPE

Impedance functions are calculated as OD matrices of logsums over modes and time-of-day periods (peak and off-peak) according to Equation 2. The calculation is based on *mode choice utilities* that have to be calculated for all modes and time-of-day periods as a first step. Then, these utilities are combined into a *composite logsum* in a second step. Both steps are described below in the subsequent sub-sections. In the context of accessibility, travel impedance is calculated for a one-directional half-tour without stops (i.e. single trip) between the origin and destination. Specifics of tours with multiple chained trips are not considered at this stage.

2.3.1. Mode Utilities

For calculation of accessibility measures, the set of modes is simplified and includes five main modes: 1=SOV, 2=HOV, 3=Walk to Transit, 4=Drive to Transit, 5=Non-Motorized. Walk to Transit (WT) and Drive to Transit (DT) utilities are based on the best transit skims implemented for the entire transit network including all modes. Mode utilities also calculated separately for each of four aggregate travel purposes: 1=Work, 2=University, 3=School, 4=Other. Segmentation by travel purpose is essential since each travel purpose is characterized by a different set of mode preferences. For example, DT is frequently chosen for the Work purpose but it is rare for School trips or Other trips. All non-work purposes are aggregated for calculation of impedances although they are separated with respect to size variables. Additional important segmentation relates to household car sufficiency, which is important because car sufficiency strongly affects mode availability and preferences. We distinguish between three household groups: 1=household without cars, 2=household with cars fewer than workers, 3=households with cars greater than or equal to workers. .

Overall, by combining 5 aggregate modes with 4 travel purposes, 3 car sufficiency groups and 2 time-of-day periods a set of 5×4×3×2=120 mode utilities was pre-calculated for all OD pairs. The components of the mode utility functions and corresponding coefficients are summarized in Table 10. The coefficients were adopted from the existing MAG 4-setp model with some simplifications and adaptations. These coefficients will be re-estimated at Phase 2 as part of the mode choice model estimation. All coefficients are generic across time-of-day periods. The distinction between peak and off peak utilities is due to different level-of-service variables (calculated for one direction).

TABLE 10: COMPONENTS AND COEFFICIENTS OF MODE UTILITIES

Variable	SOV	HOV	WT	DT	NM
Work travel purpose:					
SOV time, min	-0.03				
HOV time, min		-0.03			
Highway distance, miles	-0.015	-0.01			-1.5
Highway distance greater than 3 miles, dummy					-999
WT weighted time, min			-0.03		
WT fare, cents			-0.002		
WT in-vehicle time less than 1 min, dummy			-999		
DT weighted time, min				-0.03	
DT fare, cents				-0.002	
DT in-vehicle time less than 1 min, dummy				-999	
Zero car household	-999	-3.0			
Cars fewer than workers	-1.5	-2.0			
Cars greater than or equal to workers		-2.5			
University travel purpose:					
SOV time, min	-0.03				
HOV time, min		-0.03			
Highway distance, miles	-0.03	-0.02			-1.5
Highway distance greater than 3 miles, dummy					-999
WT weighted time, min*			-0.03		

Variable	**SOV**	**HOV**	**WT**	**DT**	**NM**
WT fare, cents			-0.004		
WT in-vehicle time less than 1 min, dummy			-999		
DT weighted time, min**				-0.03	
DT fare, cents				-0.004	
DT in-vehicle time less than 1 min, dummy				-999	
Zero car household	-999	-2.0			
Cars fewer than workers	-1.5	-1.0			
Cars greater than or equal to workers	0	-1.5			
School travel purpose:					
SOV time, min	-0.05				
HOV time, min		-0.05			
Highway distance, miles	-0.06	-0.04			-1.5
Highway distance greater than 3 miles, dummy					-999
WT weighted time, min*			-0.03		
WT fare, cents			-0.006		
WT in-vehicle time less than 1 min, dummy			-999		
DT weighted time, min**				-0.03	
DT fare, cents				-0.004	
DT in-vehicle time less than 1 min, dummy				-999	
Zero car household	-999	-1.0		-5.0	2.0
Cars fewer than workers	-1.5	0		-5.0	2.0
Cars greater than or equal to workers	0	-0.5		-5.0	2.0
Other travel purpose					
SOV time, min	-0.03				
HOV time, min		-0.03			
Highway distance, miles	-0.03	-0.02			-1.5
Highway distance greater than 3 miles, dummy					-999
WT weighted time, min*			-0.03		
WT fare, cents			-0.004		
WT in-vehicle time less than 1 min, dummy			-999		
DT weighted time, min**				-0.03	
DT fare, cents				-0.004	
DT in-vehicle time less than 1 min, dummy				-999	
Zero car household	-999	-3.0		-5.0	
Cars fewer than workers	-1.5	-2.0		-5.0	
Cars greater than or equal to workers	0	-2.5		-5.0	

*WT weighted time includes in-vehicle time and out-of-vehicle time with weight equal to 2.5. Out-of-vehicle time includes initial wait, transfer wait, access walk, transfer walk, egress walk, and 4 min penalty for each transfer.

**DT weighted time additionally includes access drive in out-of-vehicle time.

2.3.2. Mode & Time-of-Day Choice Logsums

After mode utilities have been calculated for each mode, purpose, car-sufficiency group, and time-of-day period they are combined into composite *OD accessibility measures*, i.e. mode & time-of-day choice logsums by Equation 2. The list of logsum measures that have to be prepared to support various accessibility measures is summarized in Table 11.

TABLE 11: LIST OF MODE & TIME-OF-DAY CHOICE LOGSUMS

Impedance	Accessibility from the given (residential) zone to:	Token
1	Workplace by all modes for all car-sufficiency groups	Work
2	University by all modes for all car-sufficiency groups	Univ
3	School by all modes for all car-sufficiency groups	Scho
4	Non-mandatory activity location by auto	Auto
5	Non-mandatory activity location by by WT	Tran
6	Non-mandatory activity location by NM (walk)	Nonm
7	Non-mandatory activity by all modes, individual travel, zero-car household	Indi_0
8	Non-mandatory activity by all modes, individual travel, cars<workers	Indi_1
9	Non-mandatory activity by all modes, individual travel, cars≥workers	Indi_2
10	Non-mandatory activity by all modes, joint travel, zero-car household	Join_0
11	Non-mandatory activity by all modes, joint travel, cars<workers	Join_1
12	Non-mandatory activity by all modes, joint travel, cars≥workers	Join_2
13	Escort accessibility, joint travel, zero-car household	Esco_0
14	Escort accessibility, joint travel, cars<workers	Esco_1
15	Escort accessibility, joint travel, cars≥workers	Esco_2
16	Workplace by auto modes for all car-sufficiency groups (auto dependency)	Wrkad
17	University by auto modes for all car-sufficiency groups (auto dependency)	Unvad
18	School by auto modes for all car-sufficiency groups (auto dependency)	Schad
19	Workplace by non-auto modes (non-auto dependency)	Wrknad
20	University by non-auto modes (non-auto dependency)	Unvnad
21	School by non-auto modes (non-auto dependency)	Schnad

Overall, 21 different OD accessibility measures are prepared to support various zonal accessibility measures needed for different sub-models of the MAG ABM. Structure of each logsum and associated parameters are summarized in Table 12. This table essentially represents a control file for the impedance (OD) part of the program that calculates accessibility measures.

TABLE 12: STRUCTURE OF MODE & TIME-OF-DAY CHOICE LOGSUMS

Token	Purpose	Car sufficiency			Modes included					Off-peak constant
		Zero cars	Cars fewer than workers	Cars equal to or greater than workers	SOV	HOV	WT	DT	NM	
Work	1=Work	0.05	0.35	0.6	1	1	1	1	1	-0.9
Univ	2=Univ	0.05	0.35	0.6	1	1	1		1	-0.5
Scho	3=Scho	0.05	0.35	0.6	1	1	1		1	-1.2
Auto	4=Other	0.05	0.35	0.6	1					0.5
Tran	4=Other	0.05	0.35	0.6			1			0.5
Nonm	4=Other	0.05	0.35	0.6					1	0.5
Indi_0	4=Other	1			1		1		1	0.5
Indi_1	4=Other		1		1		1		1	0.5
Indi_2	4=Other			1	1		1		1	0.5
Join_0	4=Other	1				1	1		1	0.5
Join_1	4=Other		1			1	1		1	0.5
Join_2	4=Other			1		1	1		1	0.5
Esco_0	4=Other	1				1			1	-0.5
Esco_1	4=Other		1			1			1	-0.5
Esco_2	4=Other			1		1			1	-0.5
Wrkad	1=Work	0.05	0.35	0.6	1	1		1		-0.9
Unvad	2=Univ	0.05	0.35	0.6	1	1		1		-0.5
Schad	3=Scho	0.05	0.35	0.6	1	1		1		-1.2
Wrknad	1=Work	0.05	0.35	0.6			1		1	-0.9
Unvnad	2=Univ	0.05	0.35	0.6			1		1	-0.5
Schnad	3=Scho	0.05	0.35	0.6			1		1	-1.2

Each impedance measure is associated with a certain aggregate travel purpose (1-4) for which the mode utilities are calculated according to the coefficients in Table 10. Then, depending on the type of accessibility measure, car sufficiency is taken into account. If a general accessibility measure is calculated that is going to be applied in the model system before the car-ownership model, the mode utilities are averaged across all car-sufficiency groups with the weight that reflects the observed proportion between different car-sufficiency groups in the region. If an accessibility measure is calculated for a specific car-sufficiency group (that means that it is going to be applied after the car-ownership model) the mode utilities for this specific group are used.

Not every mode is included in each logsum. The set of modes is restricted for two reasons. The first reason is that some modes are not observed for some of the trip purposes. For example, Drive to Transit (DT) is relevant for work trips only. The second reason is that certain modes are made unavailable in order to calculate a specific (mode-restricted) type of accessibility needed for a particular behavioral model. For example, mode-specific accessibilities that are used in the car-ownership model are based on a single representative mode each. Accessibilities that describe individual activities should logically exclude HOV. Accessibilities that describe joint activities naturally exclude SOV. Accessibilities that describe auto dependency include only modes that need an auto (SOV, HOV, and DT). Accessibilities that describe auto non-dependency include only modes that do not need an auto (WT and NM).

Finally, to complete the logsum calculation across time-of-day periods, a bias constant for off-peak period is specified (the peak period is used as the reference alternative with zero bias). This constant is set to replicate the observed proportion of trips in the peak period vs. off-peak.

2.4. LIST OF ZONAL ACCESSIBILITY MEASURES ADOPTED FOR MAG ABM

The set of *zonal accessibility measures* incorporated in the Phoenix ABM is summarized in Table 13. The variety of measures stems from the combination of different size variables segmented by the underlying activity type with different impedance measures segmented by trip purpose and person/household type. The impact of various accessibility measures prepared for the models that were estimated at Phase 1 will be discussed in detail in the subsequent sections in the context of model estimation results. Such models as car ownership (mobility attributes), work and schooling from home, and coordinated daily activity-travel pattern are very good illustrations for zonal accessibility measures with some components that relate to OD accessibility measures. Such models as usual workplace and school location are based on OD accessibility measures.

TABLE 13: ZONAL ACCESSIBILITY MEASURES

Measure	Size variable		Impedance measure		Model in which applied
	No	Token	No	Token	
1	12	Whom1	1	Work	Work from home
2	13	Whom2	1	Work	Work from home
3	14	Whom3	1	Work	Work from home
4	15	Whom4	1	Work	Work from home
5	16	Whom5	1	Work	Work from home
6	17	Shom1	3	Scho	Schooling from home
7	18	Shom2	3	Scho	Schooling from home
8	19	Shom3	2	Univ	Schooling from home
9	11	AllNM	4	Auto	Car ownership
10	11	AllNM	5	Tran	Car ownership
11	11	AllNM	6	Nonm	Car ownership
12	11	AllNM	7	Indi_0	Coordinated Daily Activity-Travel Pattern
13	11	AllNM	8	Indi_1	Coordinated Daily Activity-Travel Pattern
14	11	AllNM	9	Indi_2	Coordinated Daily Activity-Travel Pattern
15	11	AllNM	10	Join_0	Coordinated Daily Activity-Travel Pattern
16	11	AllNM	11	Join_1	Coordinated Daily Activity-Travel Pattern
17	11	AllNM	12	Join_2	Coordinated Daily Activity-Travel Pattern
18	5	Shop	10	Join_0	Joint tour frequency
19	5	Shop	11	Join_1	Joint tour frequency
20	5	Shop	12	Join_2	Joint tour frequency

Measure	Size variable		Impedance measure		Model in which applied
21	6	Main	10	Join_0	Joint tour frequency
22	6	Main	11	Join_1	Joint tour frequency
23	6	Main	12	Join_2	Joint tour frequency
24	7	Eati	10	Join_0	Joint tour frequency
25	7	Eati	11	Join_1	Joint tour frequency
26	7	Eati	12	Join_2	Joint tour frequency
27	8	Visi	10	Join_0	Joint tour frequency
28	8	Visi	11	Join_1	Joint tour frequency
29	8	Visi	12	Join_2	Joint tour frequency
30	9	Disc	10	Join_0	Joint tour frequency
31	9	Disc	11	Join_1	Joint tour frequency
32	9	Disc	12	Join_2	Joint tour frequency
33	4	Esco	13	Esco_0	Allocated tour frequency
34	4	Esco	14	Esco_1	Allocated tour frequency
35	4	Esco	15	Esco_2	Allocated tour frequency
36	5	Shop	7	Indi_0	Allocated tour frequency
37	5	Shop	8	Indi_1	Allocated tour frequency
38	5	Shop	9	Indi_2	Allocated tour frequency
39	6	Main	7	Indi_0	Allocated tour frequency
40	6	Main	8	Indi_1	Allocated tour frequency
41	6	Main	9	Indi_2	Allocated tour frequency
42	7	Eati	7	Indi_0	Individual tour frequency
43	7	Eati	8	Indi_1	Individual tour frequency
44	7	Eati	9	Indi_2	Individual tour frequency
45	8	Visi	7	Indi_0	Individual tour frequency
46	8	Visi	8	Indi_1	Individual tour frequency
47	8	Visi	9	Indi_2	Individual tour frequency
48	9	Disc	7	Indi_0	Individual tour frequency
49	9	Disc	8	Indi_1	Individual tour frequency
50	9	Disc	9	Indi_2	Individual tour frequency
51	10	Atwo	7	Indi_0	Individual sub-tour frequency
52	10	Atwo	9	Indi_2	Individual sub-tour frequency

The 52 zonal accessibility measures are combined of 19 size variables (numbered and tokenized in Table 8 and Table 9 above) and 15 impedance measures (numbered and tokenized in Table 11 and Table 12 above). There are 6 impedance measures (16-21) that are used only as OD accessibilities. Multiple examples of impacts of the first 17 zonal accessibility measures on different aspects of travel behavior can be found in subsequent sections on models that were estimated at Phase 1. The other accessibility measures will be used at Phase 2.

2.5. EXAMPLES OF ACCESSIBILITY MEASURES

In this section we illustrate the essence of zonal accessibility measures in several examples. In these examples, the accessibility measures are mapped as calculated by Equation 1 but before taking the logarithm. Since the logarithm is a monotonically increasing function and the mapping is done using an equal number of features option, this does not affect the results and only changes the absolute scale.

To illustrate the most common structural features of zonal accessibilities we present an example of three mode-specific measures (9-11) that share the same size variable – all non-mandatory attractions (size variable 11) but differ with respect to the travel modes used to reach these activity locations. Figure 3 shows the size variable independent of accessibility. Figure 4 shows measure 9, auto accessibility to all non-mandatory attractions, while Figure 5 shows transit accessibility to all non-mandatory attractions, Figure 6 shows =non-motorized accessibility to non-mandatory attractions. All presented maps correspond to the MAG area although the accessibility measures were prepared and all models were estimated for both MAG and PAG areas.

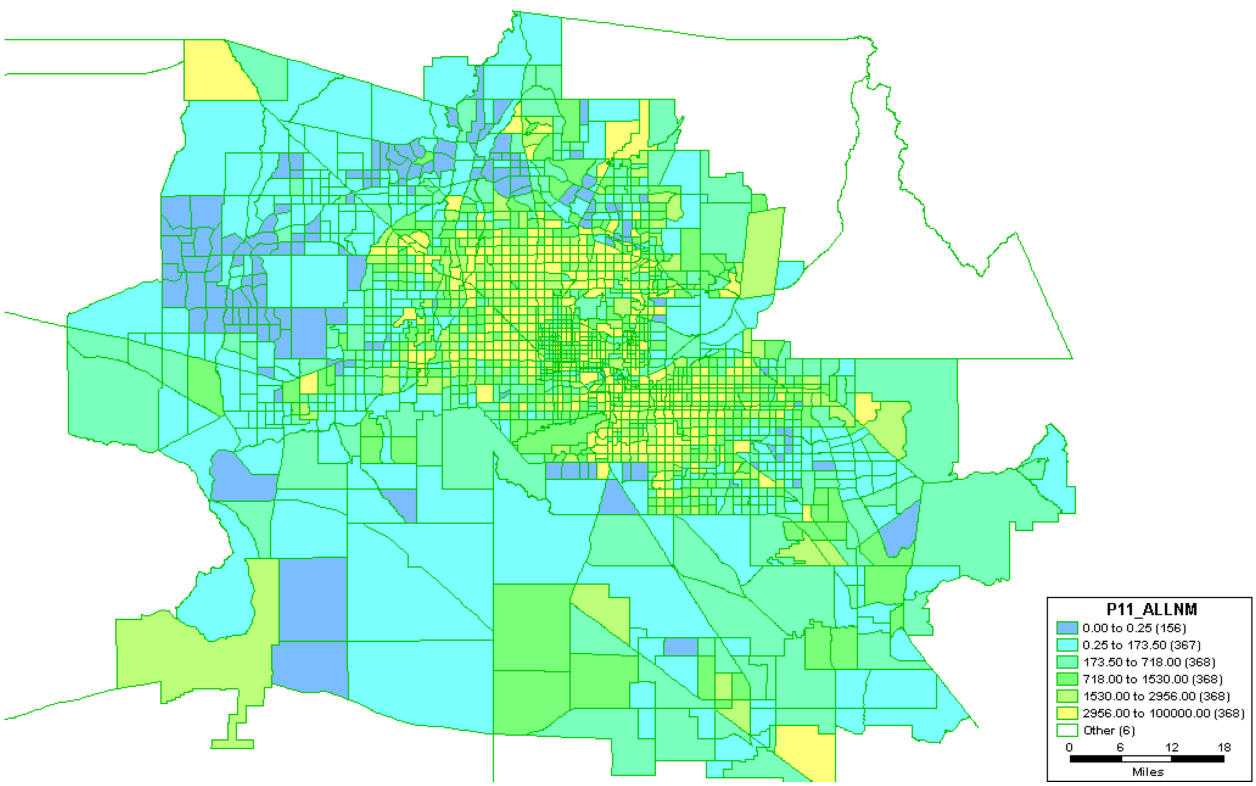

FIGURE 3: ALL NON-MANDATORY ATTRACTIONS – SIZE VARIABLES

As can be seen in Figure 3, non-mandatory attractions (which are highly correlated with retail employment) are mostly concentrated in central urbanized areas and the inner ring. However, there are many large shopping malls in the outer ring as well. The size variables themselves are checkered with two neighboring zones having sometimes extremely low and high values.

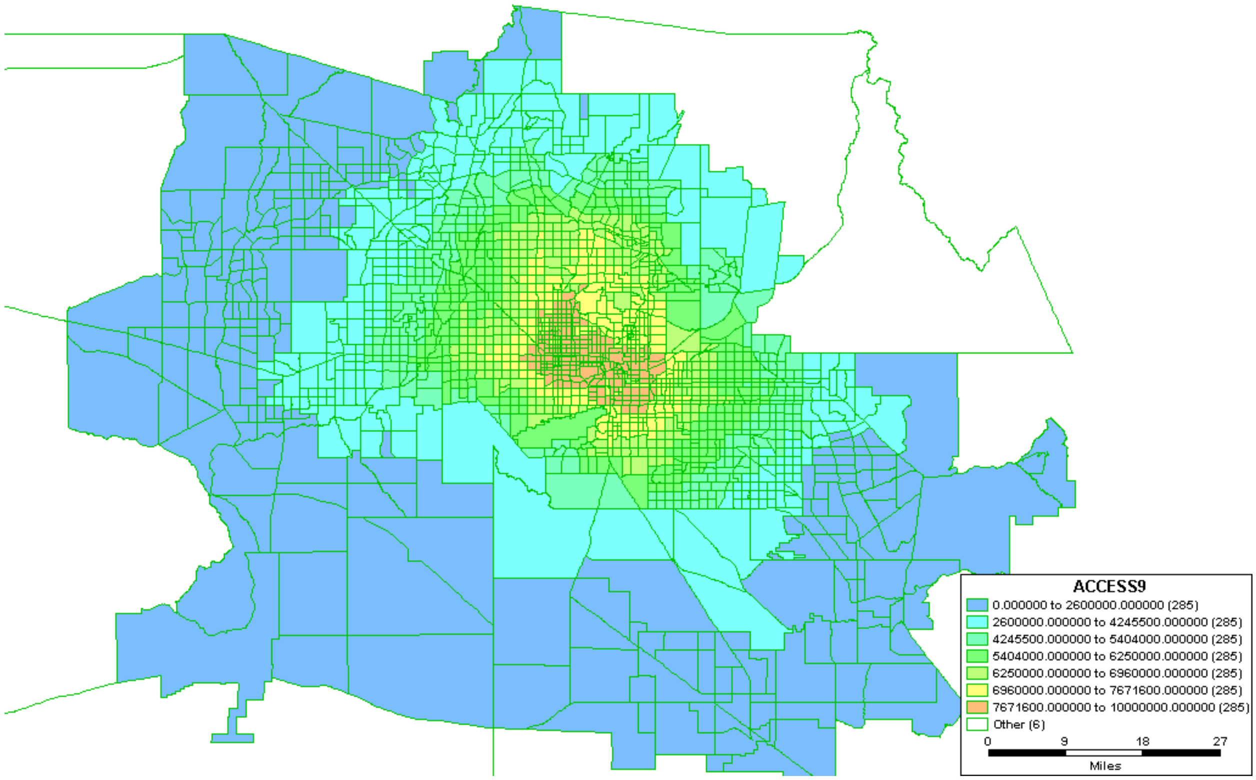

FIGURE 4: ALL NON-MANDATORY ATTRACTIONS – AUTO ACCESSIBILITY

When auto accessibility is considered, it normally tends to smooth the pattern as well illustrated in Figure 4. Since the auto accessibility is pretty uniform across the region there cannot be small "islands" i.e. zones with accessibility that is significantly different from the neighboring zones. As a result, a typical concentric pattern arises. Interestingly, the zones with the highest accessibility to non-mandatory attractions in the central area may be zones that have few attractions within them; however, for those who live there it is convenient to travel to many places around that have large size variables.

On the contrary, peripheral zones with relatively low accessibility might be zones with a large internal size variable. However, this is not enough to compensate for the long travel associated with visiting attractions in other zones. In this sense, the accessibility measures reflect variety of opportunities within reach.

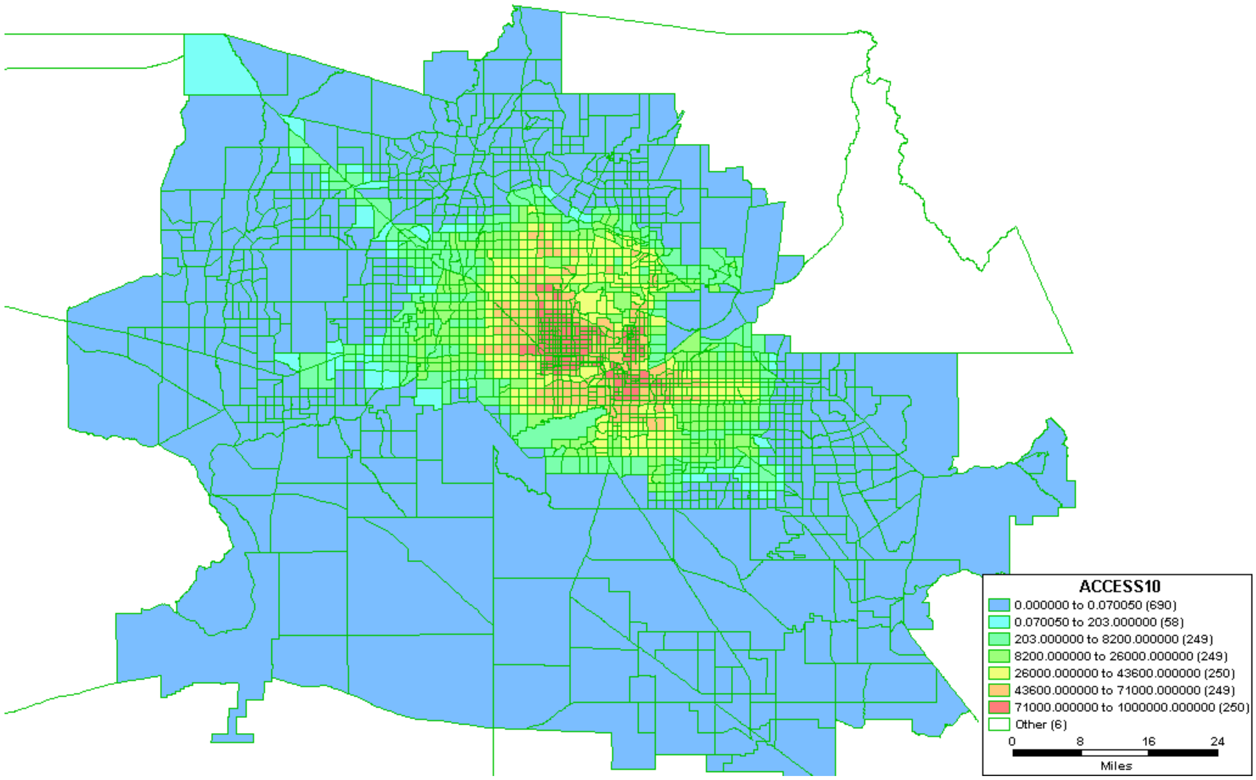

FIGURE 5: ALL NON-MANDATORY ATTRACTIONS – TRANSIT ACCESSIBILITY

Transit accessibilities to the same attractions in Figure 5 look different from auto accessibilities although the general concentric pattern is similar. The inner areas of high accessibility are generally smaller than the auto accessibilities, reflecting that, in general, transit accessibility is not provided as ubiquitously as auto accessibility. However, it can be also seen that the concentric pattern is combined with a radial pattern that is characterized by higher accessibilities along the main transit corridors.

Non-motorized accessibilities (shown in Figure 6) are limited to a 3-mile radius and highly sensitive to walk distance within the 0-3 mile range. Logically, non-motorized accessibilities exhibit a checkered pattern that is the closest to the original zonal size variables. Contrary to motorized accessibilities (whether by auto or transit), non-motorized accessibilities are not uniformly distributed and can have clusters and islands, especially in urbanized areas.

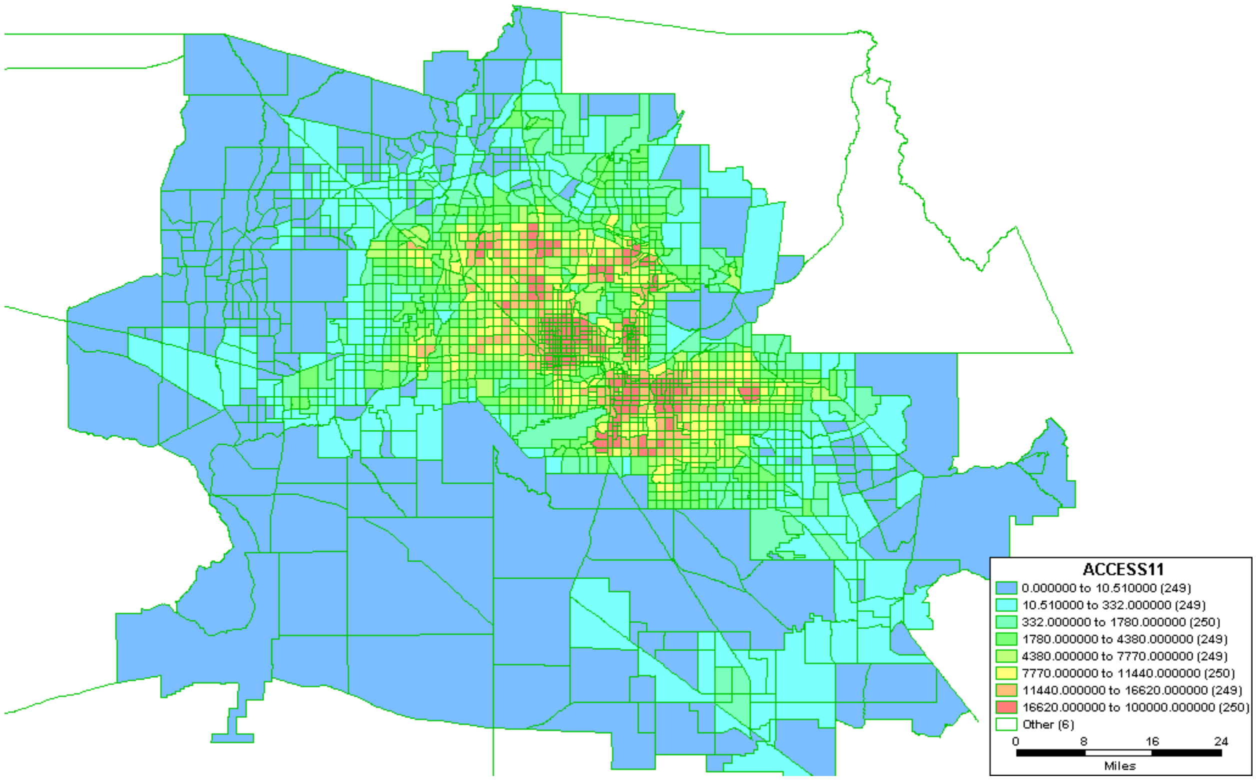

FIGURE 6: ALL NON-MANDATORY ATTRACTIONS – NON-MOTORIZED ACCESSIBILITY

In order to obtain additional insight into accessibility measures, we present an additional set of maps showing the impact of car sufficiency. These measures are based on the same size variable (11) that includes all non-mandatory attractions but also includes all travel modes. The difference is due to car sufficiency, which has a strong impact on mode availability and preferences. Measure 12 is calculated for zero-car households (Impedance 7) and presented in Figure 7. Measure 13 is calculated for households with number of cars fewer than number of workers (low car sufficiency, Impedance 8) and presented in Figure 8. Measure 14 is calculated for households with number of cars greater than or equal to number workers (high car sufficiency, Impedance 9) and presented in Figure 9.

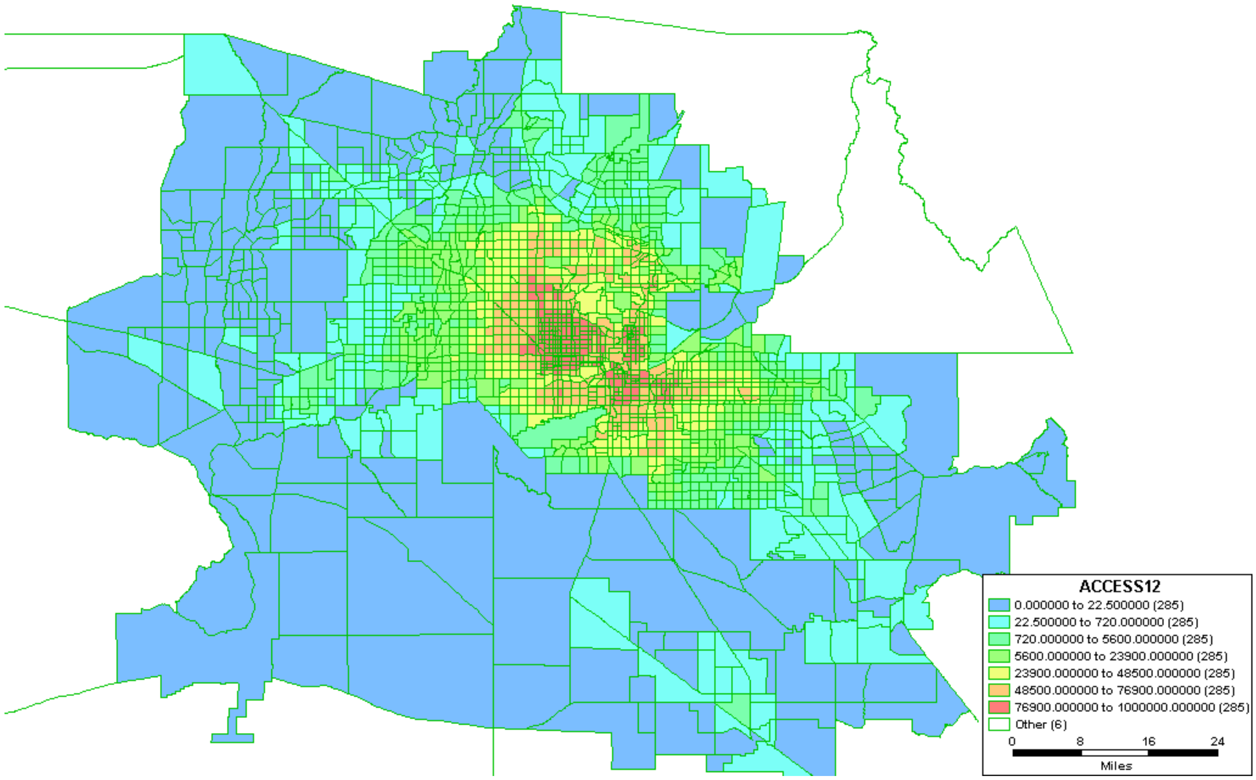

FIGURE 7: ALL NON-MANDATORY ATTRACTIONS – ZERO-CAR HOUSEHOLDS

Accessibilities for households without cars presented in Figure 7 are driven by transit and non-motorized accessibilities discussed above. As the result the pattern is somewhat mixed and non-uniform with some clusters and islands. It can also be seen that the accessibility is reasonable only for households residing in the inner ring while it is very low for the outer rings. A household in the outer ring would need a car even for shopping and discretionary activities even if there is no need for car to commute to work (for example, if all household workers work from home or if there is a household of retirees).

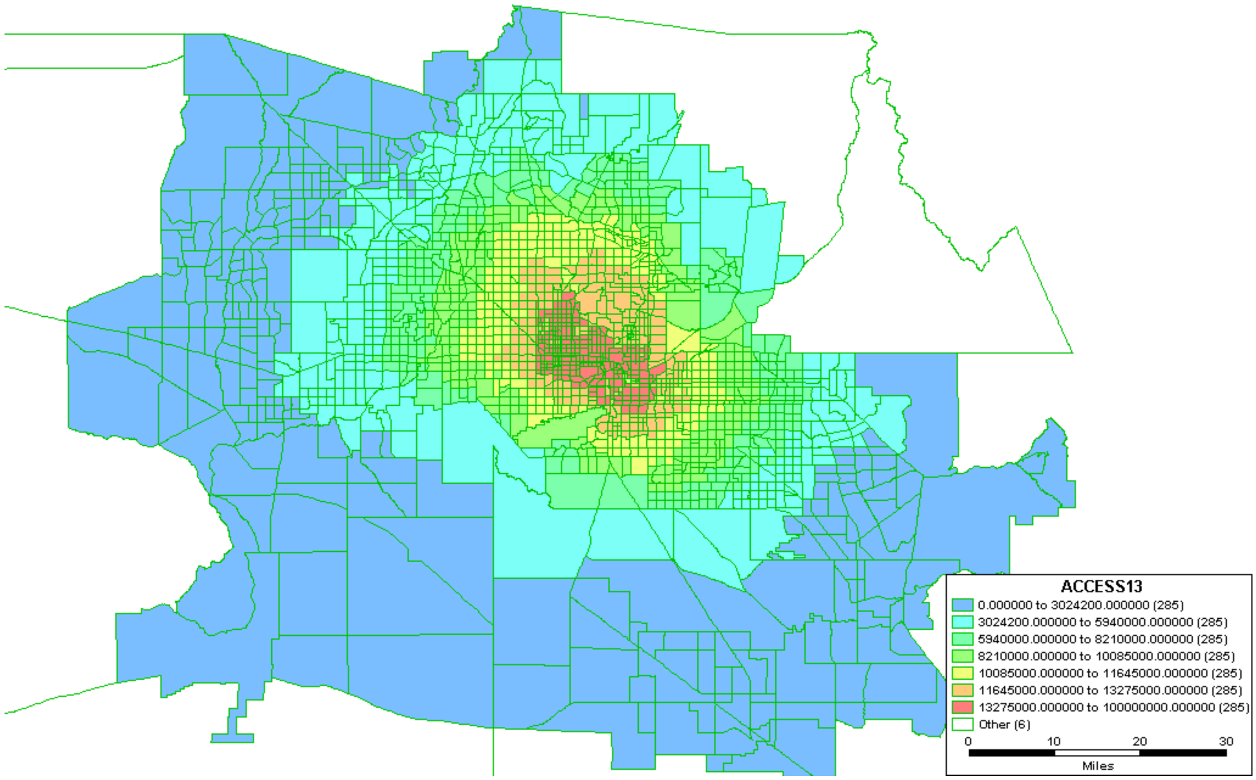

FIGURE 8: ALL NON-MANDATORY ATTRACTIONS – LOW CAR SUFFICIENCY

The accessibility maps for low and high car sufficiency households presented in Figure 8 and Figure 9 respectively are the most similar in shape. Even if the number of cars is fewer than number of workers in the household, the accessibility is largely driven by the auto accessibility, i.e. the most frequent mode used for shopping, household maintenance, eating out, and other discretionary activities would be auto. Hence, the pattern is smooth and concentric as the auto accessibility pattern discussed above.

There are some differences in the overall scale that are not seen on these maps because of the "equal number of features" principle. However, these quantitative differences will affect individual activity patterns and frequencies. In general, higher car ownership results in higher frequencies for all non-mandatory activities.

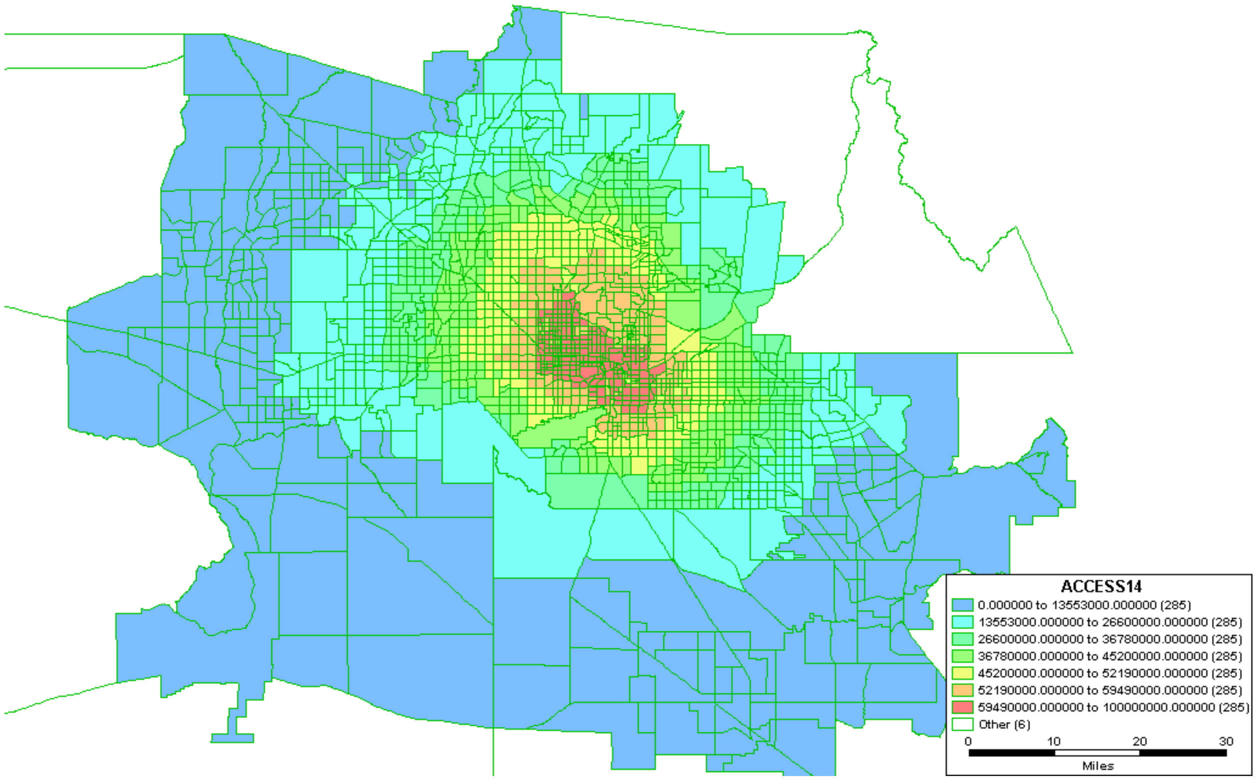

FIGURE 9: ALL NON-MANDATORY ATTRACTIONS – HIGH CAR SUFFICIENCY

CHAPTER 3.

RESIDENTIAL TYPE AND LOCATION CHOICE FOR UNIVERSITY STUDENTS

3.1. CHOICE FRAMEWORK AND CONTEXT FOR UNIVERSITY STUDENTS

3.1.1. New Approach to Modeling University Students in MAG ABM

As was described in the MAG ABM design document, university students are modeled in a principally new way compared to many other travel models. In the MAG ABM, university students are first generated based on university enrollment and then located by residence type and place of residence. This approach requires several new choice models to be estimated for university students. These models can be summarized in the hierarchical structure as shown in Figure 10.

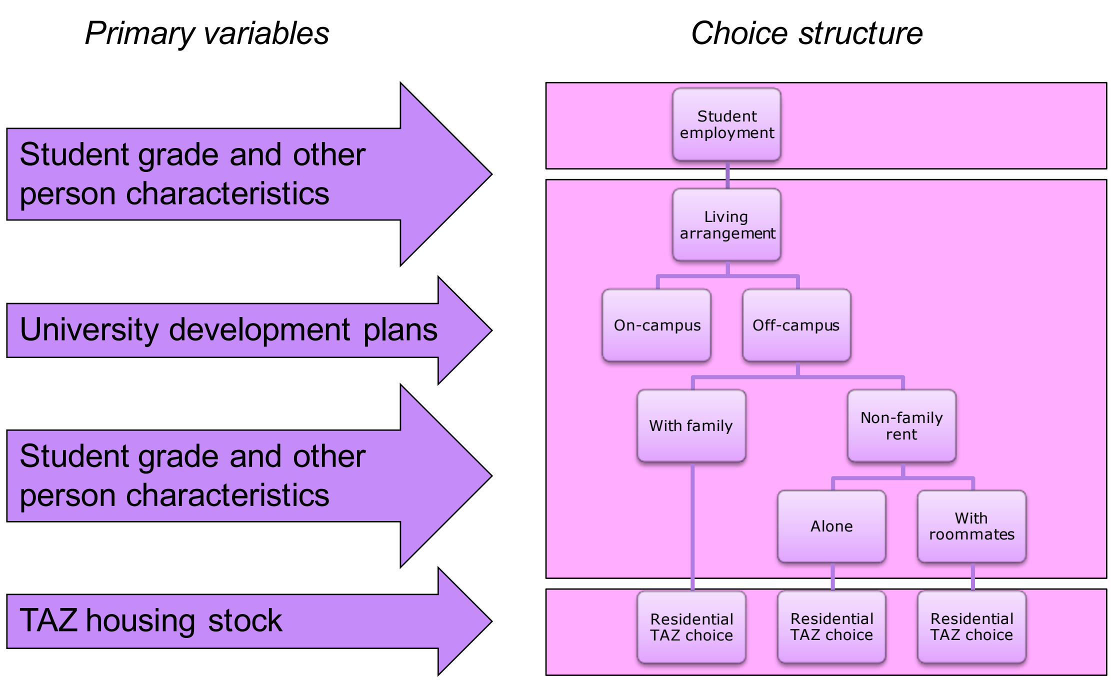

FIGURE 10: HIERARCHY OF CHOICE MODELS FOR UNIVERSITY STUDENTS

There are three main choice levels, each estimated as a separate model, although the entire choice structure can be considered as a nested model. The upper level choices are used as explanatory variables for lower-level choices. The lower-level choice logsums are used in the upper-level choices where possible without excessively complex calculations. The following three main choice dimensions were defined:

1) *An employment choice model.* This is a model of four possible choice alternatives: 1=on-campus employment only, 2=off-campus employment only, 3=both on- and off-campus employment, and 4=not employed.
2) *A living arrangement model*: This is a model that currently includes the following four choice alternatives: 1=live on-campus, 2=live off-campus with family, 3=live off-campus with roommates, 4=live off-campus alone. The choices can be nested as shown in Figure 10 but the nested structure was not confirmed statistically.
3) *A residential location model*: This is a standard residential location choice model that has an multinomial logit (MNL) structure with TAZs as alternatives.

All models have been specified to include only available explanatory variables regarding the student grade and other person characteristics. This caused some problems in the estimation described below and required some simplifications, but the developed models are fully compatible with the population synthesis and other models implemented in Phase 1. This also explains why certain variables, such as car-ownership, cannot be used in these models. While income is very important in many travel models, it is ambiguous for students, especially in non-family households; thus, it was not used. Many further refinements are reserved for future phases, but would require additional data collection (through an additional survey or otherwise) to estimate more elaborate model specifications and forms.

3.1.2. ASU Student Survey Sample Description

A survey of ASU students was conducted in 2007 to measure travel demand to and from ASU and obtain rich socio-economic and demographic information about the respondents. A little over 2000 students responded to the survey and provided information about their socio-economic and demographic characteristics, travel characteristics, and class and work schedules and locations. Table 14 offers a statistical description of the sample of student respondents.

TABLE 14: 2007 ASU STUDENT SURVEY SAMPLE DESCRIPTION

Characteristic	Student Employed On-Campus		Student Not Employed On-Campus		All Students	
	No	%	No	%	No	%
Total Students	592	29	1444	71	2036	100
Living On Campus	54	9.1	124	8.6	178	8.7
Living Off Campus	522	88.2	1306	90.4	1828	89.8
Off-campus Living Arrangement:						
Live alone	254	48.7	787	60.3	1041	56.9
Live with family	185	35.4	353	27.0	538	29.4
Live with roommates	83	15.9	166	12.7	249	13.6
Student Level:						
Freshman	21	3.5	161	11.1	182	8.9
Sophomore	52	8.8	225	15.6	277	13.6
Junior	71	12.0	367	25.4	438	21.5
Senior	111	18.8	408	28.3	519	25.5
Graduate	337	56.9	283	19.6	620	30.5
Primary Campus:						
Tempe Campus	532	89.9	1188	82.3	1720	84.5
Phx Downtown Campus	21	3.5	58	4.0	79	3.9
Polytechnic Campus	21	3.5	72	5.0	93	4.6
West Campus	18	3.0	126	8.7	144	7.1

Characteristic	Student Employed On-Campus		Student Not Employed On-Campus		All Students	
	No	%	No	%	No	%
Gender Distribution:						
Male	218	37.4	562	39.1	780	38.6
Female	365	62.6	874	60.9	1239	61.4
Age Distribution:						
15-19 years	37	6.3	221	15.3	258	12.7
20-24 years	235	39.7	655	45.4	890	43.7
25-29 years	165	27.9	244	16.9	409	20.1
30-34 years	74	12.5	120	8.3	194	9.5
35-39 years	37	6.3	62	4.3	99	4.9
40 years and above	44	7.4	142	9.8	186	9.1
Employed Off-Campus:	116	19.6	955	66.1	1071	52.6
Travel Time to Campus (Reported):						
0-5 minutes	42	7.1	71	4.9	113	5.6
>5-10 minutes	74	12.5	119	8.2	193	9.5
>10-15 minutes	100	16.9	202	14.0	302	14.8
>15-20 minutes	103	17.4	222	15.4	325	16.0
>20-30 minutes	113	19.1	323	22.4	436	21.4
>30-45 minutes	101	17.1	313	21.7	414	20.3
>45 minutes	59	10.0	194	13.4	253	12.4
Travel Time to Campus (Skims):						
0-5 minutes	0	0	0	0	0	0
>5-10 minutes	31	6.0	33	2.6	64	3.6
>10-15 minutes	188	36.5	320	24.9	508	28.3
>15-20 minutes	96	18.6	197	15.4	293	16.3
>20-30 minutes	132	25.6	444	34.6	576	32.0
>30-45 minutes	54	10.5	240	18.7	294	16.4
>45 minutes	14	2.7	49	3.8	63	3.5
Distance to Campus (Skims):						
0-1 mile	38	7.4	19	1.5	57	3.2
>1-3 miles	117	22.7	198	15.4	315	17.5
>3-5 miles	77	15.0	141	11.0	218	12.1
>5-10 miles	115	22.3	262	20.4	377	21.0
>10-15 miles	71	13.8	246	19.2	317	17.6
>15-20 miles	39	7.6	172	13.4	211	11.7
>20-30 miles	40	7.8	179	14.0	219	12.2
>30 miles	18	3.5	66	5.1	84	4.7

Of the 2,036 total students who responded to the survey, 29 percent worked on-campus while the remainder did not hold an on-campus job. Less than 10 percent of the students (regardless of on-campus employment status) actually lived on-campus; 90 percent of the students lived off-campus, a number consistent with the dorm capacity provided by ASU at this time. Of those living off-campus, it was found that more than one-half overall reported living alone. This is not likely to be representative of the true living arrangements for the student population, as the sample is over-representative of graduate level students who tend to live alone. About 30 percent of the student respondents in this survey are graduate students, while in reality, just about 20 percent of the ASU student body comprises graduate students. However, it is important to note that a sizeable number of students in each level responded to the survey, thus providing a representation of the cross-section of students attending school at ASU. More than 80 percent of the students report the Tempe campus as their primary campus of enrolment.

About 30 percent of the student respondents reported living with family and about one-half of that figure reported living with roommates. More females than males responded to the survey, with a little over 60 percent of the respondents being female. The sample over-represents females as the true percent of females in the ASU student population is just about 52 percent. With respect to age distribution, the sample respondents appear to mirror the true population age distribution. More than one-half of students employed on-campus are graduate level students, and it is therefore not surprising to see a slightly older age profile for this group. For the students not employed on campus (only 20 percent of whom are graduate students), the age profile appears consistent with what one would expect for a college student body. Just about 60 percent of this student group is 24 years or younger. Overall, more than one-half of the sample is 24 years or younger and about three-quarters of the sample is under 30 years of age.

The analysis of the survey responses shows that about one in five of the students employed on-campus are also employed off-campus. For the student group that is not employed on-campus, this percentage rises considerably to about two-thirds. Overall, more than one-half of the student respondents report working at an off-campus location. This may be consistent with the true profile of the students as ASU is a large commuter university with many students living off-campus and working part-time to support their way through school.

The survey respondents were asked to provide their usual commute time (one-way) to the primary campus where they attend classes. The reported travel time distribution is shown in the table. In addition, the home locations of the respondents were geocoded to the TAZ level, and travel time and distance skims were appended to the survey records to obtain travel time and distance distributions from the skims. In general, the distributions appear reasonable, although there is considerable difference between the reported usual travel time distribution and the travel time distribution implied by the skims. It appears that students with short commutes under-estimate their commute times, while those with long commutes over-estimate their usual commute times. Whereas the skims suggest that about four percent of the students have travel times of 10 minutes or less, the reported values suggest that 15 percent of students have travel times of 10 minutes or less. In the high categories, the skims suggest that 20 percent of the students have commutes 30 minutes or more, but the reported values suggest that 33 percent of the students have one-way commutes of 30 minutes or more. In the survey, respondents were not asked to provide distance as it was felt that distances could be computed based on address information. Using the skims, it is found that about 20 percent of the students live within three miles of campus, another 20 percent lives 5-10 miles from campus, and just about five percent of the students live more than 30 miles away.

Although the sample size is rather small and the sample over-representative of graduate students and females, the database is nonetheless a rich data source for estimating models of living arrangement (household type and composition) and residential location. The models are envisioned to be largely applicable to the student body as a whole by virtue of the relationships estimated in disaggregate choice models. The models described below require age category, gender of student, and some other variables that have to be defined in model system application as an input. Typically, universities estimate total future enrollment by grade, but not gender or age. Predetermined distributions for these variables (conditional upon grade) will be derived from the available ASU database.

3.2. EMPLOYMENT TYPE CHOICE FOR UNIVERSITY STUDENTS

This is a model of four possible choice alternatives: 1=on-campus employment only, 2=off-campus employment only, 3=both on- and off-campus employment, and 4=not employed. The estimation results are presented in Table 15. We realize that these four choices are correlated because they have common attributes, but a nested logit model did not work and so we had to resort to the MNL for this first effort. We suggest trying more sophisticated specifications in Phase 2 (such as a cross-nested logit). We assume that this model is "first" because employment is a socio-economic characteristic that has a strong impact on the subsequent choices of living arrangement and residential location.

TABLE 15: EMPLOYMENT TYPE CHOICE MODEL FOR UNIVERSITY STUDENTS

Variable	Description	Coefficient	t-stat
On-Campus Employment Only:			
constant		-0.7402	-5.2
grad	Indicator for affiliation of the student -- Graduate Student	1.7767	12.5
underc	Indicator for affiliation of the student -- Freshman or Sophomore	-0.8323	-4.2
agect1	Indicator variable for age category less than 25 years old	0.4618	2.7
Off-Campus Employment Only:			
constant		0.7130	8.0
underc	Indicator for affiliation of the student -- Freshman or Sophomore	-0.8366	-6.0
agect1	Indicator variable for age category less than 25 years old	0.3757	2.9
On- and Off- Campus Employment:			
constant		-2.0740	-8.2
grad	Indicator for affiliation of the student -- Graduate Student	1.5900	6.5
underc	Indicator for affiliation of the student -- Freshman or Sophomore	-0.8790	-2.9
male	Indicator variable for male respondent	-0.4195	-2.0
agect1	Indicator variable for age category less than 25 years old	0.8950	3.5
Unemployed (reference alternative with zero utility)			
Model Stats:			
R-squared		0.1921	
R-squared Adjusted		0.1905	
Number of observations		2006	
Log-likelihood		-2246.6	

In general the results look reasonable. Graduate students have a significantly higher probability to be employed, while freshmen and sophomore students have a significantly lower probability to be employed. The employment advantage of graduate students is especially prominent for on-campus employment (where they are naturally involved in research and teaching activities) and less prominent for off-campus employment (that is a more a function of familiarity with the area and job search as well possibility to adjust the schedule of classes). However, the employment disadvantage of freshmen and sophomore students is uniform across on-campus and off-campus employment options. Interestingly, younger students of age under 25 years exhibited a higher propensity to be employed versus the older students. This variable is of course applied on top of the affiliation variable, thus it partially mitigates the effect of student's grade since these two variables are correlated..

3.3. CHOICE OF LIVING ARRANGEMENTS FOR UNIVERSITY STUDENTS

This is a model that currently includes the following four choice alternatives: 1=live on-campus, 2=live off-campus with family, 3=live off-campus with roommates, 4=live off-campus alone. The choices can be nested as shown in Figure 10 but at the estimation effort in Phase 1 the nested structure was not confirmed statistically. We tried a nested logit model of employment choice and living arrangement choice, but no success was achieved in estimating such a nested logit model. So, we had to separate the two models and have a separate MNL for living arrangement. So, we have an MNL; however, employment choice enters as an explanatory variable. Thus, in the MAG ABM system, we have to simulate employment choice in Model 1, and then use that indicator as an explanatory variable in the living arrangement model. Table 16 presents the MNL model for choice of on-campus and off-campus living arrangements.

TABLE 16: LIVING ARRANGEMENT CHOICE MODEL FOR UNIVERSITY STUDENTS

Variable	Description	Coefficient	t-stat
Living On-Campus (reference alternative with zero utility)			
Living Off-Campus With Family:			
constant		2.5006	20.8
on_emp	Indicator variable for on-campus employment status	-0.6438	-5.9
agect1	Indicator variable for students age less than 25	-0.9520	-7.7
Living Off-Campus With Roommates:			
constant		0.7146	4.8
under	Indicator for affiliation of the student -- Freshman or Sophomore	-0.8780	-6.3
male	Indicator variable for male respondent	0.4735	4.3
of_emp	Indicator variable for off-campus employment status	0.3917	3.6
agect1	Indicator variable for students age less than 25	0.3317	2.3
Living Off-Campus Alone:			
constant		0.4825	4.0
under	Indicator for affiliation of the student -- Freshman or Sophomore	-1.5998	-6.8
male	Indicator variable for male respondent	0.4250	3.0
Model Stats:			
R-squared		0.1981	
R-squared Adjusted		0.1975	
Number of observations		2006	
Log-likelihood		-2227.7	

As the constants for off-campus alternatives suggest, living with family has an edge over living off-campus with roommates, which in turn, has a slight edge over living off-campus alone. This progression in magnitude is consistent with expectations given that ASU has a large commuter student body with many students continuing to live with family well into their college years.

Students employed on-campus are less likely to live off-campus with family, although the direction of causal relationship is unclear. Do students stay in the dorm because they choose to work on-campus, or do students work on-campus because they are staying in the dorm? In any event, there is a positive association between living and working on-campus that in the current model system structure is modeled as employment as the prior choice and living arrangements as the subsequent choice.

Freshmen and sophomore students show a lower propensity to live off campus in a non-family arrangement (either with roommates or alone with the latter being the least probable case). This is likely a result of the tendency for students who are relatively young in their college years to stay on-campus and build bonds with other students, participate in school activities outside the classroom, and have their meals at on-campus eating facilities. In addition, the ASU campus policy heavily encourages freshman and sophomores to stay on-campus and build a sense of community prior to potentially leaving and staying off-campus. ASU on-campus facilities are largely targeted at this demographic. This effect is complemented by a strong negative impact of young age on probability to live with their families. A certain positive impact of young age on probability of living off-campus with roommates partially compensates for a large negative coefficient for freshmen and sophomore students.

Off-campus employment is logically positively associated with living off-campus with roommates. Male students are more likely to live in off-campus alone or with roommates, which means that females prefer dorms or staying with the family, all else being equal.

3.4. CHOICE OF RESIDENTIAL LOCATION

This is a standard residential location choice model that has a MNL structure with TAZs as alternatives. We have developed a model for all students, and then two separate models for two separate market segments – those that live with family, and those that live alone or with roommates. We currently recommend using the single model for all students as it is more sensitive to a variety of variables, and includes living arrangement as an explanatory variable. It also includes transit availability as an explanatory variable. Thus in the MAG ABM system we simulate living arrangement in model 2 and then use that as an explanatory variable to simulate residential location choice. Table 17 provides model estimation results for the residential location choice model for students living off-campus.

This model is applicable to all students residing off-campus. The results are found to be quite reasonable. As this is a location choice model, with individual zones serving as the location choices, there is no constant in the model. For model estimation purposes, 30 zones were randomly sampled (including the chosen zone) to comprise the choice set for each individual. In future refinements of the model, more effective sampling of zonal alternatives will be undertaken to constitute the choice set. The utility function includes two types of variables: 1=travel impedance variables, 2=zonal size variables that characterize the attractiveness of the zone as residential place for students.

TABLE 17: RESIDENTIAL LOCATION CHOICE MODEL FOR UNIVERSITY STUDENTS

Variable	Description	Coefficient	t-stat
Model for all students living off-campus:			
par_3	SOV Travel Time	-0.0510	-11.7
lagedist	Age less than 25 indicator × Highway Distance	-0.0247	-4.0
mfhh_per	Percentage of Multi-Family Households	-0.4669	-5.0
ln_retem	Logarithm of (1 + Retail Employment)	0.2666	15.4
hinc_ind	Indicator -- Fewer households belong to the bottom two income quintiles compared to the upper three income quintiles	0.4027	7.0
tr_thres	Indicator for poor transit or lack of transit -- 1 if transit travel time > 45 or if transit is not available	-0.1878	-2.6
roaldist	Indicator variable for Roommates/Staying Alone × Highway Distance	-0.1177	-14.5
R-squared		0.2211	
R-squared Adjusted		0.2210	
Number of observations		1758	
Log-likelihood		-4657.4	
Model for students living of-campus with family:			
par_3	SOV Travel Time	-0.0498	-11.0
mfhh_per	Percentage of Multi-Family Households	-0.9053	-6.7
ln_retem	Logarithm of (1 + Retail Employment)	0.2829	12.9
hinc_ind	Indicator -- Fewer households that belong to the bottom two income quintiles compared to the upper three income quintiles	0.7054	8.6
fem_dist	Female indicator × Highway Distance	-0.0146	-2.2
R-squared		0.1445	
R-squared Adjusted		0.1444	
Number of observations		1009	
Log-likelihood		-2935.8	
Model for students living of-campus alone or with roommates:			
par_3	SOV Travel Time	-0.1034	-15.4
lagedist	Age less than 25 indicator X Highway Distance	-0.1093	-7.7
ln_retem	Logarithm of (1 + Retail Employment)	0.2755	9.5
R-squared		0.3488	
R-squared Adjusted		0.3487	
Number of observations		749	
Log-likelihood		-1658.9	

With respect to the travel impedance variables, as expected, zones that are farther away (in terms of travel time) from the ASU campus that the student attends are less desirable as residential locations. The negative impact of travel time (distance) is relatively stronger for younger students who have less flexible schedules and higher frequency of commuting. Logically, poor transit service also works as a negative factor since university students are dependent on transit and have a relatively low car ownership.

Another interesting and logical finding is that the negative impact of distance is relatively stronger for students living alone or with roommates. This means that there is an added positive utility of staying with the family (no need to pay the rent!) that results in higher tolerance to longer commuting. In other words, for students who grew up in Phoenix, living with family is an attractive option even if longer commuting is involved, but students who came to Phoenix for college from a different region, will primarily search for an apartment close to the campus. A possible improvement for this model at Phase 2 is to include a university mode choice logsum as the impedance measure.

With respect to the zone size variables, zones with a generally high income and high density retail establishments are more prone to be chosen as residential locations by students. In interesting effect was found with respect to the household type. In general, zones with a higher percentage of multi-family households are less preferred. This is contrary to the stereotype of students' choice in dense urbanized areas. However, in dense urban areas there is usually a huge gap in rent prices between the single-family and multi-family households that pushes students to seek small studio-type apartments. This is not a significant factor for the Phoenix metropolitan area.

These findings are all consistent with expectations and provide the sensitivity necessary to test how spatial residential distribution of ASU students may vary depending on the type of housing stock in a zone, retain presence in a zone, and distance and travel time from ASU campuses.

Several additional details regarding integration of these models with the core population synthesis procedure will have to be finalized at Phase 2. Students living on campus or off-campus alone or off-campus with roommates will be considered as a separate segment of the group quarter population. However, students living with family should be properly "embedded" in residential households generated by the population synthesizer. The current design assumes that the number of university students in each TAZ defined by the student residential location model will be used as a person-level constraint in the residential population synthesis. This model will be applied not only for ASU but also for all other colleges and universities, thus, generating the entire student population in the region. In order to ensure consistency between the number of residential households and number of university students used as controlled variables in the population synthesis, the student residential location model will be applied in a constrained fashion using a reasonable threshold for a maximum average number of students per household derived from the Census data.

CHAPTER 4.

CHOICE OF USUAL WORKPLACE

4.1. MODEL STRUCTURE

The model for usual workplace is applied for each worker represents a two-level choice. At the upper level a worker makes a *binary workplace type choice between working from home and working at an out-of-home location* (either permanent or variable). At the lower level, those who have an out-of-home workplace choose their *usual workplace location* (currently, at the TAZ level). The entire model can be thought of as a nested choice structure although choices at two levels were estimated sequentially rather than simultaneously. . These two choices are connected through an accessibility-to-jobs measure segmented by occupation in the upper-level model, which represents a simplified workplace location choice logsum. A possible behavioral interpretation of the adopted structure is that work from home vs. work out of home is largely a strategic lifestyle decision made prior to the consideration of possible workplace locations. This is of course a gross simplification of the reality where many factors influencing the decision to work from home and possible out-of-home workplaces are closely intertwined. Also, in reality many of these choices are made (or at least constrained) by the employer rather than by the worker himself.

The workplace type choice model was estimated on a pooled dataset including workers from Phoenix and Tucson. However, some model constants were tested for significance of differences between the two regions. The usual workplace location model was estimated for each area separately because the size and shape of the region as well as the spatial distributions of population and employment all have a strong impact on travel behavior.

4.2. WORKPLACE TYPE CHOICE

4.2.1. Choice Structure

The work from home choice model predicts if a worker's usual work place is home or if he/she works out of home. The model was estimated as a binary MNL using the ALOGIT software. This model is one of the first applied in the model chain and is applied before the usual out-of-home workplace location choice model. The workplace type choice model is applied to all workers and it includes general accessibilities to work locations, household characteristics and worker characteristics as explanatory variables.

This model is of special importance for certain types of policies as well as for long-term planning applications in general. This is due to a direct impact of workplace type on intensity of commuting. The observed tendencies are characterized by a steep increase in the share of workers who work from home. This is due to the tremendous progress made in communication technologies and because of the structural shifts in occupation towards more professional types. The current version of the model captures several cross-sectional effects. It is envisioned that for model application for future years, certain scenarios with a generally higher share of telecommuters compared to the base year should be considered. The corresponding adjustment of the model can be made by means of the alternative-specific constant for work from home.

4.2.2. Estimation Dataset

In the NHTS 2008 household travel behavior survey, there are 4,324 observed worker records for both full-time and part-time workers. Table 18 below shows the distribution of workers in surveyed households by worker status, gender, age group, household income category, education level, and occupation type.

TABLE 18: DISTRIBUTION OF WORKERS BY WORKPLACE TYPE IN NHTS 2008

Worker characteristics		Work from home				Total
		Yes		No		
		Number	%	Number	%	
All workers		593	14%	3,731	86%	4,324
Status	Full-time Worker	348	10%	3,050	90%	3,398
	Part-time Worker	245	26%	681	74%	926
Gender	Male	315	14%	1,889	86%	2,204
	Female	278	13%	1,842	87%	2,120
Age	35 years or younger	68	10%	631	90%	699
	36 years to 45 years	93	10%	803	90%	896
	46 years to 55 years	175	14%	1,109	86%	1,284
	56 years to 65 years	166	16%	896	84%	1,062
	Older than 65 years	91	24%	292	76%	383
Household income	$24,999 or Less	50	15%	283	85%	333
	$25,000 to $49,999	104	12%	781	88%	885
	$50,000 to $74,999	110	13%	742	87%	852
	$75,000 to $99,999	111	13%	717	87%	828
	$100,000 or more	181	15%	1,012	85%	1,193
	Unknown	37	16%	196	84%	233
Education level	Less than Bachelors	274	12%	1,986	88%	2,260
	Bachelors or higher degree	318	16%	1,715	84%	2,033
	Unknown	1	3%	30	97%	31
Occupation	Sales or marketing	109	24%	350	76%	459
	Clerical administrative or retail	84	10%	732	90%	816
	Production, construction, manufacturing, or transport	77	17%	376	83%	453
	Professional, managerial, or technical	186	11%	1,470	89%	1,656
	Person care and services	86	14%	541	86%	627
	Other	32	18%	142	82%	174
	Unknown	19	14%	120	86%	139

Overall, the dataset shows 14% of all workers working from home and 89% of workers traveling to work locations. The data shows that 26% of part-time workers work from home, whereas only 10% of full-time workers are working from home. The percentage of individuals working from home increases from 10% for 35-year-old or younger individuals to 24% for 65-year-old or older workers. There is a logical tendency for higher educated workers to work from home more frequently as well as a significant variation across occupation categories. Interestingly, there was not a systematic impact of either gender or income found.

The NHTS observations were joined with general zonal accessibilities described above to create the estimation file. Specifically, accessibility to jobs for the corresponding occupation type were expected to have an impact on choice to work from home.

4.2.3. Utility Structure

The utility of working out of home was chosen as the reference case and set to zero. Thus all explanatory variables were included in the utility of working from home. The utility (U_{in}) of choosing to work from home for an individual (*n*) residing in zone (*i*) can be written in the following general way:

$$U_{in} = \alpha + \delta \times A_i + \sum_k \beta_k \times N_{nk} \qquad \textbf{Equation 3}$$

where,

k	=	list of person and household characteristics,
α	=	constant for choosing to work from home to estimate,
A_i	=	accessibility to jobs from zone (*i*),
N_{nk}	=	values for person or household characteristics for individual (*n*).
δ, β_k	=	other coefficients to estimate.

4.2.4. Main Explanatory Variables

The following variables were examined and proved to be significant in the utility functions:

- Accessibility to jobs by occupation category
- Household income group:
  - Low income (less than $25,000)
  - Medium-Low income ($25,000-$49,999)
  - Medium income ($50,000-$74,999)
  - Medium-High income ($75,000-$99,999)
  - High income ($100,000 or more)
- Household Composition:
  - Presence of a Non-Working Adult
  - Presence of a Preschool Child (for Female Workers)
- Person Characteristics:

- Work Status – Full-time vs. Part-time.
- Gender – Female vs. Male
- Education Level
- Age group
- Occupation / Job Category

- Tucson dummy (to explore systematic differences between the Phoenix and Tucson regions).

4.2.5. Model Estimation Results

The estimation results for workplace type choice (work from home) are summarized in Table 19. Since all variables were included in the utility for work from home, the coefficient should be analyzed as positive or negative impacts on propensity to work from home.

TABLE 19: MODEL ESTIMATION RESULTS FOR WORKPLACE TYPE CHOICE (WORK FROM HOME)

Variable		Coefficient	t-stat
Constants	General	-0.851	-2.46
	Tucson	-0.034	-0.33
Status	Full Time Worker	-1.178	-11.04
Gender	Female	-0.346	-3.43
Household composition	Female Worker with Preschool Child	0.382	1.68
	Non-Working Adults in the HH	-0.192	-1.54
Occupation	Sales or marketing	0.765	5.89
Age Group	Age <= 35 years	-0.230	-1.31
	35 years to 44 years (reference)		
	45 years to 54 years	0.332	2.34
	55 years to 64 years	0.348	2.37
	Age 65years or older	0.432	2.38
Household Income group	$49,999 or Less	-0.090	-0.63
	$50,000 to $74,999 (reference)		
	$75,000 to $99,999	0.160	1.07
	$100,000 or more	0.267	1.95
Education Level	Less than High School Educated	-0.398	-0.95
	High School completed (reference)		
	Bachelor's or Some College degree	0.295	2.28
	Master's or higher degree holder	0.300	1.89
Accessibility	Accessibility to Employment Locations	-0.069	-2.22
Model stats	Number of Observations	4,324	
	Likelihood with Constants only	-1728.4776	
	Final likelihood	-1601.4239	
	Rho-Squared (0):	0.4657	
	Rho-Squared (constant):	0.0735	

4.2.6. Behavioral Findings and Interpretations

In general, the model estimation yielded logical results with the following main behavioral findings and interpretations:

- The general constant for work from home is large and negative reflecting a fact that working from home is still by far the less frequent alternative. The insignificant additional constant for Tucson suggests that there is no principal difference between the regions in this choice context.
- The coefficient on Full-time worker is negative as expected. It means that part-time workers are more likely to work-from-home on a regular basis.
- There are two coefficients for females – one directly on female and another capturing the affect when a preschool child is present in the household. The later coefficient is positive and fully offsets the affect of negative female coefficient. This means that females without preschool children are less likely to work-from-home and females with preschool children are more likely to work-from-home.
- Workers in Sales and Marketing occupation are more likely to work from home. These job types have more flexibility and options to work remotely (by phone, E-mail, or Internet) compared to other job types.
- Workers from higher-income households are more likely to work from home as compared to lower income groups. This might be a reflection on higher and managerial positions in addition to occupation. We decided to adopt several coefficients for income variables even though they proved to be not extremely significant statistically since the coefficient values had a logical sign and magnitude.
- There is a clear progression within age group which shows that older age workers are more likely to work from home compared to younger workers. Older workers more frequently hold positions which have greater flexibility of work schedules as compared to younger workers.
- Workers with college education are more likely to work-from-home. This probably is due to the nature of work and positions for individuals with a bachelor or higher degree (more professional and intellectual).
- In households with non-working adults, the workers are less likely to work-from-home. This is quite probably due to non-workers involvement in child care and other household errands that makes the workers less interested in staying at home.
- Accessibility to jobs of the corresponding occupation category logically has a strong negative impact on working from home. It is an interesting question if the decision to find a work from home was driven by the lack of relevant jobs within a reasonable commuting distance or those who had already decided to work from home reside in areas distant from the jobs (i.e. is work from home a privilege or necessity?). This causality cannot be resolved by a cross-sectional analysis. In the MAG ABM system the underlying behavioral assumption is that the lack of accessible jobs may cause the person to work from home.

4.3. CHOICE OF USUAL WORKPLACE LOCATION ZONE

4.3.1. Choice Structure

The usual work location choice model predicts the usual work location for full-time and part-time workers who work out of home. The model was estimated in a MNL form using the ALOGIT software. This model is preceded by the binary model for workplace type (work from home) choice which identifies individuals working from home. The work location choice model is applied only to workers who do not work from home. The model considers all TAZs in the region that have jobs of the corresponding occupation type as alternatives for the given individual. It includes zonal employment (as the size variable), OD accessibilities (mode & time-of-day choice logsums), general zonal accessibilities, distance terms, household characteristics, and worker characteristics as explanatory variables.

The specific advanced features of this model include segmentation of the size variable by occupation and a greater level of details in describing individual effects on the travel impedance. Since there are a large number of alternatives (TAZs), it is not possible to include all alternatives in the estimation dataset. The model was estimated using a sampling-by-importance strategy where 40 TAZs were randomly selected for each record based on the size variable and simplified distance-based impedance function. Each original worker record was duplicated 10 times before sampling, with the weight of each record set to 0.1. The 40 random draws for each resulted record were implemented independently. This way the estimation dataset was enlarged and the Monte-Carlo variability associated with random draws was reduced to minimum. This approach is roughly equivalent to selecting 400 alternatives for the choice set.

The model was estimated separately for the Phoenix and Tucson regions. It was found in previous research that models for location choices are highly dependent on the size and shape of the region as well as the spatial structure of population and employment.

4.3.2. Estimation Dataset

In the NHTS 2008, there are 2,947 observed worker records in the Phoenix and Tucson regions including both full-time and part-time workers after excluding workers who work from home and workers with either unknown workplace TAZ or workplace outside the modeled area. Table 20 below shows distribution of the working adults in surveyed households by worker status, gender and income group. In general, this is a large enough and representative dataset with the necessary variation across household and person characteristics.

In order to create the estimation file, the survey observations after duplication and sampling were joined with the alternative-specific (i.e. sampled TAZ-specific) variables that include OD accessibility (mode & time-of-day choice logsum), distance from home to work TAZ, employment of the occupation category corresponding to the worker occupation in TAZ (see Table 9), and general zonal accessibilities from TAZ (see Table 13).

TABLE 20: DISTRIBUTION OF WORKING ADULTS

Characteristic	Phoenix (MAG)		Tucson (PAG)		Total	
	Number	%	Number	%	Number	%
Worker Status:						
Full-time	1,643	81.9%	744	79.0%	2,387	81.0%
Part-time	286	14.3%	147	15.6%	433	14.7%
Unknown	76	3.8%	51	5.4%	127	4.3$
Gender:						
Male	980	48.9%	466	49.5%	1,446	49.1%
Female	1,025	51.1%	476	50.5%	1,501	50.9%
Income group:						
Less than 25K	105	5.2%	62	6.6%	167	5.7%
25K to 50K	371	18.5%	230	24.4%	601	20.4%
50K to 75K	401	20.0%	182	19.3%	583	19.8%
75K to 100K	416	20.7%	165	17.5%	581	19.7%
More than 100K	623	31.1%	252	26.8%	875	29.7%
Unknown	89	4.4%	51	5.4%	140	4.8%
Total	2,005	100%	942	100%	2,947	100%

4.3.3. Main Explanatory Variables

The following variables were examined and proved to be significant in the utility functions:

- Total employment of the occupation category corresponding to the worker's occupation ("relevant employment"). It was used as the size variable.
- Composite Mode & Time-of-Day Choice Logsum
- Impedance distance-decay between the home and potential work destinations including the following terms:
  - Linear distance
  - Natural logarithms of distance
  - Square root of distance
  - Distance squared
  - Distance cubed
- Household income group interacted with distance terms (categories can be aggregated):
  - Low income (less than $24,999)
  - Low medium income ($25,000 - $49,999)

  - Medium income ($50,000 - $74,999)
  - Medium High income ($75,000 - $99,999)
  - High income ($100,000 and more)
- Person characteristics interacted with distance terms:
  - Work Status – Full-time vs. Part-time.
  - Gender – Female vs. Male; for females, additionally, presence of a child in the household was tested.
- Zonal accessibility indices from the work TAZ to potential destinations. These terms are introduced to account for possible at-work sub-tours as well as competition between work location TAZs. These terms help bring the model that is estimated in an unconstrained fashion to the way how this model is applied with constraining.
  - Accessibility to Total Employment from the workplace TAZ (one of zonal accessibility measures 1-5 described above that is relevant according to the worker occupation)
  - Accessibility to non-mandatory attractions form the workplace TAZ (one of the zonal accessibility measures 51-52 described above that is relevant by the actual car availability)

The zonal accessibility measures take the form of destination choice logsums and represent a result of summation of attractions across all destinations. High total-employment accessibility means that the TAZ is located in the cluster of jobs of the relevant occupation category. Hence, this TAZ has to compete with the nearby TAZs and might have a higher balancing factor in a constrained model. High non-mandatory/retail accessibility means that the TAZ is located in a cluster of retail jobs and might have an additional attractiveness.

4.3.4. Utility Structure

The utility (U_{ijn}) of choosing a work destination (*j*) for an individual (*n*) in zone (*i*) has the following form:

$$U_{ijn} = \ln\left(S_j\right) + \alpha \times L_{ij} + \sum_k \left(\delta^k \times A_{ij}^k\right) + \sum_{mz} \left(\beta^{mz} \times D_{ij}^m N_n^z\right) + C_{jn} \qquad \textbf{Equation 4}$$

where:

S_j	=	size variable for location zone (*j*),
L_{ij}	=	composite mode & time-of-day choice logsum between zone pair (*ij*),
A_{ij}^k	=	accessibility terms by type (*k*),
D_{ij}^m	=	distance terms by type (*m*) (linear, log, squared, cubed and square root),
N_n^z	=	Indicator on person/household characteristic (*z*) for individual (*n*),
C_{jn}	=	correction term introduced to compensate for the sampling error,
$\alpha, \delta^k, \beta^{mz}$	=	coefficients to estimate.

The correction term is introduced in the model estimation to represent the difference between the sampling probability and frequency for each alternative. This correction factor is explained below. The correction factors are included in the model estimation only and are not applied in the model application.

A combination of distance terms is used in the utility in such a way that the composite distance utility function is monotonically decreasing within the maximum chosen work distance range (72 miles). Table 21 shows the observed frequency of distance to work location for 2,947 workers in the dataset.

TABLE 21: FREQUENCY DISTRIBUTION FOR DISTANCE FROM HOME TO WORK

Bin (miles)	Frequency – Phoenix	Frequency - Tucson	Total Frequency
≤5	492	270	762
5-10	439	287	726
10-15	348	176	524
15-20	283	114	397
20-25	180	54	234
25-30	122	24	146
30-35	64	6	70
35-40	34	8	42
40-45	21	0	21
45-50	7	1	8
50-55	8	2	10
55-60	1	0	1
60-65	2	0	2
65-70	1	0	1
70-75	2	0	2
75-80	1	0	1
Total	2,005	942	2,947

The same data in a form of distance distributions are presented in Figure 11 below. It can be seen that the size of region has a significant impact on the commuting distance. The entire distribution for Tucson is shifted towards shorter distances compared to Phoenix.

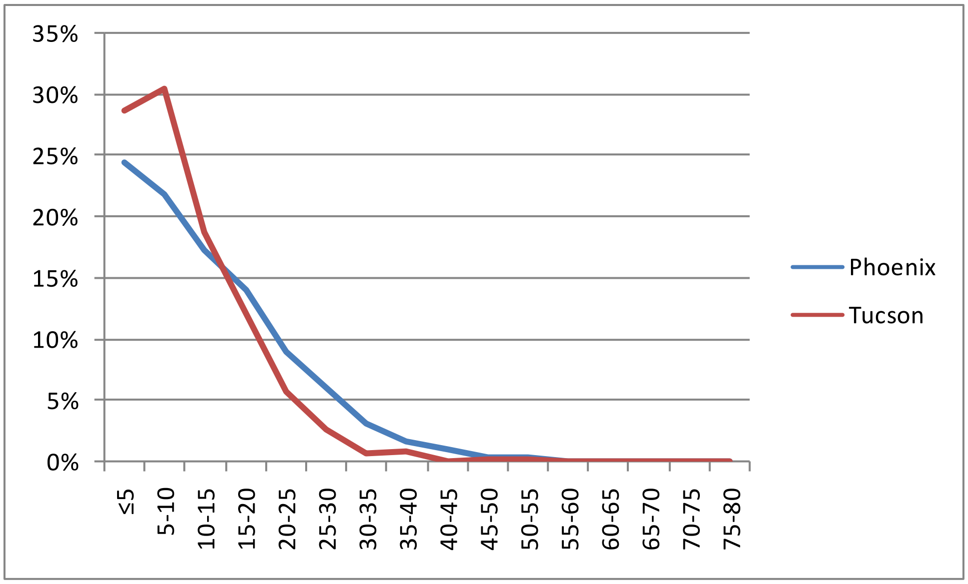

FIGURE 11: DISTRIBUTION FOR DISTANCE FROM HOME TO WORK

4.3.5. Model Estimation Results

The final estimation results for usual workplace location model for both Phoenix and Tucson regions are summarized in Table 22 below. The models for Phoenix and Tucson are intentionally presented back to back in order to see the similarities and differences. In general, the same factors and effects proved to be statistically significant in both regions. However, the actual values of coefficients as well as some details of the variable transformation proved to be specific to the region.

TABLE 22: MODEL ESTIMATION RESULTS FOR USUAL WORK LOCATION CHOICE

Variable		Phoenix		Tucson	
		Coefficient	t-stat	Coefficient	t-stat
Size variable	Log of total employment of the relevant occupation	1.000		1.000	
Accessibility from work	Access to Work Locations from Destination	-0.034	-1.84	-3.037	-9.48
Base impedance measures	Composite mode & time-of-day choice logsum	0.500	Fixed	0.171	3.56
	Intra-zonal dummy (home and work in the same TAZ)	0.581	8.30	1.583	13.83
	Linear Distance	0.015	0.85	-0.140	-4.03
	Log (1+ Distance)	-0.672	-8.22	-0.083	-0.53
	Distance Squared	-0.001	-1.42	0.002	1.55
	Distance Cube	-0.0001	-0.12	-0.0001	-0.87
Additional impedance effects for part-time worker	Linear Distance	-0.124	-4.27	0.143	3.33
	Log (1+ Distance)	0.191	1.36	-1.258	-6.26
	Distance Squared	0.002	2.57	-0.002	-2.52
	Distance Cube	-0.0002	-1.80		
Additional impedance effects for females	Linear distance	-0.055	-6.53	-0.004	-0.81
	Distance Squared	0.001	3.45		
	Distance Cube	-0.0003	-3.23		
Additional impedance for female with preschool child	Extra Distance over 40 miles – Linear	-0.048	-0.98		
	Linear Distance			-0.025	-2.02
Additional impedance effects by household income groups	Linear distance, Income <=50K	-0.020	-7.32		
	Linear distance over 20 miles, Income <=50K			-0.077	-3.64
	Linear distance, Income >=100k	0.048	5.41	-0.010	-0.85
	Squared distance, Income >=100k	-0.002	-5.78	0.001	2.05
	Cubed distance, Income >=100k	0.000	5.34		
Goodness-of-fit stats	Number of Observations	18,860		7,340	
	Likelihood with Constants only	-67807.493		-25389.043	
	Final likelihood	-62385.388		-24267.297	
	ρ^2 w.r.t. zero	0.082		0.0472	
	ρ^2 w.r.t. constants	0.080		0.0442	

4.3.6. Behavioral Findings and Interpretations

The estimated workplace location choice models are much richer than standard gravity models for work trip distributions and contain many more parameters. Below is an explanation of the following main effects:

- The coefficient on the composite mode & time-of-day choice logsum is positive and in the unit interval as required by theory. However, the logsum coefficient for the Phoenix region was asserted because the estimated value was very weak and statistically insignificant. It might be simpler for policy analysis with a model for entire Phoenix-Tucson region to enforce the same logsum coefficient for both regions. This model modification will have to be considered at Phase 2.
- The accessibility variables are used as additional size terms, but may not be used in the model application. These terms do not affect the model application results since they are absorbed by the balancing factors (shadow prices) introduced in the constraining procedure. However, they help the unconstrained model estimation, as indicated by the negative value for accessibilities - which means that the unconstrained destination choice model would overestimate TAZs located in major employment clusters where TAZs compete with each other. Another possible accessibility term that could help account for this competition is accessibility to labor force by place of residence. However, it did not prove significant in the current models.
- A composite distance function (or Distance-Decay factor) has been defined as a combination of linear, logged, squared and cubed distance terms with different coefficients. This term should be analyzed as a composite term and the coefficient (positive or negative) of individual terms should not be looked at separately. For example, the coefficient on linear distance is positive but it does not mean that workers choose distant locations as work places. One should look at the combined effect of all terms. Figure 12 and Figure 13 below show the base distance-decay factor (or the composite distance term) for the reference case (i.e. Full-time worker, male. and Medium Income group) and marginal impacts for other segments for Phoenix and Tucson regions, respectively. The functions are presented in utility units. Further, all marginal factors are combined to produce final distance-decay functions for each person & household type as presented in Figure 14 and Figure 15 for Phoenix and Tucson respectively. The final combined functions are monotonously decreasing in within the maximum observed work distance range. These functions are discussed below in more detail.
- Part-time workers are more sensitive to commute distance than full-time workers, and their sensitivity increases with longer distances. It is behaviorally appealing that part-time workers would avoid long commuting with a disproprtional travel time comapred to the work activity duration.
- Females are less likely to travel longer distances compared to males. This could be due to household responsibilities and child care at home. Logically, the sensitivity of a female worker to distance increases if there is a preschool child in the household.
- Low-income workers are more sensitive to commuting longer distances while higher-income workers are less sensitive. This is a known behavioral effect that is rooted in two main factors. The first factor relates to the versatility of the labor force, which is inversely proportional to income. The second factor relates to residential self-choice,

where higher-income households tend to reside in suburban areas while lower-income households tend to reside in dense urban areas.

The distance-decay functions illustrate the impact of person and household attributes on sensitivity to travel distance. All functions are presented in utility units as a function of distance. Based the typical MNL sensitivity, it can be said that a positive (negative) value of 1.0 might roughly result in doubling (halving) the probability to chose the workplace, all else being equal. Large negative values over 5.0 practically eliminate the probability for TAZ to be chosen as the workplace.

The distance effects are estimated for the base case - a male full-time worker (FTW) from a medium-income household (medinc). There are five additional, additive effects on top of the base distance-decay function that include the following: 1=the effect of part-time work (PTW), 2=the effect of female gender, 3=the additional effect of female with a preschool child, 4=the effect of low income (Lowinc), 5=the effect of high income (Highinc). The corresponding graphs are presented in Figure 12 for Phoenix and Figure 13 for Tucson.

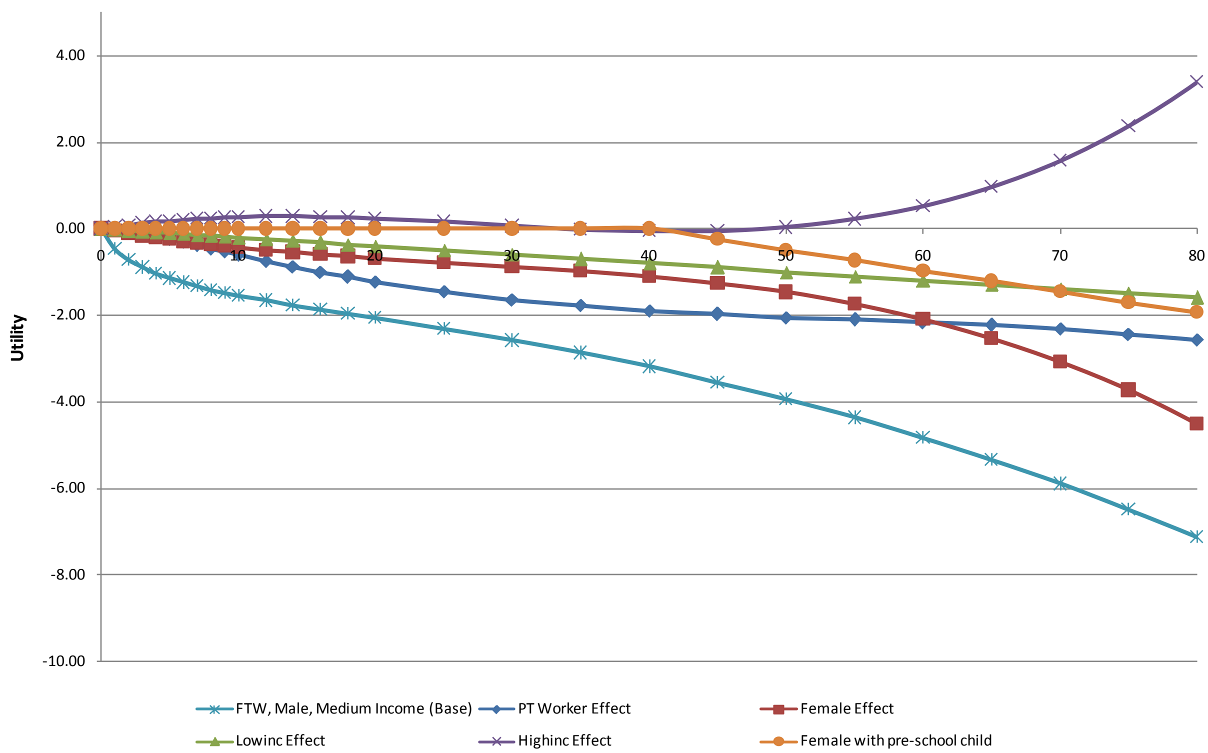

FIGURE 12: BASE DISTANCE-DECAY FUNCTION AND ADDITIONAL EFFECTS FOR PHOENIX REGION

It can be seen from Figure 12 for the Phoenix region that the base curve is monotonically decreasing, part-time workers have an additional distance-averse effect, females have additional distance-averse effect that becomes especially strong after 60 miles, females with a preschool child have an additional distance-averse effect after 40 miles, low incomes have somewhat distance-averse effect (although not extremely prominent), while high incomes are much more tolerable to long distances after 40 miles compared to the base case. It should be noted that all effects after 50 miles are based on a very small number of observations in the NHTS, 2008. Thus, it is important to validate these functions against the CTPP data in model application.

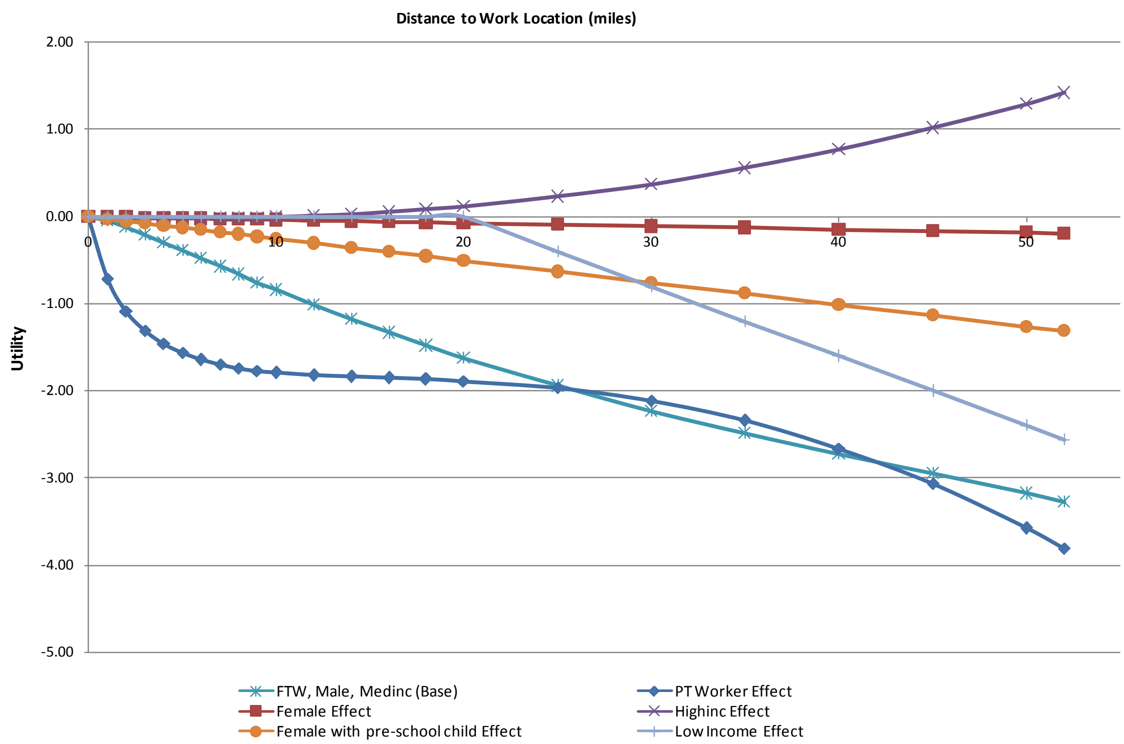

FIGURE 13: BASE DISTANCE-DECAY FUNCTION AND ADDITIONAL EFFECTS FOR TUCSON REGION

The corresponding graphs for the Tucson region presented in Figure 13 show tendencies similar to those observed in the Phoenix region. The most notable differences include the fact that the part-time distance-decay curve for Tucson is much steeper, reflecting the smaller size of the region and shorter average commuting distance. However, it should be noted that the entire baseline distance effect in the Tucson model is mitigated by the smaller coefficient for the mode & time-of-day choice logsum. Most of the effects in the Tucson region manifest themselves earlier on the spatial scale (after 20-30 miles). Another interesting detail specific to the Tucson region is the minimal impact of gender (unless the household has a preschool child). This might be a consequence of the shorter average commuting distance in Tucson compared to Phoenix.

The baseline and additional effects are additive in the utility function. Thus, for each particular person and household the relevant combination should be considered. By combining all possible person and household types we arrive at 18 segments (2 worker statuses × 3 gender-child states × 3 income groups). The corresponding composite distance-decay curves for all segments are presented below in Figure 14 for Phoenix and Figure 15 for Tucson.

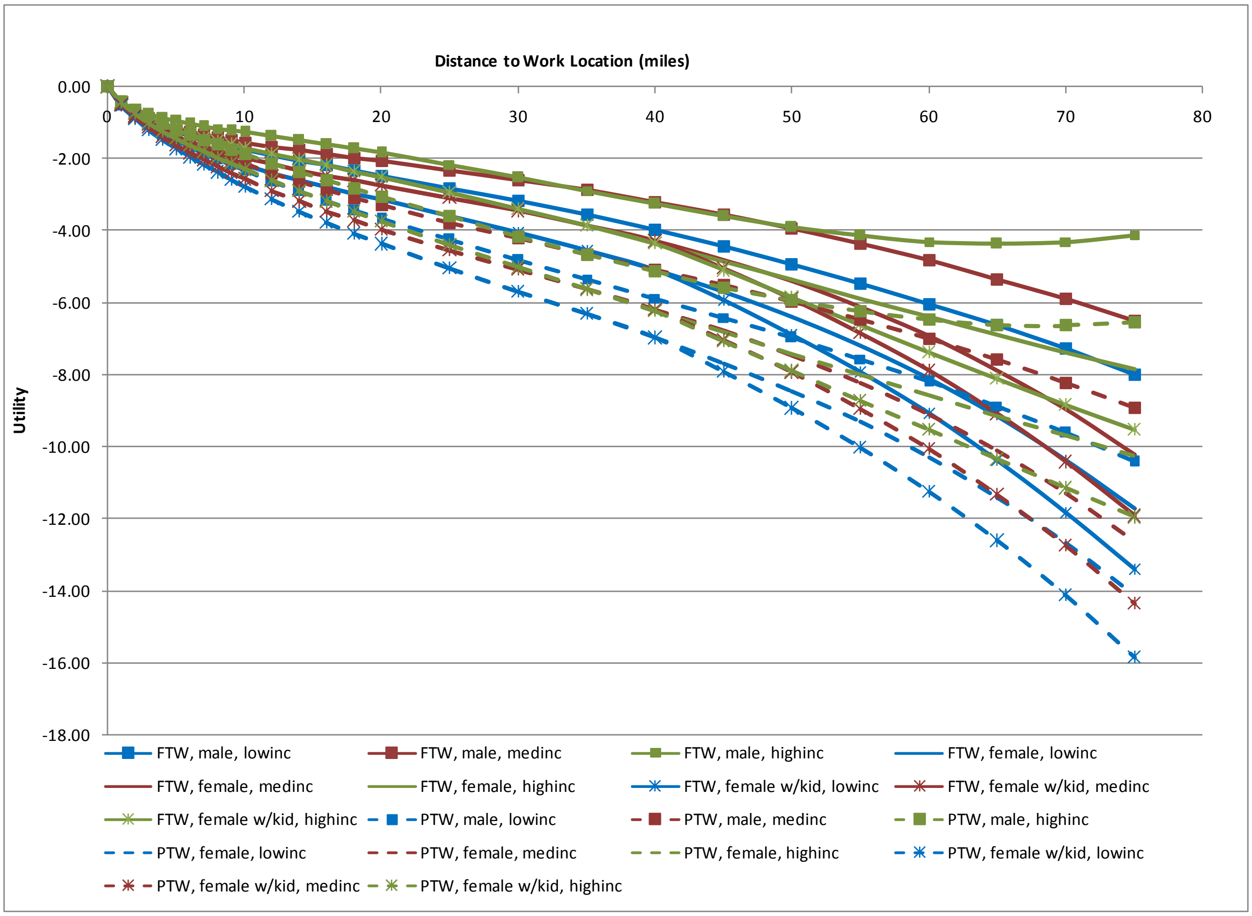

FIGURE 14: DISTANCE-DECAY FUNCTIONS BY PERSON & HOUSEHOLD TYPE FOR PHOENIX REGION

It can be seen from Figure 14 for Phoenix that there is a very wide range of person and household types with respect to sensitivity to commuting distance. Logically, the most sensitive individual category is a part-time female worker with a preschool child from a low-income household. For this individual type any distance beyond 20 miles is improbable since the distance-decay function reaches a value of negative 4.0. On the opposite side of the spectrum is a full-time male worker from a high-income household. For this individual, longer commuting is quite probable and the distance-decay function approaches the same value of negative 4.0 only after 50 miles. The other individual worker types lie in between these two extremes.

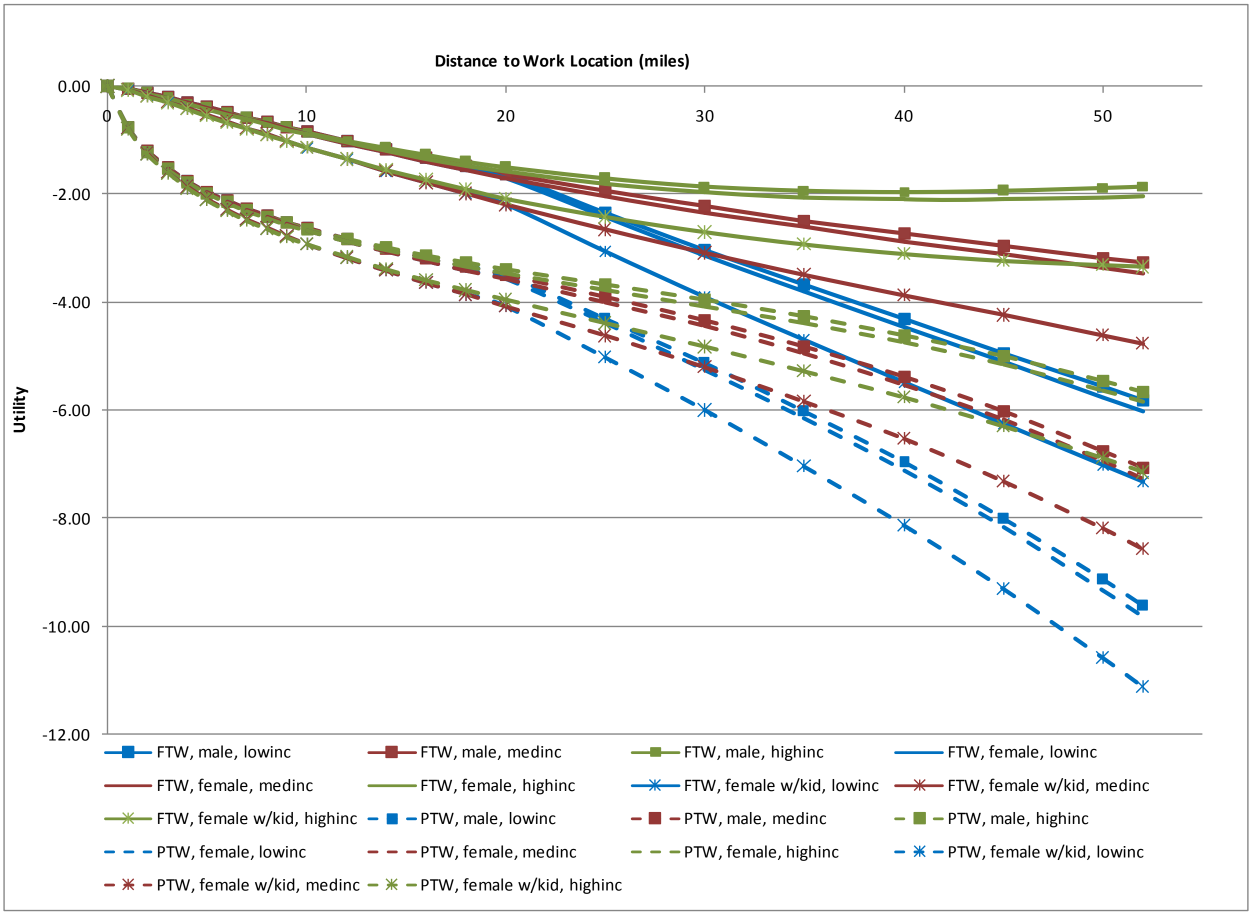

FIGURE 15: DISTANCE DECAY-FUNCTIONS BY PERSON & HOUSEHOLD TYPE FOR TUCSON REGION

The same final composite distance-decay functions are presented for 18 worker types for the Tucson region in Figure 15. The same extreme cases as discussed above for the Phoenix region can be mentioned. A part-time female worker from a low-income household with a preschool child reaches a value of negative 4.0 at 20 miles. Conversely, a full-time male worker from a high-income houshold is quite tolerable to commuting distance up to the end of the observed range (50 miles). Since there is no strong impact of gender in the Tucson region, almost the same tolerance to long commuting is observed among full-time female workers from high-income househlds.

4.3.7. Sampling Correction Factors for MNL Choice Model

The methodology for theoretically correct calculation of the sampling correction factors for a MNL model is outlined below. This method was applied for all location choices that require sampling of alternatives including usual workplace choice discussed above and usual school location choice discussed below. Introduce the following notation:

$i \in C$	=	unique alternatives from the full set
$i \in D \subset C$	=	unique alternatives from the sample
$q(i)$	=	selection probability (probability to be drawn)
n_i	=	selection frequency in the sample
N	=	sample size
V_i	=	utility of a choice alternative
$P(i)$	=	choice probability

Note that the selection frequencies in the sample over unique alternatives are totaled to the sample size:

$$\sum_{i \in D} n_i = N \, .$$

However, the number of unique alternatives in the sample D can be any number between 1 and N inclusive.

The choice probability with sampling correction factors can be calculated by the following formula:

$$. P(i) = \frac{\exp\left[V_i + \ln\left(\frac{n_i}{N \times q(i)}\right)\right]}{\sum_{j \in D} \exp\left[V_j + \ln\left(\frac{n_j}{N \times q(j)}\right)\right]} = \frac{\left(\frac{n_i}{N \times q(i)}\right) \times \exp(V_i)}{\sum_{j \in D} \left(\frac{n_i}{N \times q(j)}\right) \times \exp(V_j)} \qquad \textbf{Equation 5}$$

Since N is a fixed number it can be cancelled out and the formula (Equation 5) can be equivalently rewritten in a simpler form:

$$P(i) = \frac{\exp\left[V_i + \ln\left(\frac{n_i}{q(i)}\right)\right]}{\sum_{j \in D} \exp\left[V_j + \ln\left(\frac{n_j}{q(j)}\right)\right]} = \frac{\left(\frac{n_i}{q(i)}\right) \times \exp(V_i)}{\sum_{j \in D} \left(\frac{n_i}{q(j)}\right) \times \exp(V_j)} \qquad \textbf{Equation 6}$$

Formula (Equation 5) assumes a utility correction factor of $\ln\left(\frac{n_i}{N \times q(i)}\right)$, while formula (Equation 6) assumes a correction factor of $\ln\left(\frac{n_i}{q(i)}\right)$. Since both formulas yield the same probabilities, the simpler correction factor from the formula (Equation 6) is normally applied in the choice context.

However, if also a logsum of this choice model is applied in some upper level choice model, then this log-sum should be calculated as a denominator of the formula (Equation 6):

$$LS^1 = \ln\left\{\sum_{j \in D} \exp\left[V_j + \ln\left(\frac{n_j}{N \times q(j)}\right)\right]\right\} = \ln\left(\sum_{j \in D}\left(\frac{n_i}{N \times q(j)}\right) \times \exp(V_j)\right) \qquad \textbf{Equation 7}$$

If the formula (Equation 6) was applied for the choice probability correction, then the log-sum takes the following form:

$$LS^2 = \ln\left\{\sum_{j \in D} \exp\left[V_j + \ln\left(\frac{n_j}{q(j)}\right)\right]\right\} = \ln\left(\sum_{j \in D}\left(\frac{n_i}{q(j)}\right) \times \exp(V_j)\right) \qquad \textbf{Equation 8}$$

Then the log-sum (Equation 8) calculated based on the formula (2) should be scaled in order to replicate the value of (Equation 7) based on the formula (Equation 5):

$$LS^2 = LS^1 - \ln N \qquad \textbf{Equation 9}$$

Thus, there are two ways to implement corrections in both choice model and logsum calculations:

- Use formula (Equation 5) for utility correction factors and use the logsum directly from the denominator of the formula (Equation 5)
- Use formula (Equation 6) for utility correction factors and then scale the logsum from the denominator by formula (Equation 9)

If we assume that all selection frequencies are equal to one ($n_j = 1$) and all selection probabilities are equal ($q(j) = q = \frac{1}{R}$, where R is the size if the full set) the formula (Equation 7) can be simplified:

$$LS^1 = \ln\left(\sum_{j \in D}\left(\frac{R}{N}\right) \times \exp(V_j)\right) = \ln\left(\sum_{j \in D} \exp(V_j)\right) + \ln\left(\frac{R}{N}\right) \qquad \textbf{Equation 10}$$

This formula can be applied for crude estimations when the original logsum was calculated without correction factors and sampling was random w/o replacement. In this case the logsum just has to be expanded by a factor equal to the full set size divided by the sample size.

CHAPTER 5.

CHOICE OF USUAL SCHOOL LOCATION

5.1. SCHOOLING FROM HOME VS. OUT-OF-HOME SCHOOL

5.1.1. Choice Structure

The schooling from home choice model predicts if a child is schooled at home and therefore, does not travel to school. The model was estimated as a binary MNL using the ALOGIT software. This model is one of the first models applied in the model chain and is applied before the usual out-of-home school location choice model. The latter model is only applied for those children who go to school out of home. The schooling-from-home choice model is applied to children between ages 0 and 15 years with separate models for preschool children (0 to 5 years) and school children (6 to 15 years). The model is not applied to driving-age school children (16 to 17 years) since there was not a single case of schooling from home for them recorded in the survey. The choice utility function includes general accessibilities to school locations, household characteristics and person characteristics as explanatory variables.

5.1.2. Estimation Dataset

In the NHTS 2008, there are 682 observed records for preschool children and 1,178 observed records for pre-driving age school children. Table 23 below shows these children in surveyed households by gender, age group, household income category and household location type. Overall, the dataset shows that around 2% of all school age children are home-schooled and 85% of preschool children stay at home.

There is no significant variation observed for the percentage of schooling from home for school children. This suggests that schooling from home for school children is not a strong function of the conventional person or household attributes but rather is subject to some factors that cannot be currently modeled explicitly. This may include child's disability and/or special parents' decision.

There is some logical variation for the percentage of schooling from (staying at) home for preschool children across the age categories as well as household income categories and urban types. Logically, for the youngest children of age 0 to 3 years there is no regular day care option and schooling from (staying at) home constitutes 100%. For children of age 4 to 5 years, schooling from (staying at) home falls to 60%. Schooling from home in rural areas is more frequent compared to urban areas reflecting on the availability of day care institutions within a reasonable distance

The survey observations were joined with general zonal accessibilities to schools of the corresponding type to create the estimation file. It was expected that the accessibility-to-school measure will provide a more objective quantification of the difficulty to find a school or day care institution that might affect the parents' decision to make child stay at home.

TABLE 23: SCHOOLING FROM HOME FREQUENCY FOR PRE-DRIVING AGE CHILDREN

Person and household characteristics	Pre-driving Age School Children					Pre-School Children				
	Schooling from Home		Go to School		Total	Schooling from/Stay at Home		Go to School/Day Care		Total
	Num.	%	Num.	%	Num.	Num.	%	Num.	%	Num.
Gender:										
Male	11	2%	586	98%	597	302	84%	58	16%	360
Female	16	3%	565	97%	581	279	87%	43	13%	322
Age Group:										
0 to 1 years						182	100%	0	0%	182
2 to 3 years						249	100%	0	0%	249
4 to 5 years						150	60%	101	40%	251
6 to 9 years	9	2%	449	98%	458					
10 to 12 years	8	2%	330	98%	338					
13 to 15 years	10	3%	372	97%	382					
Income group:										
Less than 25K	1	1%	125	99%	126	64	88%	9	12%	73
25K to 50K	7	3%	242	97%	249	134	84%	26	16%	160
50K to 75K	7	4%	171	96%	178	109	86%	18	14%	127
75K to 100K	7	3%	252	97%	259	118	86%	19	14%	137
More than 100K	4	1%	309	99%	313	139	83%	29	17%	168
Unknown	1	2%	52	98%	53	17	100%	0	0%	17
Household location:										
Urban	20	2%	893	98%	913	400	83%	82	17%	482
Rural	7	3%	258	97%	265	181	91%	19	10%	200
Total	27	2%	1,151	98%	1,178	581	85%	101	15%	682

5.1.3. Main Explanatory Variables

The following variables have been statistically examined and proved to be significant:

- Accessibility to School Locations for Elementary and Mid School Enrollment
- Household income group:
  - Low income (less than $25,000)
  - Medium Low income ($25,000-$49,999)

  - Medium income ($50,000-$74,999)
  - Medium High income ($75,000-$99,999)
  - High income ($100,000 or more)
- Household Composition variables including:
  - Presence of Non-Working Adults or Retirees
  - Presence of Part-Time Workers
  - Presence of Pre-driving Age School Child
  - Maximum Educated Person in the Household –bachelor's degree or higher
- Child age,
- Household Location Variables:
  - Tucson dummy to explore differences between Phoenix and Tucson,
  - Urban vs. Rural area type.

5.1.4. Utility Structure

In the binary choice between schooling from home and out of home, the out-of-home school is considered as the reference alternative with the utility set to zero. The utility (U_{in}) of schooling from home for an individual (*n*) in zone (*i*) incorporates all explanatory variables and can be written in the following way:

$$U_{in} = \alpha + \delta \times A_i + \sum_k \beta_k \times N_{nk}$$ **Equation 11**

where:

α	=	constant for schooling from home,
A_i	=	accessibility to schools from residential zone (*i*), and
N_{nk}	=	person and household characteristics (*k*) for individual (*n*).
α, β_k	=	coefficients to estimate.

5.1.5. Model Estimation Results for Preschool Children

The estimation results for schooling-from-home choice model are summarized in Table 24 for preschool children. Since all variables were included in the utility for schooling from home the values of coefficients should be analyzed as positive or negative impacts of the corresponding variables on schooling from home.

TABLE 24: ESTIMATION RESULTS FOR SCHOOLING-FROM-HOME CHOICE MODEL FOR PRESCHOOL CHILDREN

Variable		Coefficient	t-stat
Constants	General constant	-0.305	-0.64
	Tucson Constant	-0.205	-0.67
Age group	Age 0 to 1 yrs	9.000	fixed
	Age 2 to 3 yrs	9.000	fixed
	Age 4 to 5 yrs (reference)		
Household Income Group	$49,999 or less	0.153	0.38
	$50,000 to $74,999 (reference)		
	$75,000 to $99,999	-0.112	-0.24
	$100,000 or more	-0.358	-0.81
Household Composition	Presence of Part Time Worker	0.466	1.12
	Presence of Non-Working Adult or Retiree	0.343	1.06
	Most Educated Person has Bachelor's or higher degree	0.525	1.47
Household Location	Urban (reference)		
	Rural	0.794	2.43
Goodness-of-fit stats	Number of Observations	673	
	Likelihood with Constants only	-284.5699	
	Final likelihood	-161.0829	
	Rho-Squared (0):	.6547	
	Rho-Squared (constant):	0.4339	

For younger ages from 0 to 3 years, large positive constants were introduced to ensure a close to 100% probability of staying at home. For these ages, we do not have observations with a day care.

5.1.6. Findings and Behavioral Interpretations for Preschool Children

The following main findings and behavioral interpretations can be mentioned with respect to the factors influencing parents' decision to leave a preschool child at home:

- The overall constant for schooling at home is negative since for two age categories there are strong positive biases; on top of the general constant there is a negative Tucson dummy though not very significant. This means that there is no significant difference between the Phoenix and Tucson regions but there is some what a tendency for less schooling from home in Tucson (probably as a consequence of urban structure).
- All observed cases for children under 4 years of age were for home schooling or staying at home. Therefore, a large positive constant was asserted for these age groups.

- Preschool children are more likely to be home-schooled if there is an adult who is either part-time worker or a non-worker (who are the most frequent caretaking adult person types).
- Education level of the adults in the household is positively correlated with probability of home schooling for children. This can be explained by the fact that educated parents might be more interested in having the child at home rather than sending him to a day care. Education level is also normally correlated with work schedule flexibility that makes this arrangement easier.
- The accessibility to school locations did not prove to be significant for pre-school children. This might be a consequence of a lack of full information about day care institutions and their actual capacity (enrollment). Thus, a good size variable could not be prepared.
- Preschool children are more likely to be home-schooled in low income households ($49,999 or less) compared to higher-income households. This could be associated with day care cost or preschool cost. Another reason might be a correlation between the household income and number of working adults. As the result of this correlation, we observe schooling from home cases more frequently in low-income households since there are caretaking non-working adults in them while the higher-income households are comprised mostly of workers.
- In rural areas, preschool children are more likely to stay at home compared to urban areas. There are fewer preschools or day cares in rural locations and the distances to such locations are longer compared to urban areas. This way, residential area type works as a proxy for the accessibility measure. In this particular case, it was difficult to construct a proper quantitative accessibility measure since the actual day-care locations are not known.

5.1.7. Model Estimation Results for Pre-Driving Age School Children

The estimation results for schooling-from-home choice model are summarized in Table 25 for pre-driving age school children. Since all variables were included in the utility for schooling from home the values of coefficients should be analyzed as positive or negative impacts of the corresponding variables on schooling from home.

TABLE 25: ESTIMATION RESULTS FOR SCHOOLING-FROM-HOME CHOICE MODEL FOR PRE-DRIVING AGE SCHOOL CHILDREN

Variable		Coefficient	t-stat
Constants	General constant	-3.139	-1.08
	Tucson constant	-0.503	-0.88
Age Group	Age 6 to 9 years	-0.531	-1.23
	Age of 10 years or older (reference)		
Household Income	$75,000 to $99,999	-0.209	-0.43
	$100,000 or More	-1.290	-2.15
Household Composition	Presence of Another Pre-driving Age School child	-0.927	-2.17
	All Adults are Workers	-1.155	-2.38
	Most Educated Person has Bachelor's or higher degree	1.743	3.46
Household Location & accessibility	Rural Household	0.243	0.48
	Accessibility to School Locations (Logged)	-0.034	-0.13
Goodness-of-fit stats	Number of Observations	1163	
	Likelihood with Constants only	-128.283	
	Final likelihood	-115.2731	
	Rho-Squared (0):	0.857	
	Rho-Squared (constants):	0.1014	

5.1.8. Findings and Behavioral Interpretations for Pre-driving Age School Children

The following main findings and behavioral interpretations can be mentioned with respect to the factors influencing parents' decision to home-school a school child:

- The overall constant for schooling at home is large and negative which shows that on an average pre-school children are less likely to be home-schooled. There is an additional negative constant for Tucson. Although it is not extremely significant statistically, it indicates on even lower propensity for schooling from home in Tucson.
- Home schooling is less frequent for children between 6 and 9 years of age as compared to older children. This means that and additional share of school children is added with age probably as the result of accumulated problems at school.
- This is a probably a reflection on the fact that it is inconvenient to have one child going to school while another one staying at home.
- School age children are less likely to be home schooled if all adults in the households are workers. This is logical although the causality might be reversed in some cases, i.e. one of the household adults may decide to leave work in order to take care of a child who cannot go to school.
- School age children are less likely to be home schooled for higher-income households ($75,000 or more). This could be reflective of the lifestyle or busy

schedule of parents in a high income group household. Also, it may be easier for higher income group households to afford private schools or special schools for children with special needs.

- In rural locations, the school age children are more likely to stay at home. It could be due to the fact that there are fewer schools in the rural areas and the distances to school locations are longer as compared to in urban areas. In the same vein, with a better accessibility to school locations, the school age children are less likely to be home-schooled. This effect will be revised at Phase 2 since in its current form, it is not statistically significant and is based on a flat area-type dummy. Unfortunately, the most logical variable in this regard – accessibility to schools – also proved statistically insignificant.

5.2. CHOICE OF SCHOOL LOCATION ZONE

5.2.1. Choice Structure

The school location choice model predicts the usual school location (TAZ) for all students who are not schooling from home. The model was estimated separately for the Phoenix region as a multinomial logit with size variables using the ALOGIT software. The analogous model for the Tucson region will be estimated at Phase 2 after the data on school enrollment have been prepared. This model is applied early in the model chain together with the usual workplace choice model for workers as part of long-term decisions that affect mid-term decisions (for example household car ownership) and all day-level travel choices. This model is segmented by four student types: 1=preschoolers, 2=kindergarten to 8th grade (referred to as elementary school for simplicity), 3=9th to 12th grade (referred to as high school for simplicity), 4= non-ASU university & college students. The model for university location choice may be needed depending on the final technical details of the residential location choice model for university students described above. One of the possible solutions is to apply the residential location choice model for major universities (ASU in Phoenix and UA in Tucson) while the university location model would be applied to all other university students including community colleges. The usual school destination choice model includes the following main explanatory variables: OD accessibilities (mode & time-of-day choice logsums) and additional distance-decay terms that characterize the travel impedance between the residential zone and school zone; zonal characteristics that are used in the attraction (size) variable as school enrollment, relevant employment in education sector (NAICS code 61), population; school attendance area (elementary and high school students in general choose schools only within the attendance areas), as well as household and student characteristics. Structurally the school location choice model is similar to the workplace location choice model described above. However, the school location choice model has different attraction size variables that had to be estimated for some student types and not assumed *a priori* (as employment for workplace choice).

5.2.2. Estimation Dataset

In the NHTS 2008 add-on, there are 1,229 observed student records with available usual school location in the Phoenix region. It included 192 preschoolers, 563 kindergarten to 8th graders, 307 9th -12th graders and 167 university students. Table 26 below shows the distribution of students in the surveyed households by income group and person characteristics.

TABLE 26: DISTRIBUTION OF STUDENTS BY TYPE

Person & household characteristics	Preschool		K to 8th		9th to 12th		University	
	Num.	%	Num.	%	Num.	%	Num.	%
Age:								
0 to 3 years	88	46%						
4 to 6 years	104	54%	68	12%				
7 to 15 years			495	88%				
Driving Age (16-17)					146	48%		
Under 25 years							82	49%
26 years or older							85	51%
If Person Works							99	59%
Does not work							68	51%
Income Group:								
Less than 30K	15	8%	52	9%	26	8%	18	11%
30K to 60K	34	18%	113	20%	59	19%	25	15%
60K to 100K	36	19%	82	15%	49	16%	33	20%
100K to 150K	43	22%	131	23%	57	19%	32	19%
More than 150K	61	32%	162	29%	106	35%	52	31%
Missing	3	2%	23	4%	10	3%	7	4%
Total	192	100%	563	100%	307	100%	167	100%

Since, there is a large number of destination alternatives it was not possible to include all alternatives in the estimation dataset. A sampling-by-importance approach (similar to the approach used for work location choice model) was used to choose alternatives set for each student. Each record was duplicated 20 times and different choice sets with 30 alternatives each were selected based on the size term and distance. This approach is roughly equivalent to selecting 20×30=600 alternatives for the choice set.

The survey observations were joined with alternative-specific variables including OD-accessibilities (mode & time-of-day choice logsums), distance, enrollment by student type, population, and employment in education sector to create the estimation file.

5.2.3. Model Segmentation and Main Explanatory Variables

The model is fully segmented by the following four student types:

- Preschoolers of age 0-6,
- Kindergarten to 8th grade school children,
- 9th grade to 12th grade, school children,
- University and college students.

The following variables were examined and proved to be significant in the utility functions:

- OD accessibility measure between the home and potential school location in a form of composite mode & time-of-day choice logsum
- Additional distance-decay functions between the home and potential school location with the following terms:
  - Linear distance,
  - Log of distance (defined as log(1+distance)),
  - Square root of distance,
  - Distance squared,
  - Distance cubed,
- Household income group interacted with distance terms (i.e. used for segmentation of the distance terms by income):
  - Low income (less than $49,999),
  - Medium income ($50,000 - $99,999),
  - High income ($100,000 and more),
- Components of the zonal attraction (size) variables:
  - School enrollment by category (elementary/middle and high),
  - University and college enrollment,
  - Population,
  - Employment use in the preschool model as a proxy for possible day care locations including Retail (NAICS Codes 44 and 45) and Health Care/Social Assistance (NAICS code 62).
  - Education Services Employment (NAICS code 61) used a s a proxy for small colleges and business school for which enrollment was not available.
- School attendance area available for kindergarten to 12th graders; this variable allows for narrowing the set of relevant locations significantly based on the household residential zone since the absolute majority of school children go to schools within this area.
- Person characteristics interacted with distance terms and size variables:
  - Age Category,
  - Work Status,
- Household composition variables (like presence of a nonworking adult in the household) that may affect school or day care location choice.

5.2.4. Utility Structure

The utility (U_{ijn}^{k}) of choosing a school location (j) for a student (n) of type (t) residing in zone (i) can be written in the following general form:

$$U_{ijn}^{t} = \ln\left(S_{j}^{t}\right) + \alpha^{t} \times L_{ij}^{t} + \sum_{k}\left(\delta^{tk} \times A_{ij}^{k}\right) + \sum_{mz}\left(\beta^{tmz} \times D_{ij}^{m} N_{n}^{z}\right) + \omega_{ij}^{t} + C_{jn} \qquad \textbf{Equation 12}$$

where:

S_{j}^{t}	=	size variable for location zone (*j*) for student type (*t*) ,
L_{ij}^{t}	=	mode & time-of-day choice logsum between zone pair (*ij*) for type (*t*),
A_{ij}^{k}	=	accessibility terms by type (*k*),
D_{ij}^{m}	=	distance terms by type (*m*) (linear, log, squared, cubed and square root),
N_{n}^{z}	=	indicator on person/household characteristic (*z*) for individual (*n*),
ω_{ij}^{t}	=	large negative penalty if zone pair (*ij*) is not in the same attendance area,
C_{jn}	=	correction term introduced to compensate for the sampling error,
$\alpha^{t}, \delta^{tk}, \beta^{tmz}$	=	coefficients to estimate.

The correction term is introduced to compensate for the sampling error in the model estimation (i.e. represent the difference between the sampling probability and final estimated probability for each alternative). This term was explained in the section on workplace choice location above in detail.

The attraction size variable (S_{j}^{t}) for zone (*j*) and student type (*t*) is in itself a linear combination of different zonal variables (*d*) such as relevant enrollment by school type, relevant employment by type, and population (S_{jd}^{t}). Interaction of the zonal variables with person/household characteristics is also considered; this results in a finer segmentation than by student type alone, allowing index (*t*) to take more than four values. This results in the following general form for the attraction size variable:

$$S_{j}^{t} = \log(S_{j1}^{t} + \sum_{d>1}\gamma_{d}^{t} \times S_{jd}^{t}) \qquad \textbf{Equation 13}$$

The coefficients (γ_{d}^{t}) on the size terms are estimated simultaneously with the other coefficients in the utility function. They are constrained to be positive in the estimation process for theoretical consistency. One of the size terms is arbitrarily chosen as the reference case with the coefficient set to 1.

A combination of distance-decay terms used in the utility is chosen in such a way that the composite distance-decay component of the utility function is monotonically decreasing within the maximum chosen school distance range. Table 27 shows the observed distance distribution from home to school location for each student type in the dataset. It can be seen that for all student categories the tail of the distribution becomes very “thin” after 20-25 miles.

TABLE 27: OBSERVED DISTANCE TO USUAL SCHOOL LOCATION BY STUDENT TYPE

Distance, miles	Student type			
	1=Preschool child	**2=K-to-8th grader**	**3=9th-to-12th grader**	**4=University student**
0-5	43	439	227	54
5-10	9	59	48	51
10-15	2	20	17	30
15-20	2	15	8	13
20-25	0	13	4	7
25-30	1	5	0	7
30-35	0	3	1	2
35-40	0	1	0	1
40-45	0	4	0	0
45-50	0	3	0	1
50-55	0	0	0	0
55-60	0	0	0	0
60-65	0	0	0	0
65-70	0	0	2	0
70-75	0	1	0	1
75+	0	0	0	0

5.2.5. Model Estimation Results for Preschool Children

Table 28 shows the estimation results obtained for usual school location choice results for pre-school children.

TABLE 28: ESTIMATION RESULTS FOR USUAL SCHOOL LOCATION CHOICE MODEL FOR PRE-SCHOOL CHILDREN

Variable		Coefficient	t-stat
Size variable	Employment specific to preschool	1.000	
	Population	0.229	-17.04
Generic travel impedance	Mode & time-of-day logsum	0.588	7.07
	Linear Distance	-0.133	-6.02
	Logged Distance	-0.766	-4.09
	Distance Square	0.001	3.28
Additional effect on travel impedance if non-worker or retiree is in household	Linear Distance	-0.065	-7.75
Goodness-of-fit stats:	Observations	3840	
	Likelihood with Constants only	-12839.115	
	Final log likelihood	-7459.715	
	Rho-Squared (w.r.t. 0)	0.4247	
	Rho-Squared (w.r.t. constants)	0.4190	

5.2.6. Main Findings and Behavioral Interpretations for Preschool Children

Below are the following main findings and corresponding behavioral interpretations:

- The coefficient on mode choice logsum is positive and in the unit interval as required by theory because it is equivalent to a nesting coefficient in a multi-level nested logit model with time-of-day & mode choice at the lower level and school location choice at the upper level.
- The size variable is a linear function of employment (specific to preschool model, i.e. including retail and person care services) as the main component and population as a complementary one. The coefficient on employment is constrained to 1.0 and the population coefficient was estimated relative to it. The coefficient on population is smaller than the employment coefficient, which shows that employment is the main attraction for choosing locations of preschool institutions. This combination of employment and population is just a proxy for day care enrollment. However, this variable is difficult to collect for the base year and predict in the future, due to the large number of small day care institutions.
- A composite distance-decay factor was specified as a combination of linear, logged, and squared distance terms with different estimated coefficients. Linear distance was also interacted with a variable indicating the presence of a non-worker or a retiree in

the household. Figure 16 shows the resulted composite distance-decay factor in utility units for the base case (i.e. without a non-worker or retiree in the household) and for households with non-worker or retiree. Households with a non-worker or retiree are more sensitive to distance for pre-school children. The possible explanation is that in households where all adults are workers, the parents might escort the preschooler to day care facilities or kindergartens that are closer to the work location (or on the way to work). Non-workers would probably tend to choose day care locations closer to home.

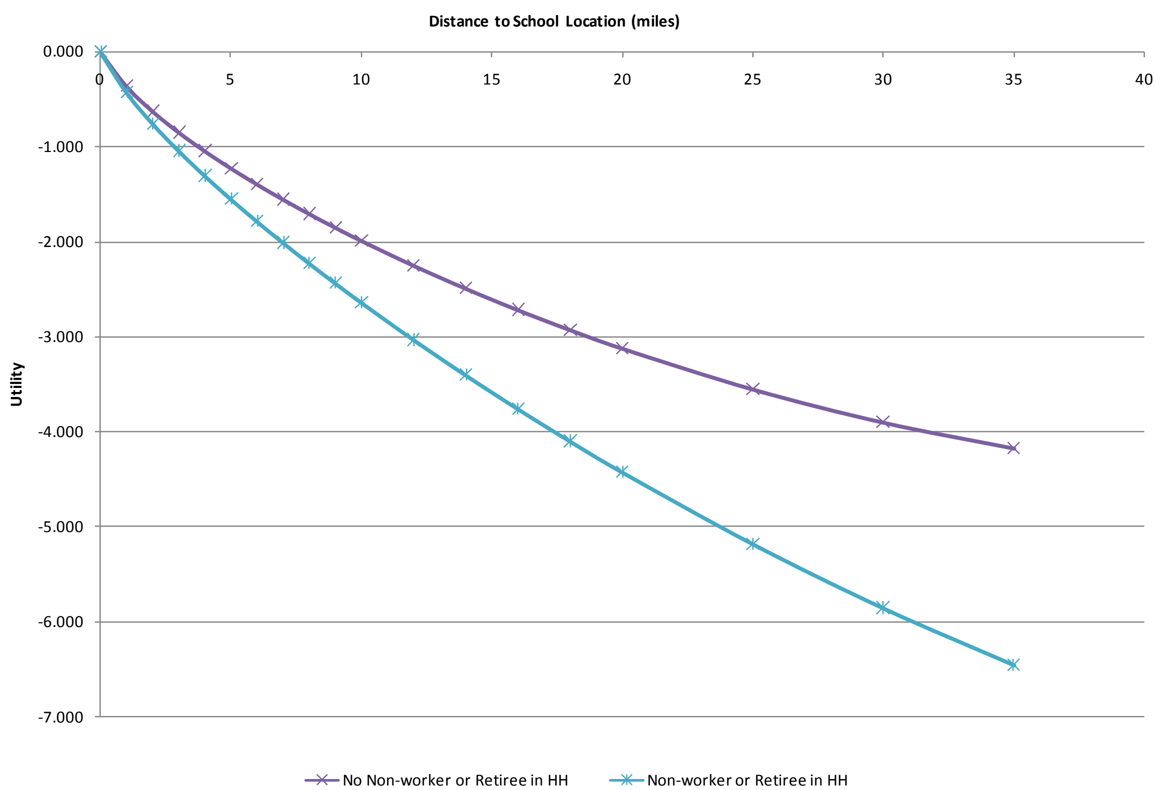

FIGURE 16: DISTANCE DECAY FACTORS FOR PRESCHOOL CHILDREN

5.2.7. Model Estimation Results for Kindergarten-to-8th Grade (K8)

Table 29 below shows estimation results for the school destination choice model for kindergarten through 8th grade students. The explanatory power of this school location model is very high – a rho-squared of 0.975. This could be because school locations for kindergarten-to-8th grade are close to home and restricted by school attendance area, so there are only a few alternatives available in addition to the chosen alternative.

TABLE 29: ESTIMATION RESULTS FOR USUAL SCHOOL LOCATION CHOICE MODEL FOR K8 CHILDREN

Variable		Coefficient	t-stat
Size variable:	School Enrollment for Kindergarten to 8th Grade	1.0000	
	Population	0.0233	-23.40
Generic travel impedance	Mode & time-of-day choice logsum	0.2295	0.66
	School is not in attendance area	-99.0000	
	Linear Distance	-1.1869	-3.56
	Logged Distance (Log(distance+1))	1.9454	1.39
	Distance Square	0.0494	4.18
	Distance Cube	-0.0007	-6.46
Additional impedance effect if age under 12 (elementary school)	Linear Distance	-0.0230	-0.40
Additional impedance effect if Bachelor's or higher degree holder in household	Linear Distance	0.1368	2.08
Goodness-of-fit stats:	Observations:	11180	
	Likelihood with Constants only:	-37235.252	
	Final log likelihood:	-910.1868	
	Rho-Squared (0):	.9757	
	Rho-Squared (constant):	.9756	

5.2.8. Main Findings and Behavioral Interpretations for Kindergarten-to-8th (K8) School Children

Below are the following main findings and corresponding behavioral interpretations:

- The coefficient on mode & time-of-day choice logsum is positive and within the unit interval as required by theoretical considerations.
- The size variable includes school enrollment and population. The coefficient for school enrollment is constrained to 1.0 as it represents the main size term. The coefficient on population is small compared to enrollment and it is kept as a proxy for some unknown private and smaller schools.
- There were no observed cases where students would go to school outside of the home school attendance area. Thus, the option of choosing a school outside of the home school attendance area was made unavailable by setting a large negative penalty. In

model application, this penalty can be somewhat relaxed to account for children going to private schools (possibly) outside the public school attendance area. Some additional data on private schools will be needed to justify this relaxation.

- A composite distance-decay function was specified as a combination of linear, log, squared and cubed distance terms with different estimated coefficients. Linear distance was also interacted with age (under 12 years which distinguishes between elementary grade and middle school grade) and households with bachelors or higher degree holders and some interesting specific effect were found. Logically, elementary school students are less likely to travel longer for school as compared to middle school students. Additionally, students from households with a generally higher level of education are less sensitive to distance to school. It might be a reflection that parents with higher education would be willing to send children to preferred schools even if a longer distance is involved. It also can be a manifestation of the fact that more educated families live in less dense areas and therefore have a longer distance to school.
- Figure 17 shows the distance-decay factor for K8 students with the additional specific effects. Younger children of age under 12 who attend an elementary school are more sensitive to distance and have a function that decreases stronger with distance compared to older children. Presence of bachelor's or higher degree holder in the household results in a slightly less sensitive function although this effect is minor compared to the age effect. All functions are monotonously decreasing (with only a small deviation for distances between 20 and 30 miles where they become practically flat) and the utility steeply decreases after 25 miles essentially making the choice of school father away unavailable.

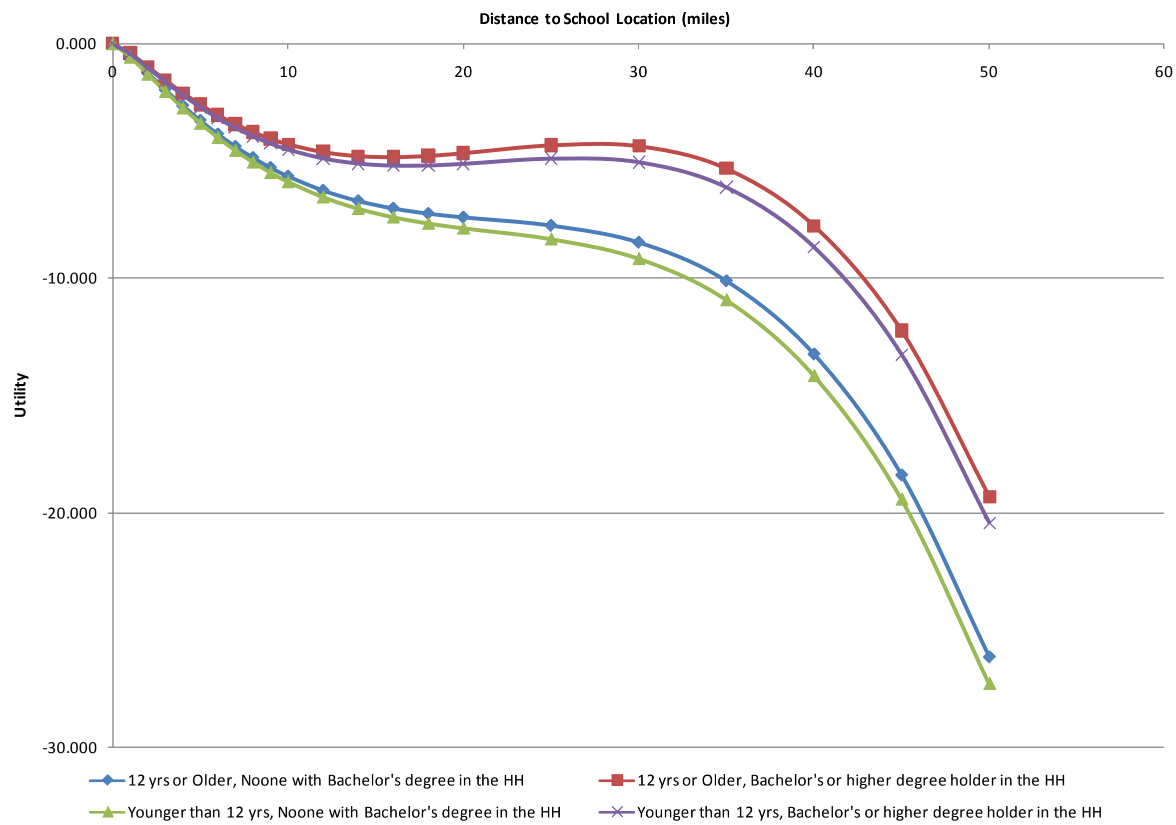

FIGURE 17: DISTANCE DECAY FUNCTIONS FOR K8 SCHOOL CHILDREN

5.2.9. Model Estimation Results for 9th-to-12th Grade School Children

The estimation results for school location choice model for 9th-to-12th grade students are shown in Table 30. The explanatory power of this school location models is also quite high (like the K8 model) – a rho-squared of 0.90.

TABLE 30: ESTIMATION RESULTS FOR USUAL SCHOOL LOCATION CHOICE MODEL FOR 9TH TO 12TH GRADE SCHOOL CHILDREN

Parameter		Coefficient	t-stat
Size variable	High School Enrollment	1.0000	
	Population	0.0129	-60.27
Generic travel impedance	Mode Choice Logsum	0.0575	0.75
	School is not in attendance area	-99.0000	
	Linear Distance	-0.4826	-10.03
Additional impedance effect is student is of driving age	Linear Distance	0.1264	3.04
Additional impedance effect if Bachelor's or higher degree holder in household	Linear Distance	0.1082	2.34
Goodness-of-fit stats	Observations:	6000	
	Likelihood with Constants only	-19438.9141	
	Final log likelihood	-1851.2125	
	Rho-Squared (0)	0.9057	
	Rho-Squared (constant)	0.9048	

5.2.10. Main Findings and Behavioral Interpretations for 9th to 12th Grade School Children

Below are the following main findings and corresponding behavioral interpretations:

- The mode and time-of-day choice logsum coefficient was affected significantly and was taking illogical negative values with addition of distance terms other than linear distance in the model. That is why only linear distance was used instead of a composite distance-decay function.
- The size variable includes school enrollment and population. The coefficient for school enrollment is constrained to 1.0 as it represents the main component. The coefficient on population is very small (even compared to the similar K8 model described above) and is a proxy for missing private and small schools.

- There were no cases where students go to school outside of the home school attendance area. Thus, the option of choosing a school outside of the home school attendance area was made unavailable by setting a large negative penalty.
- Linear distance was also interacted with such variables as driving age (16-17) for the student and presence of a bachelor or higher degree holder in the household. Driving age students are likely to travel longer distances for school as compared to pre-driving age students since they already can drive. Students from households with a generally higher education level are less sensitive to distance to school. As in the case for K8 described above, it might reflect on that parents with higher education are more selective about the schools and would be willing to send a child to a preferred school even if it is farther away from home.
- Figure 18 shows the combined effects of student's age and maximum education level in the household on distance to school for 9th to 12th grade students.

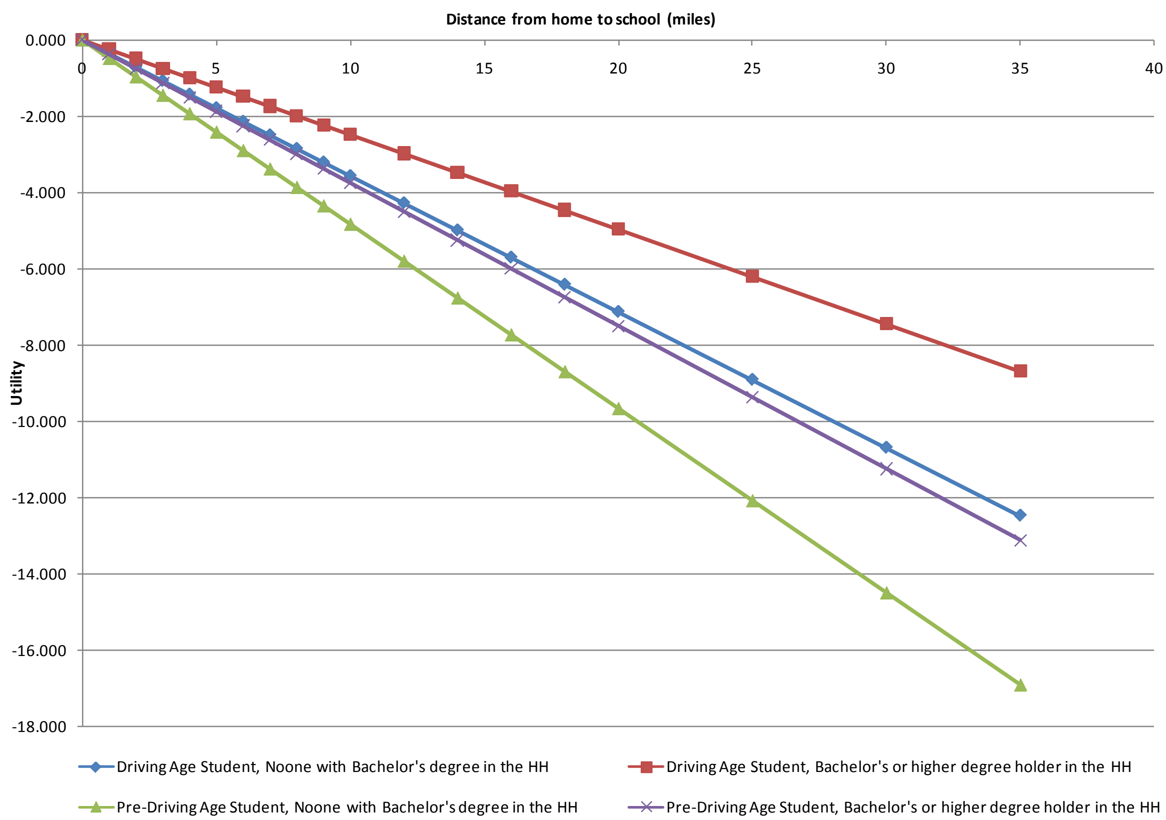

FIGURE 18: DISTANCE-DECAY FUNCTIONS FOR 9TH TO 12TH GRADE SCHOOL CHILDREN

5.2.11. Model Estimation Results for University Students

The estimation results for school location choice model for university and college students are shown in Table 31.

TABLE 31: MODEL ESTIMATION RESULTS FOR USUAL SCHOOL LOCATION CHOICE FOR UNIVERSITY STUDENTS

Variable		Coefficient	t-stat
Size variable	University Enrollments	1.0000	
	Educational Services Employment – typical student*	0.5472	-3.62
	Educational Services Employment – not typical student	3.4032	16.24
	Other College Enrollment - Income $99,999 or Less	2.1885	10.71
	Other College Enrollment – Income $100,000 or More	0.8372	-1.63
Generic travel impedance	Mode & time-of-day choice logsum	0.6655	3.74
	Logged Distance (log(Distance+1))	1.9520	8.22
	Square Root of Distance	-2.3131	-14.63
	Distance Squared	0.0007	4.89
Additional impedance effect if student works	Linear Distance	0.0406	6.84
Goodness-of-fit stats	Observations	3340	
	Likelihood with Constants only	-9759.637	
	Final log likelihood	-7729.123	
	Rho-Squared (0)	0.2208	
	Rho-Squared (constant)	0.2081	

* 25 years or younger, and does not work

5.2.12. Main Findings and Behavioral Interpretations for University Students

Below are the following main findings and corresponding behavioral interpretations:

- The coefficient on mode & time-of-day choice logsum is positive and in the unit interval as requested by the theory. Due to a high correlation between the logsum and distance terms, the linear distance term was eventually dropped from the composite function.
- The size variable is a linear function of university enrollment, other college enrollment and educational services employment. The coefficient on university enrollment is constrained to 1 and other coefficients are estimated relative to this. The impact of enrollment is further segmented by income group ($100K+ or more vs. less than $100K), work status (work or not) and age group (25 or younger vs. older than 25). A typical student was defined based on work status (does not work) and age of 25 years or less.

- Logically, other colleges are found to be more attractive than major universities for lower-income students (less than $100,000 of household income) as compared to other students. Educational services employment served a proxy for (smaller) schools other than universities and colleges. These include small colleges, training schools, business schools, driving schools, or other evening classes. These schools are less attractive for typical students (25 years or younger who do not work) and more attractive to older students and/or workers.
- A composite distance-decay function has been specified as a combination of logged, square root, and squared distance terms with different estimated coefficients. This function is monotonously decreasing within the maximum observed distance between home and school. The distance effect was further segmented by work status (work or not). Figure 19 shows the distance-decay factor for both segments within the maximum observed home to school distance range. Interestingly, working students have a relatively higher tolerance to longer distances. This result that is seemingly counter-intuitive can be explained by the fact that for many working students, university is also their workplace (see analysis of the ASU students in the section on student residential and employment choices above).

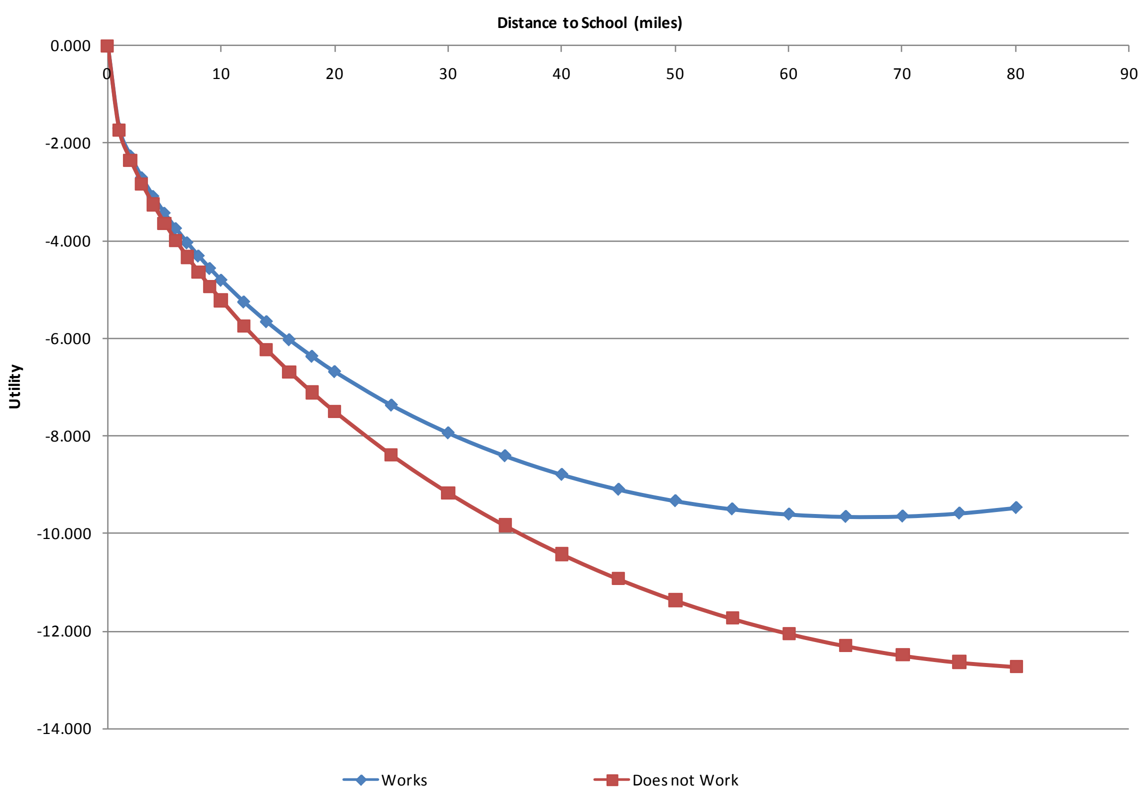

FIGURE 19: DISTANCE-DECAY FUNCTIONS FOR UNIVERSITY STUDENTS

CHAPTER 6.

CHOICE OF INDIVIDUAL MOBILITY ATTRIBUTES

6.1. GENERAL CHOICE FRAMEWORK AND DIMENSIONS

The design proposed for the MAG ABM includes a set of special sub-models for different person and household mobility attributes including household car ownership, person transit pass, person free parking eligibility, and person toll transponder. The ABM system is structured such that mobility attributes are conditional upon the residential location, household composition, and long-term choices of work and school locations, and precede short-term decisions such as frequency of daily travel.

Mobility attributes have a subsequent effect on the day-level and tour-level decisions and represent powerful explanatory variables for mode choice and in some cases, for destination choice. Mobility attributes are also important to model certain policies such as transit pass subsidies.

Unfortunately, the NHTS 2008 did not collect information regarding all mobility attributes that were needed. The survey only provided information about household car ownership and transit pass use if a transit trip was actually made on the assigned survey day. It did not collect information on free parking eligibility, and there are currently no facilities which require toll transponders in either the Phoenix and the Tucson regions. With respect to free parking eligibility and toll transponder models, we recommend borrowing choice models developed elsewhere; we have created placeholders in the MAG ABM structure for these models for the time being. We also suggest additional surveys to consider parking choice (a good prototype survey is currently being implemented in San Diego to support the ABM development).

In the subsequent section we present a model that jointly predicts both car ownership (for the household) and transit pass ownership (for each person in the household). Since transit pass ownership was deduced from the transit trip records in NHTS 2008, the model may have a systematic bias with respect to transit pass ownership. In particular many more people may have some type of transit pass but not use transit on the survey day; this may be more common for non-workers and for non-mandatory trips in general. We believe, however, that we capture most of the transit commuters (workers and students) properly. The model constants for transit pass can be calibrated to achieve a target number of transit pass holders in the region that can be derived from the recent transit on-board survey. We reserve this issue for Phases 2 and 3 that include a more substantial model calibration effort.

6.2. JOINT MODEL OF HOUSEHOLD AUTO OWNERSHIP AND PERSON TRANSIT PASS

6.2.1. Choice Structure

The Household Auto Ownership and Person Transit Pass model predicts the number of autos (including motorcycles, vans, and trucks for personal use) available to a household and transit pass ownership (approximated by transit use for at least one trip on the given day) for up to four adults in the household simultaneously. The model was estimated in a nested logit form using the ALOGIT software. In this model, household car ownership and transit pass ownership are dependent variables derived from the activity needs of the household based on household characteristics, characteristics of persons within the household, and travel environment measured by various accessibility measures described above. This model is applied after the work, university, and school location choices have been made. Consequently it includes auto, transit, and non-motorized OD accessibilities to the known mandatory activities locations (at the person level) and general zonal accessibilities to non-mandatory activities (at the household level by place of residence) as explanatory variables.

The model is formulated as a joint choice of household auto ownership and person transit use. Five auto ownership alternatives (0 cars, 1 car, 2 cars, 3 cars, and 4 or more cars) and two person transit use alternatives (Yes, No) were defined. The joint choice structure has the following 150 combinatorial alternatives with 5 auto ownership choices for each household, combined with 2 transit pass use choices for each adult in the household as shown in Table 32.

TABLE 32: COMBINED CAR-OWNERSHIP AND TRANSIT PASS ALTERNATIVES

Number of adults in household	Number of alternatives		
	Car ownership	Transit pass	Combined
1	5	2	5×2=10
2	5	2×2=4	5×4=20
3	5	2×2×2=8	5×8=40
4	5	2×2×2×2=16	5×16=80
Total			150

The first 10 alternatives are only available to a one-adult household, next 20 are available only to two-adult households, next 40 are available to three-adult households, and the last 80 are available to four-or-more-adult households.

The model was first nested based on auto-ownership categories and then the lower-level alternatives that that relate to person transit pass holding were nested within the auto-ownership alternatives as shown below in Figure 20 for an example of 2-adult household. At the upper level of the nested structure, the auto-ownership choices are split between zero cars and one or more cars. The choice of owning a car or not owning one is the most significant auto ownership decision and is, therefore, placed at the highest level in the nested structure. At the next level, the choice of one or more cars is further split into 1 car and 2 or more car choices.

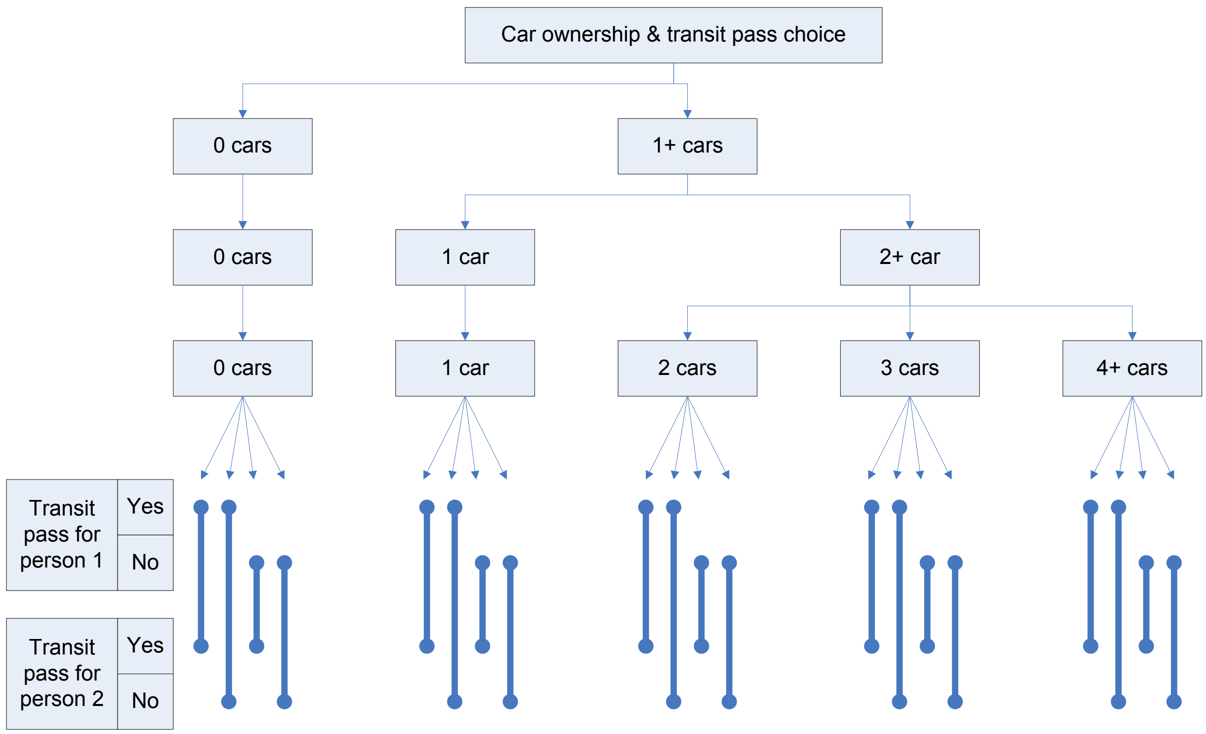

FIGURE 20: NESTED STRUCTURE OF JOINT CHOICE OF AUTO OWNERSHIP AND TRANSIT PASS (2-ADULT HOUSEHOLD)

6.2.2. Relative Car Sufficiency

Car sufficiency is an important measure that relates the number of cars to number of drivers or workers in a household. This measure is used in the subsequent sections for model segmentation as well as formation of alternative-specific variables for car ownership. For example, most of the household composition variables are stratified using relative car sufficiency rather than car ownership. More specifically, car sufficiency is calculated as the difference between number of cars in the alternative and the number of drivers in a household. Car-sufficiency is set to insufficient, sufficient, and over sufficient if the value is negative, zero, and positive respectively, for each car-ownership alternative, depending on the number of drivers or workers in the household. Also, households with zero cars are singled out as a special category. The car sufficiency categories are illustrated in Table 33 below. The choice of car-sufficiency base (drivers or workers) depends on the modeling context and both definitions can be used in the same model.

TABLE 33: RELATIVE CAR SUFFICIENCY

No. of drivers (or workers)	Auto-ownership alternatives				
	0 cars	1 car	2 cars	3 cars	4+ cars
1	Zero cars	Sufficient (cars equal to drivers)	Over-sufficient (cars greater than workers)	Over-sufficient (cars greater than workers)	Over-sufficient (cars greater than workers)
2	Zero cars	Insufficient (cars fewer than drivers)	Sufficient (cars equal to drivers)	Over-sufficient (cars greater than workers)	Over-sufficient (cars greater than workers)
3	Zero cars	Insufficient (cars fewer than drivers)	Insufficient (cars fewer than drivers)	Sufficient (cars equal to drivers)	Over-sufficient (cars greater than workers)
4+	Zero cars	Insufficient (cars fewer than drivers)	Insufficient (cars fewer than drivers)	Insufficient (cars fewer than drivers)	Sufficient (cars equal to drivers)

6.2.3. Estimation Dataset

The estimation dataset included 4,452 households from the NHTS 2008 that were surveyed on a regular weekday and have a completed travel diary for all adult household members. Table 34 below summarizes the surveyed households by number of adults in household, auto ownership and number of adult transit pass users. The table shows a logical relationship between car ownership and transit pass use. In particular, zero-car households are logically characterized by the highest percentage of transit pass users. The survey records were joined with a wide range of pre-calculated mandatory and non-mandatory accessibility measures to form the estimation dataset.

Several additional statistics for the four household car-sufficiency groups with respect to workers are presented in Table 35 below. The two-way household distributions are tabulated with car-sufficiency groups crossed with other household variables one at a time to explore the impacts of these variables. It can be seen that household income is strongly correlated with car sufficiency. Specifically, low-income households have the highest share of zero-car cases. Also, interestingly, the presence of a preschool child is correlated with higher car sufficiency (presumably to ensure the child care).

Mandatory and non-mandatory activity accessibilities, as was described above in detail, are logsum measures calculated for mode & time-of-day choice and destination choice models. Mandatory OD accessibilities reflect the actual workplace and/or school location for each worker and student in the household, while non-mandatory zonal accessibilities reflect the general accessibility of the household to all potential non-mandatory attractions from the residential zone.

TABLE 34: NUMBER OF ADULT TRANSIT PASS USERS AND AUTO OWNERSHIP

Number of Adults in Household	Number of Cars	Number of Adult Transit Pass Users					Total
		0	1	2	3	4+	
1	0	123	28	0	0	0	151
	1	996	15	0	0	0	1,011
	2	136	1	0	0	0	137
	3	31	0	0	0	0	31
	4+	13	1	0	0	0	14
	Total	1,299	45	0	0	0	1,344
2	0	22	9	3	0	0	34
	1	515	18	4	0	0	537
	2	1,465	23	5	0	0	1,493
	3	360	12	1	0	0	373
	4+	136	0	1	0	0	137
	Total	2,498	62	14	0	0	2,574
3	0	5	0	3	1	0	9
	1	31	5	0	0	0	36
	2	111	16	1	0	0	128
	3	150	8	2	0	0	160
	4+	70	0	0	0	0	70
	Total	367	29	6	1	0	403
4	0	1	0	0	0	0	1
	1	6	2	0	0	0	8
	2	12	2	1	1	0	16
	3	34	5	1	0	0	40
	4+	63	2	1	0	0	66
	Total	116	11	3	1	0	131
Total	0	151	37	6	1	0	195
	1	1,548	40	4	0	0	1,592
	2	1,724	42	7	1	0	1,774
	3	575	25	4	0	0	604
	4+	282	3	2	0	0	287
	Total	4,280	147	23	2	0	4,452

It can be seen that in general car ownership and transit pass ownership are negatively correlated. For example, there are very few cases of transit pass holding for households where number of cars is greater than or equal to number of workers.

TABLE 35: OBSERVED STATISTICS FOR HOUSEHOLD CAR SUFFICIENCY

Household category	Number of Households by Car Sufficiency					Percentage of Households by Car Sufficiency				
	Zero Cars	Cars fewer than workers	Cars equal to workers	Cars greater than workers	Total	Zero Cars	Cars fewer than workers	Cars equal to workers	Cars greater than workers	Total
By Household Income:										
Less than $24,999	116	14	128	550	808	14%	2%	16%	68%	100%
$25,000 to $49,999	31	26	321	836	1,214	3%	2%	26%	69%	100%
$50,000 to $74,999	8	19	237	491	755	1%	3%	31%	65%	100%
$75,000 to $99,999	3	3	203	365	574	1%	1%	35%	64%	100%
$100,000 or more	2	7	233	502	744	0%	1%	31%	67%	100%
Unknown	35	3	59	260	357	10%	1%	17%	73%	100%
Total	195	72	1,181	3,004	4,452	4%	2%	27%	67%	100%
By Number of Workers:										
0	167	0	0	1,642	1,809	9%	0%	0%	91%	100%
1	24	0	552	939	1,515	2%	0%	36%	62%	100%
2+	4	72	629	423	1,128	0%	6%	56%	38%	100%
Total	195	72	1,181	3,004	4,452	4%	2%	27%	67%	100%
By Number of Driving-Age Persons:										
1	151	0	382	811	1,344	11%	0%	28%	60%	100%
2	34	49	684	1,807	2,574	1%	2%	27%	70%	100%
3+	10	23	115	386	534	2%	4%	22%	72%	100%
Total	195	72	1,181	3,004	4,452	4%	2%	27%	67%	100%
By Number of Preschool Children:										
0	192	59	1,047	2,786	4,084	5%	1%	26%	68%	100%
1	3	8	94	129	234	1%	3%	40%	55%	100%
2+	0	5	40	89	134	0%	4%	30%	66%	100%
Total	195	72	1,181	3,004	4,452	4%	2%	27%	67%	100%
By Number of Pre-driving Age School Children:										
0	187	59	929	2,614	3,789	5%	2%	25%	69%	100%
1	6	9	145	240	400	2%	2%	36%	60%	100%
2+	2	4	107	150	263	1%	2%	41%	57%	100%
Total	195	72	1,181	3,004	4,452	4%	2%	27%	67%	100%

6.2.4. Main Explanatory Variables

The following variables have been examined and proved to be significant in the utility functions:

- Household Car sufficiency w.r.t Workers,
- Tucson dummy (both for car ownership and transit pass use),
- Household composition variables:
  - Number of driving-age household members
  - Presence of Preschool Children and Pre-driving Age school Children
  - Ratio of workers (full time and part time) to driving age household members
  - Ratio of pre-driving age school children to driving age household members
  - Ratio of retirees under age 80 to driving age household members
  - Ratio of retirees of age 80 and older to driving age household members
- Household income group:
  - Low income (less than $25,000)
  - Medium Low income ($25,000-$49,999)
  - Medium income ($50,000-$74,999)
  - Medium High income ($75,000-$99,999)
  - High income ($100,000 or more)
- Zonal accessibility indices from residential zones to potential destinations:
  - Non-motorized accessibility to non-mandatory activities
  - Difference between auto accessibility and transit accessibility to non-mandatory activities
- Zonal density indices:
  - Population density
  - Retail employment density
- Household Dwelling Type:
  - Detached Home
  - Non-detached Home
- Residential area type:
  - Urban (1 million population)
  - Rural
- Household mandatory activity auto dependency indices (explained below) :

  - Workers' mandatory activity auto dependency
  - Students' mandatory activity auto dependency
- Person dummies in transit pass use utility:
  - Person Type 1-5
  - Gender
  - Age Group
  - Work location type (work from home vs. work out of home).

The zonal accessibility measures for non-mandatory activities are in the form of destination choice logsums and represent a result of summation of the corresponding attractions across all destinations weighted by travel impedance. For this model, they are specifically segmented by mode (auto, transit, and walk) as explained above in the section on accessibility measures.

The household mandatory activity auto dependency variable is calculated using the difference between the mode choice logsum for auto-dependent modes (SOV, HOV and Drive-to-Transit) and the mode choice logsum for auto-independent modes (walk to transit and non-motorized), stratified by person type (worker versus student). The logsums are computed based on the household TAZ and the work TAZ (for workers) or school TAZ (for students). The household auto dependency is obtained by summing individual auto dependencies of each person type (worker versus student) in the household.

6.2.5. Utility Structure

The utility function ($U_{ct_1t_2...t_sn}$) for a combined car-ownership (c) and transit-pass-use alternative ($t_p = 0,1$) for each adult member (p) of household (n) can be written in the following general way:

$$U_{ct_1t_2...t_sn} = \sum_k \beta_{cs}^k \times N_n^k + \sum_{p=1}^{s} \sum_m \gamma_{t_p}^m \times N_{pn}^m \qquad \textbf{Equation 14}$$

where:

$p = 1,2,..,s$	=	adult household members,
N_n^k	=	variables (k) for household (n) including accessibilities,
N_{pn}^m	=	variables (m) for person (p) from household (n),
β_{cs}^k	=	coefficients to estimate by car ownership (c) and car sufficiency (cs),
γ_t^m	=	coefficients to estimate for transit pass use,

This form of utility function is based on a parsimonious component wise structure that is employed for all choice models with a large number of combinatorial alternatives. Only 5 household car-ownership components and 5 person-type-specific transit pass ownership components have to be estimated, while the utilities for each of the 150 alternatives are composed of the relevant car-ownership and transit-pass ownership terms (with a small number of interaction terms).

6.2.6. Model Estimation Results

The car ownership estimation results are summarized in Table 36 for coefficients segmented by car ownership and Table 37 for coefficients segmented by cur sufficiency and transit pass use.

TABLE 36: AUTO OWNERSHIP AND TRANSIT USE (MODEL COEFFICIENTS SEGMENTED BY AUTO OWNERSHIP)

Variable	Number of Cars									
	0		1		2		3		4+	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
General constants for car ownership:										
1 Adult	-13.965	-1.12			-1.928	-2.66	-3.109	-4.52	-3.925	-4.76
2 Adults	-15.241	-1.08	-1.151	-2.81			-0.454	-1.27	-2.110	-3.80
3 Adults	-12.155	-1.01	-0.624	-1.11	-0.110	-0.30			-0.339	-0.59
4+ Adults	-14.724	-1.04	-1.355	-1.43	-0.210	-0.42	-0.327	-0.70		
Tucson dummies for car ownership:										
1 Adult	2.002	1.13			0.201	0.79	0.089	0.20	-0.633	-0.93
2 Adults	-0.143	-0.12	-0.332	-1.85			-0.118	-0.91	0.439	2.30
3 Adults	0.465	0.19	0.448	0.84	0.344	1.29			0.141	0.45
4+ Adults			0.696	0.67	0.039	0.06	0.262	0.57		
Income effects on car ownership:										
Less than $24,999	4.451	1.33	1.327	3.53			-0.265	-1.18	-0.808	-2.05
$25,000 to $49,999	0.252	0.19	0.708	2.94			-0.111	-0.79	-0.234	-0.98
$50,000 to $74,999										
$75,000 to $99,999	-0.707	-0.33	-1.042	-3.14					0.032	0.14
$100,000 or More	-2.373	-0.80	-1.091	-2.99			0.243	1.94	0.533	2.63
Dwelling type impact on car ownership:										
Detached Home	-5.375	-1.32	-0.861	-3.39			0.285	1.55	0.361	1.31
Auto Dependency for Mandatory Travel impact on car ownership:										
Auto Dependency for Workers	-2.865	-0.92	-0.247	-1.21			0.090	1.13	0.090	1.13
Nesting coefficients for upper-level car ownership nests:										
Upper level	0.437		0.437							
Lower level			0.771		0.771					

TABLE 37: AUTO OWNERSHIP AND TRANSIT USE (MODEL COEFFICIENTS SEGMENTED BY AUTO SUFFICIENCY AND TRANSIT PASS)

Variable	Car sufficiency	Coefficient	t-stat
Car-ownership impacts:			
Presence of Pre-driving Age (0-15) children	0 cars	-1.1054	-0.75
Ratio of Workers to Driving Age Household Members	0 cars	-5.1125	-1.18
	Cars < Adults	-1.4416	-4.88
Ratio of 65 to 79 years old to Driving Age Household Members	0 cars	-2.7210	-1.11
	Cars > Adults	2.3434	3.98
Ratio of 80+ yrs to Driving Age Household Members	0 cars	0.2483	0.33
	Cars > Adults	-1.0117	-2.98
Non-Motorized Accessibility to Non-Mandatory Attractions from Residential Zone	0 cars	0.9592	1.10
	Cars < Adults	0.0921	2.91
	Cars = Adults		
	Cars > Adults	-0.1496	-4.32
	Cars - Adults = 1	-0.0825	-1.53
	Cars - Adults > 1	-0.1134	-2.09
Auto Accessibility-Transit Accessibility (to Non-Mandatory Attractions from Residential Zone)	0 cars	-0.0736	-0.50
Mandatory Travel Auto Dependency for University Students	0 cars		
	Cars < Workers	-0.7834	-0.83
	Cars = Workers		
	Cars > Workers	0.0653	0.22
Person transit pass use impacts:			
Constant		-5.1024	-18.97
Tucson Dummy		0.1640	1.11
Female with Preschool Child (Person Dummy)		-0.4316	-1.26
Full Time Worker (Person Dummy)		-0.0586	-0.29
Part Time Worker (Person Dummy)		-0.7051	-2.07
University Student (Person Dummy)		0.2245	0.68
Retired (Person Dummy)		-0.3730	-1.69
Driving Age (16+) Student (Person Dummy)		-0.3213	-0.77
Age less than 35 yrs		0.4109	1.88
Household Income Less than $24,999		0.6004	3.66
Transit accessibility to non-mandatory attractions		0.1048	4.91
Interaction between Transit Pass Use and Car Sufficiency	Cars = Adults – 4	1.5859	3.31
	Cars = Adults - 2 Cars = Adults – 3	1.3606	5.90
	Cars = Adults – 1	0.9862	6.01
	Cars = Adults Cars = Adults + 1		
	Cars > Adults + 1	-0.2723	-0.53
Two adults as transit users (dummy for each pair)		1.1313	6.81
Goodness-of-fit stats:			
Number of observations		4242	
Likelihood with constants only		-5219.5836	
Final log likelihood:		-4730.6005	
Rho-Squared w.r.t. Zero		0.6157	
Rho-Squared w.r.t. Constants		0.0937	

6.2.7. Main Findings and Behavioral Interpretations

The following main findings and corresponding behavioral interpretations are suggested for household auto ownership and person transit pass use:

- The number of driving-age household members or adults (age of 16 years and older) has a strong impact on household car ownership. In each household adult category (1, 2, 3, and 4+), the choice that corresponds to a sufficient number of autos for the total number of driving age adults is set as the reference alternative. Insufficient and over- sufficient car ownership alternatives have negative bias constants, and the more over-sufficient or insufficient the alternative, the larger is the negative constant. This finding is consistent with the expectation that people are more likely to have a sufficient number of cars than to have either more or less cars than adults.
- The ratio of workers to driving age household members has a negative coefficient for 0 and insufficient car ownership choices. Households tend to own a sufficient number of cars primarily to ensure that workers have enough cars to commute to work.
- The presence of pre-driving age children has a negative coefficient for zero cars. This shows that households with children need some cars because children frequently have to be escorted to school and other activities.
- Retirees are divided into two groups: 1=under age of 80, and 2=80 and older. The car ownership of these two groups shows opposite patterns. Younger retirees share the same auto ownership patterns as those of the workers. For older retirees, the coefficient is positive for zero cars and negative for alternatives with more cars than adults. Older retirees tend to drive less compared to younger persons.
- Income has a positive impact on car ownership across a wide range of income categories. Logically, higher-income households are more likely to own more cars when compared to lower income households, all else being equal.
- Households living in detached houses are more likely to own more cars. Detached homes are usually located in suburban areas which are accessible by auto only. Additionally, detached houses normally provide large parking space for multiple cars.
- Non-motorized accessibility has a negative influence on household car ownership. This variable represents the ease of travel by walking and biking for non-mandatory purposes. The coefficient is positive for 0 car ownership, suggesting that the more accessible a household is to non-mandatory activities by walking or biking, the more likely the household is to not own any autos. On the other hand, the coefficient is negative for the difference between the auto and transit accessibilities suggesting that households with auto access that is relatively better than transit access to non-mandatory destinations are more likely to own one or more cars. However, this particular effect was not statistically significant.
- The mandatory travel auto dependency variable represents how much household members' work and school tours are dependent on the auto mode. This variable has negative coefficients for 0 or 1 car and positive coefficient for 3 or more cars for

workers' mandatory tour dependency. Also, with growing mandatory tour dependency for students, the household is more likely to own cars more than workers so that there is a car available for student to use.

- Workers and retirees are less likely to own transit passes, whereas university students are more likely to own transit passes. Younger individuals (under 35 years), low income groups or households with better transit accessibility are more likely to own transit passes. Transit pass ownership logically has a negative correlation with car sufficiency with respect to adults. Also, if there is one adult with a transit pass in the household, it is more likely for another adult to also hold a transit pass. The last effect represents an interesting family lifestyle phenomenon.
- The model was developed and estimated for both Phoenix and Tucson regions. It can be seen that most of the Tucson-specific constants are not extremely significant statistically. However, we keep a full set of Tucson-specific constants for calibration purposes.
- The model currently does not account for the difference between transit pass cost and equivalent single ticket price embedded in the mode choice utilities and logsums. The corresponding "bulk" discounts will be considered for inclusion in the model at Phase 2. This feature can be essential for certain policies that promote transit use.

CHAPTER 7.

CHOICE OF COORDINATED DAILY ACTIVITY-TRAVEL PATTERN BY HOUSEHOLD MEMBERS

7.1. CHOICE STRUCTURE

The Coordinated Daily Activity Pattern (CDAP) model predicts a daily activity pattern (DAP) type for all household members taking into account intra-household interactions and joint travel. The model was estimated in a nested logit form using the ALOGIT software. The alternatives in the model are formed based on the number of household members with a choice of one out of three DAP types defined a mutually exclusive and collectively exhaustive way:

- *Mandatory pattern (M)* that includes at least one out-of-home mandatory activity (work, university, or school) with any additional non-mandatory activities,
- *Non-mandatory pattern (N)* travel that excludes mandatory activities; this pattern is further subdivided into:
  - *Non-mandatory travel pattern (T)* that involves at least one out-of-home maintenance or discretionary activity.
  - *Stay-at-Home pattern (H)* that involves no travel, including working or schooling at home, being sick, or being out of town.

This trinary choice for each household member (M, T, or H) is further combined with a binary choice for the entire household with respect to joint travel (either to have at least one fully joint tour or not).

Up to a maximum of 5 members are chosen from the household based on a hierarchical role if the household size is greater than 5, described below. Joint travel is defined as a fully-joint tour in which two or more household members participate fully in all activities on the tour (escorting tours and other partially-joint travel arrangements without participation in the activities are not included in this model). Fully joint tours involve only non-mandatory activities and corresponding travel purposes.

Independent variables in the model include a wide range person, household, and residential zone characteristics. Among them, person type (described below) and activity-specific zonal accessibility terms play a special role. The most important aspects of this approach is to capture the impact of person type in combination with other household, person, and accessibility variables on the person propensity to travel to work or other activities, effects of intra-household interactions on these choices, and also the corresponding propensity to engage in joint activities. In the model chain, the CDAP model is applied after the models for work at home, schooling at home, work & school location choices, and auto ownership/transit pass ownership; therefore CDAP includes explicit OD accessibilities to mandatory activities (at the person level) and general accessibilities to non-mandatory activities (at the household level) as explanatory variables.

Statistical analysis in the Columbus, Atlanta, and San Francisco Bay Area regions, has shown that there is an extremely strong correlation between DAP types of different household members, especially for joint N and H types that have the potential to be jointly utilized if several household members choose the same pattern (for example, for a family event or taking vacations together). For this reason, the DAP for different household members cannot be modeled independently without introducing significant error in both individual activity patterns and household-level activity patterns. The CDAP model handles the DAP types for different household members simultaneously, and is one of the signature features of the CT-RAMP structure.

The current choice structure includes all possible combinations by individual DAP types for up to five household members in an explicit way. For a larger household with six or more members, five representative members are explicitly considered based on their person type, and the remaining persons (which constitute less than 1.7% of the population) are sequentially modeled conditional upon the choices made by the five representative members. The rules for choosing the five representative members are as follows:

- First, the household members are prioritized (highest to lowest) based on the person type in the following order (reflecting on the possible impact on the DAP choice of the other household members):
    - Full-time worker
    - Part-time worker
    - Pre-school Child
    - Pre-Driving Age School Child
    - Driving Age School Child
    - Non Working Adult
    - Retiree
    - University Student
- Secondly, younger children get priority when choosing between 2 or more children from same person type group.

The MAG CDAP model is a choice structure with 691 alternatives across different household sizes (1-5) including 363 alternatives with no joint travel and 328 alternatives with joint travel. For each household size, the set of choices are defined as combination of individual DAP types for all household members and joint travel pattern. The formation of available choices for each household size category is summarized in Table 38.

TABLE 38: CHOICE ALTERNATIVES FOR CDAP MODEL

Household Size	Alternatives without Joint Travel	Alternatives with Joint Travel	All Alternatives
1	3	0	3
2	3×3=9	3×3-(2×2+1)=4	13
3	3×3×3=27	3×3×3-(2×3+1)=20	47
4	3×3×3×3=81	3×3×3×3-(2×4+1) =72	153
5 or more	3×3×3×3×3=243	3×3×3×3×3-(2×5+1)=232	475
Total	363	328	691

Note that there are fewer alternatives with joint travel compared to alternatives without joint travel. Alternatives without joint travel include all combinations of trinary choices of DAP for all household members. Alternatives with joint travel include only these DAP combinations that result in at least two household members having an active travel pattern. For example, for a 2-person household, alternatives with joint travel exclude all joint patterns where one of the persons stays at home. Similarly, for a 3-person household, alternatives with joint travel exclude all joint patterns where two of the persons stay at home. In a general case with S persons, there are $2\times S+1$ cases to be excluded.

The choice structure of the CDAP model for the MAG ABM is shown in Figure 21. For simplicity, an example of a two-person household is shown. A generalization for a large household is straightforward.

- A binary *sub-choice of (indicator on) joint activity & travel episode* is included. This captures the impact of joint tours on other travel decisions earlier in the decision-making chain that in previous versions of CT-RAMP such as Atlanta and for the San Francisco Bay Area. In particular, the time-of-day choice model for work tours benefits from this indicator since workers frequently adjust their schedules to accommodate a joint activity episode. Each household activity pattern alternative where at least two members have non-home patterns is considered with and without joint travel as two different alternatives. At this modeling level, we do not distinguish between single and multiple joint tours. These details are added further down the model chain by means of a joint tour frequency model, which predicts the exact number of joint tours by purpose.
- An *intermediate nesting level* is introduced to account for principal differences between Mandatory and Non-Mandatory patterns. In all previous CDAP formulations, the main trinary choice at the person level (1=Mandatory day, 2=Non-mandatory travel day, 3= staying at Home) was modeled by a MNL model. This created some IIA effects that were difficult to explain. For example, for school children, dense urban environment induced more non-mandatory activity patterns compared to children living in suburban areas. However, with the MNL structure, this trade-off was not limited to non-mandatory and home patterns but also affected the frequency of mandatory patterns. The proposed nested structure accounts for these effects and gives a more reasonable structure of the trade-offs. The dichotomy of patterns associated with the joint tour indicator can also be incorporated as the lower-level nest in this nested structure.

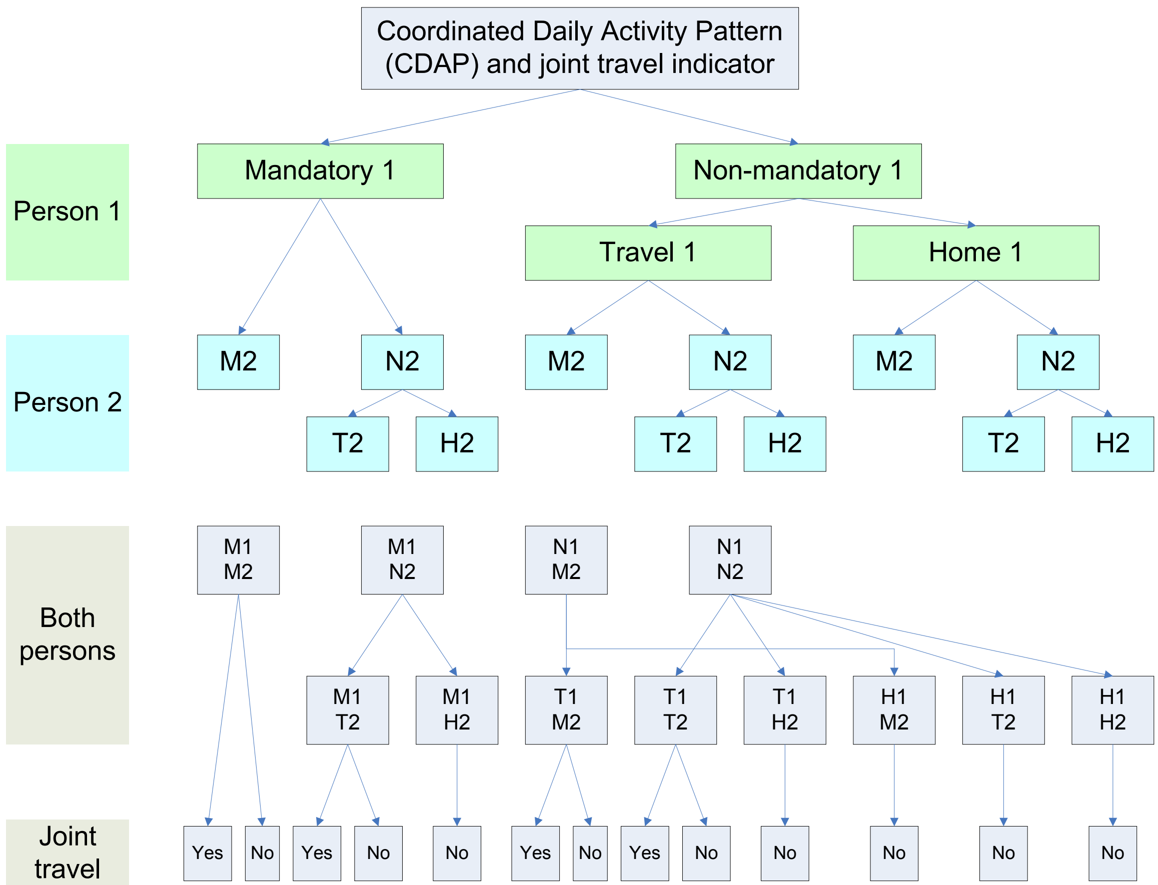

FIGURE 21: CHOICE STRUCTURE OF CDAP (EXAMPLE FOR 2 PERSONS)

7.2. ESTIMATION DATASET

The NHTS 2008 add-on for Phoenix and Tucson regions includes 5,067 households surveyed on a regular weekday travel. However, some household members were missing in 615 households. Therefore, the estimation dataset for CDAP model includes only 4,452 households with complete person information. Table 39 below shows the observed frequency of individual DAP types (1-3) by person type (1=8). It can be seen that person type has a very strong impact on the choice of DAP. Full-time workers, university students, and school children logically have a mandatory pattern as the dominant choice. On the contrary, non-workers and retirees have a very low probability of a mandatory pattern. Part-time workers and preschool children are somewhere between these two extremes.

TABLE 39: OBSERVED FREQUENCY OF DAP TYPES BY PERSON TYPE IN NHTS 2008

Person Type	Absolute Frequency				Relative Frequency		
	Total	**Mandatory**	**Non-Mandatory**	**At Home**	**Mandatory**	**Non-Mandatory**	**At Home**
1=Full Time Worker	3,398	2,653	579	166	78%	17%	5%
2=Part Time Worker	926	440	416	70	48%	45%	8%
3=University Student	324	267	45	12	82%	14%	4%
4=Non-Worker U65	1,831	69	1,417	345	4%	77%	19%
5=Retiree	2,356	18	1,856	482	1%	79%	20%
6=Driving Age School Child (16-17)	261	225	25	11	86%	10%	4%
7=Pre-driving Age School Child (6-15)	1,178	996	131	51	85%	11%	4%
8=Preschool Child U6	682	262	238	182	38%	35%	27%
Total	10,956	4,930	4,707	1,319	45%	43%	12%

The distribution of DAP types by person type observed in the NHTS 2008 add-on for MAG/PAG regions was compared to the distributions observed in the other regions were CT-RAMP ABMs were developed including Atlanta (ARC) with a 2-day survey implemented in 2001 for 8,000 households, San-Francisco Bay Area (MTC) with a 2-day survey implemented in 2000 for 15,000 households and San-Diego (SANDAG) with a 1-day survey implemented in 2007 for 3,600 households). The comparisons are presented in Figure 22 below.

It can be concluded that the NHTS 2008 in general provides a reasonable dataset that is very much in line with the main patterns of travel behavior observed in the other regions. With the exception for preschool children, for all other person types, the frequency of the travel-inactive home pattern in the NHTS 2008 survey proved to be amongst the lowest across all regional surveys. This in general indicates on a good record of out-of-home activities and trips. It should be mentioned, that cross-region comparisons should be taken with caution since the differences between regional populations and transportation systems may come into play. Also, there are some minor definitional discrepancies between the different surveys. It is however a remarkable general resemblance across the regions when comparisons are implemented at the level of DAP types by person type.

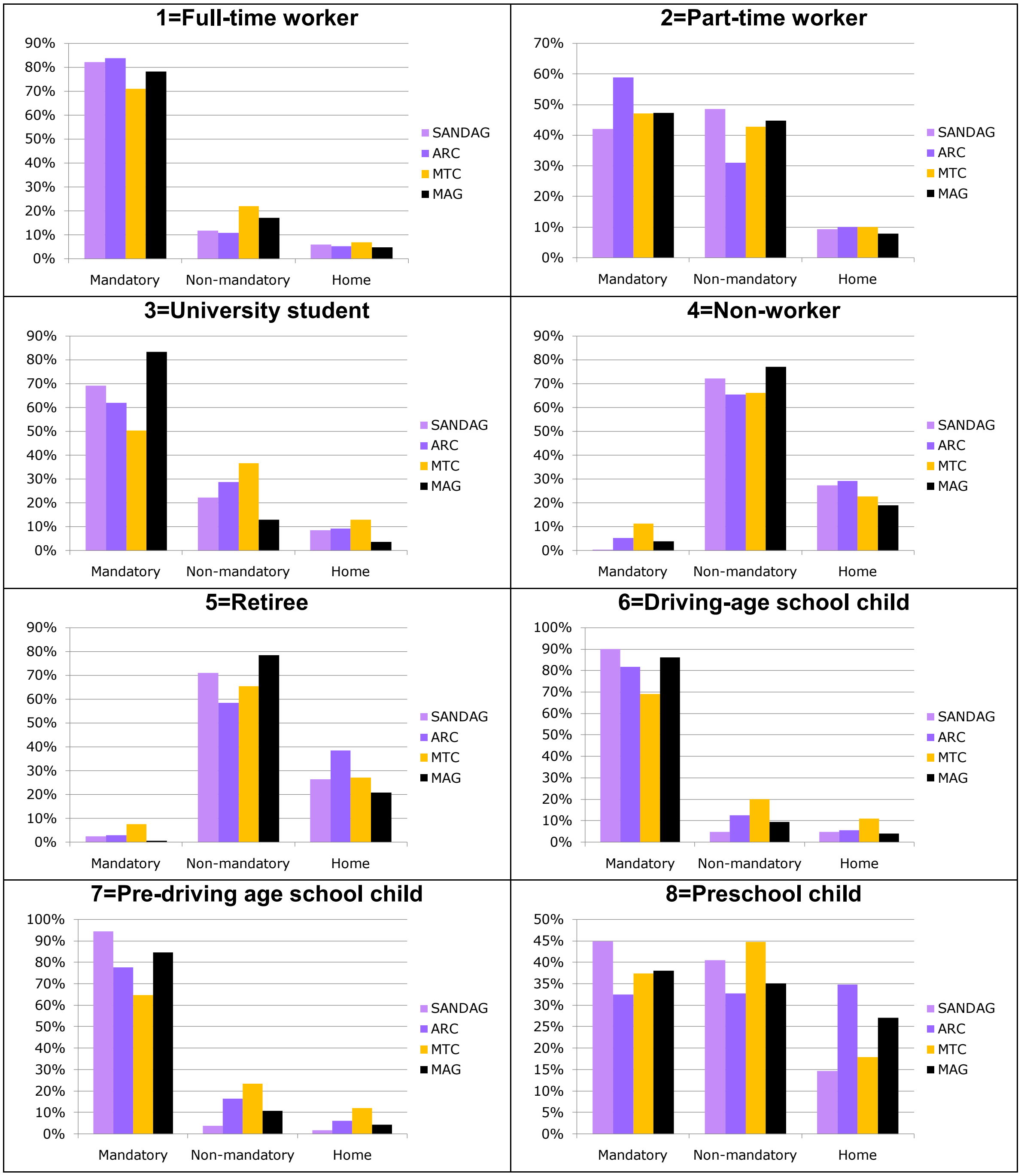

FIGURE 22: CROSS-REGION COMPARISON OF OBSERVED DAP DISTRIBUTIONS

The survey observations were joined with the OD mandatory and zonal non-mandatory accessibilities to create the estimation file. Mandatory accessibilities reflect the actual workplace and/or school location for each worker and student in the household, while non-mandatory accessibilities reflect the general accessibility to all potential non-mandatory destinations from the household residential zone.

7.3. MAIN EXPLANATORY VARIABLES

The following variables have been examined and proved to be significant in the utility functions:

- Person type
- Household Size
- Household income group:
    - Low income (less than $25,000)
    - Medium Low income ($25,000-$49,999)
    - Medium income ($50,000-$74,999)
    - Medium High income ($75,000-$99,999)
    - High income ($100,000 or more)
- Car Sufficiency with respect to workers by the following categories:
    - Zero cars
    - Cars fewer than workers
    - Cars equal to workers
    - Cars greater than workers
- Person Age Group
- Gender
- Tucson dummy to explore possible differences between the Tucson and Phoenix regions
- Accessibility measures from the household residential zone to Non-Mandatory attractions in a form of destination choice logsums as described above in the section on accessibility measures,
- OD accessibility measures in a form of mode & time-of-day choice logsum to work and school locations as described above in the section on accessibility measures,
- Housing type:
    - Detached dwelling unit
    - Non-detached dwelling unit
- Usual Workplace Type (home vs. out-of-home):
- Alternative-specific variables that characterize availability of household members for joint travel:
    - Number of Adults with Mandatory pattern,
    - Number of Adults with Non-Mandatory travel pattern,
    - Number of Pre-driving Age Children with Mandatory pattern,
    - Number of Pre-driving Age Children with Non-Mandatory travel pattern,
    - Dummy for a special case if all household adults stay at home.

The joint travel utility is dependent on the combination of DAPs for each alternative. Some of the corresponding utility variables cannot be pre-calculated in the estimation file due to a large number of alternatives. They are programmatically calculated "on-the–fly" for each alternative in the estimation process. The example in Table 40 below illustrates how some of these alternative-specific variables are calculated for a 5-person household based on DAPs for three different alternatives. The same programmatic method is used in the model application.

TABLE 40: CALCULATION OF ALTERNATIVE-SPECIFIC VARIABLES IN CDAP MODEL

Input and output variables		Coordinated Daily Activity Pattern		
		Alt 1	Alt 2	Alt3
DAP for each person by person type (input):				
Person 1	Full Time Worker	M	NT	H
Person 2	Pre-school child	NT	NT	M
Person 3	Pre-driving age school child	M	H	M
Person 4	Pre-driving age school child	M	H	M
Person 5	Non-working adult	NT	H	H
Alternative-specific variables calculated on-the-fly (output):				
Number of Adults with Mandatory Pattern		1	0	0
Number of Adults with Non-Mandatory Pattern		1	1	0
Number of Children with Mandatory Pattern		2	0	3
Number of Children with Non-Mandatory Pattern		1	1	0
All adults are at home (dummy)		0	0	1

7.4. UTILITY STRUCTURE

Alternatives of the CDAP type choice model correspond to all possible combinations of all individual trinary choices of DAP types $i = 1,2,3$ crossed with a binary choice of joint travel $j = 0,1$. All household persons are numbered from 1 to H where H corresponds to the household size. Every person has a unique number $h = 1,2,...,H$ within the household and the corresponding person type p_h. The set of the entire-household DAP alternatives is specific to the household size and denoted as $\Omega_H = \{i_1, i_2,...,i_h,...,i_H, j\}_H$. The observed part of the function for each alternative is specified to have the following general form:

$$U_{i_1...i_H,j} = \left(\sum_{h=1}^{H} V_{i_h p_h}\right) + \left[\sum_{h_1=1}^{H} \sum_{h_2=h_1+1}^{H} W_{(i_{h_1}=i_{h_2})p_{h_1}p_{h_2}}\right] + \sum_{h_1=1}^{H} \sum_{h_2=h_1+1}^{H} \left[\sum_{h_3=h_2+1}^{H} Z_{(i_{h_1}=i_{h_2}=i_{h_3})p_{h_1}p_{h_2}p_{h_3}}\right] + j \times \left(\sum_{i=1}^{3}\sum_{p=1}^{8} J_{ip}\right) \quad \textbf{Equation 15}$$

where:

$V_{i_h p_h}$ = individual component of choice of the DAP type i by the

		household member h of the person type p,
$W_{(i_{h_1}=i_{h_2})p_{h_1}p_{h_2}}$	=	pair-wise component of joint choice of DAP type i by household members h_1 and h_2 of the person types p_1 and p_2 consequently,
$Z_{(i_{h_1}=i_{h_2}=i_{h_2})p_{h_1}p_{h_2}p_{h_3}}$	=	triple-wise component of joint choice of DAP i by household members h_1, h_2, and h_3 of the person types p_1, p_2, and p_3 consequently.
J_{ip}	=	joint travel component (contribution) of joint choice of DAP i by household members of the person type p.

The individual choice utility component is specified to have the following form:

$$V_{i_h p_h} = \sum_{k \in K} c_{kip} x_{kh} \qquad \textbf{Equation 16}$$

where:

$k \in K$	=	a set of individual, household, and zonal attributes,
x_{kh}	=	value of the k attribute for the h person,
c_{kip}	=	coefficient for the k attribute in the i alternative utility that is assumed to be specific to the person type p but generic across persons h.

The pair-wise choice utility component is specified as:

$$W_{(i_{h_1}=i_{h_2})p_{h_1}p_{h_2}} = w_{ip_1p_2} \times \begin{cases} 1, & \text{if } i_{h_1} = i_{h_2} = i, p_{h_1} = p_1, p_{h_2} = p_2 \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{Equation 17}$$

where:

$w_{ip_1p_2}$	=	coefficient (added utility) for a dummy variable that corresponds to joint choice of the DAP type i by two household members of types p_1 and p_2.

The triple-wise choice utility component is assumed to have the following form:

$$Z_{(i_{h_1}=i_{h_2}=i_{h_3})p_{h_1}p_{h_2}p_{h_3}} = z_{ip_1p_2p_3} \times \begin{cases} 1, & \text{if } i_{h_1} = i_{h_2} = i_{h_3} = i, p_{h_1} = p_1, p_{h_2} = p_2, p_{h_3} = p \\ 0, & \text{otherwise} \end{cases} \qquad \textbf{Equation 18}$$

where:

$z_{ip_1p_2p_3}$	=	coefficient (added utility) for a dummy variable that corresponds to joint choice of the DAP type i by three household members of types p_1, p_2, and p_3

.

The joint travel utility component is added only if the alternative includes a joint travel episode ($j = 1$). It is assumed to have the following form:

$$J_{ip} = \lambda_{ip} \times H_{ip}$$ **Equation 19**

where:

λ_{ip} = coefficient (added utility) of a joint travel episode for a person of type p with the chosen DAP type i .

H_{ip} = number of household members of type p with the chosen DAP type i or some categorical transformation of that like "all household members of certain type and with certain pattern".

The model overall has a parsimonious structure with only 24 individual utility components of the form (2) to estimate for 3 individual DAP types and 8 person types with the addition of several joint DAP constants and joint travel components. Thus, most of the person variables (like age or gender), household variables (like income or car ownership) and zonal variables (like area type or accessibility) were tested in the individual utility component except for specific accessibilities for joint travel that are based on carpooling.

7.5. MODEL ESTIMATION RESULTS

The CDAP estimation results are summarized in two tables below. Table 41 includes coefficients for variables that relate to different person types in individual DAP components and pair-wise and triple-wise intra-household interactions between DAPs or different household members. In this table all coefficients are segmented by person types (columns) and DAP type (references in rows along with the variable itself).

Table 42 includes the other variables and joint travel utility components.

TABLE 41: CDAP ESTIMATION RESULTS (COEFFICIENTS SEGMENTED BY PERSON TYPE)

Utility Terms	FW- Full Time Worker		PW-Part Time Worker		US- University Student		NW- Non-Worker		RT- Retiree		SD- Driving School Child		SP- Pre-Driving School Child		PS- PreSchool Child	
	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat
Constants																
Mandatory	3.8665	7.32	2.0903	4.97	1.5297	1.79	-3.5673	-3.59	-6.1962	-4.98	4.9456	4.04	2.5678	4.05	2.9675	3.49
Non-Mandatory	-0.6808	-1.12	0.2505	0.36	-0.5393	-0.17	-1.1569	-3.38	-1.1158	-4.26	-1.5879	-0.54	-2.7365	-1.38	-0.9779	-0.78
Home all day (Reference)	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	
Tucson Dummy																
Mandatory	0.2717	1.49			-1.5947	-2.80	0.4900	1.09			-1.7729	-2.38			-0.5552	-1.49
Non-Mandatory									-0.5418	-4.43						
Age																
Age 0-1, Mandatory															-0.7847	-1.60
Age 4-5, Mandatory															0.7551	1.75
Age 0-1, Non-Mandatory															-0.1611	-0.52
Age 6-9, Mandatory													1.3330			
Age 6-9, Non-Mandatory													1.2455			
Age < 40 yrs, Mandatory	0.5310	2.43	0.8167	2.41												
Age over 80, Home all day									0.5713	4.33						
Household Income																
Mandatory																
$24,999 or Less	0.2629	1.22	0.5124	1.84									-0.7552	-1.98		
$25,000 to $49,999	0.2629	1.22	0.5124	1.84												
$50,000 to $74,999																
$ 75,000 to $99,999											-2.0946	-2.39				
$100,000 or More	-0.0573	-0.285							1.3067	1.06	-2.0946	-2.39				
Home All day																
$24,999 or Less							0.0273	0.13							0.3680	1.26

Utility Terms	FW- Full Time Worker		PW-Part Time Worker		US- University Student		NW- Non-Worker		RT- Retiree		SD- Driving School Child		SP- Pre-Driving School Child		PS- PreSchool Child	
	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat
$25,000 to $49,999							0.0273	0.13							0.3680	1.26
$50,000 to $74,999																
$ 75,000 to $99,999							-0.2690	-0.92	-0.1563	-0.85					-0.5368	-1.77
$100,000 or More	-0.4789	-2.06	0.3507	1.05			-0.6115	-2.18	-0.1563	-0.85					-0.5368	-1.77
Gender																
Female, Mandatory	0.3152	1.31	0.0117	0.04			-1.0678	-2.29								
Female, Non-Mandatory	0.5873	2.75							-0.3862	-2.92						
Female, At-Home			0.3967	1.26												
Car Sufficiency																
Mandatory																
Zero Cars			-2.8568	-1.61												
More Cars than Workers			0.8267	3.05	2.7495	2.85	0.6331	0.94			1.3874	1.31				
Non-Mandatory																
Zero Cars	-0.7240	-0.80														
More Cars than Workers	-0.1199	-0.81			0.9494	1.16	0.5924	3.13	0.3286	1.80	1.4116	1.19				
Job Category for Workers																
Mandatory																
Production, construct, manufacturing, or transport	0.2213	0.80	0.5293	1.10												
Person care and services	-0.4510	-1.85	-0.3045	-0.99												
Accessibility and Others																
Mandatory																
Accessibility to Work	0.0671	0.65														
Usual Work Place is Home	-4.2567	-5.98	-2.4495	-5.34												
Schooling from home													-5.7218	-4.67	-3.8471	-4.59
Bachelor's or Higher Degree	-0.1020	-0.59														

Utility Terms	**FW- Full Time Worker**		**PW-Part Time Worker**		**US- University Student**		**NW- Non-Worker**		**RT- Retiree**		**SD- Driving School Child**		**SP- Pre-Driving School Child**		**PS- PreSchool Child**	
	Coeff	**T-Stat**	**Coeff**	**T-Stat**	**Coeff**	**T-Stat**	**Coeff**	**T-Stat**	**Coeff**	**T-Stat**	**Coeff**	**T-Stat**	**Coeff**	**T-Stat**	**Coeff**	**T-Stat**
Non-Mandatory																
Non-Mandatory (Individual) Accessibility	0.1073	2.46	0.1267	2.59	0.0552	0.23	0.1414	7.21	0.2173	13.66	0.0490	0.24	0.1370	0.98	0.0482	0.54
Dwelling Type																
At Home																
Detached HH													-0.5946	-1.19		
Two Person Interactions																
Mandatory																
Full Time Worker	0.5785	2.72														
Part Time Worker	0.3102	1.24	0.7267	1.33												
University Student					2.2701	2.39										
Driving School Child	0.9273	2.90									2.8566	1.61				
Pre-Driving School Child	0.3691	2.76			0.9621	2.56					0.8662	1.90	1.5392	4.58		
PreSchool Child	0.8793	3.32			1.1620	2.55							0.3470	1.48	0.8584	2.33
Non-Mandatory																
Part Time Worker			0.5694	1.36												
Driving School Child	0.8729	1.46					1.4743	2.23			1.6222	0.63				
Pre-Driving School Child	0.1547	0.60	0.0791	0.23							1.8647	2.98	3.8607	7.18		
PreSchool Child			0.3646	1.32			0.6019	2.61	1.1559	1.06			1.7974	4.45	2.3367	5.83
Home All Day																
Full Time Worker	1.2718	2.82														
NonWorker	0.8035	2.50														
Retiree	1.0957	2.91	1.4075	3.03			0.9021	3.89	0.6643	2.69						
Driving School Child	0.7423	0.57	2.8647	2.22			2.5346	2.85	2.6268	1.92						
Pre-Driving School Child	0.4994	0.75	1.2101	1.60			1.5437	3.79			1.4183	1.22	2.0914	2.29		
PreSchool Child	0.7064	1.44	0.5180	1.03			0.9396	2.70	1.9782	1.79			1.2172	2.15	2.4358	5.58

Utility Terms	FW- Full Time Worker		PW-Part Time Worker		US- University Student		NW- Non-Worker		RT- Retiree		SD- Driving School Child		SP- Pre-Driving School Child		PS- PreSchool Child	
	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat	Coeff	T-Stat
Three Person Interactions																
Mandatory																
FWxFW	-1.5238	-2.77					1.7441	3.45								
FWxPW			-0.4236	-0.45									0.2633	1.25	0.2633	1.25
PWxNW													1.7257	1.15	1.7257	1.15
NWxKD													-1.3670	-0.91	-1.3670	-0.91
KDxKD																
Non-Mandatory																
FWxFW													0.3176	0.63	0.3176	0.63
FWxPW													-0.4727	-0.96	-0.4727	-0.96
FWxNW							-1.1354	-0.98					0.2887	1.10	0.2887	1.10
NWxNW							-1.2163	-0.98								
NWxKD													-0.2825	-1.08	-0.2825	-1.08
KDxKD													-1.9766	-3.67	-1.9766	-3.67
Home All Day																
FWxNW													-0.6443	-0.75	-0.6443	-0.75
FWxKD													-0.7205	-0.74	-0.7205	-0.74
NWxNW							2.9964	2.16								
NWxKD													-0.6088	-1.20	-0.6088	-1.20
KDxKD													-0.7392	-1.02	-0.7392	-1.02

TABLE 42: CDAP MODEL ESTIMATION RESULTS (OTHER VARIABLES)

Utility Terms	Mandatory		Non-Mandatory		At Home		Joint	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Same Pattern for All Household Members								
Three Person households	0.7205	2.40	-0.6946	-2.57	-0.5224	-0.85		
Four Person Households	-0.0614	-0.19	-0.7062	-1.30	0.9068	0.89		
Five Person Households	-0.6213	-1.33	-0.4907	-0.60				
Joint Travel								
Constant							-3.8022	-10.62
Zonal Non-Mandatory Accessibility							0.0500	
Work Accessibilities for Persons with Mandatory DAP							0.0361	0.74
# Adults (P1-6) with Non-Mandatory DAP							1.4877	10.36
# Adults (P1-6) with Mandatory DAP							0.1982	1.44
# Children (P7-8) with Non-Mandatory DAP							1.4499	8.87
# Children (P7-8) with Mandatory DAP							0.5024	5.15
If All Adults are at Home (Dap=3)							-1.0000	
Income								
$24,999 or Less							-0.4517	-2.15
$25,000 to $49,999								
$50,000 to $74,999								
$ 75,000 to $99,999								
$100,000 or More							-0.3746	-2.32
Car Ownership								
Zero Cars							-1.3836	-1.79
Fewer Cars than Workers								
Cars Equal to Worker								
More Cars than Workers							0.2388	1.53
Nesting Coefficients								
Upper Level Nest: Mandatory vs. Non-Mandatory	0.9551		0.9551					
Lower Level Nest: Joint vs. No-Joint							0.6616	
Goodness-of-fit stats								
Number of Observations	4452							
Likelihood with Constants only	-9574.0161							
Final likelihood	-6763.525							
ρ^2 w.r.t. zero	0.4392							
ρ^2 w.r.t. constants	0.2922							

7.6. MAIN FINDINGS AND BEHAVIORAL INTERPRETATIONS

The main findings from the estimation results with the highest statistical significance and corresponding behavioral interpretations are discussed below. They are grouped by the following main explanatory variables and effects:

- *Person type:* The person type specific constants indicate that, all else been equal, full-time workers and school children are most likely to have mandatory patterns; and, non-workers and retirees are least likely to carry out mandatory activities. These constant reflects the most fundamental differences in person behavior and generally follow the observed frequency patterns described above. There are several statistically effects specific to the Tucson region for such person types as university students, retirees, and driving-age school children but in general the model transferability is very high.

- *Gender:* Interaction of person type with female dummy shows that among full-time workers, females are less likely to stay at home (since they have a higher frequency of both mandatory and non-mandatory travel patterns), while among retirees, non-workers, and part-time workers, females are more likely to stay at home. This is an interesting finding that contradicts the stereotype that females always spend more time at home. In reality, it depends on the person type and role as well as some new tendencies like telecommuting. For full-time workers, flexible work schedules and telecommuting may be more frequent for males than females.

- *Age:* Among very young children (under age 6), the chances of going to school increases with age. Among children of age 6 to 15 yrs, the likelihood of going to school for children under 10 years is more than for children over 10 years. This may reflect an increasing likelihood of participation in other activities that conflict with school as age grows as well as some cases of dropping from school. Retirees older than 80 yrs are more likely to stay at home compared to retirees younger than 80 years. Workers under 40 years are more likely to have a mandatory pattern which could be due to less flexible schedule and less frequent telecommuting by workers of younger age compared to older workers.

- *Car Ownership/Sufficiency:* In households with more cars than workers, non-working adults as well as retirees and university students are more likely to travel for non-mandatory activities and have a non-mandatory travel day at the expense of staying at home. As the results, in these households, Full-time workers are less likely to have only non-mandatory pattern because non-workers or other family members are more likely to take care of maintenance activities if a car is available. Part-time workers and university students are more likely to have a mandatory tour if there are more cars than workers in the household. On the contrary, in a zero-car household, part-time workers exhibit a significantly lower frequency of mandatory patterns. This might be a reversed causality between the DAP frequency and car ownership in some case, i.e. Part-time workers and students that have more compressed schedules would possess less cars. To resolve this issue, the set of long-term models should be enhanced by an additional model for usual work arrangements (commuting and telecommuting frequency). In order to estimate this model, the corresponding data items have to be collected that are currently not in the NHTS 2008.

- *Household Income:* In general, household income has a positive effect on travel patterns and negative effect on staying at home (in other words, income and mobility are positively correlated). Full-time workers, non-workers and retirees are less likely to stay at home in higher income groups households. Persons from higher income group might engage themselves more often in out-of-home discretionary activities. There is an additional interplay between mandatory and non-mandatory travel patterns across income categories.

Workers in low-income households are more likely to travel for mandatory activities. The workers from high income households might have options to telecommute more often than workers from low income households due to the nature of the job.

- *Accessibility:* Full-time workers are more likely to travel to mandatory activities with better accessibility to work location (full-time workers who have a longer and less convenient commute may try to have a compressed work week). Better accessibilities to non-mandatory destinations improve the chances of making non-mandatory travel (at the expense of staying at home) for all person types.
- *Usual Work Location:* Workers are much less likely to travel for mandatory activities if their usual work location is home. Also, workers who reported not having any usual work location are less likely to have mandatory travel. The effects are stronger for full-time workers as compared to part-time workers.
- *Two-way Interactions:* The two-way interaction terms by person type combinations were estimated for the cases when two persons in the household have identical pattern types (MM, TT or HH). All possible interactions were tried in the estimation, except for mandatory patterns involving non-workers and retirees (mandatory pattern is infrequent for them), and combinations with unobserved cases.
    - All estimated two-way interactions are positive and many of them are strong and statistically significant. This confirms the strengths and importance of intra-household interactions in predicting travel patterns.
    - For mandatory (M) pattern, some of the largest interactions are found among school children (SD and SP) and preschoolers (PS). The interactions are also positive between workers and among workers (particularly, full-time workers) and children.
    - For non-mandatory travel (T) pattern, the largest positive interactions are among pairs of children ages 6 and above. For younger children (age younger than 6), significant positive interactions are found with adults (particularly, part-time workers and non-workers) and other children. These interactions are largely a function of escorting children to doctors as well as taking children by the caretaking parent for shopping and other errands.
    - For at home (H) pattern, largest interactions are between children of similar age group (i.e., pre-driving age child with pre-driving age child and driving age school child with driving age school child), and between non-worker and pre-driving age children. Strong interactions are also between retirees and retirees/non-working adults, and significant interactions are found between retirees and workers. At home patterns are also significant between non-working adults and children. These interactions correspond to family events, coordinated planning of vacations and trips out of town. Also, for families with multiple children it is common for them to get sick together.
- *Three–way Interactions:* These interaction terms (MMM, TTT or HHH) were considered only for selected person type combinations because there are hundreds of possible three-way combinations.
    - Combination of three full time workers showed a negative interaction term. This term is applied on top of the three positive two-way interaction terms in the triple. It mitigates the strong additive impact of two-way interactions for three workers going to work/school.

  - Combination of three children showed a negative interaction term for TTT and HHH pattern. Again, this is applied on top of the three positive two-way interaction terms in the triple by mitigating the strong cumulative impact of two-way interactions.

- *Same DAP for all household members:* The estimates prove to be all negative for non-mandatory patterns. The strength of the negative coefficient increases with household size except for at home patterns. However, for at home pattern, the coefficients are not significant for household size 4. These coefficients will offset the affect of two-way and three-way interaction terms for larger households. (Note: the number of two-way interaction terms increase significantly with household size. A 3-person household has 3 terms, a 4-person household has 6 terms, and a 5-person household has 10 terms).

- *Joint travel:* The CDAP model also predicts whether joint travel occurs at a household level. Further on the exact number of joint travel tours and person participation in them is modeled in a spate model. At this stage, it is important to capture the general propensity for a household to have a joint activity-travel episode as function of the household composition and DAP types chosen by different household members. The following main factors affecting joint travel were found:

  - There is a strong negative constant on joint travel alternatives that means that household members are less likely to have a joint tour than not to have one. It is in line with the general frequency of fully joint tours on a regular weekday that is less than 25%.

  - For a household member with a mandatory pattern, the chances of participating in joint travel are higher with a better accessibility to work/school location. This is a logical consequence of time-space constraints. Workers and student who have to commute long distances simply may not have time to participate in an additional tour from home with other household members.

  - The probability of joint travel in a household grows with the number of adults or children with a non-mandatory pattern. This is also an expected result since household members who do not have a burden of out-of-home mandatory activity but also do not stay all day at home are the first candidates to participate in joint out-of-home activities.

  - Lowest income (less than $25K) households and highest income (more than $100K) are less likely to have a joint travel episode. This is an interesting result that was not expected. There, however, can be an explanation why joint travel with household members is more of a middle-class phenomenon. Low-income households may not be able to afford such activities as visiting a cinema theater or restaurant jointly. High-income households do not have this financial constraint but it is known that individualism in behavior also grows with income.

  - Members of a household with the number of cars greater than number workers are more likely to have joint tours. This can be explained by the fact that an automobile carpool is the dominant mode for joint travel. Moreover, many households tend to have large cars, vans, or SUVs specifically for joint activities. This is probably even more prominent on weekends but also manifests itself on weekdays.

CHAPTER 8.

CONCLUSIONS AND RECOMMENDATIONS

8.1. GENERAL CONCLUSIONS FROM THE ESTIMATED MODELS

In general, the estimated models are characterized by logical results and behavioral richness in terms of explanatory variables and effects. Overall, the NHTS 2008 provided good data to support estimation of several advanced models with a complicated choice structure and non-trivial utility expressions. The survey underwent substantial data processing and data cleaning to construct the estimation files. It included building tours from elemental trips, building person daily activity-travel patterns from tours, and building entire-household joint activity patterns from person patterns and joint travel episodes. In this process, the survey records were joined with various level-of-service variables provided by MAG and a wide spectrum of developed accessibility measures. As the result, a fully-functional database for estimation of an advanced ABM was created that will serve the needs of subsequent model estimation at Phase 2 and 3.

The estimated models include many innovative features and some of them have an advanced structure with a large number of choice alternatives. They required a programmatic setup for building estimation files and forming utility expressions that was developed by PB and specifically enhanced for the MAG ABM development project. The Coordinated Daily Activity-travel Pattern (CDAP) model described in the current report represents a good example that illustrates the estimation techniques needed for development of an advanced ABM.

The models were estimated for both Phoenix and Tucson regions in order to lay a foundation for an ABM for the extended Phoenix-Tucson region. In general, a high level of transferability was observed with the same main explanatory variables playing the same role in both regions. Only in several cases, notable for work and school location choices as well as for some constants in the car-ownership model a decision was made to segment coefficients by region.

Some particular data deficiencies and limitations that were discovered in the data-processing are summarized in the subsequent section below.

8.2. RECOMMENDATIONS FOR SUBSEQUENT STAGES

The data-cleaning and data-processing work completed at Phase 1 laid a foundation for the entire model estimation effort. However, several additional data cleaning steps are needed to create consistency between the NHTS 2008 and land-use data for both the Phoenix and Tucson regions. A better consolidation of trips ends by travel purpose with the regional land-use database is highly recommended. This step at Phase 1 revealed many conflicts including work trips to a zone with a zero employment for the relevant person occupation, school trips to a zone with a zero enrollment for the relevant person school type, shopping trips to a zone with a zero retail employment, etc. All these cases were reported to MAG and many of them were resolved by MAG staff at Phase 1. However, a significant number of conflicting data items still remains and could not be fixed within the short time window between the delivery of NHTS 2008 (March 2010) and end of Phase 1 (July 2010). Further consolidation of the trip database would be very useful for fixing multiple geo-coding errors as well as problems with the land-use data.

Adding special events to the CDAP model will require data consolidation as well. Currently only about 30 participations in special events were identified and coded in the NHTS 2008 database. The preliminary expansion of these trips falls very short of the total participation of the regional population in special events. The consolidation of the NHTS 2008 and recent Survey of Special Events is needed to estimate and implement the innovative model that incorporates special events in individual daily activity-travel pattern as outlined in the Model Design and Development Plan document. Another possible enhancement for the CDAP model that can be considered for Phase 2 is to explicitly account for telecommuting as a distinctive pattern (rather than blend it with other reasons for staying at home).

Although MAG provided detailed data on employment by NAICS codes that, it was revealed in the model estimation that some additional attraction variables would be useful. A new effort to collect data such as commercial floor area, land area for public parks and open green areas, number of beds for hotels, number of seats for theaters, etc could benefit the ABM tremendously.

The NHT 2008 data has not been yet consolidated with the recent transit on-board survey. This is an important step that can help to enrich the small sub-sample of transit trips in NHTS 2008 that would be helpful for estimation of mode choice and other tour-level and trip-level choice at Phases 2 and 3. Additionally, it would help to calibrate the transit pass component in the model that predicts individual mobility attributes.