FY17 Task Orders

final report

prepared for

Metropolitan Washington Council of Governments
National Capital Region Transportation Planning Board

prepared by

Cambridge Systematics, Inc.

with

Gallop Corporation

June 2017

www.camsys.com

final report

FY17 Task Orders

prepared for

Metropolitan Washington Council of Governments
National Capital Region Transportation Planning Board

prepared by

Cambridge Systematics, Inc.
3 Bethesda Metro Center, Suite 1200
Bethesda, MD 20814

with

Gallop Corporation

date

June 2017

Table of Contents

List of Tables

List of Figures

1.0 Background

In fiscal year (FY) 17, the work program comprised two task orders: 17.1 – Meeting Support and Ad Hoc Assignments and 17.2 – Implementation of Short Term Model Improvements. It was a collaborative effort between Transportation Planning Board (TPB) staff and the Cambridge Systematics (CS) team staff to support this years' work program. This report compiles and transmits the major deliverables of the latter of these two task orders and focuses on delivery of material contributed by the CS team.

1.1 Basis for Short-Term Improvements

The FY17 work program derived from the strategic plan for model development that was developed beginning during the FY15 work program. The strategic plan for model development focused resources in the first few years on updating and enhancing the current MWCOG model. The priorities for these enhancements were determined though the effort in FY15 to understand the modeling needs of the MWCOG region and to consider the limitations of the current model to address those needs. It was through this effort that the topics of enhancing non-motorized modeling, enhancing transit modeling, and enhancing handling of managed lanes emerged as the short-term improvement priorities.

1.2 Short-Term Improvements: Non-Motorized Modeling

The need for enhanced non-motorized modeling derives from increased interest among model users in exploring such improvements and in estimating the transportation benefits that may come from so doing. During the FY16 work program, CS recommended a path forward for improving the trip-based model in the short-term by enhancing the binary modal splits at the trip generation stage with use of disaggregate model estimation using 2007/8 household travel survey data and the existing database of information related to built-environment and non-motorized facilities. The work that was thus undertaken in the FY17 work program is described primarily in Section 3.0 of this report.

1.3 Short-Term Improvements: Mode Choice, Transit Assignment

Work to migrate to the Cube Voyager Public Transport (PT) module as the handler of transit path-building and assignment began several years ago. During the FY16 and FY17 work programs, this work advanced to the stage of implementation. The newly delivered model set, developed as part of the FY17 work program, now relies on PT.

Incorporating PT supported the ability to explore alternative mode choice model structures, including having more of the transit submode choice logic reside within the transit assignment step. The FY17 work program embarked on model estimation using a newly constructed model estimation dataset that brought together data from the 2007/2008 Household Travel Surveys, regional transit on-board surveys, and two geo-focused supplemental household travel survey efforts. The development of this estimation data set is covered in Section 2.0 of this report. The use of the data to estimate a new mode choice model and the incorporation of PT into the model set are covered in Section 4.0 of this report.

1.4 Short-Term Improvements: Traffic Assignment (Managed Lanes)

During the FY16 work program, a task order provided insight on best practices in managed lane modeling for regional travel demand forecasting models, an understanding of the existing modeling framework, and proposed improvements to the modeling methodology. The associated task report made a recommendation to implement segmentation of highway assignment using value of time. This approach is consistent in objective with best practices and was deemed possible to achieve within the set timeframe.

The FY16 work program also explored possible improvements of the functions to enhance the highway assignment results of the model. Evidence was presented that the conical form of the volume delay functions used in the existing MWCOG model may be one of the reasons why the MWCOG model underestimates congested travel speeds on freeway facilities, as compared with observed traffic speed data and as reported by other studies in the region. It was thus suggested to replace the conical functions with the modified BPR functions for freeway facilities.

Thus, the FY17 work program took on both of these short-term improvements. The associated effort is described primarily in Section 5.0 of this report.

2.0 Development of Observed Data

This section summarizes activities that Cambridge Systematics (CS) has performed related to merging and review of MWCOG's household survey data and transit on-board survey data for use in the upcoming model estimation.

Survey data reviewed in this task included the following:

- 2007/08 Household Travel Survey (HTS)
- 2011 Geo-Focused HTS
- 2012 Geo-Focused HTS
- 2008 Metrorail transit on-board survey (TOS)
- 2008 Regional bus TOS
- 2007/08 MARC TOS
- 2005 Virginia Rail Express (VRE) TOS

2.1 Survey Descriptions

Household Travel Surveys

The geography covered by the 2007/08 Household Travel Survey includes the MWCOG modeled area, while the 2011 and 2012 geo-focused surveys cover only portions of the MWCOG region. After reviewing the data, we plan on merging all three datasets for use in mode choice model estimation. While the geo-focused surveys come from a different year than the regional HTS, using the geo-focused surveys is deemed beneficial since using more data in model estimation should yield more precise estimates of key model parameters. We do not believe the difference in survey were collection dates will have a material impact on the model.

Because the geo-focused surveys are specific only to portions of the MWCOG region, we believe it will be important to weight the data appropriately, since households located in the geo-focused areas may have different travel patterns than other households. The processes by which the surveys will be reweighted are discussed later in this report section.

Transit On-Board Surveys

We reviewed the documentation associated with the four transit on-board surveys, as listed above. Based on our review, we understand that the expanded sample of transit trips included in the four surveys is a close approximation of the full set of transit trips made on a typical weekday. There are a couple of exceptions to this. First, one or two of the small bus service providers in the region were not included in the regional bus survey. Second, some overlap exists across the surveys where trips use multiple transit services for the same trip (e.g., bus-to-Metrorail transfer trip). Preliminary review suggests these issues are manageable.

After review of the data, we determined that the VRE transit on-board survey (TOS) is not usable for the purposes of disaggregate model estimation, due to the absence of production and attraction zonal information. However, each of the other three surveys contains all of the necessary information for model estimation, with the exception of regional bus survey records where the access mode was coded as another form of transit (the only access modes that will be used in the mode choice model are walk and drive access). Such bus records will be dropped from the dataset.

The transit on-board data was merged with the HTS data.[1] The HTS transit data and the TOS data were reweighted so that the expanded data from the merged dataset is representative of the overall population of transit trips. Because the VRE TOS is not used, special procedures were used in reweighting to account for this. The procedures for reweighting are described in more detail later in this report section.

2.2 HTS Merging and Weighting

In order to reweight the HTS data, the 2007/08 HTS can be partitioned into three areas. The first area is partitioned so that its geography overlaps with the 2011 geo-Focused Survey, the second area is partitioned so that its geography overlaps with the 2012 Geo-Focused Survey, and the third area is the left over survey records that do not overlap with either Geo-Focused survey. The geography included in the 2011 and 2012 geo-focused surveys do not overlap.

For the purposes of this report section, the geography associated with the 2011 Geo-Focused Survey will be referred to as Geography 1 and the geography associated with the 2012 Geo-Focused Survey will be referred to as Geography 2. Table 2.1 shows the weighted and unweighted household totals in each survey for Geographies 1 and 2. The expansion factors used to weight the survey records were unmodified from what we received in the data file transmissions from MWCOG. The weighted household records in the overlapping geographies are very close for both Geographies 1 and 2 (though they do not match exactly).

Table 2.1 Overall Total Weighted Household Summaries

Description	Items	Geo-Focused Survey	Regional Survey
Geography 1		Geo-Focused HTS 2011	2007/08 HTS
	Total Weights	119,789	115,260
	Unweighted Household Totals	2,179	678
Geography 2		Geo-Focused HTS 2012	2007/08 HTS
	Total Weights	106,287	114,208
	Unweighted Household Totals	2,706	598

The three household travel surveys can be merged easily by appending the geo-focused survey records to the 2007/08 HTS records. However, the expansion factors, as received from MWCOG, are no longer relevant once this is done (since the geo-focused records are double-counted). Therefore, two sets of weights (Weight 1 and Weight 2) were calculated to reweight household records of the three survey data sets. These are described below.

For Weight 1, geo-focused expansion factors were first rescaled so that the sum of the factors matched the sum of HTS weights in the respective overlapping geography. Next, the resulting weights for the geo-focused records were factored by the geo-focused unweighted household shares within the geography. The

[1] Note that this is a common practice when constructing datasets for mode choice model estimation in regional travel demand models.

HTS expansion factors for the corresponding geography were factored by the HTS unweighted household shares within the geography. The equations below show explicitly the re-weighting factors used for Weight 1.

$$W_{1,GF,g,i} = ExpFac_{GF,g,i}\left(\frac{\sum_i ExpFac_{HTS,g,i}}{\sum_i ExpFac_{GF,g,i}}\right)\left(\frac{N_{GF,g}}{N_{GF,g}+N_{HTS,g}}\right) \quad (2.1)$$

$$W_{1,HTS,g,i} = ExpFac_{HTS,g,i}\left(\frac{N_{HTS,g}}{N_{GF,g}+N_{HTS,g}}\right) \quad (2.2)$$

Here, $W_{1,GF,g,i}$ is the final weight 1 for geo-focused record *i* in geography *g*, $W_{1,HTS,g,i}$ is the final weight 1 for HTS record *i* in geography *g*, $ExpFac$ refers to the original expansion factors in the datasets, $N_{GF,g}$ is the total number of household records in the geo-focus survey for geography *g*, and $N_{HTS,g}$ is the total number of household records in the HTS for geography *g*.

For Weight 1, the expansion factors are unchanged for HTS records in geographies that do not overlap with the geo-focused survey geographies.

For Weight 2, instead of first rescaling geo-focused expansion factors to match HTS expansion factors, HTS expansion factors were rescaled to match geo-focused expansion factor totals. Like was done for Weight 1, all the records in the geo-focused geography were then rescaled using unweighted household totals. Lastly, for Weight 2, the HTS records in non-overlapping geographies also needed to be reweighted so that the overall weights in the merged household survey matched the original HTS expansion factor totals. See the equations below:

$$W_{2,GF,g,i} = ExpFac_{GF,g,i}\left(\frac{N_{GF,g}}{N_{GF,g}+N_{HTS,g}}\right) \quad (2.3)$$

$$W_{2,HTS,g,i} = ExpFac_{HTS,g,i}\left(\frac{\sum_i ExpFac_{GF,g,i}}{\sum_i ExpFac_{HTS,g,i}}\right)\left(\frac{N_{HTS,g}}{N_{GF,g}+N_{HTS,g}}\right) \quad (2.4)$$

$$W_{2,HTS,h,i} = ExpFac_{HTS,h,i}\left(\frac{\sum_i ExpFac_{HTS,i}-\sum_g\sum_i ExpFac_{GF,g,i}}{\sum_i ExpFac_{HTS,i}-\sum_g\sum_i ExpFac_{HTS,g,i}}\right) \quad (2.5)$$

Here, $W_{2,HTS,h,i}$ is the final Weight 2 for the non-overlapping component of the HTS.

For both weighting procedures, the sum of all weights in the merged dataset matches the sum of HTS expansion factors in the original dataset. For weight 1, the sum of all weights in the merged dataset within both Geography 1 and Geography 2 matches the sum of HTS expansion factors in the original dataset within Geography 1 and Geography 2, respectively. For weight 2, the sum of all weights in the merged dataset within Geography 1 matches the sum of geo-focused 2011 expansion factors in the original dataset and the sum of all weights in the merged dataset within Geography 2 matches the sum of geo-focused 2012 expansion factors in the original dataset.

2.3 Merged HTS Data Summaries

Table 2.2 summarizes the weighted distribution of households by household size, workers, income level, and vehicles, and compares the results between the new weights and the original 2007/08 HTS. The distributions using the new weights slightly deviate from original distributions from the 2007/08 HTS and there is almost no difference between the distributions found between the two new sets of weights.

Table 2.3 and Table 2.4 tabulate total weights for the two overlapping areas between 2007/08 HTS and two geo-focused surveys. Weight 1 matches the 2007/08 HTS weights exactly. After reviewing the summaries, it is clear that the geo-focused surveys share certain characteristics that are slightly different than the similar geographies in 2007/08 HTS. For instance, both geo-focused surveys have larger household sizes on average, both have fewer workers on average, both have more low income and more high income households (and fewer middle income households), and both have more zero-vehicle households.

Presumably, the original 2007/08 HTS weights controlled for some of these household level variables, and by merging the geo-focused survey datasets, those controls may have been broken. However, the degree to which this may be an issue for mode choice model estimation is likely very limited. This is partly because the mode choice models will control for income directly, and income is correlated to different degrees with each other variable. Moreover, on a regionwide basis, the differences in variable distributions is very minor. The main benefit of merging the data is that the added observations will benefit the precision of model estimates.

Table 2.2 Overall Total Weighted Household Summaries

HH_Size	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2
1Psn	655,955	656,185	660,893	-0.7%	-0.7%
2Psn	711,410	711,536	717,802	-0.9%	-0.9%
3Psn	385,361	385,185	382,571	0.7%	0.7%
4Psn	586,555	586,376	578,015	1.5%	1.4%
Total	2,339,281	2,339,281	2,339,281		

Total Weighted Household Summaries by Workers

HH_Workers	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2
0Wrk	410,859	410,611	402,858	2.0%	1.9%
1Wrk	941,432	941,683	942,455	-0.1%	-0.1%
2Wrk	844,854	844,861	851,074	-0.7%	-0.7%
3+Wrk	142,137	142,127	142,895	-0.5%	-0.5%
Total	2,339,282	2,339,282	2,339,282		

Total Weighted Household Summaries by Income Level

HH_Inc_Level	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2
$ 0-49.99K	515,943	516,128	507,496	1.7%	1.7%
$ 50-99.99K	799,937	800,222	811,043	-1.4%	-1.3%
$ 100-149.99K	662,401	662,563	666,627	-0.6%	-0.6%
$ 150K	361,000	360,370	354,116	1.9%	1.8%
Total	2,339,281	2,339,282	2,339,282		

Total Weighted Household Summaries by Vehicles Available

HH_Vehs_Av	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2
0 Veh	140,675	140,857	136,162	3.3%	3.4%
1 Veh	733,280	733,111	735,026	-0.2%	-0.3%
2 Veh	944,585	944,670	947,370	-0.3%	-0.3%
3+ Veh	520,741	520,643	520,725	0.0%	0.0%
Total	2,339,281	2,339,281	2,339,283		

Table 2.3 Total Weighted Household Summaries in Geography 1

Total Weighted Household Summaries by Size, Geography 1

HH_Size	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2011 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
1Psn	40,735	42,335	41,363	-1.5%	2.4%	42,131	-3.3%	0.5%
2Psn	35,888	37,299	41,968	-14.5%	-11.1%	35,332	1.6%	5.6%
3Psn	15,284	15,885	13,347	14.5%	19.0%	16,511	-7.4%	-3.8%
4Psn	23,353	24,270	18,582	25.7%	30.6%	25,813	-9.5%	-6.0%
Total	115,260	119,789	115,260			119,788		

Total Weighted Household Summaries by Workers, Geography 1

HH_Workers	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2011 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
0Wrk	21,601	22,450	18,795	14.9%	19.4%	23,357	-7.5%	-3.9%
1Wrk	53,218	55,309	51,364	3.6%	7.7%	55,908	-4.8%	-1.1%
2Wrk	35,731	37,135	39,348	-9.2%	-5.6%	35,965	-0.7%	3.3%
3+Wrk	4,710	4,895	5,753	-18.1%	-14.9%	4,558	3.3%	7.4%
Total	115,260	119,789	115,260			119,788		

Total Weighted Household Summaries by Income Level, Geography 1

HH_Inc_Level	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2011 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
$ 0-49.99K	34,763	36,129	28,563	21.7%	26.5%	38,134	-8.8%	-5.3%
$ 50-99.99K	40,278	41,861	45,309	-11.1%	-7.6%	40,234	0.1%	4.0%
$ 100-149.99K	25,556	26,560	29,027	-12.0%	-8.5%	25,438	0.5%	4.4%
$ 150K	14,662	15,239	12,361	18.6%	23.3%	15,983	-8.3%	-4.7%
Total	115,260	119,789	115,260			119,788		

Total Weighted Household Summaries by Vehicles Available, Geography 1

HH_Vehs_Av	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2011 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
0 Veh	15,144	15,739	12,385	22.3%	27.1%	16,631	-8.9%	-5.4%
1 Veh	47,556	49,425	47,407	0.3%	4.3%	49,472	-3.9%	-0.1%
2 Veh	38,133	39,631	39,220	-2.8%	1.0%	39,280	-2.9%	0.9%
3+ Veh	14,427	14,994	16,248	-11.2%	-7.7%	14,405	0.2%	4.1%
Total	115,260	119,789	115,260			119,788		

Table 2.4 Total Weighted Household Summaries in Geography 2

Total Weighted Household Summaries by Size, Geography 2

HH_Size	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2012 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
1Psn	33,251	30,945	37,561	-11.5%	-17.6%	30,058	10.6%	2.9%
2Psn	33,407	31,090	33,719	-0.9%	-7.8%	31,026	7.7%	0.2%
3Psn	19,332	17,991	18,479	4.6%	-2.6%	18,167	6.4%	-1.0%
4Psn	28,218	26,261	24,449	15.4%	7.4%	27,037	4.4%	-2.9%
Total	114,208	106,287	114,208			106,288		

Total Weighted Household Summaries by Workers, Geography 2

HH_Workers	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2012 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
0Wrk	24,282	22,598	19,087	27.2%	18.4%	23,666	2.6%	-4.5%
1Wrk	46,052	42,858	48,929	-5.9%	-12.4%	42,267	9.0%	1.4%
2Wrk	38,011	35,375	40,614	-6.4%	-12.9%	34,840	9.1%	1.5%
3+Wrk	5,863	5,456	5,578	5.1%	-2.2%	5,515	6.3%	-1.1%
Total	114,208	106,287	114,208			106,288		

Total Weighted Household Summaries by Income Level, Geography 2

HH_Inc_Level	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2012 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
$ 0-49.99K	27,548	25,637	25,301	8.9%	1.3%	26,099	5.6%	-1.8%
$ 50-99.99K	35,506	33,043	41,581	-14.6%	-20.5%	31,794	11.7%	3.9%
$ 100-149.99K	26,306	24,481	27,060	-2.8%	-9.5%	24,326	8.1%	0.6%
$ 150K	24,849	23,126	20,266	22.6%	14.1%	24,069	3.2%	-3.9%
Total	114,208	106,287	114,208			106,288		

Total Weighted Household Summaries by Vehicles Available, Geography 2

HH_Vehs_Av	Weight 1	Weight 2	HTS	Pct Diff Weight 1	Pct Diff Weight 2	2012 GF HTS	Pct Diff Weight 1	Pct Diff Weight 2
0 Veh	8,654	8,054	6,900	25.4%	16.7%	8,414	2.9%	-4.3%
1 Veh	44,247	41,178	46,142	-4.1%	-10.8%	40,789	8.5%	1.0%
2 Veh	40,460	37,653	42,156	-4.0%	-10.7%	37,305	8.5%	0.9%
3+ Veh	20,848	19,402	19,010	9.7%	2.1%	19,780	5.4%	-1.9%
Total	114,208	106,287	114,208			106,288		

To ensure the reliability of reweighted HTS, two analyses have been conducted. First, total weighted households have been summarized by jurisdiction (Table 2.5) and the ratio of that by 2007 ACS number are also calculated for each jurisdiction. Note that the 2007/2008 HTS column represents previous comparisons done by MWCOG with the survey.

In several cases, the number of households, by jurisdiction, decreased in the HTS sample we received. This is evident in the case of Spotsylvania County, where no new survey records from the geo-focused surveys were added to the dataset, but the total sample size was lower than MWCOG's comparison 82 versus

85 households) and household weights for the county were lower in our sample than the 2007 ACS and MWCOG's comparison (33,000 households versus 42,000). Ultimately, we decided that the discrepancy for Spotsylvania County is not one that requires action, since the county is on the fringes of the region.

A discrepancy also exists for the City of Falls Church, where about 150 records from the geo-focused surveys were added to the household dataset. This results in the weighted household total for the jurisdiction being about 40 percent higher than 2007 ACS. Because the City of Falls Church is very small in comparison to the region, we decided this discrepancy was not critical.

For other jurisdictions, reweighted household totals match 2007 ACS well in Table 2.5.

Table 2.5 Total Weighted Household Summaries by Jurisdiction

Jurisdiction	Unweighted HHs	Sum of WEIGHT1	Sum of WEIGHT1 /ACS	Sum of WEIGHT2	Sum of WEIGHT2 /ACS	2007/2008 HTS Samples	2007/2008 HTS HHS	HTS/ACS	2007 ACS
Anne Arundel County	726	190,801	1.01	191,108	1.01	764	190,802	1.01	188,874
Arlington County	882	91,004	0.99	90,690	0.99	749	91,200	1.00	91,529
Calvert County	102	29,255	1.00	29,302	1.01	104	29,255	1.00	29,141
Carroll County	393	60,510	1.03	60,608	1.03	416	60,559	1.03	58,783
Charles County	566	51,237	1.05	49,573	1.01	232	50,062	1.02	49,001
City of Alexandria	730	64,414	1.04	63,725	1.03	408	61,429	0.99	61,822
City of Fairfax	22	8,250	1.02	8,263	1.02	22	8,067	1.00	8,063
City of Falls Church	170	6,143	1.39	5,717	1.30	22	4,427	1.01	4,405
City of Fredericksburg	45	9,028	1.01	9,043	1.01	5	9,028	1.01	8,976
City of Manasass Park	7	4,087	1.03	4,093	1.03	67	3,807	0.96	3,973
City of Manassas	62	11,649	0.98	11,668	0.99	46	11,919	1.01	11,830
Clarke County	54	5,331	0.99	5,340	0.99	55	5,317	0.99	5,384
District of Columbia	2,713	246,475	0.98	246,206	0.98	1658	252,124	1.00	251,039
Fairfax County	1,991	360,443	0.98	360,949	0.99	1489	367,961	1.00	366,243
Fauquier County	85	22,936	0.99	22,973	0.99	86	22,936	0.99	23,243
Frederick County	707	83,537	1.02	84,618	1.03	363	82,603	1.01	81,861
Howard County	518	99,271	1.00	99,430	1.01	540	99,393	1.01	98,866
Jefferson County	111	18,693	1.00	18,723	1.01	113	18,693	1.00	18,626
King George County	56	8,227	1.00	8,241	1.00	56	8,228	1.00	8,248
Loudoun County	746	94,949	1.10	93,558	1.08	383	92,204	1.06	86,607
Montgomery County	2,401	343,621	1.00	344,517	1.00	1593	342,382	1.00	343,540
Prince George's County	1,855	296,383	1.00	297,060	1.00	1475	294,326	0.99	297,614
Prince William County	703	123,668	1.01	124,332	1.01	436	124,264	1.01	122,984
Spotsylvania County	82	33,420	0.80	33,474	0.80	85	41,436	1.00	41,602
St. Mary's County	106	36,551	0.99	36,610	0.99	108	36,562	0.99	36,841
Stafford County	151	39,397	1.00	39,461	1.00	160	39,398	1.00	39,419
Grand Total	15,984	2,339,281	1.00	2,339,281	1.00	11435	2,348,382	1.00	2,338,512

In Table 2.6, the weighted distribution of households by household size, workers, income level, and vehicles available is compared against 2007-2011 5-year ACS Public Use Microdata Sample (PUMS). The patterns of all distribution are similar.

Table 2.6 Total Weighted Household Summaries: ACS and HTS

Total Weighted Household Summaries by Size

HH_Size	2007-2011 ACS 5YR PUMS	2007_2008HTS	Weight 1	Weight 2
1psn	26.94%	28.25%	28.04%	28.05%
2psn	31.25%	30.68%	30.41%	30.42%
3psn	16.56%	16.35%	16.47%	16.47%
4+psn	25.25%	24.71%	25.07%	25.07%
Total	100.00%	100.00%	100.00%	100.00%

Total Weighted Household Summaries by Workers

HH_Workers	2007-2011 ACS 5YR PUMS	2007_2008HTS	Weight 1	Weight 2
0Wrk	16.69%	17.22%	17.56%	17.55%
1Wrk	41.24%	40.29%	40.24%	40.26%
2Wrk	33.82%	36.38%	36.12%	36.12%
3+Wrk	8.25%	6.11%	6.08%	6.08%
Total	100.00%	100.00%	100.00%	100.00%

*Aggregated data including no-family households and GQ population

Total Weighted Household Summaries by Income Level

HH_Income	2007-2011 ACS 5YR PUMS	2007_2008HTS	Weight 1	Weight 2
$ 0-49.99K	27.43%	21.69%	22.06%	22.06%
$ 50-99.99K	30.78%	34.67%	34.20%	34.21%
$ 100-149.99K	19.66%	28.50%	28.32%	28.32%
$ 150K	22.12%	15.14%	15.43%	15.41%
Total	100.00%	100.00%	100.00%	100.00%

Total Weighted Household Summaries by Vehicles Available

HH_Vehs_Av	2007-2011 ACS 5YR PUMS	2007_2008HTS	Weight 1	Weight 2
0veh	8.97%	5.82%	6.01%	6.02%
1veh	31.85%	31.42%	31.35%	31.34%
2vehs	37.38%	40.50%	40.38%	40.38%
3+vehs	21.80%	22.26%	22.26%	22.26%
Total	100.00%	100.00%	100.00%	100.00%

Ultimately, since both Weight 1 and Weight 2 suggest very similar results, it likely does not matter which of the two is used for mode choice model estimation. We proceeded with use of Weight 1 for model estimation.

2.4 Merging and Weighting TOS and HTS Data

For mode choice model estimation, trip data from the transit on-board surveys and trip data from the merged HTS were merged to form a new dataset. This section discusses the basic steps used in this process.

Approach

The basic approach used to merge the TOS and HTS trip records follows the steps outlined below:

1. The sum of expansion factors for each of the TOS surveys was computed. These values were called *control totals*, and were used in later steps.
2. The three transit surveys determined to be suitable for use in mode choice model estimation (excluding the VRE survey) were merged. A new weight was appended to the merged dataset, which rescaled the expansion factors on the data so that the sum of the new weight equaled the number of records. Call this new weight *TOS interim weight*.
3. For transit data in the HTS, a new weight was appended to the HTS transit records in a similar way to the new weight generated for the TOS data. The expansion factors were rescaled so that the sum of the new weight was equal to the number of records. This new weight was called the *HTS interim weight*.
4. HTS transit data was segmented by mode, considering the following modes:
    a. Bus
    b. Metrorail
    c. VRE commuter rail
        i. VRE HTS trip records were identified by examination of mode codes (VRE is a commuter rail mode) and production and attraction location proximity to VRE stations.
    d. Non-VRE commuter rail
5. The HTS data was merged with the TOS data.
6. The final weight was computed for the transit data.
    a. For non-transit HTS the expansion factors from the existing merged HTS data were unchanged.
    b. For each segment of transit data (bus, Metrorail, VRE, non-VRE commuter rail), the two new weights computed above (i.e., *TOS interim weight* and *HTS interim weight*) were rescaled so the new weight summed to the corresponding *control total* computed above for that mode.

More specifically, the following equations were used:

$$IW_{TOS,m,i} = ExpFac_{TOS,m,i}\left(\frac{N_{TOS}}{\sum_{m,i} ExpFac_{TOS,m,i}}\right) \quad (2.6)$$

$$IW_{HTS,m,i} = ExpFac_{HTS,m,i}\left(\frac{N_{HTS}}{\sum_{m,i} ExpFac_{HTS,m,i}}\right) \quad (2.7)$$

$$FW_{TOS,m,i} = IW_{TOS,m,i}\left(\frac{\sum_i ExpFac_{TOS,m,i}}{N_{TOS,m}+N_{HTS,m}}\right) \tag{2.8}$$

$$FW_{HTS,m,i} = IW_{HTS,m,i}\left(\frac{\sum_i ExpFac_{HTS,m,i}}{N_{TOS,m}+N_{HTS,m}}\right) \tag{2.9}$$

Here, $IW_{TOS,m,i}$ is the TOS interim weight for transit mode *m* and record *i*, $IW_{HTS,tr,i}$ is the HTS interim weight for HTS transit record *i*, $ExpFac$ refers to the original expansion factors in the datasets (for HTS, the resulting expansion factors after merging the three HTS datasets), $N_{TOS,m}$ is the total number of trip records in the TOS surveys for transit mode *m*, and $N_{HTS,m}$ is the total number of non-VRE transit trip records in the HTS for transit mode *m*. $FW_{TOS,m,i}$ is the final weight for TOS record *i* and $FW_{HTS,m,i}$ is the final weight for HTS transit record *i*.

During the process of reweighting, several sets of trip records were dropped from various datasets. This was done if the coded access mode was not consistent with the model that will be estimated (e.g., missing access mode code). This was particularly an issue with the HTS data and the bus TOS data. This approach is consistent with the approach MWCOG used to generate bus mode targets for their 2010 calibration efforts.

Weighted Trip Summaries

By weighting trip records on the basis of transit mode, it ensures the sample used in mode choice model estimation reflects accurately the types of transit used by travelers. It is worth noting that if HTS weights as controls (rather than TOS weights), similar final weights would have ultimately resulted. The one exception to this is for VRE records, where the original HTS weights summed to about three times larger than the observed number of VRE trips of the TOS.

Table 2.7 summarizes the merged and TOS datasets in terms of raw number of trip records and expanded trip totals. The majority of transit records come from the TOS surveys, except for VRE trips. Moreover, the expanded trip totals of the merged transit data exactly match the TOS expansion totals, by design.

Table 2.7 HTS Versus TOS Data Summaries

	Counts		Expanded Trips	
Mode	**Merged Data**	**TOS**	**Merged Data**	**TOS**
Auto	94,455	0	16,799,969	n/a
Bus	14,249	12,003	369,537	369,537
Metrorail	73,270	69,040	770,754	770,754
VRE Commuter Rail	89	0	7,245	7,245
MARC Commuter Rail	849	763	26,451	26,451
Total Transit	183,362	81,806	1,173,987	1,173,987

Other Details

The Metrorail TOS income categories are not consistent with the other surveys or the travel demand model. In particular, the travel demand model income categories include:

- $0-50,000
- $50,000-100,000
- $100,000-150,000
- $150,000 or more

The Metrorail TOS income categories, however, consistent of the following four categories:

- $0-36,000
- $36,000-65,000
- $65,000-100,000
- $100,000 or more

As a result, the first and third income categories for the Metrorail survey can be considered to be from income categories 1 and 2 for the model, respectively. Income category 2 of the Metrorail survey is split about evenly between the model's first two income categories, and income category 4 of the Metrorail survey encompasses all of income categories 3 and 4 of the model.

To deal with this issue in model estimation, we analyzed income data for the U.S. from 2010 Census and for the Washington D.C. region from the ACS (5-year, 2006-2010). Based on our analysis (see Figure 1), we estimated that of households in the income range of $36,000 to $65,000 in the Washington region, 47 percent had incomes lower than $50,000 and 53 percent had incomes higher than $50,000. Further, we estimated that of households in the income range of $100,000 or more, 46 percent had incomes lower than $150,000 and 54 percent had incomes higher than $150,000.

Figure 2.1 Washington Area Income Group Estimates for Metrorail TOS

US 2010 Census Data

Household Income Range	Total Households (in 1000s)
$35,000 to $39,999	5,965
$40,000 to $44,999	5,679
$45,000 to $49,999	4,987
$50,000 to $54,999	5,133
$55,000 to $59,999	4,240
$60,000 to $64,999	4,459
$65,000 to $69,999	3,612
$70,000 to $74,999	3,793

- Of $35-40K, assume 20% with Income < $36K
- Then, 93% of HHs in $35-50K range are also in $36-50K range

- 65% of HHs in $50-75K range are also in $50-65K range

DC Area 2006-10 ACS Data

Household Income Range	Total Households
$35,000 to $50,000	179,375
$50,000 to $75,000	293,842
$100,000 to $150,000	362,011
$150,000 to $200,000	198,107
$200,000 or more	220,536

- 0.93 * 179,375 = 165,508 ($36-50K)
- 0.65 * 293,842 = 191,384 ($50-65K)
- 165,508 / (165,508+191,384) = 47%
- 191,384 / (165,508+191,384) = 53%

- 198,107 / (198,107+220,536) = 46% ($100-150K)
- 220,536 / (198,107+220,536) = 54% ($150K or more)

Based on this information, we developed an approach for model estimation that will not bias the coefficient estimates related to the income variables. We will replicate all trip records from income categories 2 and 4 in the Metrorail survey. We will treat the first replication of income 2 trip records as being from the model's income category 1 and give a weight to these records that is 0.47 times the existing weight (as noted above). We will treat the second replication of the record as being from the model's income category 2 and give it a weight that is 0.53 times the existing sample weight. For Metrorail trip records from income category 4, the first replication in the estimation dataset will be given a weight that is 0.46 times the existing weight, and the second replication of will be given a weight that is 0.54 times the existing weight.

Conceptually, this approach is superior to other approaches, because when the model is estimated, the likelihood values of the record falling into either income category is explicitly computed and weighted appropriately. Another approach would be to randomly assign each record into one or the other income categories, but this leads to an element of randomness in the model estimates themselves (depending on how records are split into income categories). Another option would be to average the utility contributions from each possible income category within a single record in model estimation (e.g., the income 1 dummy variable equals 0.47 and the income 2 dummy variable equals 0.53), but this has the possibility of biasing results.

3.0 Non-Motorized Model Enhancements/Estimation

This section documents non-motorized model estimation results for the MWCOG travel demand model. Normally, discrete choice models undergo the following steps:

1. Disaggregate estimation (trip/person/household level)
2. Disaggregate validation/calibration (this step is sometimes omitted)
3. Aggregate validation/calibration.

This report section addresses the first step. The second step is beyond the scope of the FY 17 effort and is omitted. The third step is discussed in Section 7.0, Model Validation.

3.1 Data for Model Estimation

The dataset used for non-motorized model estimation is a merged dataset from three household travel surveys conducted in the MWCOG region:

- 2007/08 Household Travel Survey (HTS)
- 2011 Geo-Focused HTS
- 2012 Geo-Focused HTS

The data consist of trip records where the key defining attributes of a trip that are used in model estimation include its production and attraction locations, purpose, and mode, as well as traveler characteristics such as household income.

Aggregate data at the Census block and TAZ levels were merged with the household survey data sets, including standard land use variables in the COG/TPB Zone file, and additional new variables representing urban design and the built environment such as land use density, diversity, and design. Table 3.1 shows variables and their descriptions in the merged data set. These variables are discussed later in this report section.

Section 2.0 of this technical report discusses these datasets, how they were merged, and how records from each were reweighted. That report section documents the procedure to create a weighting factor to represent the general population in the region after merging the HTS, geo-focused surveys, and transit on-board surveys. The final weight after all merging, TRANWGHT, was used later in our tabulations, and is actually an expansion factor to factor the survey sample to the whole population.

To generate a final set of weights used for model estimation, it is critical that the sum of weights equal the number of observations in the dataset. This is important because it impacts the size of the statistics generated in the model's estimation, including the magnitude of t-statistics. In the case of our dataset, the sum of weights should match the original number of observations.

Table 3.1 Data and Variables for the Non-Motorized Travel Model Development

Variable	Name	Description
Household Survey Data		
HHin_1to4	Household Income	Range 1-4: '1'/<$50K, '2'/$50K-$99.999K, '3'/$100K-149.999K, '4'/>$150K (2007 dollar)
HHwk_1to4	Number of Workers	Range 1-4: '1'/zero workers, '2'/1 worker, '3'/2 workers, '4'/3+ workers
HHva_1to4	Number of Vehicles in HH	Range 1-4: '1'/0-veh HHs, '2'/1-veh HHs, '3'/2-veh HHs, '4'/3+veh HHs
HHSZ_1to4	Household Size	Range 1-4: '1'/1 Psn HHs, '2'/2 Psn HHs, '3'/3 Psn HHs, '4'/4+Psn HHs
TAZ-Level Variables		
HH	Households	Households
HHPOP	Household Population	Household Population
GQPOP	Group Quarters Population	Group Quarters Population
TOTPOP	Total Population	Total Population
TOTEMP	Total Employment	Total Employment
INDEMP	Industrial Employment	Industrial Employment
RETEMP	Retail Employment	Retail Employment
OFFEMP	Office Employment	Office Employment
OTHEMP	Other Employment	Other Employment
JURCODE	Jurisdiction Code	Jurisdiction Code
LANDAREA	Gross Land Area	Gross Land Area
HHINCIDX	Household Income Index	Ratio of zonal HH median income to regional median HH income in tenths (e.g., a value of '10' implies a ratio of 1, meaning the TAZ's ratio equals the regional ratio)
ADISTTOX	Airline Distance	Airline distance to the nearest external station in whole miles
TAZXCRD	TAZ X-coordinate	TAZ X-coordinate
TAZYCRD	TAZ Y-coordinate	TAZ Y-coordinate
POP_10	Floating Population	One-mile "floating" population
EMP_10	Floating Employment	One-mile "floating" employment
AREA_10	Floating Area	One-mile "floating" area
POPDEN	Population Floating Density	One-mile "floating" population density
EMPDEN	Employment Floating Density	One-mile "floating" employment density
POPCODE	Population Code	Population density code (ranging 1-7, based on POPDEN) *

Variable	Name	Description
EMPCODE	Employment Code	Employment density code (ranging 1-7, based on EMPDEN) *
ATYPE	Area Type	Area Type Code (1-6)
HBWParkCos	HBW Parking Costs	8-Hour Parking cost for HBW trips in 2007 cents
HBSParkCos	HBS Parking Costs	1-Hour parking duration for HBS trips in 2007 cents
HBOParkCos	HBO Parking Costs	2-Hour parking duration for HBO trips in 2007 cents
NHBParkCos	NHB Parking Costs	2-Hour parking duration for NHB trips in 2007 cents
HB_TermTim	HB Trips Terminal Time	Home based trips terminal time (1-5 minutes)
NHB_TermTi	NHB Trips Terminal Time	Non-home based trips terminal time (1-5 minutes)
MetroShort	% TAZ within Short Walk Distance Metro	% of TAZ that is w/in short walk distance (0.5mi) to Metrorail
MetroLong	% TAZ within Long Walk Distance Metro	% of TAZ that is w/in long walk distance (1.0mi) to Metrorail
AMShort	% TAZ within Short Walk Distance AM	% of TAZ that is w/in short walk distance (0.5mi) to AM transit
AMLong	% TAZ within Long Walk Distance AM	% of TAZ that is w/in long walk distance (1.0mi) to AM transit
OPShort	% TAZ within Short Walk Distance Off-peak	% of TAZ that is w/in short walk distance (0.5mi) to OP transit
OPLong	% TAZ within Long Walk Distance Off-peak	% of TAZ that is w/in long walk distance (1.0mi) to OP transit
MCDistrict	Mode Choice Market Segment	Mode Choice Geographic Market Segment **
NO3WAYINTS	Number of 3-way Intersections	Number of 3-way intersections in TAZ
NO4WAYINTS	Number of 4-way Intersections	Number of 4-way intersections in TAZ
NUM_CULS	Number of Cul-de-sacs	Number of cul-de-sacs in TAZ
TOTAL_INTS	Total Intersections	Number of total intersections in TAZ
LANDACTIVI	Land Use Activity	Land use activity (Sum of total Pop. and total Emp)
POPSHARE	Population Share	Population/ (Population+Employment)
EMPSHARE	Employment Share	Employment/ (Population+Employment)
SIMPSONIDX	Simpson's diversity index	Simpson's diversity index (an index of the different elements in the zone, in this case, population and employment, with 0.5 representing equal distribution and 1 indicating homogeneous land use in a zone)

Variable	Name	Description
ENTROPYIDX	Entropy index	Entropy (measuring homogeneity of land use in a given area, with a value of 0 representing homogeneous land use and 1 indicating evenly distributed land uses)
numStops	Number of Stops/Stations	Number of stops/stations within the TAZ
minDistanc	Distance to the Nearest Transit Stop/Station	Distance to the nearest stop/station (ft)
ITZFD_34Q	Intersection Floating Density	Intersection TAZ floating density: 3- or 4-leg intersections within 1/4 mile
ITZFD_34O	Intersection Floating Density	Intersection TAZ floating density: 3- or 4-leg intersections within 1 mile
ITZFD_CSQ	Intersection Floating Density	Intersection TAZ floating density: cul-de-sac intersections within 1/4 mile
ITZFD_CSO	Intersection Floating Density	Intersection TAZ floating density: cul-de-sac intersections within 1 mile
STZFD_Q	Transit Stop Floating Density	Stop floating density within a quarter mile
STZFD_O	Transit Stop Floating Density	Stop floating density within one mile
Block-Level Variables		
BKPOP10	Population	2007 Population
BKEMP10	Employment	2007 Employment
BKLAND_AREA	Land Area	Land Area (sq mi)
BKLANDACTIVITY	Land Activity	Sum of total Population and total Employment
BKPOPSHARE	Population share	Population share of total land activity
BKEMPSHARE	Employment share	Employment share of total land activity
BKSIMPSONIDX	Simpson Diversity Index	Simpson Diversity Index
BKENTROPHYIDX	Entropy	Entropy
BKnumStops	Number of Transit stops/stations	Number of transit stops/stations
BKminDistance	Distance to the Nearest Transit Stop/Station	Distance to the nearest transit stop/station (ft)
BKPOP25	Floating Population	1/4 mile floating population
BKEMP25	Floating Employment	1/4 mile employment
BKAREA25	Floating Area	1/4 mile land area
BKPOPDEN	Population Floating Density	1/4 mile floating population density
BKEMPDEN	Employment Floating Density	1/4 mile employment density
BKIBKFD_34Q	Intersection Floating Density	Intersection block floating density: 3- or 4-leg intersections 1/4 mile
BKIBKFD_34O	Intersection Floating Density	Intersection block floating density: 3- or 4-leg intersections 1 mile

Variable	Name	Description
BKIBKFD_CSQ	Intersection Floating Density	Intersection block floating density: cul-de-sac intersections 1/4 mile
BKIBKFD_CSO	Intersection Floating Density	Intersection block floating density: cul-de-sac intersections 1 mile
BKSBKFD_Q	Transit Stop Floating Density	Stop floating density within a quarter mile
BKSBKFD_O	Transit Stop Floating Density	Stop floating density within one mile

Source: 2007/08 Household Travel Survey (HTS), 2011 Geo-Focused HTS, and 2012 Geo-Focused HTS. COG/TPB.
* areaType_file.dbf, p. 108 of User's Guide for the COG/TPB Travel Demand Forecasting Model, Version 2.3.66, Volume 1 of 2: Main Report and Appendix A (Flowcharts), February 13, 2017
** P. 170 of User's Guide

3.2 Model Structure

The mode choice model is structured as a binary logit (BL) model of the choice of non-motorized vs motorized mode used on a particular trip. The dependent variable, mode, is a binary variable:

- 1 = Non-motorized
- 0 = Motorized

Binary modal splits were estimated at both production and attraction ends.

A binary logit model is a special case of a multinomial logit model. In this binary model, the non-motorized mode is assigned a utility function where utilities are defined as follows:

$$U_{im} = \beta_1 X_{1,im} + \beta_{2m} X_{2,im} + \varepsilon_{im} \tag{3.1}$$

Here, β_1 is a vector of parameters to be estimated that are generic across all modes (e.g., typically, cost or time parameters in the mode choice context), β_{2m} is a vector of parameters to be estimated that are alternative-specific (e.g., constants), $X_{1,im}$ and $X_{2,im}$ are variables associated with trip *i* and mode *m*, and ε_{im} is a independent and identically distributed error term (assumed to have Gumbel distribution in the case of MNL model).

Separate models were estimated for each of five trip purposes that will be applied separately in the COG/TPB travel demand model. These include the following:

- Home-based work (HBW)
- Home-based shopping (HBS)
- Home-based other (HBO)
- Non-home-based work (NHW)
- Non-home-based other (NHO)

Table 3.2 shows the unweighted frequency distribution of walk trip records by trip purposes and household income categories in the merged dataset, while the weighted distribution is included in Table 3.3. Similarly, unweighted and weighted distributions of bike trips by trip purposes and household income categories are shown in Table 3.4 and 3.5, respectively.

Table 3.2 Walk Trip Distributions by Trip Purposes and Household Income Categories from the Household Travel Surveys (Unweighted)

	Household Income Categories (2007 dollars)				
Trip Purpose	Under $50K	$50K-$99.999K	$100K-$149.999K	>$150K	Grand Total
HBW	163	302	221	151	837
HBS	543	562	413	282	1,800
HBO	1,045	1,242	1,393	976	4,656
NHW	374	1,008	1,210	1,095	3,687
NHO	749	715	689	589	2,742
Grand Total	2,874	3,829	3,926	3,093	13,722

Table 3.3 Walk Trip Distributions by Trip Purposes and Household Income Categories from the Household Travel Surveys (Weighted)

	Household Income Categories (2007 dollars)				
Trip Purpose	Under $50K	$50K-$99.999K	$100K-$149.999K	>$150K	Grand Total
HBW	17,643	31,138	23,909	13,282	85,972
HBS	65,306	63,263	41,871	22,125	192,564
HBO	120,294	169,548	234,773	99,139	623,754
NHW	40,783	113,899	153,923	111,290	419,895
NHO	87,110	92,737	80,110	54,043	314,001
Grand Total	331,136	470,585	534,587	299,878	1,636,186

* TRANWGHT was used to expand the survey data.

Table 3.4 Bike Trip Distributions by Trip Purposes and Household Income Categories from the Household Travel Surveys (Unweighted)

	Household Income Group (2007 dollars)				
Trip Purpose	Under $50K	$50K-$99.999K	$100K-$149.999K	>$150K	Grand Total
HBW	33	88	96	106	323
HBS	21	34	35	29	119
HBO	40	114	98	107	195
NHW	11	21	26	32	90
NHO	15	37	18	38	108
Grand Total	120	294	273	312	999

Table 3.5 Bike Trip Distributions by Trip Purposes and Household Income Categories from the Household Travel Surveys (Weighted)

	Household Income Group (2007 dollars)				
Trip Purpose	Under $50K	$50K-$99.999K	$100K-$149.999K	>$150K	Grand Total
HBW	3,346	9,098	12,599	12,600	37,643
HBS	2,349	3,495	4,422	1,620	11,886
HBO	5,808	15,008	11,557	10,529	42,902
NHW	365	1,691	4,074	2,905	9,035
NHO	1,997	3,065	1,633	2,638	9,333
Grand Total	13,866	32,357	34,285	30,291	110,799

* TRANWGHT was used to expand the survey data.

In Table 3.5, the regional control total for bike trips (110,799) is close to, but does not match, the regional control total in a recent TPB staff memo (91,699).[2] This difference is attributed to the fact that the values in Table 3.5 are factored by TRANWGHT.

3.3 Variables and Assumptions

This section discusses key variables that were tested in the non-motorized model specifications. Land use and urban form variables at the production and attraction ends of the trips include measures such as residential and employment density, land use mix and diversity, and urban design. Some of the variables are derived from the basic zonal/Census-block-level variables such as density and diversity variables.

The "floating" method is used to calculate land use density, design, and transit stops/station variables at the TAZ and Census block level, with two levels of buffers:

- One-mile buffer
- Quarter-mile buffer

The "floating" method calculates the density of a certain activity around a desired location, which is often measured on the basis of a buffer area surrounding the desired point location. The buffers were defined and computed from the centroid of each TAZ/Census Block. It is hypothesized that non-motorized modal share would be positively related to the variables of population and employment density.

The density of transit stops/stations is one measure of accessibility to transit. Since transit services tend to be provided in dense areas, this variable tends to be positively correlated with land use density variables such as population and employment density.

[2] Ronald Milone to Files, Mark Moran, and Dzung Ngo, "Household Travel Survey (HTS) Files for Transmittal to CS," Memorandum, (February 6, 2017).

Diversity Variables

Entropy is calculated as follows:

$$E_i = -\sum_j p_j * \ln(p_j)/\ln(J) \tag{3.2}$$

Where p_j is the proportion of land use in the j^{th} land use category and J is the total number of different land use type classes in the area. In this case, land uses include residential and employment by types (retail, office, industrial, and other). Entropy measures the degree of homogeneity of land use in a given area, with a value of 0 representing homogeneous land use and 1 indicating evenly distributed land uses. The higher an entropy value at a TAZ or block, the more diverse the land uses are. It is hypothesized that non-motorized modal share would be positively related to the entropy values.

Simpson's diversity index was measured as follows:

$$\lambda_i = 1 - \sum_m {p_{im}}^2 \tag{3.3}$$

Here, m is an index of the different elements in the zone, in this case, population and employment, and p_{im} is the share of the total population and employment for that element. For instance, if population and employment had an equal number in the zone, the share of each would be 0.5 and the diversity index would be 0.5. If there was no employment, however, then the population share would be 1 and the diversity index would be 0.

Urban Design Variables

Urban design variables represent the built environment and are measured in terms of density of intersections by types. Floating density of 3- or 4-leg intersections is used as a proxy measure for urban design; urban environments tend to have more 3- or 4-leg intersections than suburban and rural areas. In contrast, the floating density of cul-de-sac intersections is used to represent suburban and rural environments, where cul-de-sac nodes are more typical than in the urban environment. It is hypothesized that non-motorized modal share would be positively related to the floating density of 3- or 4-leg intersections and negatively related to the floating density of cul-de-sac intersections.

3.4 Estimation Results

Binary logit non-motorized models were estimated for each trip purpose, including separate estimations using production-end variables and attraction-end variables. The TAZ-level and block-level variables were tested separately, in order to see which level of variables performs better. Tests were conducted for different model specifications; more than one hundred model estimations were performed. Final model estimation results can be found for over forty models estimated in two companion Excel workbook files, one for productions and the other for attractions. Table 3.6 through Table 3.10 show the final TAZ-level model estimate results for productions and attractions by trip purpose.

Major findings of the estimation for the non-motorized modal share models for productions include the following:

- In general, the TAZ-level and block-level model estimation results are quite similar, with only a few cases where the block-level models are slightly better than the TAZ-level models. This is an unexpected result as we would expect that the block-level variables would offer more accurate measurements of urban design and the built environment than the TAZ-level variables. It can be hypothesized that the floating density method of measuring variables tends to have a smoothing effect and may reduce aggregation errors.

- In most cases, estimated parameters for urban design and built environment variables such as density, diversity and design were consistent with our hypotheses on their significance in explaining the non-motorized modal shares. In many cases, these estimated coefficients have expected signs and significance, including coefficients for population and employment floating density variables, entropy and Simpson index, intersection floating density for 3- or 4- legs and cul-de-sac (1 miles in most cases, ¼ miles for a few cases), and transit stop floating density (1 mile in most cases, ¼ miles for a few cases).

- In the non-motorized modal share models for productions, home-based work (HBW) trips are likely to have higher non-motorized shares where there are high employment, more diverse land uses, more 3- or 4-leg intersections, and fewer cul-de-sac streets. All these results are consistent with our expectations. In another model specification, density of transit stops/stations also showed significance in explaining the non-motorized modal shares, but its introduction reduced the significance and magnitude of population density in the model. This result can be attributed to the multicollinearity between population density and density of transit stops/stations. The two model specifications (with and without density of transit stops/stations) have similar goodness-of-measures. Therefore, the choice of the two model specifications would depend on the desired functionality and interests of evaluating policies. For HBW trips, TAZ-level models slightly outperformed block-level models in terms of goodness-of-fit but population density variables in the TAZ-level models are less significant than those in the block-level models.

- For home-based shopping (HBS) trips, non-motorized modal shares for productions show results similar to HBW trips in the types of significant explanatory variables and expected signs of these variables. Different from HBW trips, HBS trips have population density variables with higher significance than employment density variables. When variables related to density of transit stops/stations are introduced, the coefficient estimates for employment density at the TAZ level turned out to have wrong and negative signs. For HBS trips, the block-level models show slightly better performances than the TAZ-level models in terms of goodness-of-fit measures.

- For home-based other (HBO) trips, non-motorized modal shares for productions have model estimation results similar to HBS trips, with smaller goodness-of-fit measures. The block-level models slightly outperformed the TAZ-level models in terms of goodness-of-fit measures.

- For non-home-based work trips (NHW), the estimated coefficient for the entropy measure is not significant at either the block or TAZ level. Instead, the coefficient estimate for the Simpson Index is significant with a negative sign for the TAZ-level models. The block-level models have slightly better goodness-of-fit measures than the TAZ-level models.

- For non-home-based other trips (NHO), entropy measure is significant at the block level but not at the TAZ level. The coefficient estimate for the Simpson Index has a negative sign for the TAZ-level models but is insignificant for the specification without density of transit stops/stations. When the variable of the density of transit stops/stations was introduced, the TAZ-level model has a wrong negative sign for the estimated coefficient for employment density. The block-level models outperform the TAZ-level models with better goodness-of-fit measures and higher significances for independent variables.

Major findings of estimation for the non-motorized modal share models for attractions include:

- In general, the model estimation results for the attraction-side variables are similar to those for the production-side variables. The block-level models slightly outperform the TAZ-level models in terms of goodness-of-fit and significances of independent variables. One major difference is that employment density variables are not significant or with wrong signs in several trip purposes (HBW, HBS, and HBO).
- For HBW trips, neither employment density nor entropy variables are significant in explaining the non-motorized modal shares at the block level. At the TAZ level, employment is not significant for the model specification without density of transit stops/station and has a wrong sign (negative) for the model specification with density of transit stops/station.
- For HBS and HBO trips, the employment density variable was removed for model estimation because of wrong signs. Entropy is significant for HBS, but not for HBO trips.
- For NHW trips, the coefficient estimate for entropy is insignificant or has the wrong sign, and the coefficient estimate for Simpson Index has a negative sign at the TAZ level. Employment density is significant at both TAZ and block levels.
- For NHO trips, both employment and entropy variables are significant at the block, but not at the TAZ level.

Given the small differences between the TAZ-level and block-level models, it is recommended to adopt the TAZ-level models for the sake of reducing resources that would otherwise be needed to generate block-level input data.

Table 3.6 Binary Logit Model for HBW (Non-Motorized versus Motorized)

Parameter	Parameter Estimate	T Test
Production		
CONSTANT	-4.735	-29.9
POPDEN	0.00001332	1.2
EMPDEN	0.00002117	9.3
ENTROPYIDX_	0.8356	5.1
ITZFD_34O	0.003072	9.3
ITZFD_CSO	-0.006501	-7
“Rho-Squared” w.r.t. Zero	.8302	
“Rho-Squared” w.r.t. Constants	.1767	
Attraction		
CONSTANT	-4.339	-31.2
POPDEN	0.00005535	4.7
EMPDEN	0.0000005819	0.3
ENTROPYIDX	0.4441	3.5
ITZFD_34O	0.001164	3.3
ITZFD_CSO	-0.002827	-2.7
“Rho-Squared” w.r.t. Zero	.7975	
“Rho-Squared” w.r.t. Constants	.0496	

Table 3.7 Binary Logit Model for HBS (Non-Motorized versus Motorized)

Parameter	Parameter Estimate	T Test
Production		
CONSTANT	-4.399	-31.6
POPDEN	0.00008908	8.8
EMPDEN	0.000004251	1.6
ENTROPYIDX_	0.7427	5.4
ITZFD_34O	0.003643	13.2
ITZFD_CSO	-0.007536	-9.8
"Rho-Squared" w.r.t. Zero	.7390	
"Rho-Squared" w.r.t. Constants	.2509	
Attraction		
CONSTANT	-4.417	-31.6
POPDEN	0.0001144	11.2
ENTROPYIDX	0.624	5.2
ITZFD_34O_A	0.00333	12.4
ITZFD_CSO_A	-0.008057	-10.5
"Rho-Squared" w.r.t. Zero	.7324	
"Rho-Squared" w.r.t. Constants	.2398	

Table 3.8 Binary Logit Model for HBO (Non-Motorized versus Motorized)

Parameter	Parameter Estimate	T Test
Production		
CONSTANT	-3.467	-50.6
POPDEN	0.00003493	5.8
EMPDEN	0.000007634	4.4
ENTROPYIDX_	0.2678	3.7
ITZFD_34O	0.002727	17.3
ITZFD_CSO	-0.002147	-5.8
“Rho-Squared” w.r.t. Zero	.6159	
“Rho-Squared” w.r.t. Constants	. .0776	
Attraction		
CONSTANT	-3.218	-62.4
POPDEN	0.00006292	10.2
ITZFD_34O	0.001495	9.5
ITZFD_CSO	-0.0002486	-0.7
“Rho-Squared” w.r.t. Zero	.5983	
“Rho-Squared” w.r.t. Constants	.0472	

Table 3.9 Binary Logit Model for NHW (Non-Motorized versus Motorized)

Parameter	Parameter Estimate	T Test
Production		
CONSTANT	-2.765	-18
POPDEN	0.0000361	4.4
EMPDEN	0.00001651	15.1
SIMPSONIDX	-0.6068	-3.6
ITZFD_34O	0.002619	11.2
ITZFD_CSO	-0.002505	-3.5
“Rho-Squared” w.r.t. Zero	.4634	
“Rho-Squared” w.r.t. Constants	.2609	
Attraction		
CONSTANT	-2.949	-18.8
POPDEN	0.00002852	3.5
EMPDEN	0.00001801	16.1
SIMPSONIDX	-0.262	-1.5
ITZFD_34O	0.002779	11.6
ITZFD_CSO	-0.003499	-4.8
“Rho-Squared” w.r.t. Zero	.4765	
“Rho-Squared” w.r.t. Constants	.2838	

Table 3.10 Binary Logit Model for NHO (Non-Motorized versus Motorized)

Parameter	Parameter Estimate	T Test
Production		
CONSTANT	-3.676	-26.6
POPDEN	0.00007209	8.8
EMPDEN	0.00001063	8.8
SIMPSONIDX	-0.2108	-1.3
ITZFD_34O	0.002564	11
“Rho-Squared” w.r.t. Zero	.6159	
“Rho-Squared” w.r.t. Constants	.1517	
Attraction		
CONSTANT	-3.122	-37.8
POPDEN	0.000008405	1
ITZFD_34O	0.002287	7.8
ITZFD_CSO	-0.006536	-9.6
STZFD_O	0.0057	9.9
“Rho-Squared” w.r.t. Zero	.6248	
“Rho-Squared” w.r.t. Constants	.1790	

4.0 Mode Choice Model Enhancements/Estimation

This report section documents mode choice model estimation results for the MWCOG travel demand model.

4.1 Dataset

The dataset used for mode choice model estimation comes from three household travel surveys conducted in the MWCOG region and three transit on-board surveys conducted of bus, commuter rail, and Metrorail transit services in the region. The data consists of trip records where the key defining attributes of a trip that are used in the model include its origin, destination, purpose, time of day, and mode. In addition, the income of the traveler is a characteristic used in the model.

Section 2.0 describes these datasets, how they were merged, and how records from each were reweighted. For additional information on the survey data, the reader is referred to that section.

4.2 Model Structure

The mode choice model is structured as a multinomial logit (MNL) model of the choice of mode used on a particular trip. The dependent variable, mode, is a discrete variable with six alternatives:

- Auto – single-occupancy vehicle (SOV)
- Auto – high-occupancy vehicle, 2 passengers (HOV2)
- Auto – high-occupancy vehicle, 3+ passengers (HOV3)
- Transit – park-and-ride access (PNR)
- Transit – kiss-and-ride access (KNR)
- Transit – walk access (WTR)

Notably, the first three alternatives are all forms of automobile mode with differing occupancy levels, and the last three alternatives are all forms of transit mode with differing access modes. Trips using walk and bicycle modes are modeled earlier in the model process, so for the mode choice model, the collection of trips to be modeled includes only trips using a motorized mode.

In a MNL model, each alternative is assigned a utility function where utilities are defined as follows:

$$U_{im} = \beta_1 X_{1,im} + \beta_{2m} X_{2,im} + \varepsilon_{im} \tag{4.1}$$

Here, β_1 is a vector of parameters to be estimated that are generic across all modes (e.g., typically, cost or time parameters in the mode choice context), β_{2m} is a vector of parameters to be estimated that are alternative-specific (e.g., constants), $X_{1,im}$ and $X_{2,im}$ are variables associated with trip *i* and mode *m*, and ε_{im} is a independent and identically distributed error term (assumed to have Gumbel distribution in the case of MNL model).

Separate MNL models were estimated for each of five trip purposes that will be applied separately in the MWCOG travel demand model. These include the following:

- Home-based work (HBW)

- Home-based shopping (HBS)
- Home-based other (HBO)
- Non-home-based work (NHBW)
- Non-home-based other (NHBO)

Nested logit models were also explored in the course of model estimation testing. As will be discussed later in this report section, the nesting parameters estimated in these models were outside the bounds that would be considered acceptable for a random utility model. Thus, the nested logit model structure was not used in the final model specifications.

Table 4.1 shows the weighted[3] frequency distribution of mode choices in the survey dataset used for model estimation. Note that the mode choice estimation dataset summaries may differ slightly from other summaries of the data due to certain records getting dropped from the estimation dataset for various reasons.[4]

Table 4.1 Mode Choice Frequency Distribution by Trip Purpose

Mode	HBW	HBS	HBO	NHBW	NHBO	Total
Trip Totals						
Drive Alone	35,610	7,967	14,667	10,234	7,931	76,409
HOV 2	3,695	5,281	12,756	1,565	5,502	28,799
HOV 3+	2,162	4,295	11,159	1,128	4,357	23,101
PNR – Transit	2,813	6	142	85	22	3,070
KNR – Transit	632	4	55	38	13	742
Walk – Transit	6,878	182	984	825	244	9,113
Total	51,790	17,736	39,762	13,876	18,070	141,234
Mode Shares						
Drive Alone	68.8%	44.9%	36.9%	73.8%	43.9%	54.1%
HOV 2	7.1%	29.8%	32.1%	11.3%	30.4%	20.4%
HOV 3+	4.2%	24.2%	28.1%	8.1%	24.1%	16.4%
PNR – Transit	5.4%	0.0%	0.4%	0.6%	0.1%	2.2%
KNR – Transit	1.2%	0.0%	0.1%	0.3%	0.1%	0.5%
Walk – Transit	13.3%	1.0%	2.5%	5.9%	1.4%	6.5%

[3] Note that Table 4.1 uses estimation weights, rather than expansion weights. Expansion weights reflect the prevalence of an observation in the total population (e.g., the number of times each observation should be replicated so that an accurate representation of the population emerges). Estimation weights are equivalent to expansion weights, except they are rescaled so that the sum of estimation weights equals the total number of observations, rather than the total population. This ensures the validity of statistics reported for estimated models.

[4] Records can be dropped for a variety of reasons including, incomplete information about the record (e.g., origin or destination) or the chosen mode is considered unavailable due to the characteristics of the model (e.g., transit was the chosen mode but a transit skim for the origin-destination pair was not generated in the skimming procedure).

4.3 Data Weighting in Model Estimation

As shown in Table 4.1, the weights used for model estimation sum to the number of observations in the dataset, rather than sizing the dataset to reflect the population. This is important because it impacts the size of the statistics generated in the model's estimation, including the magnitude of t-statistics.

Using weights in model estimation is typically a requirement only when a choice-based sample is used for data collection, though they are sometimes used more generally in other cases. As a general matter, all logit estimation software comes with the ability to estimate models using Exogenous Sampling Maximum Likelihood (ESML) and Weighted ESML (WESML) methods, with essentially no cost to doing so. Typically, if simple random samples or stratified samples are used, the results of using the ESML or WESML estimators will be negligible, perhaps with the exception of alternative specific constants. With a choice-based sample, typically the prevalence of certain choices in the dataset is highly skewed (e.g., transit is overrepresented by a factor or 2 or 3 or more), and this can impact model estimates, if proper weight are not used.[5] The WESML estimates are obtained by maximizing the following (log) likelihood function:

$$LL = \sum_i w_i \ln\left(\frac{\exp(V_{ik})}{\sum_j \exp(V_{ij})}\right) \tag{4.2}$$

Here, w_i is the sampling weight corresponding to record *i*, and V_{ik} is the systematic utility associated with alternative *k* for observation *i*. The difference between this likelihood function and the ESML likelihood function is the presence of w_i. In the case of ESML, w_i is equal to 1.0 for all records.

4.4 Latent Variables

One key element that was identified to ensure the updated MWCOG travel model was sensitive to managed lane policies was the incorporation of different value of time (VOT) segments into the highway assignment model component. Since it is typically desirable for the time and cost sensitivities used in highway assignment to match mode choice model sensitivities (or at least be close), it was necessary to incorporate VOT segmentation into the mode choice model. The idea behind VOT segmentation is that some travelers have a high VOT while others have a low VOT, and this impacts the behavior. It is clear then that VOT is a traveler attribute, however, we do not observe VOT in our survey dataset. Instead, VOT must be treated as a latent variable.

Upon examination of the survey datasets we had available to us, we identified two other important variables that were not observed in the dataset or were somehow censored. The first such variable was the income variable in the Metrorail transit on-board survey (TOS). While income was recorded for each respondent, the income categories in the survey did not match the income categories used in the other survey or used by the model. The second such variable was the vehicle occupancy variable for auto passenger records from the household travel survey. In such cases, we know that the traveler used the auto mode and we know at least one other person traveled in the automobile (since there must be a driver of the vehicle), but we could not distinguish between occupancy of two and occupancy of three or more, which are distinct modes that are considered in the mode choice model.

[5] See discussion in Ben-Akiva and Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, p. 239.

Thus, we have three latent variables:

- Income for Metrorail TOS records
- Vehicle occupancy for auto passenger records
- VOT for all records

For each of these three variables, some or all information needed by the model is not observed. This can be treated in an MNL model through latent variable techniques. This section describes the approach and details for each of these variables.

Latent Variable MNL Model

It is clear that any variable can take on any of some set of possible values. A latent variable is one that we do not observe its value. However, we can still make statements about the variable. In particular, we can define the set of possible values that the variable can have. For example, vehicle occupancy of auto passengers can be any integer value of 2 or higher. In this case, our model is only sensitive to the difference between the variable being 2 or more than 2, so we can simplify the set of possible values to 2 or 3+.

Next, the distribution of each variable needs to be described in some way. Very generally, we can describe the distribution of each variable as follows:

$$X' = h(X, \gamma) + \eta \tag{4.3}$$

$$\eta \sim D(0, \sigma) \tag{4.4}$$

Here, X' is the latent variable, X is our set of observed variables γ is a vector of unknown parameters, η is a random error term for the latent variable, which has distribution from some distribution family D (e.g., normal distribution), with mean zero and standard deviation σ.

So in this very general case, it is possible to make the assumption that part of our unobserved variable can actually be explained by our set of observed data, and $h(X, \gamma)$ represents this relationship. This is the most general form of the latent variable model. In our case, we assume that $h(X, \gamma)$ is known. In particular, we assume that $h = 0$ (i.e., the latent variable does not depend on other observed data).

Furthermore, since each of our latent variables is a discrete variable, η is a multinomial error term. And, as will be documented below, the distribution of each variable is assumed to be known based on the information we have, and thus, is an input to model estimation. In other words, while the general form of the latent variable model allows the distributional parameters of each latent variable to be estimated directly, we did not do that, and instead developed that information from other data.

To understand how the model is estimated, first consider the standard, weighted MNL (log) likelihood function shown in equation 2 above. In order to incorporate our discrete latent variables, we change the likelihood as follows:

$$LL = \sum_i w_i \ln\left(\sum_s \rho_s \frac{\exp(V_{isk})}{\sum_j \exp(V_{isj})}\right) \tag{4.5}$$

The difference here, compared to equation 2, is the additional summation over s and the introduction of the term ρ_s. Here, s indexes the discrete outcomes of each latent variable, and ρ_s is the probability associated with observing that outcome for the variable.

To illustrate, suppose we observe a trip with mode equal to park-and-ride (PNR). Also suppose that the probability that this trip has low, medium, and high VOT is 10, 60, and 30 percent, respectively. Each outcome of VOT is associated with a different utility function. So based upon the utility function parameters, we can compute the probability of the trip choosing PNR mode if the trip has low VOT, medium VOT, and high VOT. For instance, suppose that under a low VOT outcome, the probability of PNR (based upon the utility function) is 1 percent, while under medium VOT and high VOT, the probability of PNR is 3 percent and 10 percent, respectively. Under these circumstances, we can compute the likelihood that this trip chooses PNR as the sum of the product of these probabilities (e.g., likelihood = 0.10 * 0.01 + 0.60 * 0.03 + 0.30 * 0.10), which is equal to 0.049. This is the value that would appear for this record inside the natural logarithm shown in equation 5.

For more information on latent variable modeling, the reader is referred to Ben-Akiva et al. (2002).[6]

Equation 5 above and the example in the preceding paragraph are applicable to a single latent variable. However, the method can be easily extended for any number of discrete latent variables by adding summations to the likelihood function across each latent variable in the model.

In terms of implementing the procedures above, since discrete distributions were used for the latent variables, we ultimately replicated the trip records in our dataset with each replication assigned distinct values of each latent variable. This was done for convenience in writing the procedures to estimate the model. The next section discuss more of the details associated with each of the three variables.

Income – Metrorail TOS

Section 2.0 discussed data merging and reweighting. At the end of that section, the income variable for Metrorail TOS data was discussed in detail. To summarize the results, the income variable in the Metrorail TOS is coded as having four possible categories:

- $0,000 – 36,000
- $36,000 – 65,000
- $65,000 – 100,000
- $100,000 or more

The model, and the other travel surveys have income coded as follows:

- $0,000 – 50,000
- $50,000 – 100,000
- $100,000 – 150,000
- $150,000 or more

[6] Moshe Ben-Akiva, Joan Walker, Adriana Bernardino, Dinesh Gopinath, Taka Morikawa, and Amalia Polydoropoulou, *Integration of Choice and Latent Variable Models*, MIT working paper, http://www.joanwalker.com/uploads/3/6/9/5/3695513/benakivawalkeretal_iclv_chapter_2002.pdf.

Thus, Metrorail's income category 1 and income category 3 are fully bounded within the model's income categories 1 and 2, respectively. However, Metrorail's income category 2 spans across the model's income categories 1 and 2, and Metrorail's income category 4 spans across the model's income categories 3 and 4. As such, the previous analysis showed that the appropriate percentage of Metrorail income category 2 records with income less than $50,000 is 47 percent, with the remaining 53 percent greater than $50,000 income. The appropriate percentage of Metrorail income category 4 records with income less than $150,000 is 46 percent, with the remaining 54 percent greater than $150,000 income. Thus, these percentages are used as the latent variable weights, ρ_s, in the likelihood function.

Auto Passengers

For travelers in the HTS choosing the auto driver mode, the vehicle occupancy was also recorded as a characteristic of the trip. For these auto travelers, it was straightforward to place each trip into either drive alone, HOV 2, or HOV 3+ modes. However, for travelers in the HTS choosing the auto passenger mode, vehicle occupancy was not recorded. Since there is always a driver in these instances, it is obvious that all such records must have a chosen mode of either HOV 2 or HOV 3+, but the specific value is unknown, and thus, is treated as a latent variable.

In order to generate probabilities associated with each outcome (HOV2 or HOV3+), the occupancy level frequencies associated with the auto driver records were used. Table 4.2 shows the distribution of drivers and passengers, based upon the auto driver records in the data.

Table 4.2 Distribution of Auto Modes

Mode	Drivers	Passengers (Imputed from Drivers)
Drive Alone	9,694,181	0
HOV2	2,362,906	2,362,906
HOV3+	1,121,674	2,795,420
Total	13,178,761	5,158,326

The values in the 'Drivers' column reflects the total number of weighted observations of auto-driver records in the original HTS datasets. The 'Passengers (Imputed from Drivers)' column reflects the total number of weighted passengers we would expect based on the vehicle occupancy levels of the auto-driver records. By definition, drive alone trips have zero passengers and HOV 2 trips have one passenger each. The passengers for HOV 3+ trips were computed by summing the observed number of passengers in each auto-driver HOV 3+ trip. In this case, the average occupancy of these trips was about 3.5 (or 2.5 passengers per trip). Based on the imputed passenger trip totals from Table 4.2, it is straightforward to derive that about 46 percent of auto passenger trips use HOV 2 mode, with the remaining 54 percent using HOV 3+ mode. Therefore, the weights of 0.46 and 0.54 were used as the latent variable weights, ρ_s, in the likelihood function.

The total number of weighted auto-passenger trips in the HTS data was 5,022,000, which is close to the imputed total obtained using auto-driver records. This suggests the imputed number of auto passengers is reasonably close to the observed number.

VOT Segmentation

As detailed in Section 5.1, a three-level VOT segmentation approach was considered (meaning that VOT could take one of three possible values). In the context of mode choice model estimation, we treated VOT as a latent variable with three possible outcomes, as discussed above.

We expected VOT level to be impacted by income level. That is, we expected the distribution of VOT to be skewed toward lower values for low income travelers and skewed toward higher values for high income travelers. Moreover, we expected HBW trips to have higher VOTs than non-HBW trips. Section 5.1 describes that we computed these distributions on the basis of a variety of data and assumptions. For instance, low income ($0-50K) HBW travelers were computed to fall into the low, middle, and high VOT categories 34, 57, and 9 percent of the time, respectively (see Table 5.6). These percentages were precisely the weights, ρ_s, that were then used in the latent variable likelihood function.

It is worth noting that using these weights is only valid if the other analyses and assumptions made in Section 5.1 are also valid. In other words, if mode choice models are estimated where the implied VOTs do not match those computed in Section 5.1, the weights would need to be recomputed and the mode choice models reestimated to be fully consistent. In theory, this would be possible, but in practice, this was not practical both for consideration of the project schedule and for technical limitations. Specifically, issues often arise in estimating mode choice models from revealed preference data, including issues related to obtaining reasonable estimates of time and cost parameters, which often results in constraints being placed on the coefficient estimates. Indeed, reasonability issues arose in model estimation of the MWCOG models, which required that relationships of certain model coefficients be constrained, as discussed later.

4.5 Variables

This section details the different variables that were tested in the mode choice model specifications. In addition to the variables listed below, a full set of alternative-specific constants (ASCs) were included in each model specification. These variables ensure that the applied model will replicate the observed mode shares from the estimation dataset.

Level-of-service variables

The level-of-service (LOS) variables include travel time, cost, and number of transit boardings. In some mode choice model estimations, travel time is broken into component parts, such as in-vehicle travel time (IVT) and out-of-vehicle travel time (OVT). In the case of our model estimation work, IVT and OVT were combined into a weighted travel time variable, as described in Koppelman and Bhat.[7] Thus, in all of the models, the relative weights of the travel time components were constrained. Table 4.3 shows the weights that were used, which is consistent with the weights used in highway and transit skimming processes.

[7] Frank S. Koppelman and Chandra Bhat, *A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models* (U.S. Department of Transportation, Federal Transit Administration, June 30, 2006), 114, http://www.ce.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf.

Table 4.3 Travel Time Weights

Travel Time Type	Mode	Weight
In-Vehicle Time	Auto modes	1.00
	Local bus	1.00
	Express bus	1.00
	Commuter rail	0.85
	Metrorail	0.90
Out-of-Vehicle Time	Auto terminal time	2.50
	Walk times for transit	2.50
	Drive access to transit	1.50
	Initial Wait (first 7 minutes)	2.50
	Other Wait	1.50

For transit boardings, all models weighted transit boardings by transit mode according to the weights used in the transit skimming processes. These weights were in equivalent units of in-vehicle time as follows:

- Local bus – 15 min
- Express bus – 13 min
- Commuter rail – 5 min
- Metrorail – 3 min

In addition to the above constraints that were used in the case of all models, several other constraints were ultimately adopted in the final model specifications. These constraints are consistent with the VOT memo analysis and skimming procedures.

The sensitivity to cost was segmented by income level in each of the home-based models.[8] The relative cost sensitivities of the income categories were constrained as follows:

- Income 1 – 1.00
- Income 2 – 0.61
- Income 3 – 0.47
- Income 4 – 0.25

The VOTs for each income category were constrained as follows:

- VOT 1 – $2.70/hr
- VOT 2 – $8.29/hr
- VOT 3 – $27.36/hr

[8] In model application, non-home-based trips will not have household characteristics associated with them, and therefore, income is unknown.

With the above constraints, a single parameter associated with LOS can be estimated. This parameter reflects the scale of the sensitivity of LOS characteristics to other, non-LOS variables in the model. Strictly speaking, the estimated parameter is the cost sensitivity of income category 1, but sensitivities associated with all other LOS variables can be derived from this parameter on the basis of the constraints detailed above. For instance, if the value of the LOS parameter were doubled, the impact of each minute of travel time and the impact of each dollar cost to the traveler's utility functions for each mode would double. However, the model would still be able to maintain the relationships between income and VOT categories above. To see this consider what the constraints above mean in the context of the coefficients of the mode choice model. Suppose that the estimated cost coefficient for income category 1 was -0.1. From this, we can compute all of the relevant LOS coefficients, as illustrated in Table 4.4.

Table 4.4 Example Calculation of LOS Parameters from Constrained Relationships

Parameter	Calculation	Implied Value
Cost – Income 1	1.00 * -0.1	-0.100
Cost – Income 2	0.61 * -0.1	-0.610
Cost – Income 3	0.47 * -0.1	-0.470
Cost – Income 4	0.25 * -0.1	-0.250
IVT – Income 1, VOT 1	(2.70/60) * 1.00 * -0.1	-0.005
IVT – Income 1, VOT 2	(8.29/60) * 1.00 * -0.1	-0.014
IVT – Income 1, VOT 3	(27.36/60) * 1.00 * -0.1	-0.046
IVT – Income 2, VOT 1	(2.70/60) * 0.61 * -0.1	-0.003
IVT – Income 2, VOT 2	(8.29/60) * 0.61 * -0.1	-0.008
…	…	…
Metrorail IVT – Income 1, VOT 1	0.9 * (2.70/60) * 1.00 * -0.1	-0.004
Metrorail IVT – Income 1, VOT 2	0.9 * (8.29/60) * 1.00 * -0.1	-0.012
…	…	…
Walk Time – Income 1, VOT 1	2.5 * (2.70/60) * 1.00 * -0.1	-0.011
Walk Time – Income 1, VOT 2	2.5 * (8.29/60) * 1.00 * -0.1	-0.035
…	…	…

It is worth noting that we tested other variable specifications where we released one or more of the constraints in order to test how well the relationships above matched the estimated parameters from the mode choice data. In general, the implied relationships between freely estimated parameters did not match the relationships specified above, and therefore, we chose to constrain them for consistency with skimming procedures and our experience in other regions. For each trip purpose, we also estimated a simple model where VOT segmentation and income segmentation was removed with travel time and cost parameters estimated freely. This model implies that all travelers for that trip purpose share a single VOT and have identical sensitivities to cost and travel time. These models were estimated mostly for informational purposes, but the results suggested lower than expected implied VOTs, which is not uncommon.

Income Variables

Indicator (or 'dummy') variables associated income level were also included in the models for home-based trips. These are alternative-specific variables that take a value of one for the indicated income category. Estimated parameters of these variables differ by mode.

These variables can be interpreted as being modifiers to the alternative specific constants. In other words, we can view each income category as having it's own set of alternative specific constants, or baseline preferences toward each mode alternative.

As a general point, only 3 of the 4 income category variables are identifiable from a model estimation standpoint. That is, one income category must be held as a reference category, from which the other income variables measure the relative effects. For the mode choice models, the 3rd income category served as this reference category. In general, a full set of income variables were tested, and ultimately only those found to have significant and reasonable impacts on the model were retained in final model specifications.

Transit Accessibility Variables

Specification

These variables were identified as being particular useful in describing mode choice behaviors in the literature.[9] Here, logsum measures are used to measure transit accessibility. A mode choice logsum concisely represents the accessibility across all modal options from one origin to one destination in a single value, and a destination choice logsum represents accessibility across all destination options for a single origin.

Formally, the accessibility logsum for a zone *i* is computed as follows:

$$A_i = \ln\left[\sum_j S_j \times \exp\left(MLS_{ij}\right)\right] \tag{4.6}$$

Here the sum is across all zones in the region. A_i is the accessibility measure for zone i and trip purpose p, S_j is the size variable for zone j, and MLS_{ij} is the mode choice logsum as defined below.

$$MLS_{ij} = \ln\left[\sum_m \delta_m \exp\left(U_{ijm}\right)\right] \tag{4.7}$$

Here, δ_m is an 0/1 indicator denoting whether mode m is to be used in the logsum calculation. For the MWCOG model, A_i is computed for either drive alone mode or walk-transit mode (i.e., the summation over *m* is the summation over a single mode only). The transit accessibility measures used in the mode choice model specifications are actually relative measures of transit accessibility, defined as follows:

$$RA_i = A_{i,WT} - A_{i,DA} \tag{4.8}$$

If $A_{i,WT}$ is undefined because zone *i* has no valid walk-transit paths generated by the skimming procedure to any other zone, then RA_i is set to zero. In such cases, a second indicator variable is also defined to avoid biasing estimation results. This second variable indicates if a zone has no transit accessibility to any other zone:

$$No\_WT\_Acc = 1 \;\; if \; A_{i,WT} \; is \; undefined, 0 \; otherwise \tag{4.9}$$

[9] See, e.g., FY2016 Report on Task 5.

The specification of the utility function shown in equation (5) consists of only LOS variables. It uses parameter values of the LOS variables similar to the constraints imposed on the LOS variables noted above. For instance, the travel time weights shown in Table 4.3 are used and the transit boarding weights used in the mode choice model are used here also. One key difference is that only a single VOT assumption is made, rather than segmenting VOT by VOT category. This was done because changes in the VOT tend to change the scale of the accessibility variable itself. These scale differences tend to dwarf the more subtle difference in accessibility measurements from one zone to another, and tend to reduce the explanatory power of the variable. For these reasons, the middle VOT level was used to make accessibility variable calculations (of $8.29/hr).

There are two final pieces needed to compute the accessibility measures. First, the size variable, S_j, was defined as total zonal employment for all trip purposes except for home-based shopping trips, where it was defined as the zone's retail employment. Second, peak period skims were used for all calculations. Arguments could be made for other assumptions, but ultimately, we do not believe different assumptions would make too much difference in the response captured in the model.

Discussion

Accessibility measures have several advantages over typical density measures or area types that are often used in mode choice models. First, they measure how well the zone is connected to other zones via the transit network, and therefore, are directly applicable to the transit accessibility. Secondly, they can be weighted on the basis of the size of each zone (as is shown below). The larger the zone, the more significant the zone's impact on the overall accessibility. By accounting for a zone's connectivity to other zones, this measure avoids some of the issues with spatial aggregation that can occur with density measures.[10]

While accessibility measures are often used in MPO travel models, they are more typically found as variables in upper level models, for instance vehicle availability or day pattern models of an activity-based model (ABM)[11]. The basis for using these variables in mode choice is as a replacement to the more typical density measures. We believe that the use of density variables may be misguided, and that transit accessibility at the origin and the destination is the important variable for transit choice. Since density variables typically mimic transit network density, we believe that density variables are really a less policy-sensitive proxy for both transit and highway accessibility.

Other Zonal Variables

Several other zonal variables were tested in the mode choice models beside accessibility. Simpson's diversity index was one such variable, measured as follows:

$$\lambda_i = 1 - \sum_m {p_{im}}^2 \tag{4.10}$$

Here, m is an index of the different elements in the zone, in this case, population and employment, and p_{im} is the share of the total population and employment for that element. For instance, if population and employment had an equal number in the zone, the share of each would be 0.5 and the diversity index would

[10] A density variable is typically measured at the zonal level, meaning characteristics of the areas immediately surrounding the zone have no impact on the zone's density, and this is the result of aggregating spatial information into zones in the first place. Of course, as long as zones are used at all, some level of spatial aggregation error will persist.

[11] This is the case for the Baltimore ABM as well as all recent CT-RAMP ABM implementations.

be 0.5. If there was no employment, however, then the population share would be 1 and the diversity index would be 0. Ultimately, we found this variable to have the most reasonable and significant impacts on the walk-transit mode at the production end of the trip.

Another variable we tested is the number of cul-de-sac intersections in the zone, measured as the count of cul-de-sac intersections. Not surprisingly, the number of cul-de-sac intersections was found to be negatively correlated with use of transit (thus, the negative signs on the coefficient estimates).

Lastly, we tested a measurement of the walk accessibility to Metrorail, measured as the percent of the land area of a zone that is within one-half mile of a Metrorail station. This variable was found to be highly significant and positively correlated with transit usage.

4.6 Incorporating Cube Public Transport Module

A major component of the FY17 work program was to implement the Cube Public Transport (PT) process to replace the existing Trnbuild (TB) process for transit skimming and assignment processes of the MWCOG model. The implementation of the PT process is a continuation of the work carried out in the FY15 and FY16 work programs[12], in which the PT process was implemented and tested for its compatibility and consistence with the existing TB process.

In the FY 17 work program, the PT process was further revised to accommodate the requirements of a new mode choice model being developed in the study. The new mode choice model adopts a "shallow" choice structure, in which the model splits person trips between highway modes and a transit mode with the "best path", without further splitting transit trips among transit submodes, since, under the revised model, transit submode estimation would occur as part of path-building, not mode choice. The PT process was accordingly revised to generate a set of "best path" skim matrices, instead of skim matrices for various transit submodes.

The revised PT process was applied to generate a set of the base-year (2007) transit skim matrices, which were used for the calibration of the new mode choice model. The skim data generated from the revised PT process were examined for their compatibility with the skim data generated from the existing TB process. Also, the impact of transit fare in determining the transit "best path" in the revised PT process was investigated. This technical memorandum summarizes the results of these examinations.

Comparison of Skim Data of PT Process and TB Process

To examine the PT skim data, a set of base-year skim data was generated using the existing TB process. Transit path-building is usually done on the basis of perceived travel times, which are often called weighted travel times, because the individual travel time components, such as in-vehicle time, out-of-vehicle time, and wait time, are weighted or factored by coefficients that represent how much impact they have on travel choice. In PT, one can build paths on the basis of a generalized cost, which includes both time and cost. The use of generalized cost in PT is described later in this memo. Figures 1a – 1d display the scatter diagrams of total skim times (unweighted and weighted times, for walk access and drive access) between these two sets of skims. It should be noted that the PT process generates the "best path" skim matrices while the TB process generates skim matrices of various transit submodes. For each O-D pair, the transit submode of the TB skim data is selected for comparison based on the modes used in the "best path" generated from the PT

[12] Gallop Corporation, Task Order 15.4, Modeling with Public Transport, Final Report (October 2015).
Cambridge Systematics and Gallop Corporation, Metropolitan Washington Council of Governments, National Capital Region Transportation Planning Board, FY 16 Task Orders, Final Report (November 2016).

process. For example, if the best path in the PT process use commuter mode, the TB skim data of the commuter mode are used to compare against the PT skim data.

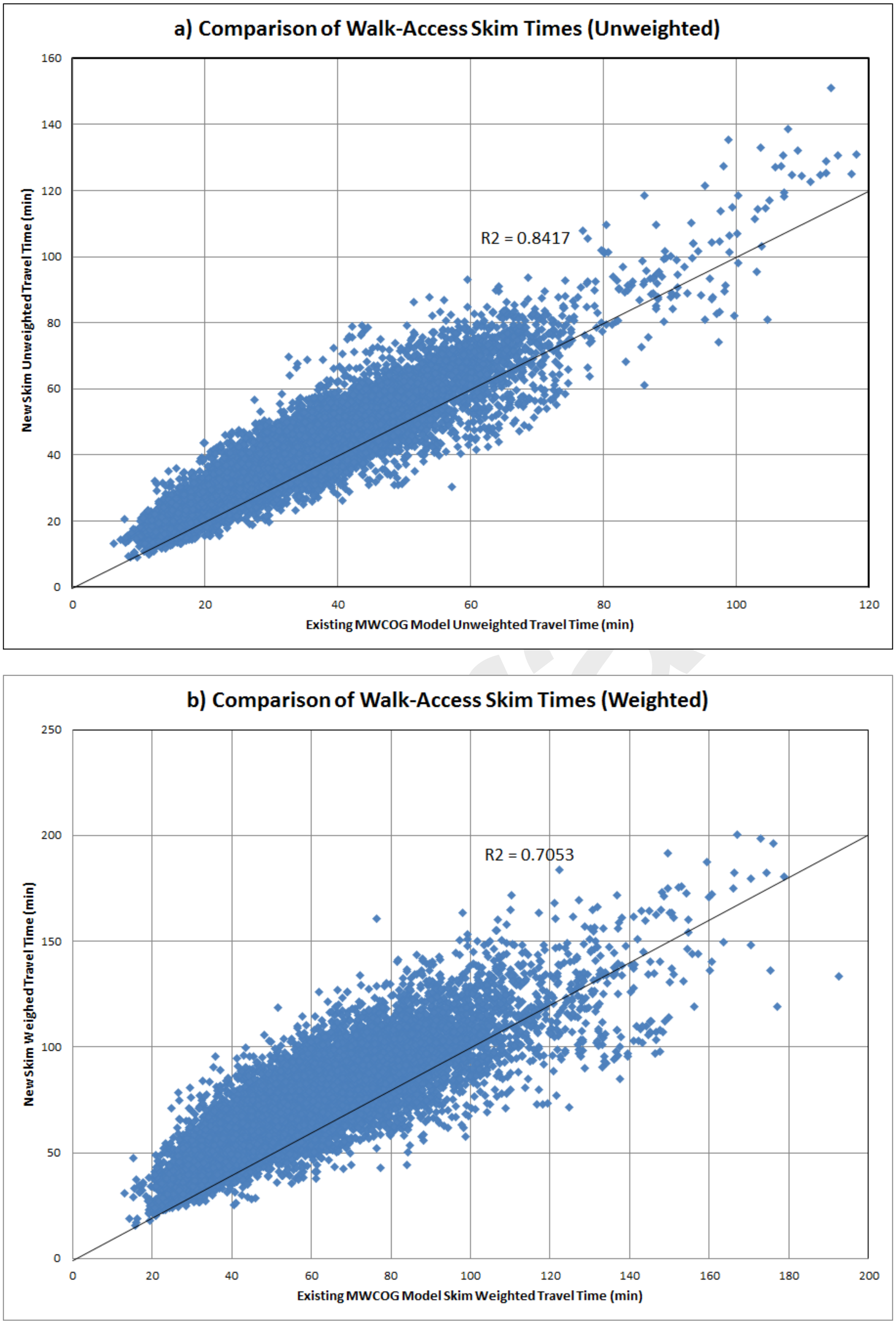
It should be noted that the PT and TB skimming processes are substantially different, with different definitions of skimming parameters. It is thus not possible to generate identical skim data from these two processes. Still, Figures 4.1a to 4.1b indicate that the two sets of processes generated fairly consistent skim times, with R^2 around 0.85 for unweight times and 0.70-0.80 for weighted times. Because of the different weighting factors used in the two processes, the results of the weighted travel times are slightly worse than those of the unweighted times.

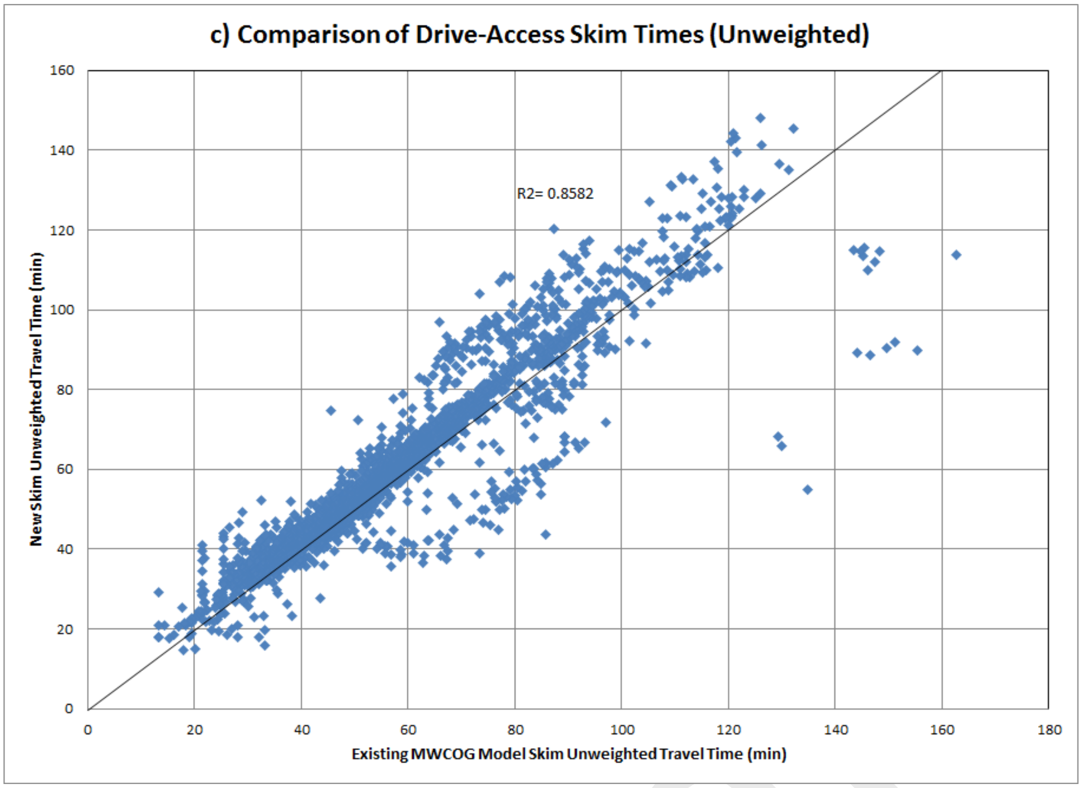
Figures 4.1c and 4.1d reveal that the drive-access skim times of the two processes are also very consistent. Most of the points are near the diagonal line of perfect correlation, except for a few outlier points. It is because for drive-access, the choices of paths from origins are limited. For example, for origins from outlying areas, the best paths usually are the ones driving to the closest Metro stations or the commuter rail stations. A further investigation of those outlier points reveal that those points are from origins in Howard County. The trips from these origins should drive to commuter bus stations and take commuter buses to the downtown area. However, the existing TB process cannot build those paths because drive times exceed the maximum drive-access time as set in the TB process. Instead the TB process generate paths to nearby bus stops and take circuitous paths with extra transfers to destinations. As shown in Figure 4.1c, the unweighted skim times from the TB process for those outlier points are excessively large, more than 120 minutes.

Figures 4.2a and 4.2b compare the skim fares of the two processes. It should be noted that for both processes, the skim fare matrices are derived from the MWCOG special fare programs (MFARE1 and MFARE2). The programs determine the transit fare of an individual O-D pair based on the fare zones of the origin and destination, the chosen mode, and the boarding and alighted stations if a Metrorail path is used. The use of the program ensures that the skim fares derived from the two processes are compatible.

Figures 4.2a and 4.2b indicate that the skim fares of the two process are fairly similar. It should be noted that the skim fares of the two processes are identical for a substantial number of O-D pairs (i.e., overlapping points along the diagonal line) because the fares on most transit modes are either with a flat fare or a step-wise distance-based fare structure.

Figure 4.1 Comparison of Skim Times between PT and TB Processes




c) Comparison of Drive-Access Skim Times (Unweighted)
R2= 0.8582
New Skim Unweighted Travel Time (min)
Existing MWCOG Model Skim Unweighted Travel Time (min)
0
20
40
60
80
100
120
140
160
180

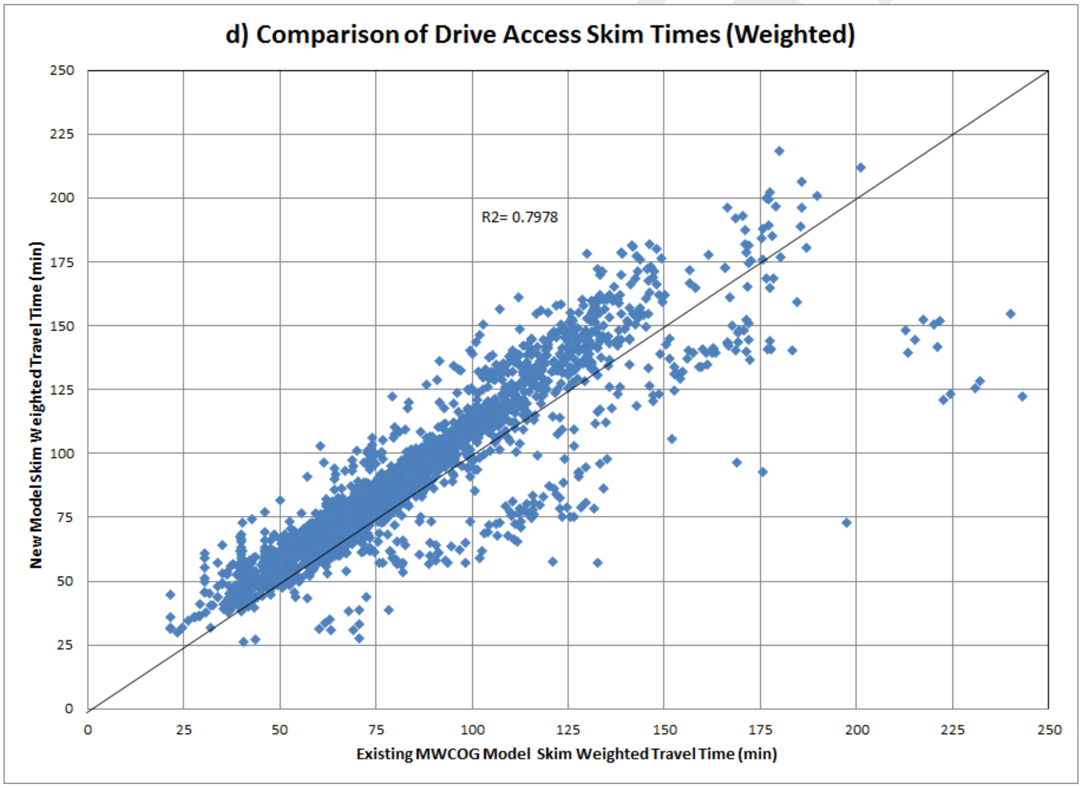

d) Comparison of Drive Access Skim Times (Weighted)
R2= 0.7978
New Model Skim Weighted Travel Time (min)
Existing MWCOG Model Skim Weighted Travel Time (min)
0
25
50
75
100
125
150
175
200
225
250

Figure 4.2 Comparison of Skim Fares between PT and TB Processes

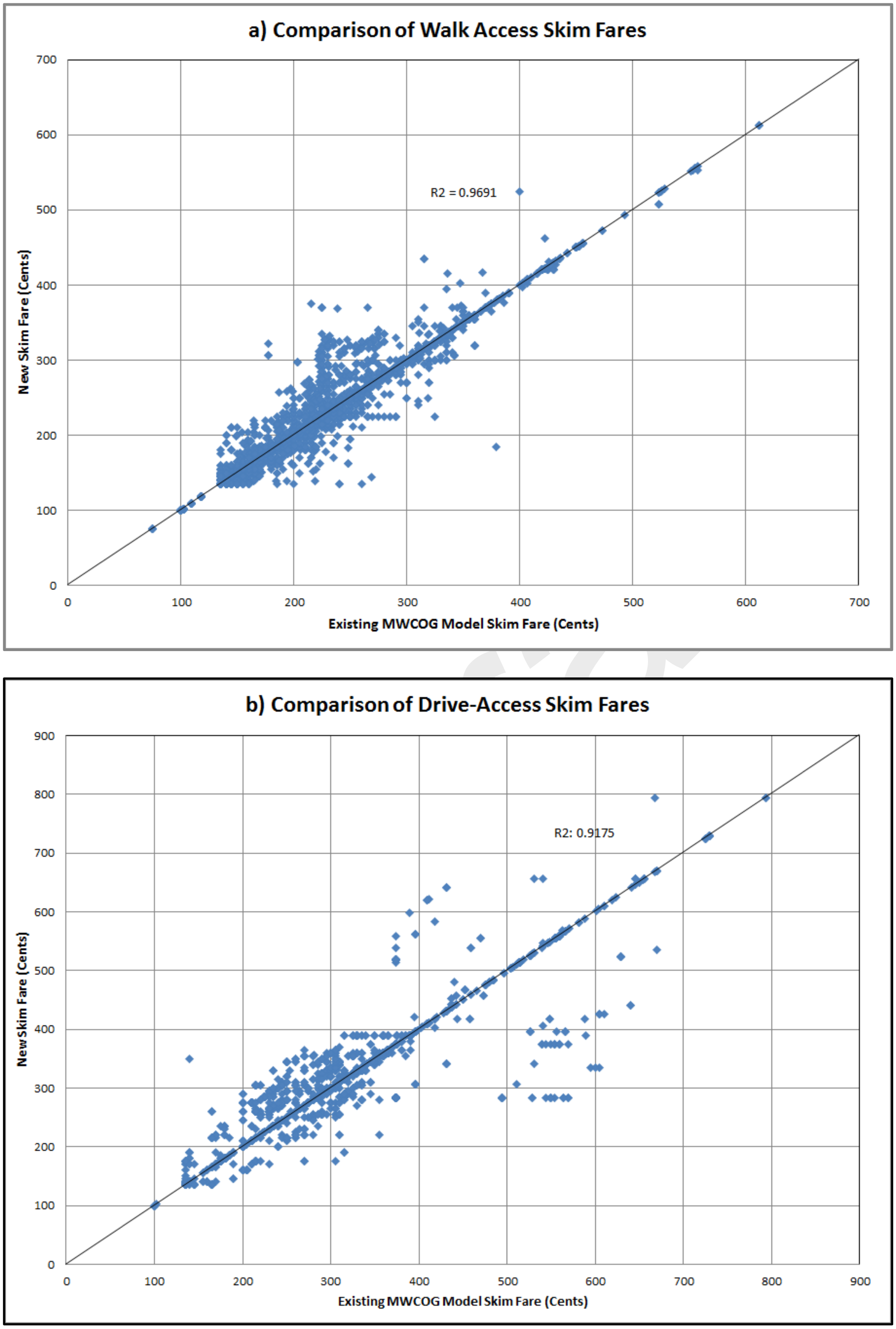


Table 4.5 summarizes the numbers of O-D pairs with transit path connections under the PT and TB processes. As indicated in the table, for walk access paths, almost 90 percent of O-D pairs have matching

transit connections or no-transit connections. The matching shares of O-D pairs for drive-access connections are slightly lower, with 82% for peak period and 76% for off-peak period.

Table 4.5 also indicates that, for drive-access paths, the numbers of O-D pairs with transit connections with the PT process are higher than those with the TB process, because the maximum drive-distance used in the PT process is longer than that used in the TB process. For AM peak walk-drive trips, the number of O-D pairs with transit connections with the PT process is similar with that with the TB connections. However, for off-peak period, the number of O-D pairs with transit connections under the TB process is significantly lower than that with the PT process (i.e., 5,189,142 vs. 5,566,819). This is due to the increase in the PT process of the maximum allowable total weighted skim time for a transit path to be built. In the TB process, the maximum total weighted travel time threshold is set to 360 minutes. By contrast, in the PT process, this is increased to 600 equivalent minutes (generalized cost) in order to allow more O-D pairs to have transit connections for the calibration of the new mode choice model. The increase of maximum threshold travel time value would have bigger impacts for the transit connections in the off-peak period, because of much less frequent service for some bus routes and commuter routes, and hence incurring substantially long wait times for those services in the off-peak period.

In order to examine the impact on the assignment results of the two skimming processes, i.e., the single "best path" PT process versus the multiple-submodes TB process, a same set of transit trip tables are used to test these two processes. The transit trip tables used for the test are generated from the base year MWCOG model using the existing TB skimming process. The trip tables of various transit submodes are assigned to the TB transit network individually, following the regular transit assignment process of the MWCOG model. Then these transit tables are combined into transit trip tables by access mode. These combined transit tables are assigned to the PT transit network with the single "best path" PT process.

Table 4.6 summarizes the assigned unlinked trips by submode derived from these two processes. The table reveals that the assignment results for these two processes are comparable with each other. However, it should be noted that this is just a preliminary test to examine how different these two skimming processes would be in the assignment results. A thorough validation of the assignment process will be performed after the new PT skimming process and the new mode choice model are calibrated and implemented.

Table 4.5 Numbers of O-D Pairs with and without Transit Path Connections under PT and TB Skim Processes

Condition	AM Peak Walk-Access	Off Peak Walk-Access	AM Peak Drive-Access	Off Peak Drive-Access
Both PT and TB Skims Exist	5,320,581	4,562,615	5,958,728	5,470,656
PT Skims Exist, but TB Skims Do Not Exist	604,270	1,004,204	2,194,134	2,849,698
PT Skims Do Not Exist, but TB Skims Exist	743,170	626,527	336,467	419,873
Neither PT Nor TB Skims Exist	7,185,263	7,659,938	5,363,955	5,113,057
Share of O-D Pairs w/ Matching Transit or No-transit	90%	88%	82%	76%
Total Connected PT O/Ds	5,924,851	5,566,819	8,152,862	8,320,354
Total Connected TB O/Ds	6,063,751	5,189,142	6,295,195	5,890,529
Difference (PT - TB)	-138,900	377,677	1,857,667	2,429,825

Table 4.6 Comparison of Assigned Trips by Transit Submode Derived from PT process and TB process

Main Transit Mode	Trips – Existing TB Procedure	Trips – New PT Procedure	Difference	% Difference
Local Bus	603,227	582,553	-20,674	3.4
Express Bus	83,562	88,680	5,118	6.1
Metrorail	997,821	1,019,597	21,776	2.2
Commuter Rail	29,535	36,942	7,407	25.1
Total	1,714,145	1,727,772	13,627	0.8

Note: Table presents unlinked transit trips.

Impact of Transit Fare on Best Path Choice of the PT Process

In the existing TB process, transit fare is not considered in the path choice process. The transit path of each of the transit submode paths (i.e., All-Bus, Metro-Only, Bus/Rail and Commuter Rail) is determined based on the perceived transit time, which is the weighted sum of all travel time components and various boarding/transfer penalties. For each submode, the effect of transit fare on path choice is limited. It is thus not necessary to consider fare in the path choice process.

In the single best-path process, with all transit modes considered together as a single transit mode, transit fare would have a bigger impact on choosing the best transit path because the fare structures of various transit modes are substantially different, e.g., flat fares for most of local bus routes and distance-based fares for Metrorail and commuter rail services. For example, a transit passenger might choose a slower local bus path with lower fare instead of taking a faster Metrorail path with high fare.

In the PT process, transit fare is considered in the path choice process. Four fare functions are considered in the process for four types of transit service: local bus, express bus, Metrorail and commuter bus. Flat fare structures are considered for local bus and express bus modes, following the WMATA bus fare policy. Distance-based fare structures are considered for Metrorail and commuter rail services. For these two types of service, data about transit fares and “along the track” station-stations distances of individual station-station pairs were collected and used to derived the distance-based fare functions. Figure 4.3 and 4.4 display the scatter diagrams of fares against station-station distances, together with the derived fare functions, for Metrorail and commuter rail respectively. For each of two services, the minimum and maximum fare levels are set based on the published fare schedules. In the case of Metrorail, a polynomial function was fit to the data:

$$y = -0.0043x^2 + 0.2484x + 0.4926$$

This function has a local maximum at 28.8 miles.

In the case of commuter rail, a linear function was fit to the data:

$$y = 0.1024x + 3.4925$$

With the consideration of transit fare in the path choice process, the best path in the PT process is determined based on **generalized cost**, which is the sum of total perceived travel time and penalties, parking cost and fare with an assumed value-of-time factor. It should be noted that transit fare is considered only in the path choice process. The final skim fare matrices are still derived from the existing special fare programs (MFARE1 and MFARE2) of the MWCOG model, given the chosen best paths of individual O-D pairs.

In order to examine the impact of considering fare in the path choice process on the resulted skim data, another set of PT skim data was generated without the consideration of fare in the path choice process. The resulting skim times of these two sets of skims (i.e., with and without fare consideration) are displayed in a set of scatter diagrams as shown in Figures 4.5a to 4.5d. These scatter diagrams reveal that the resulting skim times with and without fare considered in the path choose process are similar with each other. In general, the resulting weighted travel times with fare considered are slightly higher than those without fare consideration. This is expected since with fare consideration, the best paths are chosen based on minimum general cost, not based on the minimum perceived travel time.

Table 3 summarizes the level of differences of chosen transit paths with and without fare consideration. The table reveals that for more than 80% of O-D pairs, the two processes generate identical chosen paths. There are less than 4% of O-D pairs with more than 20% difference in unweighted skim times between the two sets of skims. For those O-D pairs with a major difference in skim data, a further investigation was conducted to examine how different the two processes were regarding chosen paths. The chosen paths of three selected O-D pairs were traced and plotted in Figures 4.6-4.8. In these figures, the paths with fare considered in the path choice process are shown in blue color, whereas the paths without fare consideration are shown in red color.

Figure 4.6 traces the walk-access paths from Zone 346 to Zone 44. With the consideration of transit fare (blue), the path choice process selects the Metrorail path with relatively long walk to the Metrorail station. The trace report indicates that the total generalized cost of the path is 117.3 minutes, slightly higher than the total perceived total time (103.3 minutes). Without fare consideration (red), the path choice process selects the path with short bus rides transferring to and from the Metrorail stations. Although the total perceived time for these two paths are similar (98.37 minutes vs. 103.30 minutes), the selected paths are different and the path without fare consideration results in extra bus transfers in the chosen paths.

Figure 4.7 shows similar results for the paths from Zone 293 to Zone 21. The chosen path with fare consideration (blue) takes a single Metrorail ride with relatively long walk distance at the origin end, whereas the path without fare consideration (red) takes two short bus rides with short walk distances at both ends. Figure 4.8 traces the differences in the drive-access paths from Zone 430 to Zone 55. Again, the path with fare consideration (blue) takes a relatively long drive to a Metrorail station, while the path without fare consideration (red) takes a bus ride and then transfers at the same Metrorail station.

In all the three cases above, the chosen paths with fare consideration seem to be more reasonable as compared with the paths without fare consideration. Nevertheless, the chosen paths are also affected by other factors like the boarding penalties and weighting factors of IVT and OVT components. These skimming parameters will further be examined and refined during the validation stage of the new mode choice model and the transit assignment process.

Figure 4.3 Observed Station-Station Metrorail Fares and Derived Fare Function

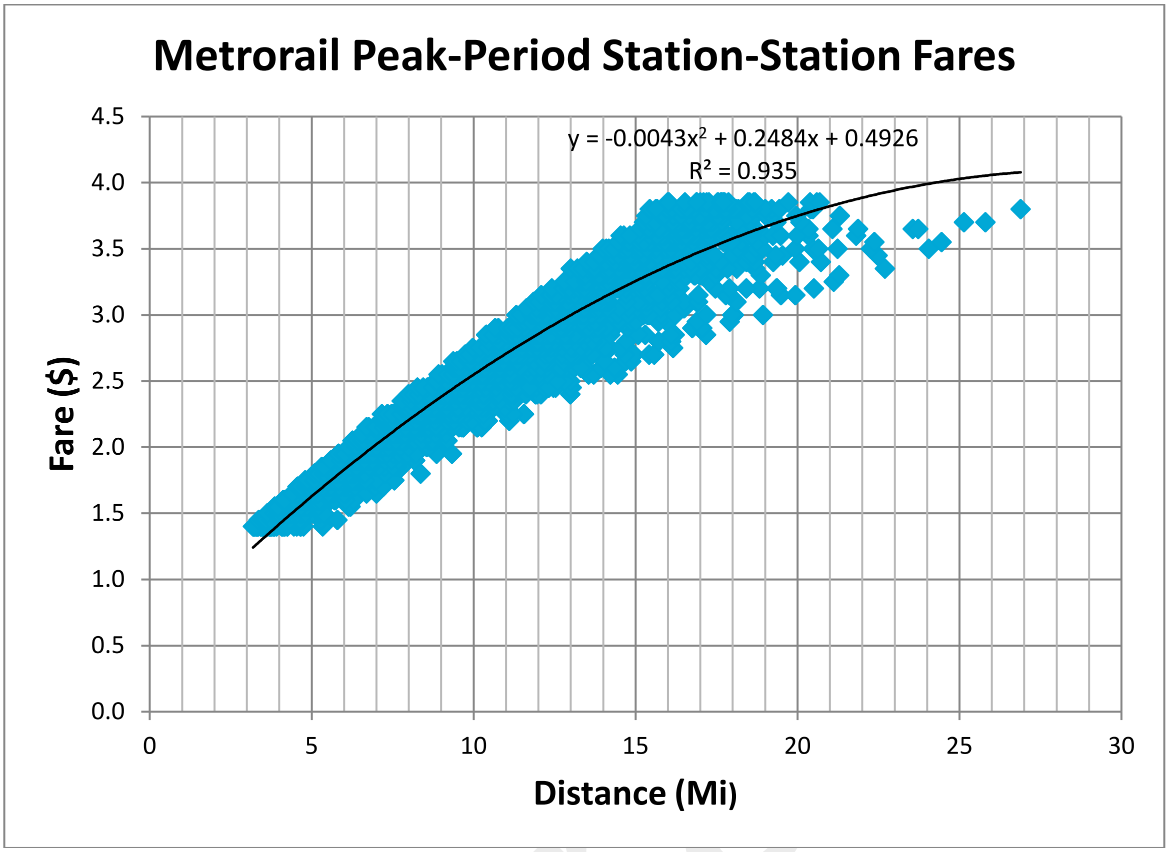


Figure 4.4 Observed Station-Station Commuter Rail Fares and derived Fare Function

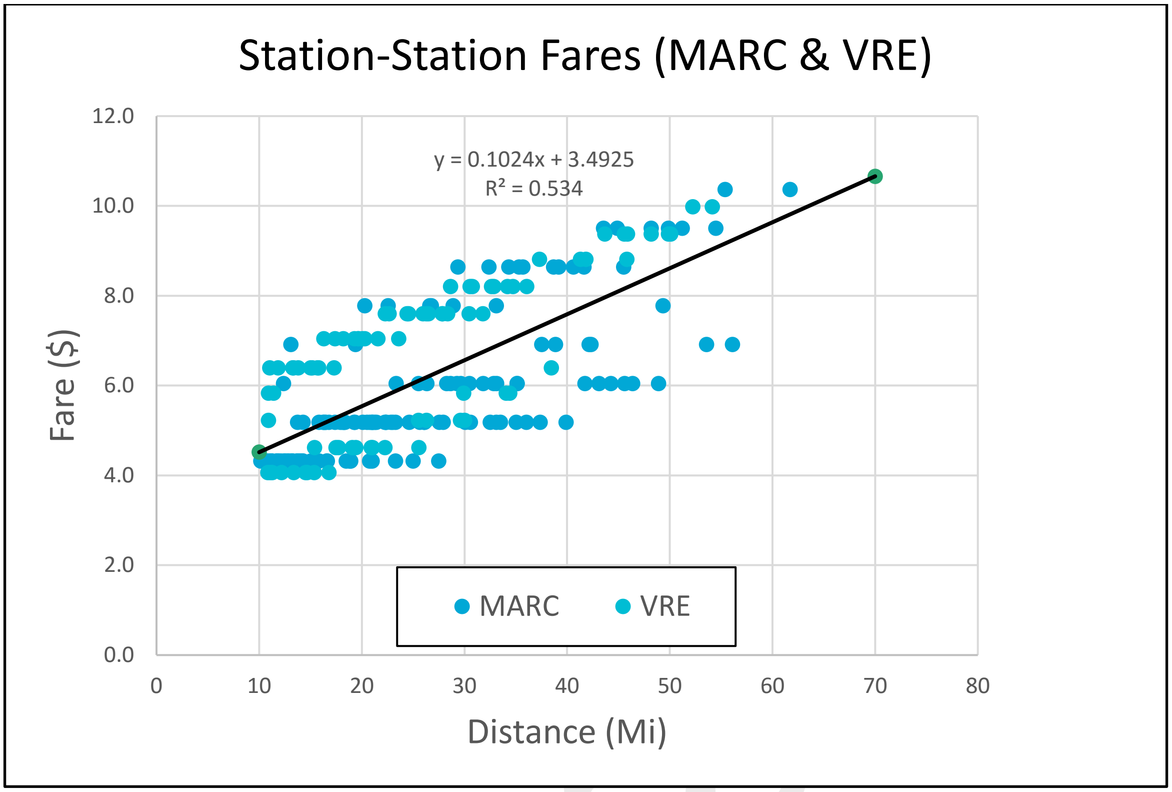
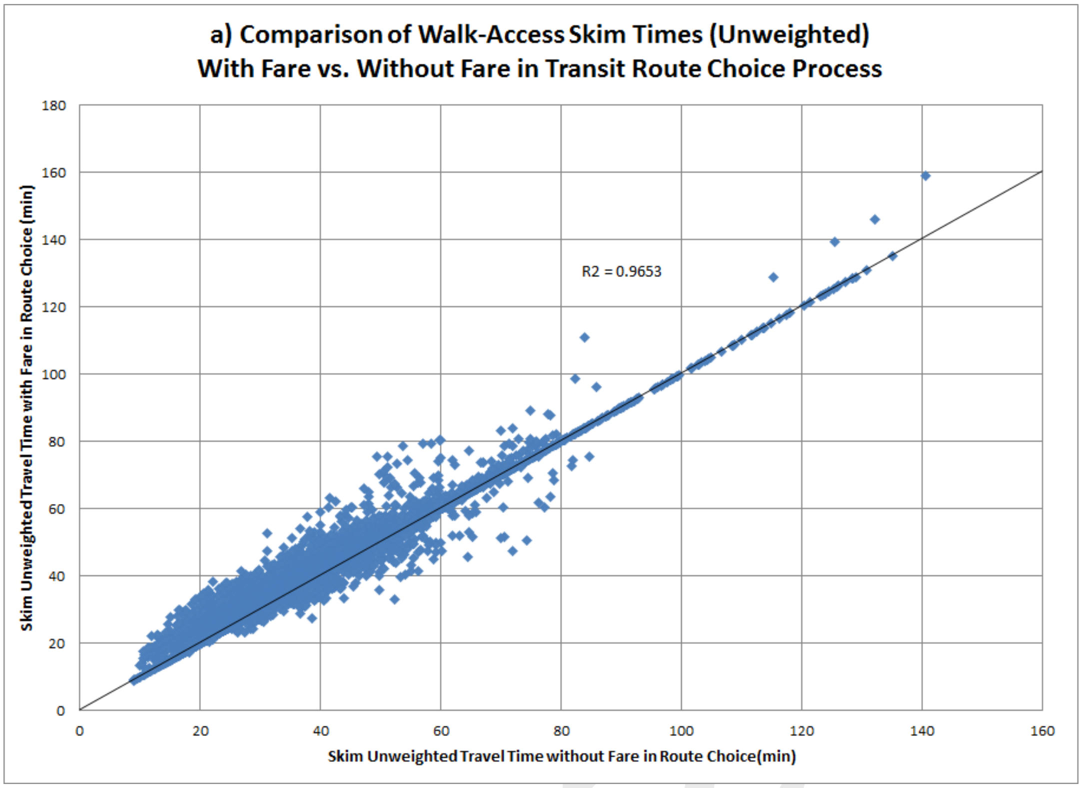
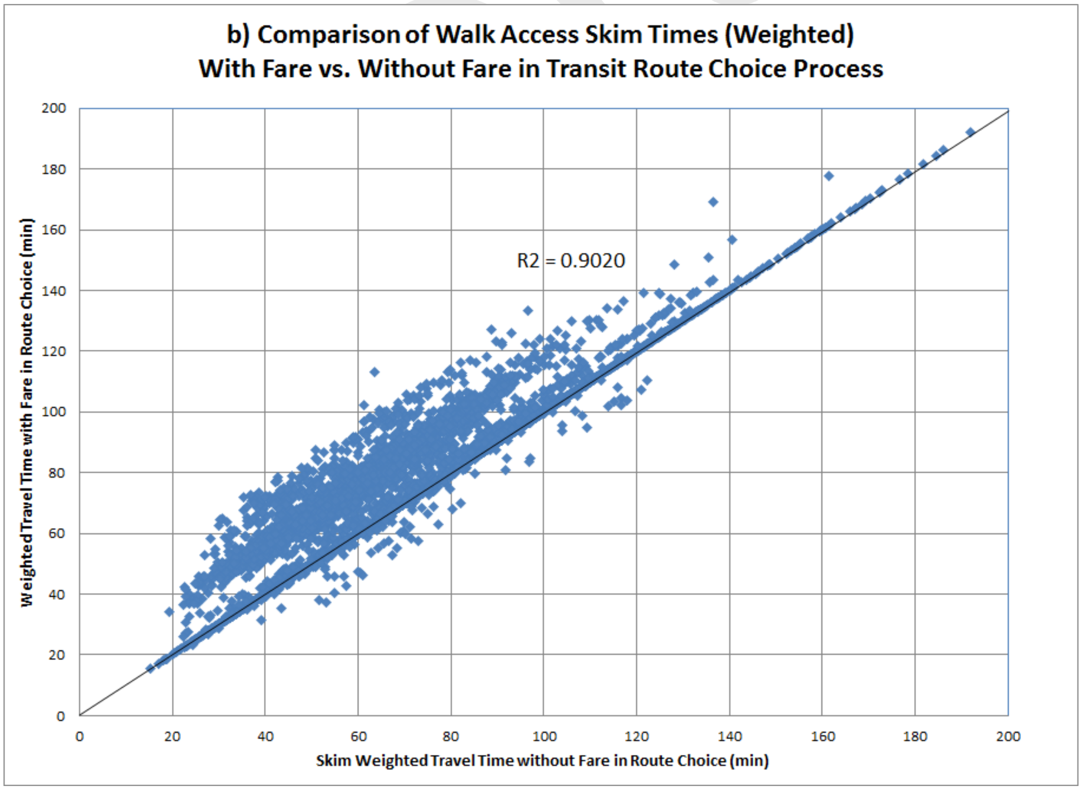
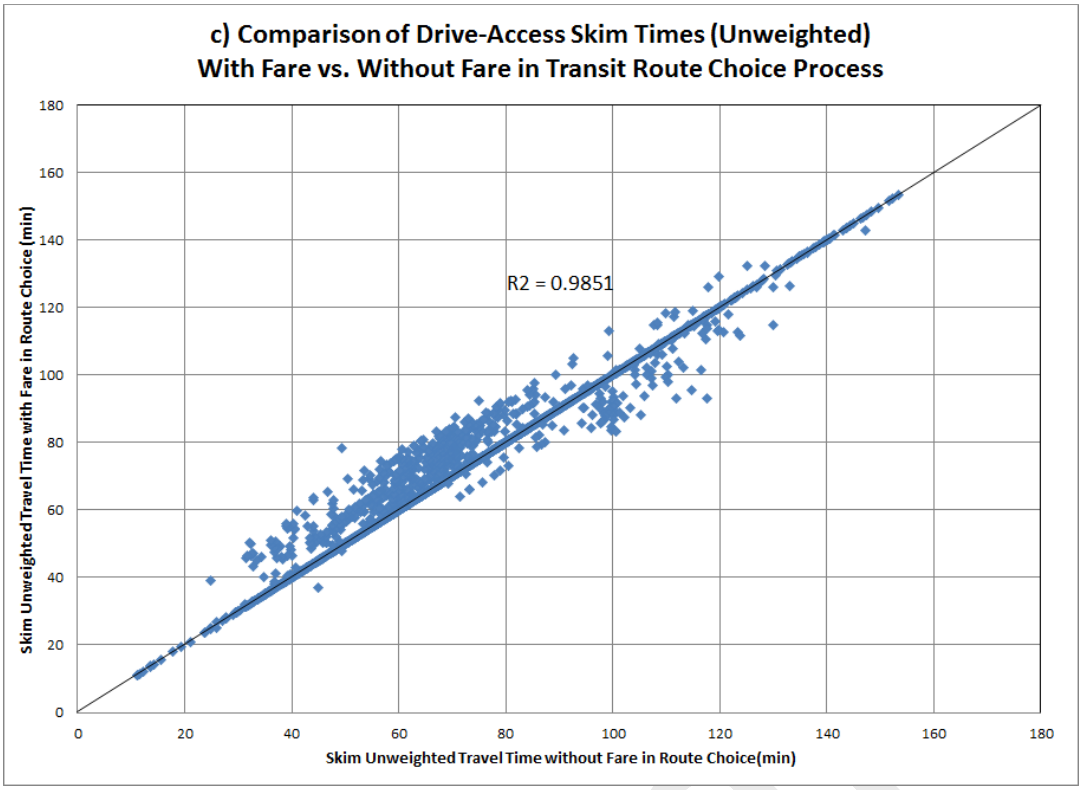


Figure 4.5 Comparison of Transit Skim Times with and without Fare Considered in the Path Choice Process






c) Comparison of Drive-Access Skim Times (Unweighted)
With Fare vs. Without Fare in Transit Route Choice Process
R2 = 0.9851
Skim Unweighted Travel Time with Fare in Route Choice (min)
Skim Unweighted Travel Time without Fare in Route Choice(min)
0
20
40
60
80
100
120
140
160
180

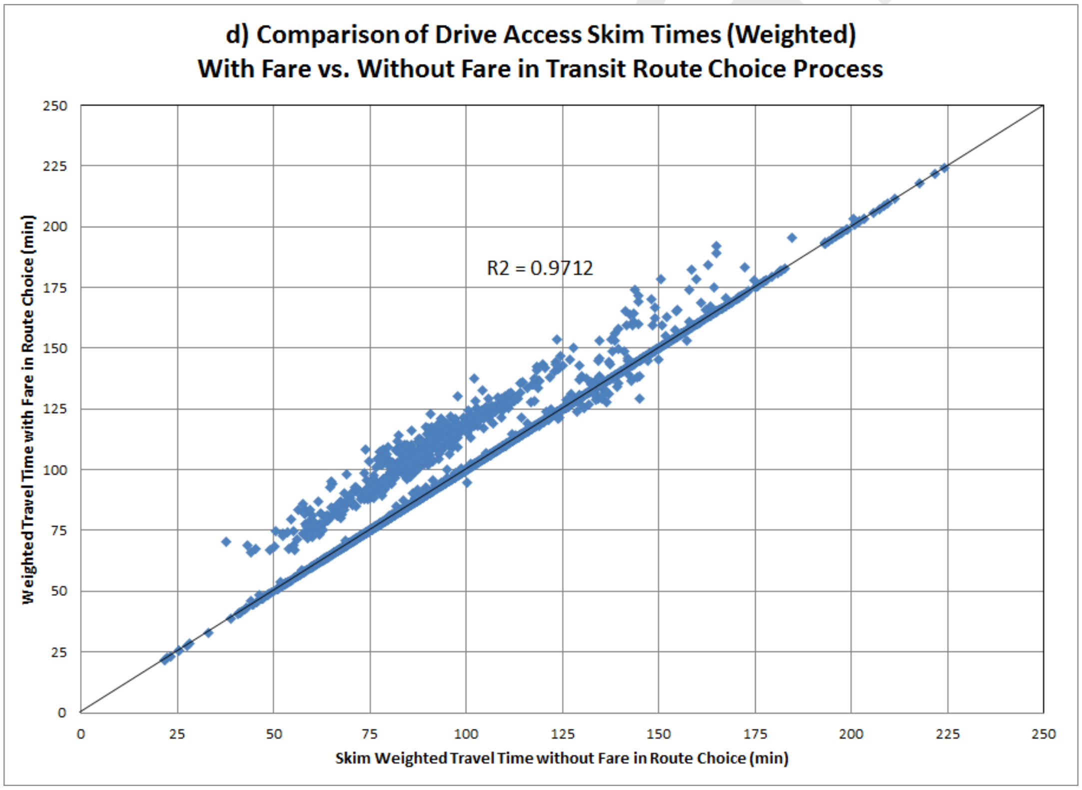

d) Comparison of Drive Access Skim Times (Weighted)
With Fare vs. Without Fare in Transit Route Choice Process
R2 = 0.9712
Weighted Travel Time with Fare in Route Choice (min)
Skim Weighted Travel Time without Fare in Route Choice (min)
0
25
50
75
100
125
150
175
200
225
250

Table 4.7 Summary of Differences in Transit Paths with Transit Fares vs. without Transit Fares in Path-Choice Process

	Walk Access Paths		Drive Access Paths	
Condition	No. of O/D Pairs	% of O/D Pairs	No. of O/D Pairs	% of O/D Pairs
Identical paths	76,768	85.3%	40,125	81.7%
Paths with < 5% difference in total unweighted travel time	3,211	3.6%	1,695	3.4%
Paths with 5-10% difference in total unweighted travel time	3,509	3.9%	2,306	4.7%
Paths with 10-20% difference in total unweighted travel time	4,328	4.8%	3,435	7.0%
Paths with 20-50% difference in total unweighted travel time	2,123	2.4%	1,561	3.2%
Paths with > 50% difference in total unweighted travel time	47	0.1%	13	0.0%
Total	89,986	100.0%	49,135	100.0%

Figure 4.6 Difference in Chosen Walk-Access Paths from Zone 346 to Zone 44 with and without Fare Consideration in Path Choice Process

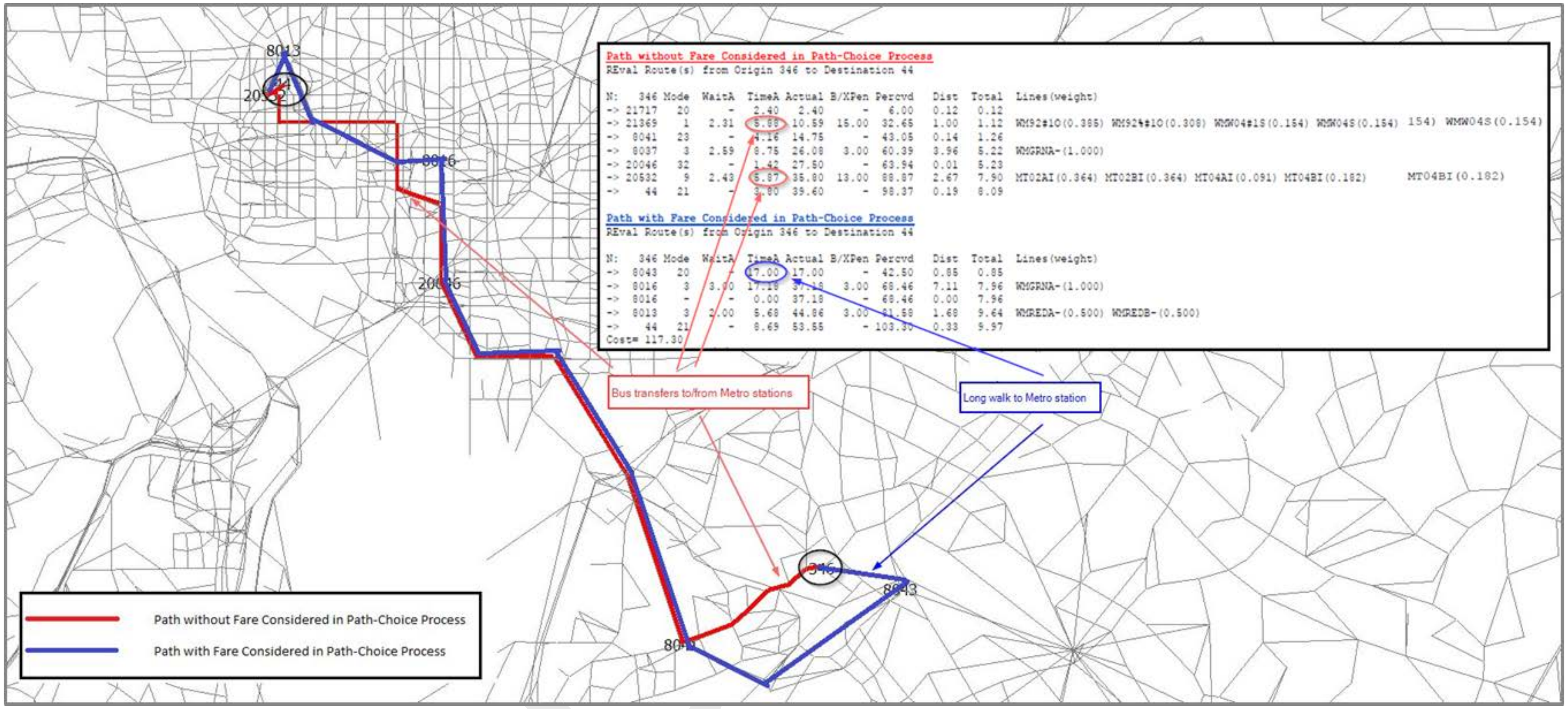


Figure 4.7 Difference in Chosen Walk-Access Paths from Zone 293 to Zone 21 with and without Fare Consideration in Path Choice Process

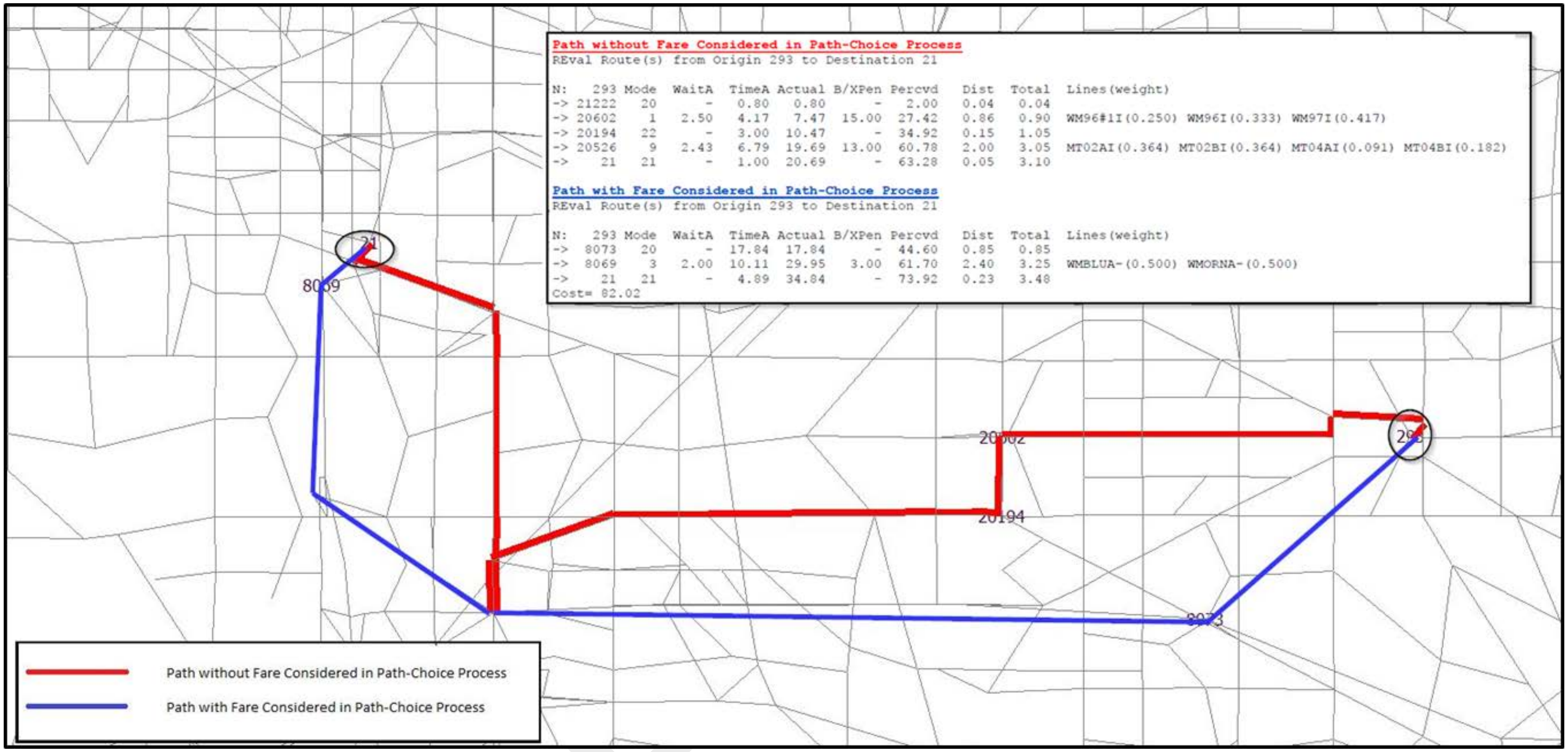
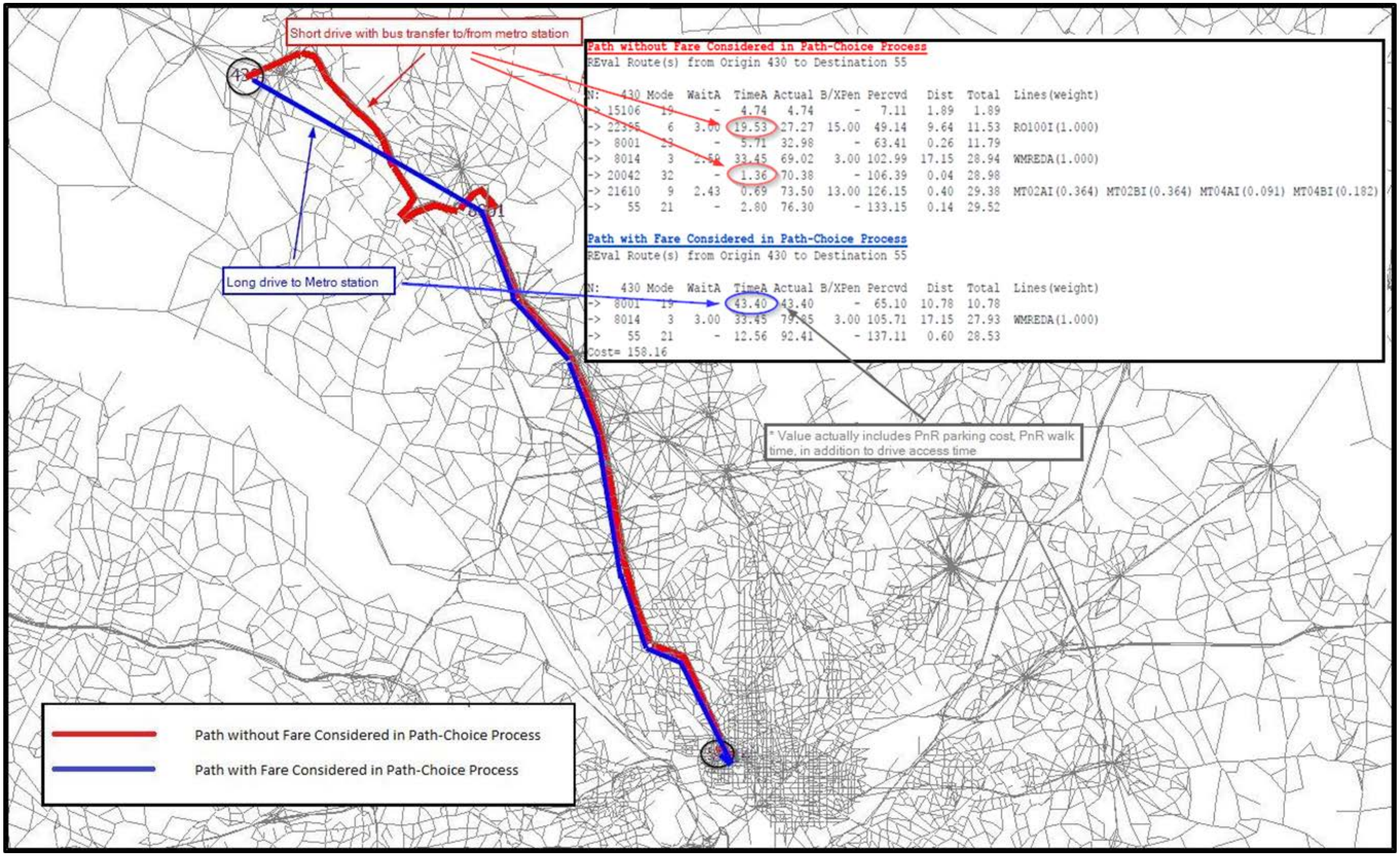


Figure 4.8 Difference in Chosen Drive-Access Paths from Zone 430 to Zone 55 with and without Fare Consideration in Path Choice Process



4.7 Mode Choice Estimation Results

The recommended models are shown in Table 4.8 through Table 4.12 and are discussed later in this section. A more extensive compilation of model estimation results can be found in a companion Excel workbook. We will first describe general findings and then discuss more detailed findings related to the recommended models.

Key general findings of model estimation include:

- Implied values of time, when unconstrained, consistently were estimated to be lower than our expectations, and in most cases, much lower than expected. This is not an unusual result: We have encountered it in other regions when estimating mode choice models from revealed preference data.[13] We believe this is because many travelers are more or less captive in their mode choices, either in reality or in their perception. Consequently, the level of service variables often exhibit a degree of correlation with one another, making it difficult to distinguish individual effects.

- The effect of relaxing constraints on the heterogeneity of VOT was not as large as the changes noted above with unconstrained values of time.[14] In the case of the HBW trip purpose for instance, the estimated model (Model 131) suggested low, medium, and high VOTs of -$0.49, $7.83, and $17.22 per hour, respectively. Compared to the assumed VOTs for these segments of $2.70, $8.29, and $27.36, the difference between low and medium[15] and medium and high[16] were not terribly different from our asserted VOTs. However, one clear issue that would emerge if these variables were not constrained is that each trip purpose would be estimated to have its own set of VOT segments. Since the highway assignment model will depend on all trip purposes being aggregated into like VOT categories, we moved forward with the constrained estimates. Nonetheless, it is reassuring to find that our assumptions are not too different from the unconstrained model.

- Low income travelers, all else being equal, tended to be more likely to use transit modes, likely a result that such travelers are less likely to have automobiles available for making trips. High income travelers showed a tendency away from transit modes.

- Transit accessibility, as expected, was found to have positive and significant effects on use of transit, particularly walk-transit at the production trip end and all transit at the attraction trip end.

- The no transit accessibility variable was found to have important impacts on the estimated models also (this variable takes a value of 1 if the zone has no walk access skims starting in the zone, and 0 otherwise). In the HBW and HBO models, for instance, when there is no transit access in a zone, we found that PNR and KNR trips were less likely to be produced in the zone. Since the variable only takes value of 1 in the most outer (largely rural) zones, this seems reasonable, even though such travelers are driving to access transit. For all trip purposes at the production end, we constrained the value of this coefficient estimate to -20 for the walk-transit mode. The walk-transit mode is actually never even

[13] For instance, we encountered these issues in Baltimore, San Antonio, and Houston, among others.

[14] Note the difference here between this comment and the previous. The first relates to the absolute value of VOT, which was estimated to be too low (HBW was found to have the most reasonable unconstrained VOTs). This second bullet refers to the relationships between low, medium, and high VOTs, ignoring their absolute values.

[15] That is, $7.83 – (-0.49) = $8.32 estimated, versus $8.29 – 2.70 = $5.59 asserted.

[16] That is, $17.22 – 7.83 = $9.39 estimated, versus $27.36 – 8.29 = $19.07 asserted.

considered a valid modal alternative in such cases anyway, since a valid walk-transit skim is needed for the mode to be considered valid, and a single valid skim would result in the no transit accessibility variable being zero instead of one. The constrained variable was added solely for illustrative purposes. On the attraction end of trips, the coefficient value of the no transit accessibility variable for all transit modes was constrained to be -3, which greatly discourages the choice of transit if the attraction end of the trip has no valid transit skims starting in the zone. Technically, it is possible to have a valid walk-transit skim ending in a zone, but none beginning in the zone, which is the only case in which this variable would ever be used. This is because for transit to be a valid alternative, there must be a valid transit skim between the production and attraction zones. If this is the case and the no transit variable at the attraction zone is 1, it means there is a valid production-to-attraction skim, but no walk-transit skims starting in the zone, which is rare. The only transit mode in this case would be KNR or PNR transit.

- Zonal diversity, as measured by Simpson's diversity index, tended to have only small impacts on mode choices. However, the variable was consistent in its impacts across different trip purposes, where we found a small positive impact of diversity associated with the production end of the trip on walk-transit trips. This is actually consistent with literature in some respects. From a policy perspective, this variable was tested as a measure of distinguishing transit oriented development (TOD) from transit adjacent development.[17] As such, it seems reasonable that TOD would have its primary impact on the home end (i.e., production end) of walk-access transit trips.

- Cul-de-sacs were found to be negatively associated with the production end of walk-transit trips and the attraction end of all transit trips in most models. In a couple of cases (HBS and HBO), cul-de-sacs were also found to be positively associated with HOV 2 and 3+ trips.

- Walk accessibility to the nearest Metrorail station was found to have substantial positive influence on transit choices, meaning better walk access resulted in higher propensity to choose the transit mode. As described earlier, the walk accessibility variable was measured as the percentage of the land area in the zone that is a within one-half mile of a Metrorail station. In addition to its direct impact on transit, we also documented dramatic shifts in the estimated coefficients of other variables when this variable was included. However, we believe this variable may be difficult to forecast into the future and would not be sensitive to many important transit policies. For these reasons, we recommend omitting this variable from the mode choice models.

- There are several cases across the estimated models where the absolute value of the t-stats associated with estimated coefficients was less than 2.0, which is the threshold associated with statistical significance at the 95 percent confidence level. There are a few reasons these variables were retained in the recommended models. Our typical approach is to consider the reasonableness of the estimated coefficient and our judgment regarding how important we think the variable is to the type of model being estimated. When a variable is critical (e.g., travel time in the mode choice context), we would never consider dropping the variable, though we would consider constraining the variable coefficient potentially. If we are testing a variable more in an exploratory fashion, we would be more discerning, looking for statistical significance and coefficient reasonableness in sign and magnitude. Another reason for retaining a variable we think is important to the choice context of the model is so that the variable exists and can be adjusted in model calibration if we find a need for such adjustment. When variables

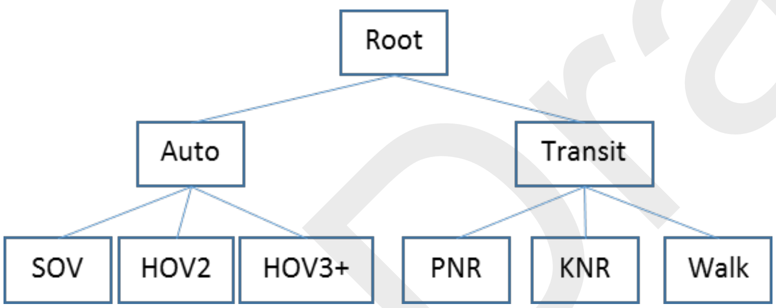
[17] See, e.g., Kamruzzaman, M., F. Shatu, J. Hine, G. Turrell (2015) Commuting Mode Choice in Transit Oriented Development: Disentangling the Effects of Competitive Neighborhoods, Travel Attitudes, and Self-Selection. *Transport Policy*, 42, 187-196.

are dropped, they are often not coded into the applied model with coefficient of zero, but are instead not included in the application at all.

- Rho-squared values[18] with respect to the constants only models for HBW and NHBW trips were 0.19 and 0.08, respectively, which is generally in the range typically found in the context of mode choice models.[19] For HBS, HBO, and NHBO models, the rho-squared values were much lower, only 0.02, 0.03, and 0.01, respectively. While these values are generally lower than what one typically observes, it is important to recognize that the model fit does not necessarily indicate how well the model will perform or whether the model will reasonably reflect the behavior of travelers. It is simply a statistical measure of how well the model explains the variation in the observed data. What is much more critical is the reasonableness of the model parameters used in applying the model. Based on our judgment, we feel that the model sensitivities suggested by each of the estimated models are reasonable. In addition, the model sensitivities will be examined as part of model validation.

- Nested logit model specifications were also tested, in addition to the MNL models we ended up with. In each nested logit model test, we found that the estimated nesting parameters were inconsistent with random utility theory, and therefore were rejected (i.e., the nesting coefficient estimate was not between 0 and 1). A couple of nesting structures were tested, but in each structure, the three transit alternatives were always placed in a single nest and all three auto alternatives appeared in other nest(s). The nesting structures that were tested are shown in Figure 4.9 and Figure 4.10.

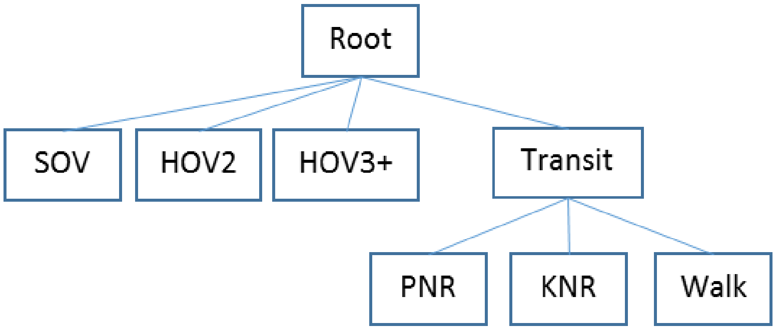
Figure 4.9 First Nesting Structure Tested



[18] Rho-squared and adjusted rho-squared values are measures of how well the model fits the data.

[19] Model fit for non-home-based trips tends to be lower than home-based trips for a couple of reasons. First, non-home-based trip models are more limited in the characteristics of the traveler that are controlled for in the model, since these trips are not linked to the home location, which is the basis for the traveler characteristics in model application. Second, non-home-based trips include a greater diversity in terms of the types of trips. For instance, a non-home-based trip could be one made to go out to lunch from work, to go to a business meeting, or to make a quick stop on the way home from work.

Figure 4.10 Second Nesting Structure Tested



In addition to the general findings listed above, the recommended models are described in the following subsection of the results.

HBW Trips

HBW mode choice estimation results are shown in Table 4.8. As described above, a single level-of-service coefficient was estimated, with all other level-of-service coefficients constrained using the constraints described earlier. The sign and magnitude of the estimated level-of-service coefficient is reasonable. What is particularly relevant is the (average) IVT coefficient implied from estimated coefficient, which was -0.038. Typically the FTA's rule-of-thumb suggests that IVT coefficient should be between -0.02 and -0.03. The magnitude estimated for the HBW model is slightly larger. Because the FTA's guidance is with respect to an IVT coefficient that is applied uniformly to all travelers and the IVT coefficient for the model here varies across VOT segment (the value of -0.038 represents a weighted average across VOT segment), we believe the FTA's rule-of-thumb may not necessarily be appropriate here.

We found several income indicator variables to have important impacts on the model. Income category 1 travelers were found to have a higher propensity for shared ride modes, KNR, and especially walk-transit. These findings are consistent with expectations. We might have expected PNR to have a higher propensity of usage for these travelers than drive alone, but the model did not support this expectation, likely for two reasons.[20] First, PNR mode often requires car ownership and low income households tend to have lower car ownership. Second, because transit service in the peak periods is often competitive with auto modes (the time period for which many HBW trips are made), higher income travelers have a higher propensity toward PNR. Income category 4 travelers were found to have a lower propensity for the walk-transit mode, which was reasonable.

The no-transit access variables in the HBW model align with the earlier, more general findings, described above (see that discussion for more detail). The transit accessibility variables were found to be very important to choice of transit mode. At the production trip end, higher transit accessibility was found to lower the propensity for drive access transit mode usage, probably because when transit accessibility is very high, it does not make sense to access transit by driving. In contrast, walk-transit propensity is increased with

[20] Note that coefficients specific to income 1 and PNR and drive alone modes are not in the table. In the case of drive alone, this is because it is the reference or base alternative, from which other alternatives are measured. For PNR, it is because the coefficient that was estimated was very small with low t-statistic, so the variable was dropped from the model. The implication is that income 1 travelers have no inherent preference for PNR relative to drive alone.

transit accessibility at the production trip end. At the attraction trip end, higher transit accessibility is associated with higher transit propensity across all transit modes, which makes sense since regardless of access mode, typically the egress mode of transit is by walking.

The land use diversity index effects were consistent with the more general findings, described above.

Cul-de-sacs were found to be negatively associated with walk-transit trips at the production end and all transit trips at the attraction end. This makes sense since the prevalence of cul-de-sacs is typically associated with auto modes.

Table 4.8 Home-Based Work Mode Choice Model Estimation Results

Model Fit Statistics	
Observations	51,928
Log Likelihood at zero	-84923.5
Log Likelihood - Constants	-51286.2
Log Likelihood - Final	-41370.1
Rho-Squared (wrt Constants)	0.193
Adj. Rho-Squared	0.193

Variable	Mode	Coefficient	t-stat
ASC	S2	-2.178	-115.2
ASC	S3	-3.268	-94.2
ASC	PNR	-0.492	-6.0
ASC	KNR	-3.982	-16.5
ASC	WTRN	2.569	21.3
Cost ($), Income 1	All	-0.318	-40.5
Cost ($), Income 2 (Constrained)	All	-0.197	n/a
Cost ($), Income 3 (Constrained)	All	-0.150	n/a
Cost ($), Income 4 (Constrained)	All	-0.080	n/a
Avg. Base IVT (min), Income 1 (Constrained)	All	-0.038	n/a
Avg. Base IVT (min), Income 2 (Constrained)	All	-0.039	n/a
Avg. Base IVT (min), Income 3 (Constrained)	All	-0.038	n/a
Avg. Base IVT (min), Income 4 (Constrained)	All	-0.038	n/a
Income 1	S2	0.156	2.8
Income 1	S3	0.573	6.5
Income 1	KNR	0.596	2.4
Income 1	WTRN	3.381	32.1
Income 4	WTRN	-1.228	-15.9
No Transit Access, Prod	PNR	-5.131	-17.4
No Transit Access, Prod	KNR	-1.409	-1.6
No Transit Access, Prod (Constrained)	WTRN	-20.000	n/a
No Transit Access, Attr (Constrained)	All Transit	-3.000	n/a
Transit Accessibility, Prod	PNR	-0.496	-18.9
Transit Accessibility, Prod	KNR	-0.152	-2.0
Transit Accessibility, Prod	WTRN	0.545	15.7
Transit Accessibility, Attr	PNR	1.206	23.0
Transit Accessibility, Attr	KNR	0.096	1.0
Transit Accessibility, Attr	WTRN	0.305	8.7
Diversity Index, Prod	WTRN	0.595	4.1
Cul-de-sacs, Prod	WTRN	-0.011	-9.4
Cul-de-sacs, Attr	All Transit	-0.010	-4.0

HBS Trips

HBS mode choice estimation results are shown in Table 4.9. Like the HBW model, a single level-of-service coefficient was estimated, with all other level-of-service coefficients constrained using the constraints described earlier. The sign and magnitude of the estimated level-of-service coefficient is reasonable. As mentioned above for HBW trips, FTA's rule-of-thumb for the IVT coefficient is between -0.02 and -0.03. For HBS trips, the average IVT coefficient, of -0.021, is consistent with this range. The fact that the estimated IVT sensitivity is less (in magnitude) than the HBW IVT sensitivity, of -0.038, is also consistent with expectations.

Income category 1 travelers were found to have a higher propensity for all transit modes, as shown by the relatively large, positive value of their coefficient estimates. This is in contrast to the HBW model, where PNR mode was not found to have a higher propensity among income category 1 travelers. The difference here largely has to do with trip purpose in our opinion. While PNR may be fairly competitive with auto modes during peak periods when HBW trips are typically made, PNR is less competitive with auto modes in off-peak periods when HBS trips are more typically made. Because of this, the higher cost sensitivity of low income travelers plays a bigger role, and PNR tends to have lower costs associated with it than auto modes. Income category 2 travelers were found to have similarly higher propensity for transit modes (than income category 3 travelers), though less so than income category 1 travelers. Income category 4 travelers were found to have a lower propensity for all transit modes. These travelers were also found to have slightly lower propensity for shared ride modes than drive alone.

The two no-transit access variables that were included in the model apply to the production trip end of walk-transit mode and attraction end of all transit modes. These coefficients were constrained, per the earlier discussion found in the general findings above. Positive sensitivities to the transit accessibility variables were found for production and walk-transit and attraction end for each transit mode. These results are consistent with the HBW model results for similar variables. In the case of the walk-transit production end variable and the KNR attraction end variable, the t-stats were low at 1.1 and 0.9, respectively. Since the coefficient signs and magnitudes were in line with our expectations, the low t-stats were not a concern.

Like the HBW model, the land use diversity index effects found in the HBS model were consistent with the more general findings, described above, with a small positive effect associated with walk-transit mode at the production end of the trip. While the t-stat of this variable was also low (at 0.9), the coefficient sign and magnitude was very similar to the HBW model, offering validation that the estimated coefficient was reasonable.

Cul-de-sacs were found to be negatively associated with walk-transit trips at the production end and all transit trips at the attraction end, like the HBW model. These coefficients also were estimated to have low t-stats (of -1.4 and -0.8, respectively), but because the sensitivities were consistent with HBW results, the variables were retained. For HBS trips, cul-de-sacs were also found to result in a higher propensity for shared ride modes. Cul-de-sacs are associated more with suburban area types, where HBS trip lengths may be longer on average. With this in mind, HBS trips may offer a place for socialization among household members in an otherwise constrained day schedule, in addition to the utility of the shopping activity itself. In more urban areas, shopping trips may be shorter, leaving other opportunities for socialization with household members.

Table 4.9 Home-Based Shopping Mode Choice Model Estimation Results

Model Fit statistics	
Observations	17,711
Log Likelihood at zero	-27806.9
Log Likelihood - Constants	-19601.0
Log Likelihood - Final	-19272.8
Rho-Squared (wrt Constants)	0.017
Adj. Rho-Squared	0.015

Variable	Mode	Coefficient	t-stat
ASC	S2	-0.337	-11.7
ASC	S3	-0.877	-24.1
ASC	PNR	5.595	1.5
ASC	KNR	-4.426	-2.0
ASC	WTRN	-0.014	0.0
Cost ($), Income 1	All	-0.260	-3.5
Cost ($), Income 2 (Constrained)	All	-0.161	n/a
Cost ($), Income 3 (Constrained)	All	-0.122	n/a
Cost ($), Income 4 (Constrained)	All	-0.065	n/a
Avg. Base IVT (min), Income 1 (Constrained)	All	-0.021	n/a
Avg. Base IVT (min), Income 2 (Constrained)	All	-0.021	n/a
Avg. Base IVT (min), Income 3 (Constrained)	All	-0.021	n/a
Avg. Base IVT (min), Income 4 (Constrained)	All	-0.021	n/a
Income 4	S2	-0.025	-0.5
Income 4	S3	-0.426	-5.5
Income 1	PNR	6.040	2.4
Income 2	PNR	1.164	0.9
Income 4	PNR	-4.780	-2.3
Income 1	KNR	3.290	2.1
Income 4	KNR	-2.179	-0.6
Income 1	WTRN	4.146	10.0
Income 2	WTRN	0.665	2.1
Income 4	WTRN	-1.848	-3.5
No Transit Access, Prod (Constrained)	WTRN	-20.000	n/a
No Transit Access, Attr (Constrained)	All Transit	-3.000	n/a
Transit Accessibility, Prod	WTRN	0.160	1.1
Transit Accessibility, Attr	PNR	4.114	2.5
Transit Accessibility, Attr	KNR	0.646	0.9
Transit Accessibility, Attr	WTRN	0.699	4.7
Diversity Index, Prod	WTRN	0.509	0.9
Cul-de-sacs, Prod	WTRN	-0.009	-1.4
Cul-de-sacs, Attr	All Transit	-0.007	-0.8
Cul-de-sacs, Prod	S2	0.001	2.6
Cul-de-sacs, Prod	S3	0.003	6.8

HBO Trips

HBO mode choice estimation results are shown in Table 4.10. Like the HBW and HBS models, a single level-of-service coefficient was estimated, with all other level-of-service coefficients constrained using the constraints described earlier. The sign of the estimated level-of-service coefficient is reasonable, but the magnitude is substantially less than either the HBW or HBS models. The implied average IVT coefficient is about -0.007, which is of lower magnitude than we typically like to see. In order to achieve reasonable sensitivities to level of service variables, we may decide to increase the magnitude of this coefficient during model calibration/validation.

The overall effects of the income indicator variables in the HBO model is very similar to the effects found for the HBS model, with the exception that the income category 2 variable for PNR mode was dropped from the HBO specification (due to low significance). The reader is referred to the HBS discussion for more details.

Like HBW and HBS models, the two no-transit access variables that apply to the production trip end of walk-transit mode and attraction end of all transit modes were constrained to -20 and -3, respectively. See the earlier discussion found in the general findings above. In addition, the PNR and KNR mode no-transit access variables specific to the production end were estimated to have a large negative coefficient, consistent with the HBW results. Since the variable only takes value of 1 in the most outer (largely rural) zones, this seems reasonable, even though such travelers are driving to access transit.

The estimated coefficients of the transit accessibility variables were found to be similar to those of the HBW and HBS models. At the production trip end, higher transit accessibility was found to lower the propensity for drive access transit mode usage, consistent with HBW results (HBS model dropped these variables). In contrast, the propensity for the walk-transit mode is increased as transit accessibility increases at the production trip end, consistent with both the HBW and HBS models. Similarly, at the attraction trip end, higher transit accessibility is associated with higher transit propensity across all transit modes, also consistent with HBW and HBS results.

Like both the HBW and HBS models, the land use diversity index was found to have a small positive effect associated with walk-transit mode at the production end of the trip. In the case of the HBO model, the effect was smaller than HBW and HBS models and the t-stat of the variable's coefficient was also low (at only 0.6). Nonetheless, because the results were consistent with the previous two models and the expectation, as described earlier in the general findings, the variable was retained.

The effects of cul-de-sacs were found to be similar to those found for the HBS model. The HBS model found negative, but not statistically significant impacts of cul-de-sacs on transit, and positive and statistically significant impacts of cul-de-sacs on shared ride trips. In contrast, the HBO results suggest statistically significant impacts for the transit variables and statistically insignificant impacts for the shared ride variables, though with the same signs in both cases. Due to the consistency in results across models, we accepted the statistically insignificant results.

Table 4.10 Home-Based Other Mode Choice Model Estimation Results

Model Fit statistics	
Observations	41,166
Log Likelihood at zero	-64281.3
Log Likelihood - Constants	-49146.6
Log Likelihood - Final	-47436.5
Rho-Squared (wrt Constants)	0.035
Adj. Rho-Squared	0.034

Variable	Mode	Coefficient	t-stat
ASC	S2	-0.016	-0.8
ASC	S3	-0.373	-16.3
ASC	PNR	2.541	4.5
ASC	KNR	-3.024	-4.3
ASC	WTRN	1.316	6.2
Cost ($), Income 1	All	-0.086	-3.5
Cost ($), Income 2 (Constrained)	All	-0.053	n/a
Cost ($), Income 3 (Constrained)	All	-0.041	n/a
Cost ($), Income 4 (Constrained)	All	-0.022	n/a
Avg. Base IVT (min), Income 1 (Constrained)	All	-0.007	n/a
Avg. Base IVT (min), Income 2 (Constrained)	All	-0.007	n/a
Avg. Base IVT (min), Income 3 (Constrained)	All	-0.007	n/a
Avg. Base IVT (min), Income 4 (Constrained)	All	-0.007	n/a
Income 1	S2	-0.191	-4.8
Income 1	S3	-0.367	-7.0
Income 1	PNR	3.047	7.9
Income 4	PNR	-3.067	-6.3
Income 1	KNR	1.829	3.8
Income 4	KNR	-0.663	-1.1
Income 1	WTRN	3.817	19.5
Income 2	WTRN	0.305	2.2
Income 4	WTRN	-2.311	-9.1
No Transit Access, Prod	PNR, KNR	-4.558	-5.1
No Transit Access, Prod (Constrained)	WTRN	-20.000	n/a
No Transit Access, Attr (Constrained)	All Transit	-3.000	n/a
Transit Accessibility, Prod	PNR	-0.505	-6.4
Transit Accessibility, Prod	KNR	-0.483	-5.0
Transit Accessibility, Prod	WTRN	0.467	7.2
Transit Accessibility, Attr	PNR	2.939	12.2
Transit Accessibility, Attr	KNR	1.162	4.8
Transit Accessibility, Attr	WTRN	0.618	9.0
Diversity Index, Prod	WTRN	0.159	0.6
Cul-de-sacs, Prod	WTRN	-0.008	-2.7
Cul-de-sacs, Attr	All Transit	-0.019	-4.6
Cul-de-sacs, Prod	S2	0.0001	0.2
Cul-de-sacs, Prod	S3	0.0003	1.0

NHBW Model

NHBW mode choice estimation results are shown in Table 4.11. Like each of the previous models, a single level-of-service coefficient was estimated, with all other level-of-service coefficients constrained using the constraints described earlier. Unlike the home-based trip models, the cost coefficient is not segmented by income category. This is because non-home-based trips are not generated at the household, and therefore, in model application, the income category will be unknown. As a result, we treat all trips as being from a single income category. The sign of the estimated level-of-service coefficient is reasonable, and the magnitude is relatively high based on our professional judgment. The implied IVT coefficient is about -0.065. For non-home-based trips, higher sensitivity to level-of-service may be reasonable, since non-home-based trips will only occur when a traveler chains multiple activities into the same tour. When trips are chained, it may become more convenient to use private modes of transport (i.e., auto), particularly in cases were zone pairs are not served particularly well by transit. These sensitivities will be scrutinized during model calibration and validation.

Like the earlier models, the two no-transit access variables that apply to the production trip end of walk-transit mode and attraction end of all transit modes were constrained to -20 and -3, respectively. See the earlier discussion found in the general findings above. In addition, the PNR and KNR mode no-transit access variables specific to the production end were estimated to have a large negative coefficient, consistent with the HBW and HBO results.

The estimated coefficients of the transit accessibility variables were not found to have as consistent effects as in the previously described models. On the production end, negative coefficients were estimated for PNR and KNR modes and on the attraction end, a positive coefficient was estimated for PNR mode, but other transit accessibility variables were dropped to estimated coefficients being unreasonable. We believe this could be because NHB trips are dissimilar to home-based trips in some ways. Indeed, the production and attraction end of trips are not as well-defined since neither end corresponds to a person's home. Similar findings are reported below for the NHBO model. Relatedly, it seems that the level-of-service sensitivities could be making up for the lack of transit accessibility sensitivity (since the level-of-service sensitivity is higher in the NHBW model compared to home-based models).

The land use diversity index was not found to have any impact in the NHBW model. Given that the home-based models saw the impact of this variable at the production/home end of the trip, it makes sense that non-home-based trips do not share this sensitivity.

Cul-de-sacs were found to have a negative impact on propensity to use transit modes, consistent with findings for the home-based models.

Table 4.11 Non-Home-Based Work Mode Choice Model Estimation Results

Model Fit statistics	
Observations	13,918
Log Likelihood at zero	-22528.3
Log Likelihood - Constants	-12014.4
Log Likelihood - Final	-11100.2
Rho-Squared (wrt Constants)	0.076
Adj. Rho-Squared	0.075

Variable	**Mode**	**Coefficient**	**t-stat**
ASC	S2	-1.784	-60.8
ASC	S3	-2.440	-56.3
ASC	PNR	-2.705	-7.7
ASC	KNR	-4.868	-11.0
ASC	WTRN	0.097	1.5
Cost ($)	All	-0.381	-18.8
Avg. Base IVT (min, Constrained)	All	-0.065	n/a
No Transit Access, Prod	PNR, KNR	-3.358	-2.7
No Transit Access, Prod (Constrained)	WTRN	-20.000	n/a
No Transit Access, Attr (Constrained)	All Transit	-3.000	n/a
Transit Accessibility, Prod	PNR	-0.374	-3.6
Transit Accessibility, Prod	KNR	-0.610	-5.0
Transit Accessibility, Attr	PNR	0.491	3.1
Cul-de-sacs, Prod	WTRN	-0.030	-5.4
Cul-de-sacs, Attr	All Transit	-0.014	-2.5

NHBO Trips

NHBO mode choice estimation results are shown in Table 4.12. Like each of the previous models, a single level-of-service coefficient was estimated, with all other level-of-service coefficients constrained using the constraints described earlier. Like the NHBW model, the cost coefficient is not segmented by income category because non-home-based trips are not generated at the household and income category is unknown in model application for these trips. The sign of the estimated level-of-service coefficient is reasonable, but like the NHBW model, the magnitude is relatively high. The implied IVT coefficient is about -0.07. Consistent with the reasoning noted above for NHBW trips, higher sensitivity to level-of-service may be reasonable. Nonetheless, the level-of-service sensitivity will be scrutinized during model calibration and validation.

Like the earlier models, the two no-transit access variables that apply to the production trip end of walk-transit mode and attraction end of all transit modes were constrained to -20 and -3, respectively. See the earlier discussion found in the general findings above. Unlike any of the earlier models, we found a positive impact of the production end no transit access variable for PNR and KNR modes. Due to the lack of a clear home-end directionality of NHB trips, we decided to accept this result even though it was in contrast to other model results, the reasoning being that the model may be picking up an effect not captured elsewhere in the model. Since very few zones have no transit accessibility and they tend to rural area type, it is unlikely that this variable will have a major impact on the applied model.

When transit accessibility variables were included in the model, they were not found to have statistically significant and/or reasonable estimated coefficients. As such, these variables were removed from the specification. As noted for NHBW trips, we believe the NHB models higher estimated level-of-service sensitivities are related to the lower sensitivities to transit accessibility variables.

Like the NHBW model, the land use diversity index was not found to have any impact in the NHBO model.

Cul-de-sacs were found to have a small positive impact on propensity to use shared ride modes, consistent with findings for the HBS and HBO models.

Table 4.12 Non-Home-Based Other Mode Choice Model Estimation Results

Model Fit Statistics	
Observations	18,135
Log Likelihood at zero	-28275.0
Log Likelihood - Constants	-20408.6
Log Likelihood - Final	-20143.6
Rho-Squared (wrt Constants)	0.013
Adj. Rho-Squared	0.013

Variable	Mode	Coefficient	t-stat
ASC	S2	-0.283	-12.1
ASC	S3	-0.915	-30.2
ASC	PNR	-3.604	-13.4
ASC	KNR	-4.361	-14.7
ASC	WTRN	-0.722	-7.6
Cost ($)	All	-0.412	-14.6
Avg. Base IVT (min, Constrained)	All	-0.070	n/a
No Transit Access, Prod	PNR, KNR	2.827	5.7
No Transit Access, Prod (Constrained)	WTRN	-20.000	n/a
No Transit Access, Attr (Constrained)	All Transit	-3.000	n/a
Cul-de-sacs, Prod + Attr	S2, S3	0.001	5.4

5.0 Assignment Enhancements

This report section presents documentation of enhancements to both the transit assignment and highway assignment approaches in the regional model. The transit assignment enhancements are primarily focused on the new use of the PT module in Cube. The traffic assignment enhancements are primarily focused on incorporating a value of time segmentation into the model to improve modeling capabilities around managed lanes. Section 5.1 presents background on the development of the value of time segmentation. Section 5.2 presents the Updates to the Transit Assignment. Section 5.3 presents the Updates to the Highway Assignment.

5.1 Value of Time Segmentation

This report section discusses the value of time segmentation incorporated into the delivered travel demand model system to support managed lane policy analysis.

Objective

The objective of the analysis in this report section was to develop a framework by which value of time (VOT) segmentation could be implemented. The analysis details the income distributions in the Washington region, assumptions about the mean VOTs for each income group and by trip purpose, assumptions about how VOTs are distributed and the sources of VOT heterogeneity, and from this, makes recommendations on how VOT segments should be generated. This work presumed that the VOT segments to be used in the travel model could not be estimated directly via the statistical estimation of the mode choice model.

VOT Analysis

Household Travel Survey

The following analysis used the 2007-08 household travel survey data, as provided by MWCOG. Several important pieces of information can be obtained from the household travel survey itself, including the distribution of incomes and workers within the region and the distribution of trips by income and purpose within the region. Table 5.1 shows the joint distribution of income and workers, along with the average number of workers per household (which is used in the subsequent analysis on mean VOTs).

Table 5.1 Distribution of Households by Income and Workers in Washington Household Travel Survey

Income	Workers				Total	Share	Worker/HH
	0	1	2	3+			
$0-50K	209,954	214,630	73,061	9,851	507,496	21.7%	0.769
$50-100K	124,419	402,247	246,957	37,421	811,043	34.7%	1.243
$100-150K	50,586	235,158	321,717	59,165	666,627	28.5%	1.584
$150K or more	17,899	90,421	209,338	36,458	354,116	15.1%	1.747
Total	402,858	942,455	851,074	142,895	2,339,281	100.0%	1.314

One key assumption made regarding VOTs is that they are homogenous for work trips and homogenous for non-work trips. That is, we group all non-work trips into a single category and assume the distribution of VOT is the same. Table 5.2 shows the distribution of trips by purpose and income category in the household travel survey. The table further segments non-work trips between home-based and non-home-based. This is for convenience in later analysis, since VOT distribution of NHB trips will not be segmented by income due to the travel model not segmenting NHB trips by income.

Table 5.2 Distribution of Trips by Income and Purpose in Washington Household Travel Survey

	Purpose	\$0-50K	\$50-100K	\$100-150K	\$150K or more	Total
Trips	HBW	496,633	1,264,762	1,286,863	777,324	3,825,582
	HBNW	1,620,512	3,705,458	3,777,362	1,947,727	11,051,059
	NHB	816,824	1,925,713	1,846,627	969,118	5,558,282
	Total	2,933,969	6,895,933	6,910,851	3,694,169	20,434,922
Share	HBW	13.0%	33.1%	33.6%	20.3%	100.0%
	HBNW	14.7%	33.5%	34.2%	17.6%	100.0%
	NHB	14.7%	34.6%	33.2%	17.4%	100.0%
	Total	14.4%	33.7%	33.8%	18.1%	100.0%

Mean VOTs

The literature suggests that VOT is related to wage rate in a region. Therefore, if it is possible to ascribe reasonable estimates of wage rate for each household income category, it should be possible to derive reasonable VOT assumptions. Willumsen (2014)[21] suggests that, for traffic and revenue studies, the relevant VOT for work trips is 50 to 80 percent of the prevailing wage rate, while for non-work trips, it is 50 to 60 percent of the prevailing wage rate. On the other hand, in a more general context, Litman (2013)[22] states that most studies find VOT for work trips to be closer to 25 to 50 percent of the prevailing wage rate.

We looked at the Bureau of Labor Statistics (BLS), from which average wage rates for the entire region can be found. Here, the BLS category for the Washington-Arlington-Alexandria region was used. The mean wage rate for the region was found to be \$31.69.[23] If we assume that VOT for HBW trips and non-HBW trips is 50 percent and 33 percent, respectively, of the prevailing wage rate, then average HBW and non-HBW VOTs are \$15.85 and \$10.56.

In order to go deeper and look at mean VOTs by income category, wage rates must be assigned to each income category. To obtain average incomes by income category, American Community Survey (ACS) data

[21] Willumsen, L. "Better Traffic and Revenue Forecasting" (2014), Maida Vale Press, ISBN: 13:978-0-9928433-0-4.

[22] Litman, T. "Transportation Cost and Benefit Analysis II – Travel Time Costs" (2013), Victoria Transport Policy Institute (VTPI), http://www.vtpi.org/tca/tca0502.pdf.

[23] See, http://www.bls.gov/oes/current/oes_47900.htm#00-0000.

(Footnote continued on next page...)

for the Washington region (5-year data, 2006-2010) and national 2010 Census data was used. From the ACS data, income distribution at a finer resolution can be obtained, as shown in Table 5.3. From the national 2010 Census, the categorical income distribution at an even finer resolution ($5,000 increments up to $200,000, a $200,000-250,000 category, and $250,000 or more) can be obtained, along with mean incomes for each category.[24] Of particular importance is the mean income for the top income category, which was nearly $400,000 for the $250,000 or more income category.

Table 5.3 Supplemental Income Data from ACS and Census

Income	Households in Washington Region[1]	Households in United States (in thousands)[2]	Average Household Income[2]	Rounded Avg. Household Income
$0-10K	74,322	9,373	$4,900	$5,000
$10-15K	41,585	7,149	$12,389	$12,500
$15-25K	86,763	14,407	$19,664	$20,000
$25-35K	101,920	13,047	$29,520	$30,000
$35-50K	179,375	16,631	$41,835	$42,000
$50-75K	293,842	21,237	$61,261	$61,000
$75-100K	249,925	13,662	$86,080	$86,000
$100-150K	362,011	14,413	$120,038	$120,000
$150-200K	198,107	5,350	$169,690	$170,000
$200K or more	220,536	4,658	$315,295	$315,000

Notes: [1] Obtained from 5-year 2006-2010 ACS data for Washington region.
[2] Aggregated from national 2010 Census data on household income.

Using the national data, average incomes that coincide with the more aggregate income categories used in the ACS can be computed, as shown in Table 5.3 (rounded average household incomes were used in subsequent calculations). Then, using the ACS households as weights, average household incomes for the four income categories of the household travel survey can be computed. These are shown in column A of Table 5.4.

Table 5.4 Mean VOTs by Income Category Derived from Wage Rates

Income	A. Mean (derived)	B. Wrkr / HH	C. Income / Wrkr	D. Wage Rate	E. HBW VOT	F. Non-HBW VOT
$0-50K	27,312	0.769	35,513	17.76	7.19	4.79
$50-100K	72,490	1.243	58,302	29.15	11.81	7.87
$100-150K	120,000	1.584	75,747	37.87	15.34	10.23
$150K or more	246,384	1.747	141,071	70.54	28.57	19.04
Average	102,552	1.314	78,059	39.03	15.81	10.54

[24] See, http://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-06.2010.html.

One issue here, however, is that wages are a personal variable whereas incomes (at least in our case) is a household variable. This is where the distribution of number of workers by income category is useful. In order to derive a personal level income for each household income group, we can simply divide by the number of workers in the household, on average (see column C of Table 5.4). Dividing by 2000[25], we obtain wage rate estimates for each household income category (column D of Table 5.4). This is not a perfect measurement, since high and low income workers may work different number of hours, on average, which would affect the wage rate calculations. This is ignored here.

Lastly, we can apply a fraction, as described earlier, to the wage rate to obtain VOT. If we assume 50 percent, as was assumed earlier when computing a regional average VOT for HBW trips, we obtain a different regional average VOT of $19.51, rather than the $15.85 above. Since the regional BLS wage data is likely more accurate than the wage calculations here, we adjust the fraction downward so that the average VOT that results is in line with the previous calculation. We use 40.5 percent of the wage rate for HBW trips and 27 percent for non-HBW trips (i.e., two-thirds of 40.5 percent). The results are shown in columns E and F of Table 5.4.

Alternatively, Willumsen (2014) suggests an elasticity of VOT with respect to per capita income of 0.7 to 0.9. Using an elasticity of 0.8 and assuming the overall average VOT for the region matches the value derived from the BLS above, Table 5.5 shows the resulting VOT distribution by income category. Overall, the VOT estimates by income category are very close to those derived above using wage rates. For the purposes of the remainder of the analysis, the VOTs derived in Table 5.4 are used.

Table 5.5 Mean VOTs by Income Category Derived from Income Elasticity

Income	Mean (derived)	Avg. HH Size	Income / Person	HBW VOT	Non-HBW VOT
$0-50K	27,312	1.842	14,828	7.52	5.01
$50-100K	72,490	2.259	32,086	12.59	8.39
$100-150K	120,000	2.713	44,233	16.16	10.77
$150K or more	246,384	2.769	88,974	29.29	19.53
Average	102,552	2.375	43,176	15.85	10.56

VOT Distributions

Because we are ultimately interested in VOT segments, it is useful to consider the distribution of VOTs within each income category. In our recent Baltimore activity-based model (ABM) application, we considered a true distribution for VOTs, which was made possible by the fact that the demand part of that model, as an activity-based model, is applied disaggregately. In Washington, the model we are updating is applied as an aggregate model, so aggregate VOT segments must be considered. However, the VOT segments can still be informed by an underlying distribution, which we propose be taken from the Baltimore application.

In the Baltimore application, the cost parameter was actually assumed to be static within each income category, while the travel time parameters were assumed to be log-normally distributed. The log-normal distribution is described by two parameters: the median and a scale parameter that describes the variance

[25] This is an assumed number of work hours per year, to convert yearly income into hourly wages.

(Footnote continued on next page...)

of the distribution. For work tours, the median IVT parameter was equal to -0.018 and for non-work tours, the median IVT parameter was -0.012.[26] The log-normal distributions were defined as having scale parameter of 0.75 for each distribution.

We have adopted identical distributions of IVT for this analysis. Using these distributions and the mean VOT calculations from the previous section, it is possible to compute cost coefficients that would apply to each income category.[27] Using those cost coefficients, we can then describe the full range of the resulting VOT distributions that result from the distributed IVT and static cost coefficients. Note that while the shapes of the distributions are identical to Baltimore's (due to the IVT distributions being identical), the mean VOTs are different, since the VOT depends on both the IVT coefficient and cost coefficient. Mean VOTs are based on the calculations provided above.

While for the current analysis we have adopted the IVT distributions from the Baltimore application, this is for convenience. During mode choice model estimation, we plan on estimating the parameters associated with IVT. The analysis in this memorandum will be used as a guide for making assertions when mode choice data does not support estimation of reasonable model coefficients.

The log-normal curve is ideal for describing the distribution of IVT parameters, because it is a density function that is strictly positive and because the majority of the density function is concentrated toward zero, but it has a long tail going out to the right.

It is worth noting that we have assumed a distribution here for the IVT coefficient (or IVT sensitivity), but not for VOT itself. Instead, the VOT distribution is derived from the IVT distribution and the cost coefficient, where VOT is equal to the IVT coefficient divided by the cost coefficient. A VOT distribution plot can be generated by first generating the x and y values associated with the IVT distribution. Then, x values (i.e., IVT values) are divided by the cost coefficient (and multiplied by conversion factor if necessary, e.g., 60 min/hr). Next, the y values (i.e., density values) must be adjusted by multiplying by the absolute value of the cost coefficient (and divided by the conversion factor if necessary).

Results and Recommendations

Results

From the above information, it is possible to set VOT breakpoints for each of the three VOT segments planned for the travel demand model, and from setting those breakpoints, we can generate a variety of statistics about the composition of each VOT segment, including mean VOT of the segment and number of trips by income category in each segment. It is also possible to compute factors by which each income category is assigned to VOT segment. In other words, what share of low income trips fall into each VOT segment.

Table 6 summarizes the final results for the VOT groups by trip purpose. VOT breakpoints of $4.00 and $15.00 are used (VOT breakpoints can be adjusted, as desired). These breakpoints result in unequal shares of trips falling into each VOT category, with the middle category having about 53 percent of HBW trips and

[26] The log-normal distribution actually takes as input the log of the median IVT parameter, which, since it is negative, must be multiplied by minus one before taking the log. The resulting log-normal distribution is always positive, and the negativity of the IVT coefficient is accommodated by simply multiplying the log-normal distribution by negative one.

[27] Note that given the median and scale parameters of a log-normal distribution, one can compute the mean of the distribution.

56 percent of HBNW and NHB trips. Table 5.6 also shows for each income group, the share of trips assigned into each VOT segment. For instance, about 34 percent of low income HBW trips are assigned to the low VOT segment, while only 1 percent of high income HBW trips are assigned to the low VOT segment. Note that NHB trips will not be segmented by income category in the model. The shares by income category of NHB trips are shown only for informational purposes. The model will assign NHB trips according to the average shares shown in Table 6 (i.e., 25.6, 55.7, and 18.7 percent).

Table 5.6 Characteristics of VOT Groups – By Trip Purpose

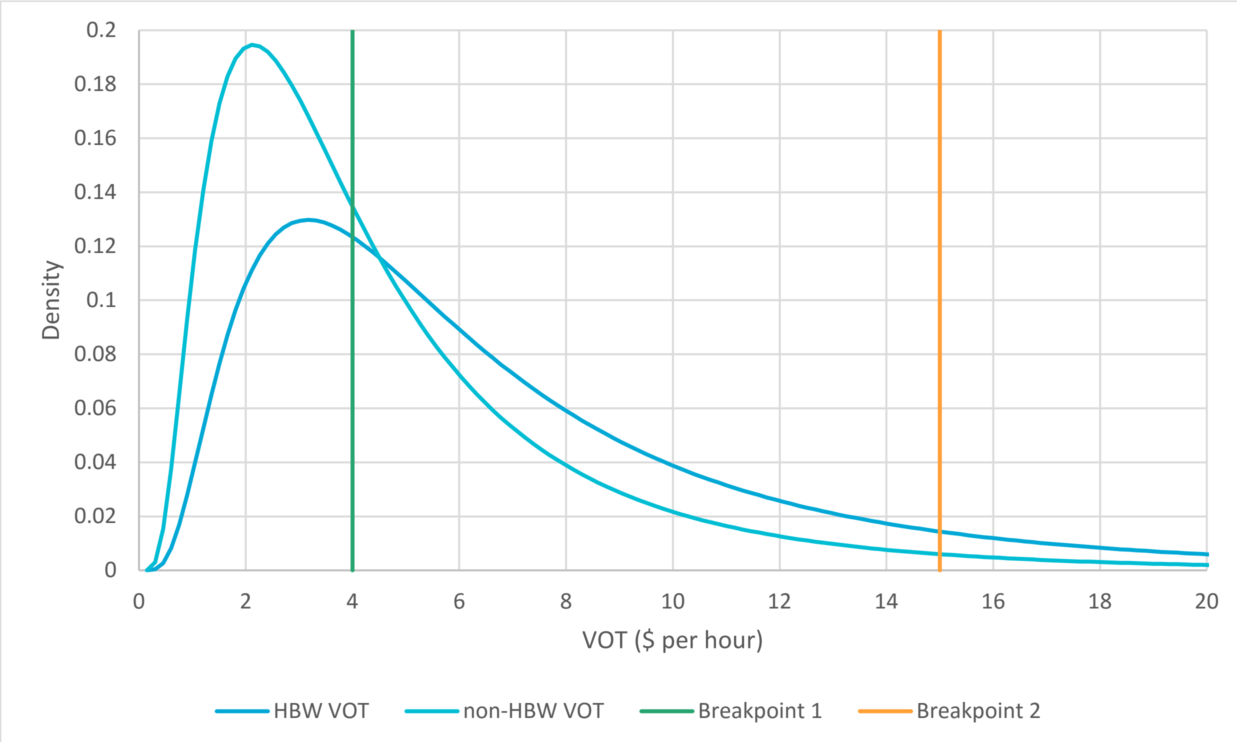
	VOT Groups	VOT Lower Bound	Share – All Incomes	Share - $0-50K	Share - $50-100K	Share - $100-150K	Share - $150K or more
HBW Trips	VOT1	0.00	12.0%	34.2%	14.3%	7.8%	1.2%
	VOT2	4.00	52.6%	57.0%	61.3%	55.7%	30.2%
	VOT3	15.00	35.4%	8.8%	24.4%	36.5%	68.6%
	Total			100.0%	100.0%	100.0%	100.0%
HBNW Trips	VOT1	0.00	25.4%	55.3%	29.9%	19.0%	4.4%
	VOT2	4.00	55.7%	41.8%	59.3%	62.2%	47.9%
	VOT3	15.00	18.8%	2.9%	10.8%	18.8%	47.7%
	Total			100.0%	100.0%	100.0%	100.0%
NHB Trips	VOT1	0.00	25.6%	55.3%	29.9%	19.0%	4.4%
	VOT2	4.00	55.7%	41.8%	59.3%	62.2%	47.9%
	VOT3	15.00	18.7%	2.9%	10.8%	18.8%	47.7%
	Total			100.0%	100.0%	100.0%	100.0%

The values in Table 5.6 are derived in different ways. The VOT bounds are asserted by the analyst and the selection procedure is described in more detail below. The overall shares shown in Table 5.6 come from aggregation of trips by VOT category resulting from the income-specific shares shown in the table. The income-specific shares in Table 5.6 were derived by examining the assumed log-normal distributions of VOT associated with the income group and comparing to the VOT breakpoints. For instance, low income households are assumed to have mean VOTs of $7.19 and $4.79 for HBW and non-HBW trips[28], respectively (see Table 5.4 and the accompanying text for the derivation of the mean VOTs). Using those mean VOTs and the standard deviation parameter assumed above[29], it is possible to describe the VOT distributions of low income households for HBW and non-HBW trips, as shown in Figure 5.1.

[28] Non-HBW trips includes both HBNW and NHB trips. As discussed above, VOT assumptions for these trip purposes were identical.

[29] The median and standard deviation parameters were discussed in the VOT Distributions section above. The median IVT parameter was assumed to be -0.018 for HBW and -0.012 for non-HBW trips and the parameter describing the standard deviation of IVT was assumed to be equal to 0.75, per work done for BMC. To obtain the density plot in Figure 1, the values of the IVT coefficient in its pdf must be converted into equivalent values of low income VOT, using the low income cost sensitivity of -0.1990, as shown in the spreadsheet. This value (-0.1990) was derived using the estimate from earlier that average VOT for low income HBW trips is $7.19.

Figure 5.1 Example VOT Distributions of Low Income Households



Based on the VOT distribution and the VOT breakpoints, it is possible to determine the share of the distributions belonging in each partition. For HBW trips, the cumulative distribution of VOT from $0 to $4 is 0.34, giving these trips a 34 percent chance of belonging to VOT category 1. Similarly, the cumulative distribution of VOT from $4 to $15 (VOT category 2) is 0.57, and the cumulative distribution of VOT from $15 and above (VOT category 3) is 0.09. For non-HBW trips, the shares are 55, 42, and 3 percent, respectively. Since the majority of trips are non-HBW (about 83 percent per Table 2), the share of trips made by low income households that fall into VOT category 1 is about 52 percent (the aggregated total across all low income household trips).

Table 5.7 shows the income distribution of trips within each VOT segment for HBW and HBNW trips. This table is mostly for informational purposes and is not used in any specific way for the model.

Table 5.7 Income Distribution of Trips by VOT Segment and Trip Purpose

	VOT Group	$0-50K	$50-100K	$100-150K	$150K or more	Total
HBW Trips	VOT1	36.9%	39.2%	21.9%	2.1%	100.0%
	VOT2	14.1%	38.6%	35.6%	11.7%	100.0%
	VOT3	3.2%	22.8%	34.7%	39.4%	100.0%
HBNW Trips	VOT1	31.9%	39.8%	25.3%	3.0%	100.0%
	VOT2	11.0%	36.1%	37.8%	15.1%	100.0%
	VOT3	2.3%	19.5%	33.8%	44.5%	100.0%

Table 5.8 shows the overall composition of the VOT groups, aggregating across trip purpose (including HBW, HBNW, and NHB trips). It includes the overall shares of trips within each VOT category, which are aggregations of the shares discussed above across trip purpose. The overall shares are aggregations across income category. These were important, as they were the basis for selecting the VOT breakpoints of $4 and $15. The breakpoints were selected to achieve about 20-25 percent of trips in both boundary VOT categories and about 50 percent in the middle category. The argument for making the middle category larger is that the VOT differentials in the middle part of the distribution are not as big, while larger differentials emerge on toward the edges. For instance, the time and cost sensitivities of individuals from the 70th percentile VOT will be more similar to the time and cost sensitivities of individuals from the 50th percentile VOT than the 90th percentile VOT. This has to do with the relatively long tails one would expect of the continuous VOT distribution (that which we are discretizing to low, medium, and high VOT categories).

Table 5.8 Characteristics of VOT Groups – Overall

VOT Group	VOT Lower Bound	Share - Overall	Mean VOT	Share - $0-50K	Share - $50-100K	Share - $100-150K	Share - $150K or more
VOT1	0.00	23.0%	2.70	51.7%	27.0%	16.9%	3.7%
VOT2	4.00	55.1%	8.29	44.4%	59.6%	61.0%	44.1%
VOT3	15.00	21.9%	27.36	3.9%	13.3%	22.1%	52.1%
				100.0%	100.0%	100.0%	100.0%

Table 5.8 also shows the mean VOTs for each VOT group. Mean VOTs can be derived for each combination of trip purpose, income category, and VOT group using the assumed VOT distributions and the VOT breakpoints. For instance, referring back to Figure 5.1, the HBW low income mean VOT for the low VOT category can be found by taking the integral of the product of VOT with the VOT probability density function over the range from $0 to $4. In this particular example, the mean VOT for HBW low income low VOT trips is $2.63. The mean VOTs reported in Table 5.8 are aggregations across income group and trip purpose.

For the travel model, income segmentation of trips is performed at the production end of the trip for home-based trips, and, thus, is reflective of the income distribution of households in the production zone. The VOT segmentation classification will likewise work as a production end segmentation scheme, to remain consistent with the process by which trips are segmented by income. Since non-home-based trips are not segmented by income category by the travel model, it is unnecessary to define the trip end for which VOT segmentation is performed.

Recommendations

The above analysis served as a framework for associating the income categories with the VOT segments incorporated into the travel demand model. This information was used in the estimation of the mode choice model. What was of particular importance in this regard was the assignment of each income category's trips into the three VOT segments. These assignment factors were used in order to properly weight trip records in model estimation, and thus, were inputs to the model estimation procedures (they could not be products of it). VOTs were allowed to move up or down during model estimation. We looked out for large differences in the relationships between the VOTs of different categories, as this could result in unreasonable discrepancies between the analysis here and the mode choice model.

The key elements of the above analysis are two-fold. First, the trip shares by purpose shown in Table 5.6 are used to assign trips into one of the three VOT categories after the trip distribution step.[30] Second, the mean VOTs shown in Table 5.8 are used to appropriately weight the highway skims and assign highway volumes for each of the three VOT segments.

5.2 Updated Transit Assignment

This section describes the implementation of the updated transit assignment process to accommodate the changes to the mode choice model as well as the conversion of transit skimming/assignment process from the existing TRNBUILD (TB) module to the Public Transport (PT) module.

The updated transit assignment process is performed by time period (peak and off peak), and by access mode (Walk, PnR and KnR). It should be noted that the updated process is performed for a single transit mode, not by transit submode separately as is done in the existing model. Also, the process is not stratified by value-of-time segmentation. Thus, the number of assignment runs in the updated process is much less than that of the existing process (i.e., 6 runs in the updated process vs. 22 runs in the existing process).

The two transit trip table files used in the assignment process, for peak and off-peak periods separately, are generated by assembling the output transit trip tables from the upgraded mode choice model Each of the trip table files consists of 3 tables, associated with the 3 access modes (i.e., walk, PnR and KnR). Similar to the existing transit assignment process, the peak transit trip tables include trips of HBW purpose only, while the off-peak trip tables include trips of all other purposes. Also, the trip tables are in "production/attraction" format.

The assignment process uses the "PT network" files and "route" files that are generated in the path-building and skimming process. This ensures that the transit paths chosen in the assignment process are consistent with the paths developed in the skimming process. Also, it would significantly reduce the computer run time of the assignment process.

[30] The trip shares are also important elements of the weighting procedures used in model estimation. This will be discussed in the model estimation documentation.

The outputs of the updated assignment process are similar to those of the existing process. The process generates two sets of output files in "DBF" format: link volume files and station-to-station Metrorail volume files. The link volume files contain the passenger volume data of all transit and non-transit links in the network. These files are the same as those output from the existing process. However, the headers of the data fields in these files are not the same as the files generated by existing process. Thus, the transit volume summary program "LINESUM.EXE" that summarizes the transit assignment results of the existing assignment process may not work with the output files of the updated assignment process, unless the program is revised to read the link volume data with the updated header names. Modifying the program was not part of the FY17 work program and the program source code was not available to the Consultant. The station-to-station volume files summarize the passenger volumes of individual station pairs of the Metrorail system.

Figure 5.2 displays the flowchart of the input and output files associated with the transit assignment process. The description of the file names is presented in Table 5.9. In the PT process, transit fare is considered in the path choice process. A set of transit fare systems for various transit modes is defined in a fare system file. A fare system defines the fare types (e.g., flat fare, distance-based fare, zone-fares, etc.), boarding fare levels, and transfer discounts for a specific transit mode group. Figure 5.3 displays the fare systems as specified in the fare system file for the AM-Peak PT process.

As shown in Figure 5.3, there are four fare systems defined in the PT process: local bus, express bus, Metrorail and commuter rail. The fare systems are defined by a number of keywords with specified numeric or string values. The following keywords are used to specify the fare systems in the PT process:

NUMBER:	Fare system number;
NAME:	Short name of the fare system;
LONGNAME:	Long name of the fare system;
STRUCTURE:	Fare structure of fare system (e.g., "FLAT" for flat fare, "DISTANCE" for distance-based fare, etc.);
SAME:	A flag that indicates if consecutive transit legs with the same fare system are considered as a single leg in the calculation of fare, with value of either "CUMULATIVE" or "SEPARATE"; CUMULATIVE => Treat consecutive legs as one leg when calculating fare; SEPARATE => Calculate the fare for each leg separately. The default value is CUMULATIVE.
IBOARDFARE:	Boarding fare incurred upon boarding the first transit leg of a trip;
FAREFROMFS:	Transfer fare (or transfer fare discounts) from other fare systems;
FARETABLE:	A set of fare points with distance- and fare-values that define the distance-based fare function;
INTERPOLATE:	A flag that specifies interpolation between coded fare points of the distance-based fare function specified following the keyword "FARETABLE".

For example, Figure 5.3 reveals that the fare system for Metrobus is defined as a flat fare structure, with boarding fare of $1.35, free transfer between Metrobus routes, and transfer fares of $0.90, $0.90, and $1.35 from express bus, Metrorail and commuter rail, respectively. On the other hand, the Metrorail fare system as defined in Figure 5.3 is a distance-based fare structure, with cumulative distance of consecutive Metrorail legs, and with a fare discount of $0.45 transferring from Metrobus or express bus. The coded values following the keyword "FARETABLE" imply that the Metrorail distance-based fare function is defined by a number of fare points with distance- and fare- values as specified in Table 5.10. The values in Table 5.10 are based on the peak-period, station-to-station Metrorail fare function that was estimated earlier and documented in an earlier memo.[31]

The values of the keywords as shown in the fare system files are specified based on the published fare information from WMATA, and other transit operators (e.g., MARC and VRE). The distance-based fare functions of Metrorail and commuter rail are derived from regression analysis based on data of published fares and calculated distances of various station-station pairs of the Metrorail and commuter rail systems.

Figure 5.4 lists the PT script of the updated assignment process. It should be noted that the transit assignment is performed only after the final iteration ("i4') of the model feedback process.

[31] Refer to Section 4.6 of this report.

Figure 5.2 Flowchart of Input and Output Files of the PT Assignment Process

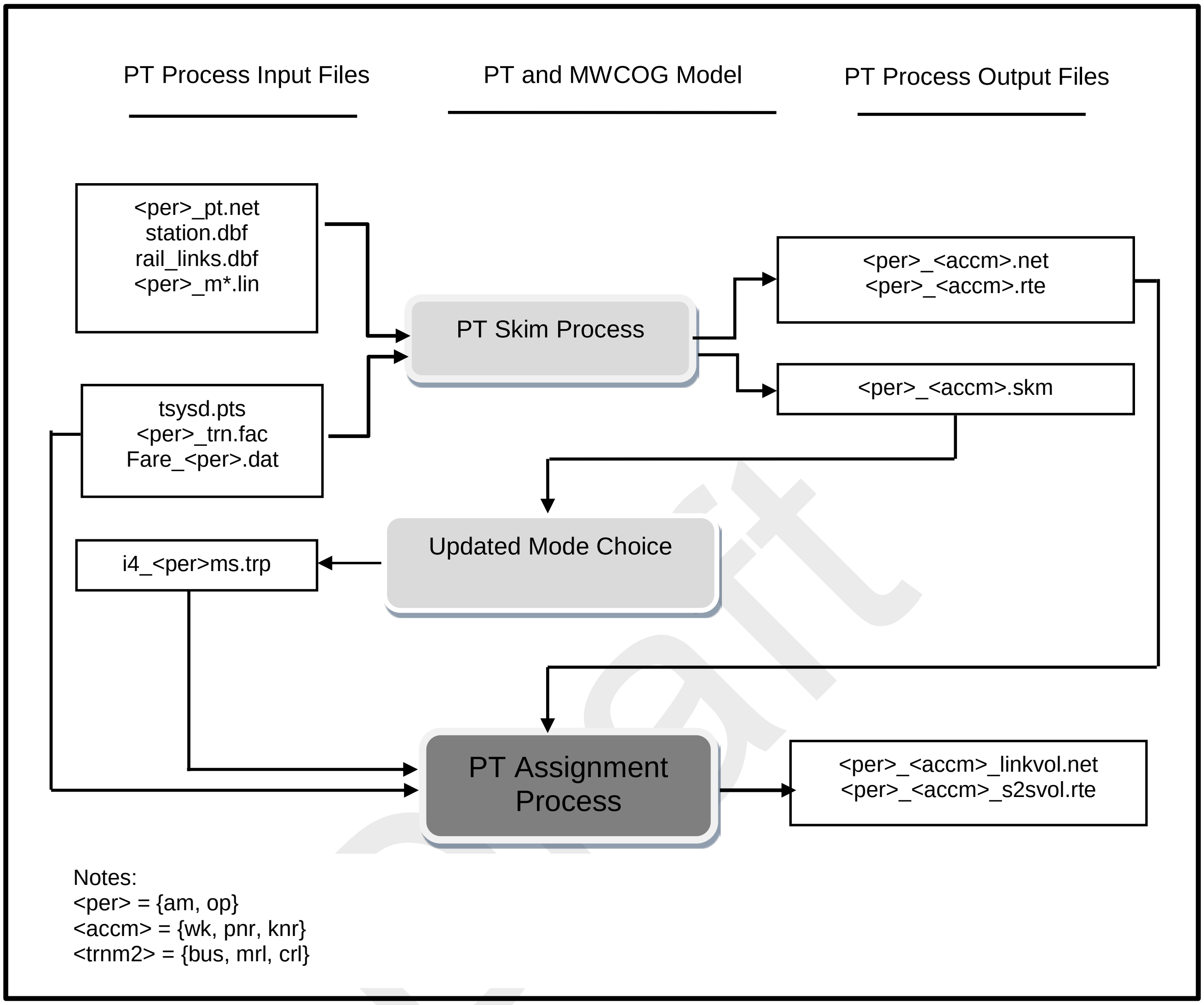


Table 5.9 Input and Output Files of Transit Assignment Process

File Name	File Description
Input Files	
tsysd.pts	Transit system file
<per>_trn.fac	Factor files
fare_<per>.dat	Fare system file
i4_<per>ms.trp	Assembled transit trip tables generated from the mode choice model
<per>_< accm>.net	PT processed network files (generated from PT skim process)
<per>_<accm>.rte	PT generated route files (generated from PT skim process)
Output Files	
<per>_linkvol.dbf	Link volume output files
<per>_s2svol.dbf	Metrorail station-station volume files

Notes: <per> = {am, op}
<accm> = {wk, dr, kr}
<m#> = {m1, m2, m3, m4, m5, m6, m7, m8, m9, m10}

Figure 5.3 Transit Fare Functions for AM Peak PT Skimming and Assignment Process

```
FARESYSTEM NUMBER=1 NAME="Metrobus Peak" LONGNAME="WMATA Metro Bus Peak",
  STRUCTURE=FLAT, SAME=SEPARATE, IBOARDFARE=1.35,
  FAREFROMFS=0,0.90,0.90,1.35
;
FARESYSTEM NUMBER=2 NAME="Express bus Peak" LONGNAME="Express bus",
  STRUCTURE=FLAT, SAME=SEPARATE, IBOARDFARE=3.00,
  FAREFROMFS=2.55,0, 2.55, 3.00
;
FARESYSTEM NUMBER=3 NAME="Metrorail Peak" LONGNAME="WMATA MetroRail Peak",
  STRUCTURE=DISTANCE, SAME=CUMULATIVE,
  FARETABLE=0,1.35, 4.75,1.35, 5.0,1.63, 7.5,2.11, 10.0,2.55, 12.5,2.93,
            15.0,3.25, 17.5,3.52, 20.0,3.74, 21.5,3.85,99.0,3.85,
            INTERPOLATE=T, FAREFROMFS=-0.45,-0.45, 0,0
;
FARESYSTEM NUMBER=4 NAME="Commuter Rail" LONGNAME="Commuter Rail",
  STRUCTURE=DISTANCE, SAME=SEPARATE, FARETABLE=0,4.75, 10.0,4.75, 15.0,5.0285,
            30.0,6.5645, 60.0,9.6365, 65.0, 10.0, 120.0,10.0,
            INTERPOLATE=T
```

Table 5.10 Distance-Fare Values of Metrorail Distance-Based AM Peak Fare Function

Distance (Miles)	Fare ($)
0.00	1.35
4.75	1.35
5.00	1.63
7.50	2.11
10.00	2.55
12.50	2.93
15.00	3.25
17.50	3.52
20.00	3.74
21.50	3.85
99.00	3.85

Source: The values in this table are based on the peak-period, station-to-station Metrorail fare function that was estimated (See Section 4.6 of this report). These values represent an approximation of the quadratic equation that was estimated.

Figure 5.4 PT Assignment Script

```
;; PT Transit Assignment Process
;; Loop by time period (AM, OP) and by access mode (WALK, PNR AND KNR)
;; Input transit trip table files: "I#_AMMS.TRP" and "I#_OPMS.TRP".
;; Each file has 3 tables for 3 access modes.
;;

LOOP PERIOD=1,2
 IF (PERIOD = 1)
  TIME_PERIOD = 'AM'
  AM_MODEL = ' '
  OP_MODEL = ';'
 ELSE
  TIME_PERIOD = 'OP'
  AM_MODEL = ';'
  OP_MODEL = ' '
 ENDIF

LOOP ACCESS = 1,3
  IF (ACCESS = 1)
   ACCESS_MODE = 'WK'
   WLK_MODEL = ' '
   PNR_MODEL = ';'
   KNR_MODEL = ';'
  ELSEIF (ACCESS = 2)
   ACCESS_MODE = 'DR'
   WLK_MODEL = ';'
   PNR_MODEL = ' '
   KNR_MODEL = ';'
  ELSE
   ACCESS_MODE = 'KR'
   WLK_MODEL = ';'
   PNR_MODEL = ';'
   KNR_MODEL = ' '
  ENDIF

RUN PGM = PUBLIC TRANSPORT
FILEI NETI = "@TIME_PERIOD@_@ACCESS_MODE@.NET"
FILEI MATI[1] = I4%_iter_%_@TIME_PERIOD@MS.TRP

FILEI ROUTEI = "@TIME_PERIOD@_@ACCESS_MODE@.RTE"
FILEO REPORTO = "@TIME_PERIOD@_@ACCESS_MODE@_ASGN_M.PRN"
FILEO LINKO = "@TIME_PERIOD@_@ACCESS_MODE@_LINKVOL_M.DBF", ONOFFS=Y
FILEO STOP2STOPO = "@TIME_PERIOD@_@ACCESS_MODE@_S2Svol_M.dbf",
   ACCUMULATE = ADJACENTBYMODE,
   NODES= 1-70000, MODES = 3, LIST=N

FILEI FAREI = INPUTS\FARE_@TIME_PERIOD@.DAT

@WLK_MODEL@ PARAMETERS EFARE=T,TRIPSIJ[1] = MI.01.1
@PNR_MODEL@ PARAMETERS EFARE=T,TRIPSIJ[1] = MI.01.2
@KNR_MODEL@ PARAMETERS EFARE=T,TRIPSIJ[1] = MI.01.3

ZONEMSG=50

ENDRUN
ENDLOOP
ENDLOOP
```

5.3 Updated Highway Assignment

The updated highway assignment process is performed with value of time (VOT) segmentation of passenger vehicle trips. For each of the three types of personal vehicles (SOV, HOV2 and HOV3+), the trip table is stratified into three VOT segments (VT1, VT2, and VT3). Thus, together with other three classes of nonpersonal vehicle trips (i.e., commercial vehicle, airport passenger vehicle, and truck), the input trip table file for the updated assignment process contains 12 trip tables (instead of 6 for the existing assignment process). Each of the 12 tables is associated with a specific mean value of time, which is used to calculate the generalized cost (or impedance) for the path choice algorithm of the assignment process.

Another update of the assignment process is the specification of the volume-delay functions. The traffic assignment results of the existing model reveal that the model tends to underestimate congested speeds on freeway facilities.[32] The conical type function for freeway and expressway facilities in the existing model are thus replaced with the BPR type function. With two parameters, the BPR function provides greater flexibility in representing travel speeds under congested conditions. Also, previous studies reveal that the BPR function performs reasonably well, as compared with the conical function, in matching the simulated traffic volumes with observed data[33]. After a series of test runs of the assignment process, a set of parameters of the BPR functions are specified as shown in the table below. The conical volume-delay functions of other facility types are kept the same as the existing model.[34] Figure 5.5 displays the volume-delay curves used in the updated traffic assignment process.

Table 5.11 Parameter Values of Updated BPR Functions

Facility Type	Alpha	Beta
Freeways	0.4	8.0
Expressways	0.6	5.0

The execution of the updated traffic assignment process basically follows that of the current model (Ver. 2.3.66). The assignment process is performed for four time periods (i.e., AM, PM, MD and NT). Also, in the updated assignment process, the special assignment procedures implemented in the existing model for the assignment of HOV/HOT traffic are kept in place. To improve the assignment of HOV/HOT traffic on HOV lanes, the current model applies two special procedures associated with the traffic assignment process: the "two-step" assignment procedure and the "HOV3+ Skims Replacement (HSR)" procedure. The "two-step" assignment process is conducted for the two peak periods (AM and PM), first for the assignment of non-HOV3+ vehicles and then for the HOV3+ vehicles. This two-step procedure is designed to ensure that the HOV3+ traffic has a greater incentive to use the HOV lanes, and hence improving HOV 3+ loadings on the

[32] See, for example, AECOM and Stump/Hausman Partnership, *Draft FY 2013 Final Report, COG Contract 12-006: Assistance with* Development *and Application of the National Capital Region Transportation Planning Board Travel Demand Model* (National Capital Region Transportation Planning Board, Metropolitan Washington Council of Governments, July 1, 2013), 6–21, http://www.mwcog.org/transportation/activities/models/review.asp.

[33] Cetin, Mecit, Asad J. Khattak, Mike Robinson, Sanghoon Son, and Peter Foytik (2012), *Evaluation of Volume-delay Functions and Their Implementation in VDOT,* Prepared for Virginia Department of Transportation; and Moses, Ren, Enock Mtoi, Steve Ruegg; and Heinrich McBean (2013), *Development of Speed Models for Improving Travel and Highway Performance Evaluation*, Final Report Prepared for Florida Department of Transportation.

[34] See, for example, p. 8-15 of Ronald Milone et al., *Calibration Report for the TPB Travel Forecasting Model, Version 2.3, on the 3,722-Zone Area System*, Final Report (Washington, D.C.: National Capital Region Transportation Planning Board, January 20, 2012), http://www.mwcog.org/transportation/activities/models/documentation.asp.

priority-use and general-use facilities. With the two-step procedure, the entire assignment process consists of 6 assignment runs as follows:

- AM peak Non-HOV3+
- AM peak HOV3+
- PM peak Non-HOV3+
- PM peak HOV3+
- Midday
- Evening

The "HOV3+ Skims Replacement (HSR)" procedure, also called the "multi-run" procedure, was implemented to accommodate the VDOT policy that the operating speeds of the HOT lanes will not be degraded due to the toll paying traffic. Under the HSR procedure, the entire MWCOG model is run twice for each alternative containing HOT lanes (i.e., "base run" and "final run"). The "base run" captures the travel times for unimpeded flow of HOV traffic on HOT lanes. The "final run" of the travel model uses the HOV skims from the "base run" and other skims from the "final run". The detailed description of the two special procedures is provided in the model user's guide[35]. **These two special procedures will be kept in the execution of the updated traffic assignment process**. However, Cambridge Systematics has previously recommended that this approach be streamlined (see Section 8.2).

The Voyager script file for the existing highway assignment process, "Highway_Assignment_Parallel.S", was revised for performing the assignment process by VOT segment. Also, the input file "Toll_Minutes.TXT" that specifies the values of time (in minutes per year-2007 dollar) for various types of vehicle trips for the assignment process needs to be revised to reflect the variation of value of time by VOT segment. Figure 5.6 presents the updated "toll_minutes" file. It should be noted that in the updated highway assignment process, the values of time are specified for individual VOT segments. The values in each segment are kept the same for all vehicle types (SOV, HOV2, and HOV3) and for all time periods. The values for the three VOT segments as shown in Figure 5.6 (i.e., 22.22 mins/$, 7.24 mins/$ and 2.19 mins/$) are specified based on the derived mean values of time for the three VOT segments (i.e., $2.70/hr, $8.29/hr and $27.36/hr).[36] This contrasts with the current model (Ver. 2.3.66), where the value of time (measured in minutes per dollar) varies by auto occupancy. For example, in the current model, for the AM peak period, the dollars per minute are 2.5, 1.5, and 1.0 for SOV, HOV2, and HOV3+ respectively.[37]

[35] Ronald Milone, Mark Moran, and Meseret Seifu, *User's Guide for the COG/TPB Travel Demand Forecasting Model, Version 2.3.66: Volume 1 of 2: Main Report and Appendix A (Flowcharts)* (Washington, D.C.: Metropolitan Washington Council of Governments, National Capital Region Transportation Planning Board, February 13, 2017), 26 & 204.

[36] See Section 5.1.

[37] Ronald Milone, Mark Moran, and Meseret Seifu, *User's Guide for the COG/TPB Travel Demand Forecasting Model, Version 2.3.66: Volume 1 of 2: Main Report and Appendix A (Flowcharts)* (Washington, D.C.: Metropolitan Washington Council of Governments, National Capital Region Transportation Planning Board, February 13, 2017), 119.

Figure 5.5 Volume-Delay Curves Used in the Updated Traffic Assignment Process

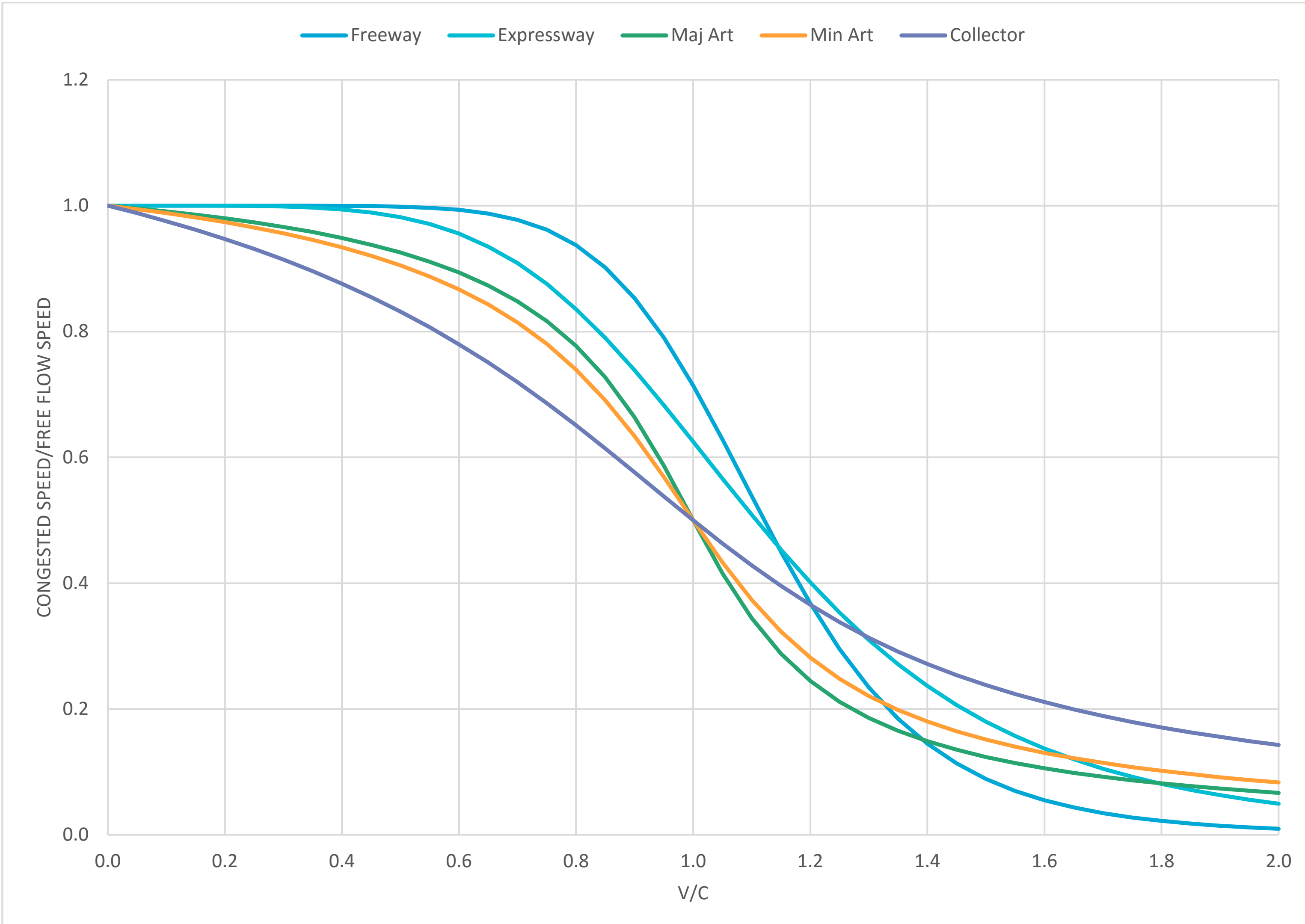


Figure 5.6 Updated Toll Minutes per Dollar Value for Highway Assignment Process

```
;     ======================================================================
;     = Equivalent Toll Minutes by Time Prd & Vehicle Type & by VOT Segment =
;     = in minutes per 2007 dollar - 5/2017                                =
;     ======================================================================
;
;    AM Peak                   Midday               PM Peak                 Night
;---------------         -----------------     ------------------     -------------------

; VOT1 Segment
  SVAMEQMVT1 = 22.22     SVMDEQMVT1 = 22.22    SVPMEQMVT1 = 22.22     SVNTEQMVT1 = 22.22 ; <--- SOVs
  H2AMEQMVT1 = 22.22     H2MDEQMVT1 = 22.22    H2PMEQMVT1 = 22.22     H2NTEQMVT1 = 22.22 ; <--- HOVs-2 Occ
  H3AMEQMVT1 = 22.22     H3MDEQMVT1 = 22.22    H3PMEQMVT1 = 22.22     H3NTEQMVT1 = 22.22 ; <--- HOVs-3+Occ

; VOT2 Segment
  SVAMEQMVT2 =  7.24     SVMDEQMVT2 =  7.24    SVPMEQMVT2 =  7.24     SVNTEQMVT2 =  7.24 ; <--- SOVs
  H2AMEQMVT2 =  7.24     H2MDEQMVT2 =  7.24    H2PMEQMVT2 =  7.24     H2NTEQMVT2 =  7.24 ; <--- HOVs-2 Occ
  H3AMEQMVT2 =  7.24     H3MDEQMVT2 =  7.24    H3PMEQMVT2 =  7.24     H3NTEQMVT2 =  7.24 ; <--- HOVs-3+Occ

; VOT3 Segment
  SVAMEQMVT3 =  2.19     SVMDEQMVT3 =  2.19    SVPMEQMVT3 =  2.19     SVNTEQMVT3 =  2.19 ; <--- SOVs
  H2AMEQMVT3 =  2.19     H2MDEQMVT3 =  2.19    H2PMEQMVT3 =  2.19     H2NTEQMVT3 =  2.19 ; <--- HOVs-2 Occ
  H3AMEQMVT3 =  2.19     H3MDEQMVT3 =  2.19    H3PMEQMVT3 =  2.19     H3NTEQMVT3 =  2.19 ; <--- HOVs-3+Occ

  CVAMEQM    =  2.0      CVMDEQM    =  2.0     CVPMEQM    =  2.0      CVNTEQM    =  2.0  ; <--- Comm Veh
  TKAMEQM    =  2.0      TKMDEQM    =  2.0     TKPMEQM    =  2.0      TKNTEQM    =  2.0  ; <--- Trucks
  APAMEQM    =  2.0      APMDEQM    =  2.0     APPMEQM    =  2.0      APNTEQM    =  2.0  ; <--- Apaxs
```

The outputs of the updated process are the same as the existing process. A combined loaded network is generated storing the link volume and operational data (e.g., speeds, v/c ratios, etc.) for the four time periods and 24-hour volume data. The combined loaded network can be used for some post-assignment analyses as the existing assignment process, including the skimming of congested speeds and accessibility analysis. For mode choice modeling, the VOT-specific paths must be skimmed using the updated VOT-specific "Toll_Minutes" file as mentioned above. For certain analyses involving non-VOT specific paths (e.g., accessibility analyses), a separate set of aggregated non-VOT specific "Toll_Minutes" parameters can be used for the skimming process. However, this may result that the skimming results of the non-VOT specific paths are not consistent with those of the VOT-specific paths. An alternative approach is to use the skimmed paths of a representative VOT segment (e.g., VOT Segment 2) for the non-VOT specific skimming analyses.

As the assignment process is updated with VOT segmentation, certain "pre-model" and "post-model" analyses related to the assignment process also need be updated. These include the select link analysis and the toll setting procedures for the HOT lane facilities. The overall analysis procedures of these analyses remain the same. Only the traffic assignment component associated with these analyses needs to be updated.

6.0 Integrated Model Application for Year 2014

An integrated model application for the Year 2014 was developed to include the enhanced components, but based on the existing Version 2.3.66 model application. All necessary model files and setups were delivered electronically. This section presents the annotated model flowcharts, highlighting the changes made to achieve the delivered model set.

7.0 Model Validation

This section compiles the adopted model validation plan (Section 7.1) and a summary of work performed towards validation as part of the FY17 work program (Section 7.2 through Section 7.4). While additional validation work remains as of the conclusion of the FY17 work program, significant progress was made and sound foundation was delivered to support continued validation efforts.

7.1 Validation Plan

This report section presents the validation plan for the Metropolitan Washington Council of Governments/National Capital Region Transportation Planning Board (TPB) model, as enhanced by the CS team under Task Order 17.2. As background, the model calibration and validation work specified in Task Order 17.2 is presented below:

- **Non-Motorized Model Enhancements** – The results of the new non-motorized model will be calibrated and validated, using the existing data such as household travel survey data and latest American Community Survey (ACS) based Census Transportation Planning Products (CTPP) data.

- **Mode Choice Model Enhancement** – The results of the new mode choice model and transit assignment processes will be validated, including, but not necessarily limited to, the same types of mode choice and transit assignment validation checks normally performed by TPB staff. Model calibration and validation of mode choice and transit assignment will be performed as a joint process to ensure consistency in parameters and weights between the two model components. Sensitivity tests will be performed as part of this task to ensure model sensitivities are appropriate for a variety of scenarios.

- **Managed Lane Modeling** – The results of the new highway assignment process will be validated, including, but necessarily not limited to, the same types of assignment validation checks normally performed by TPB staff. Particular attention will be paid to assigned volumes on toll roads and managed lanes. While there is no observed data on traffic volumes by value of time segment, the effects of different values of time are most pronounced when priced roadways are used.

With this in mind, and considering the specific model components that have been developed under task 17.2, the model validation process will consist of the following steps:

1. Summary of results for model components unchanged by the work in Task Order 17.2
2. Non-motorized model validation
3. Mode choice model validation
4. Transit assignment validation
5. Highway assignment validation

Note that the mode choice validation and transit assignment validation (steps 3 and 4) must be done as an integrated process.

This report section does not provide specific standards, benchmarks, or guidelines for the various comparisons between model results and observed data, or for the proposed sensitivity tests. It is good model validation practice not to set "pass/fail" standards for model validation; this practice is consistent with past TPB model validation efforts.

The base year of the updated model is 2014. Therefore, the observed data that will be used to compare the model results to will reflect the best estimates of 2014 conditions.

Summary of Results for Model Components Unchanged by the Work in Task Order 17.2

The model components that are not being changed directly by the work in Task Order 17.2 include trip generation, trip distribution, and time of day. Some elements of these components might be indirectly affected—for example, while the trip distribution process is unchanged, the inputs related to highway and transit network skims are affected. Other elements might require revisions during the validation/calibration process (for example, time of day factors).

The general objective is that the model changes made in Task Order 17.2 should not adversely affect the validity of model components that are not being changed. Some validation checks related to comparisons of model results to observed data, such as traffic counts, may show slightly higher differences, but these differences should not be large, and it is expected that some checks may show improvements in model results compared to observed data. While the intention here is not to try to correct all pre-existing model validation issues, we should discuss with TPB staff and make note of these issues so that they can be considered in the examination of the new model validation results.

The unchanged model component that we need to pay the most attention needs to is trip distribution. It is important to recognize the impacts of the trip distribution results on subsequent model steps. For example, incorrect modeled origin-destination travel patterns would make it difficult to get transit ridership correct in various corridors and subregions, and it would be hard to produce a good match between modeled volumes and traffic counts on screenlines if the amount of travel crossing the screenlines were inaccurate.

The following checks related to the model components not directly being changed in Task Order 17.2 will be performed prior to validation of the revised model, depending on data availability:

- Comparison of modeled trips per household by trip purpose to data from the household survey
- Comparison of modeled average trip lengths by trip purpose to data from the household survey
- Comparison of modeled origin-destination trips by trip purpose at the district level to data from the household survey
- Comparison of modeled home based work origin-destination trips at the district level to data from the recent American Community Survey (ACS) data set
- Comparison of modeled trips by time period by trip purpose to data from the household survey
- Comparison of modeled vehicle miles traveled (VMT) by time period by trip purpose to data from traffic counts

If these checks indicate the need, adjustments will be made to the trip distribution step.

Non-Motorized Model Validation

The newly created non-motorized model component will be checked through comparisons of the results of this model component to observed data. The main data source is likely to be the household survey data set;

data from the ACS may be used for checks of home based work mode shares. Any available data on walk and bicycle counts at the facility level would be difficult to use since non-motorized trips are not being assigned to networks.

The main validation checks for this component will be comparisons of non-motorized mode shares by trip purpose to targets created based on the household survey/ACS data. The targets should be based on geographic subregion (e.g., counties/cities) and/or area types. It is possible that the targets may have to be more aggregate for some trip purposes with lower non-motorized mode shares.

Once the targets have been set, the non-motorized model results (mode shares) will be compared to the targets. Key model parameters will be changed to improve the match between the model results and the targets. Parameters that may be calibrated include coefficients of land use type variables (e.g., densities, diversity index), income level indicators, and constant terms.

Mode Choice Model Validation

As noted above, mode choice model validation must be done as an integrated process with transit assignment validation. It is important to get the transit boardings correct at some level, not just the transit mode shares.

The main mode choice model check will be comparisons of the modeled mode shares to target shares based on observed data. The data sources for the targets will be the household survey data set, transit on-board survey data, the ACS data (for home based work trips), and transit ridership counts. The target shares should be segmented by trip purpose, geographic, and demographic segments. The number of segments should be enough to produce confidence that the mode choice model accurately reflects travel behavior for important travel markets while considering the limitations of the observed data, especially sample sizes.

The following segmentation variables are recommended:

- Trip purpose – The set of trip purposes in the model (5)
- Geographic segments (8)
  - D.C. to D.C.
  - Inner Maryland counties (Prince George's, Montgomery) to D.C.
  - Inner Virginia jurisdictions (Arlington, Alexandria, Fairfax) to D.C.
  - Other to D.C.
  - Inner/Outer Maryland counties to Inner/Outer Maryland counties
  - Inner/Outer Virginia counties to Inner/Outer Virginia counties
  - Between Maryland and Virginia/West Virginia
  - All other
- Income levels – The set of income levels used in the model (4)

It is certain that the full set of 960 cells in the target matrix (six modes by 160 (5x8x4) segments) will have to be combined due to sparsity of the observed data. This will be done by combining cells with few observed data points and somewhat similar mode shares with "adjacent" cells. (In this case "adjacent" refers to cells with the next higher or lower income level, similar or adjacent geographic definitions (e.g., inner Maryland to D.C. and inner Virginia to D.C.), and relatively non-disparate trip purposes (e.g., NHBW and NHBO). It is expected that the final number of cells will be less than 100.

Each cell target represents the number of trips using a particular mode for a particular trip purpose within a particular geographic area, for travelers of a particular income level. The initial (pre-combination) targets will be set using the following process:

1. Aggregate the total number of trips for each segment defined by trip purpose, geographic segment, and income level from the trip distribution model results.

2. Determine the transit trips by segment, by:

   a. Splitting the total number of regional transit boardings into boardings by access mode (park-and-ride, kiss-and-ride, and walk) using the percentages of boardings by access mode from the on-board survey data

   b. Dividing the total number of regional boardings into unlinked trips by purpose, using the percentages of boardings by trip purpose from the on-board survey data

   c. Producing linked transit trips for each trip purpose by dividing the result of step (b) by the observed transit transfer rate for the trip purpose, from the on-board survey data set

   d. Dividing the transit trips by purpose from step (c) into trips by geographic segment using the origin-destination information for trips by purpose in the on-board survey data set

   e. Dividing the transit trips by purpose and geographic segment from step (d) into trips by income level using the income information for trips by purpose and geographic segment in the on-board survey data set

3. Subtract the transit trips by segment from step 2 from the total trips by segment in step 1 to produce auto trips by segment. (If there are any negative results, they will likely be eliminated when segments are combined in step 5.)

4. Divide the total auto trips by segment into trips by submode (SOV, HOV2, HOV3+) using the percentages of auto submode trips by segment from the household travel survey.

5. Examine segment targets to determine which segments should be combined. This will be based on an examination of cells with low incidence for specific modes and which "adjacent" cells may be similar enough to combine.

Once the initial targets are set, targets with low magnitudes will be combined with "adjacent" cells as discussed above. Then the mode choice model will be run, and the model results compared to the targets. Initially, this will be done with congested skims as a starting point but without feedback. In cases where the model results differ substantially from the targets, the model parameters will be examined and revised (calibrated) as appropriate. The parameters to be examined will include the modal constants, but it may be appropriate to calibrate other parameters, such as the coefficients for income dummy variables, variables related to land use and density, and perhaps even the level of service variables (for example, if the targets show that demand for a mode is underestimated or overestimated for particular geographic segments that are correlated with trip length). The mode choice model will be rerun with the revised parameters, and a new comparison made. This process will be repeated until a satisfactory match between model results and the targets is achieved.

After the comparison between the mode choice model results and the targets is satisfactory, the transit assignment will be run, and the results compared to the transit boarding counts. The transit assignment validation process is described in more detail in the next section. This examination may indicate further calibration changes to the mode choice model (or perhaps the transit path building parameters). For example, if the model is underestimating transit in certain markets or corridors, model parameters correlated with travel in the corridors may be adjusted. After any such changes, new comparisons of the mode choice results to the mode choice targets must be reviewed.

An important component of mode choice model estimation is sensitivity testing. In these tests, specific inputs to the mode choice model are changed, and the resulting effect on mode shares is examined. We will work with TPB staff to define specific sensitivity tests for mode choice, but they could include something like the following examples:

- Changing transit fares for particular transit user segments (or across the board)
- Changing service frequencies for selected transit services
- Changing auto operating or parking costs
- Changing income level distributions
- Assuming different land use patterns by changing values for variables such as the diversity index

After agreeing with TPB staff on the sensitivity tests to be performed, the tests will be done, and any necessary calibration changes associated with the results of these tests will be performed.

Transit Assignment Validation

As discussed in the previous section, transit assignment validation is integrated with mode choice validation. After the mode choice model has been calibrated to achieve a satisfactory comparison to the validation targets, the transit assignment results will be examined.

The main validation check for transit assignment results is comparison of the assigned transit volumes to boarding counts. Generally, boarding counts are obtained for bus service at the route level and rail service at the station level. However, it is not reasonable to expect a good match between modeled and observed boardings for every bus route and rail station in a system as large and complex as is operated in the MWCOG region. There are many overlapping routes on both the bus and rail systems, and there are several essentially equivalent paths for many origin-destination pairs. For example, a trip on Metrorail from Franconia-Springfield on the Blue Line to L'Enfant Plaza on the Yellow Line could transfer at any one of six stations where the two lines overlap. Furthermore, the path builder assumes average tradeoffs among level of service variables (for example, between in-vehicle and wait time), but individual travelers with non-average preferences might choose different paths than those selected by the path builder, and so the model aggregates the number of transit paths used compared to what actual travelers do.

It is therefore common to compare modeled transit boardings to observed data at a more aggregate level. Bus routes are often grouped into route groups, for example by aggregating routes within the same corridor. Rail stations may be grouped by line or line segment, as has been done in past TPB model validation efforts.

The station groups used previously by TPB staff can be used again, or we can work with TPB staff to define appropriate new segments.

We will examine the comparisons of modeled transit boardings to counts and make any further mode choice calibration changes that are indicated, as described in the previous section. The mode choice and transit assignment comparisons will be performed iteratively until satisfactory matches between model results and observed data are achieved for both model components.

Highway Assignment Validation

TPB staff have used the following highway assignment validation checks in previous model validation efforts:

- Comparison of modeled to observed VMT by county/city
- Comparison of modeled volumes to traffic counts for a set of screenlines
- Computation of percentage root mean square error (RMSE) between modeled link volumes and traffic counts by facility type.

These checks are all worthwhile and should be included in the current validation effort. In addition, TPB staff may want to consider the following additional checks:

- Comparison of modeled to observed VMT by facility type
- Comparison of modeled volumes to traffic counts for managed lanes/toll roads
- Computation of percentage root mean square error (RMSE) between modeled link volumes and traffic counts by volume group.

Typically, the highway assignment validation process may indicate that changes are needed in the volume-time functions or network attributes such as speeds and capacities. The checks of volumes on toll roads and managed lanes may reveal some changes that are needed in the value of time segmentation or the speed assumptions on these facilities.

A worthwhile sensitivity test would be to make changes in assumed toll levels to examine the sensitivity of toll road/managed lane use to price (noting that some managed lanes are not priced, e.g., HOV lanes). Additional calibration adjustments may needed if the assignment process is too sensitive or not sensitive enough to changes in price.

7.2 Non-Motorized Model Validation

The estimated non-motorized models take the place of the existing procedures for non-motorized trip estimation in the trip generation estimation process of the TPB travel demand model. As detailed in the model validation plan, the non-motorized model was validated through comparisons of observed data to model results. The primary observed data source was the household travel survey. Observed non-motorized modal shares were tabulated by trip purpose and area type at both the production and attraction trip ends, using the merged household survey datasets (which included the 2007/08 household travel survey and 2011 and 2012 geo-focused household travel surveys).

The 2014 model set provided by the MWCOG was used to implement the model and conduct model calibration and validation. The 2014 model results were tabulated for non-motorized trip modal shares by trip purpose and area type and the model estimates were compared with the observed modal shares. Adjustment factors were applied to the non-motorized trip estimations to ensure the estimated non-motorized modal shares were consistent with the observed non-motorized modal shares. Table 7.1 shows the observed non-motorized shares by trip purpose and area types at the zonal trip production level and Table 7.2 shows the model application results after making adjustments.

Table 7.1 Observed Non-Motorized Modal Shares by Trip Purpose and Area Type – Zonal Productions

Trip Purpose	Area Type						Total
	1	2	3	4	5	6	
HBW	20.2%	5.4%	2.5%	1.2%	0.6%	0.7%	3.2%
HBS	50.8%	18.9%	3.1%	2.2%	1.2%	0.2%	6.7%
HBO	31.2%	16.7%	9.3%	7.6%	5.6%	3.7%	9.4%
NHW	54.4%	17.8%	5.1%	2.0%	3.3%	2.2%	21.0%
NHO	36.9%	17.2%	4.2%	4.2%	3.7%	1.9%	9.9%
Total	41.3%	14.9%	5.4%	4.7%	3.3%	2.1%	9.0%

Data Source: 2007/08 Household Travel Survey (HTS), 2011 Geo-Focused HTS, and 2012 Geo-Focused HTS. COG/TPB

Table 7.2 Estimated Non-Motorized Modal Shares by Trip Purpose and Area Type – Zonal Productions

Trip Purpose	Area Type						Total
	1	2	3	4	5	6	
HBW	21.0%	5.7%	2.5%	1.2%	0.5%	0.7%	3.2%
HBS	54.4%	19.3%	3.2%	2.3%	1.2%	0.2%	7.1%
HBO	32.0%	17.2%	9.5%	7.8%	5.6%	3.7%	10.3%
NHW	57.0%	18.1%	5.0%	2.0%	3.0%	2.7%	21.8%
NHO	40.7%	16.7%	4.4%	4.0%	4.1%	1.1%	9.2%
Total	44.0%	15.3%	5.5%	4.7%	3.4%	2.0%	9.4%

Data Source: Calibration Model Run

Overall, the differences between the estimated and observed shares are small. The largest differences between modeled and observed modal shares appears in the area type 1 segment, where non-motorized mode shares are the highest, over 20 percent for all trip purposes. However, even within the area type 1 segment, the differences between model and observed are never larger than a 3-4 percentage points.

Table 7.3 and Table 7.4 summarize observed and modeled (following adjustments) the non-motorized shares by trip purpose and area type at the zonal attraction level, respectively. Overall, the differences between the estimated and observed shares are small, just as observed at the production level. Again, the area type 1 segment's differences between modeled and observed are the largest, but they are quite minor.

Table 7.3 Observed Non-Motorized Modal Share by Trip Purpose and Area Type – Zonal Attractions

Trip Purpose	Area Type						Total
	1	2	3	4	5	6	
HBW	5.5%	4.0%	1.6%	1.8%	0.8%	1.6%	3.2%
HBS	31.8%	13.9%	1.7%	4.6%	1.1%	0.9%	6.7%
HBO	16.4%	14.0%	6.7%	9.9%	6.2%	6.0%	9.4%
NHW	56.2%	18.1%	5.0%	2.0%	3.0%	2.8%	21.0%
NHO	40.2%	16.8%	4.4%	3.9%	4.0%	1.2%	9.9%
Total	23.7%	12.8%	4.1%	6.7%	4.1%	3.8%	9.0%

Data Source: 2007/08 Household Travel Survey (HTS), 2011 Geo-Focused HTS, and 2012 Geo-Focused HTS. COG/TPB.

Table 7.4 Estimated Non-Motorized Modal Shares by Trip Purpose and Area Type – Zonal Attractions

Trip Purpose	Area Type						Total
	1	2	3	4	5	6	
HBW	5.7%	4.0%	1.6%	1.8%	0.8%	1.6%	3.4%
HBS	33.8%	14.1%	1.7%	4.6%	1.1%	0.9%	6.6%
HBO	17.2%	14.1%	6.8%	9.9%	6.2%	6.0%	9.9%
NHW	57.0%	18.1%	5.0%	2.0%	3.0%	2.7%	21.8%
NHO	40.7%	16.7%	4.4%	4.0%	4.1%	1.1%	9.2%
Total	23.3%	12.9%	4.1%	6.8%	4.0%	3.6%	9.2%

Data Source: Calibration Model Run

7.3 Mode Choice Model Validation

As discussed in Section 7.1, Model Validation Plan, segments were defined for mode choice model validation along three dimensions: trip purpose, geography, and income level.

Section 7.1 discusses how validation targets for each segment were generated using a combination of survey datasets and observed transit boardings. The transit boarding information was considered a more reliable measure of transit trips than transit trips from the surveys.

The first step in the process of developing the validation targets involves the aggregation of the total number of trips for each segment defined by trip purpose, geographic segment, and income level from the trip distribution model results. It should be noted that since the trip generation and trip distribution components of the model were not revised during the model update, the trip distribution model results are used to create the targets for the total number of trips by segment. This reflects that a main objective of the model is to produce demand outputs such as highway volumes and transit ridership, in many cases at the facility level.

This is also reflected in the methods used in the validation of the assignment, components, where the base year model results are compared to counts of highway volumes and transit ridership.

While it is difficult to estimate the "right" number of origin-destination auto trips to produce the set of available observed highway volumes[38], the transit trip outputs of the mode choice component are clearly related to the assigned transit volumes. In other words, to get the transit assignment right, the transit trip tables have to be good.

With that in mind, it is more important to produce the correct number of transit trips than to focus on transit mode shares. So, as described in Section 7.1, the transit mode choice targets reflect the number of observed transit boardings, adjusted to reflect transit transfers and segmented by trip purpose, income level, and geography. Since the total number of trips in the mode choice outputs equal the outputs of trip distribution (and trip generation), the mode choice validation targets must reflect both the observed transit trips and the total number of trips that are inputted to mode choice. This means that the mode choice targets for auto trips must equal the difference between the total trips from trip distribution and the transit trip targets derived from the ridership numbers.

As detailed in Section 7.1, it was necessary to combine some segments due to data sufficiency considerations. For HBS, HBO, NHBW, and NHBO trip purposes, the defining data consideration tended to be observed transit trips in these segments, since transit trip numbers are rather low for these trip purposes, particularly drive access trips to transit (either PNR or KNR). The following combinations were used:

- HOV2 and HOV3+ modes are combined.
- Geographic segments 1 through four were combined.
- The two middle income segments were combined for home based work (HBW).
- For HBW, all income segments were combined for geographic segments 7 and 8.
- The home based shopping (HBS) and home based other (HBO) purposes were combined.
- For all trip purposes except HBW, PNR and KNR access to transit modes were combined.
- For geographic segment 7, the two non-home based purposes, non-home based work (NHBW) and hon-home based other (NHBO) were combined.

These combinations reflect an attempt to retain as much heterogeneity in the targets, while ensuring sufficiency of the underlying data. For instance, HOV2 and HOV3+ modes are very similar (in terms of how and why travelers pursue them) and the HTS data did not distinguish between the two modes for auto passengers anyway. For these reasons, it made sense to combine the two modes. Geographic segments one through four all represent trips attracted to Washington D.C., and it was found the mode shares tended to be similar for each of the four segments. Similar reasoning was used in each of the other groupings.

Table 7.5, Table 7.6, and Table 7.7 show the targets generated for each of the different segments for HBW, HBS and HBO, and NHBW and NHBO trips, respectively.

[38] Trip table estimation methods are available to produce trip tables that reflect the set of traffic counts, these methods are not sensitive to the factors that affect transportation demand in forecast scenarios and are not part of the demand modeling process.

Table 7.5 Mode Choice Model Validation Targets – HBW Trip Purpose

Income	Mode	Geo 1 to 4	Geo 5	Geo 6	Geo 7	Geo 8	Total
<$50K	SOV	42,963	192,497	162,387	n/a	n/a	**397,847**
	HOV2+3	7,040	31,543	26,609	n/a	n/a	**65,193**
	KNR	4,568	1,810	1,195	n/a	n/a	**7,572**
	PNR	16,249	6,438	4,251	n/a	n/a	**26,938**
	WTrn	56,832	22,519	14,869	n/a	n/a	**94,220**
	Total	**127,651**	**254,808**	**209,312**			**591,771**
$50-149.9K	SOV	255,951	613,399	561,090	n/a	n/a	**1,430,440**
	HOV2+3	41,941	100,514	91,942	n/a	n/a	**234,397**
	KNR	11,843	1,480	1,732	n/a	n/a	**15,055**
	PNR	42,130	5,265	6,161	n/a	n/a	**53,557**
	WTrn	147,356	18,416	21,548	n/a	n/a	**187,321**
	Total	**499,222**	**739,074**	**682,473**			**1,920,769**
>$150K	SOV	154,085	273,891	301,886	n/a	n/a	**729,863**
	HOV2+3	25,249	44,881	49,468	n/a	n/a	**119,598**
	KNR	10,302	548	1,149	n/a	n/a	**12,000**
	PNR	36,650	1,950	4,089	n/a	n/a	**42,688**
	WTrn	128,187	6,820	14,302	n/a	n/a	**149,308**
	Total	**354,473**	**328,090**	**370,894**			**1,053,457**
All Income Levels	SOV	**452,999**	**1,079,788**	**1,025,363**	215,807	28,300	**2,802,256**
	HOV2+3	**74,230**	**176,938**	**168,020**	35,363	4,637	**459,188**
	KNR	**26,713**	**3,838**	**4,076**	2,410	2,821	**39,858**
	PNR	**95,029**	**13,653**	**14,501**	8,574	10,034	**141,792**
	WTrn	**332,375**	**47,754**	**50,719**	29,989	35,096	**495,933**
	Total	**981,346**	**1,321,972**	**1,262,679**	**292,143**	**80,887**	**3,939,028**

Note: Bolded cells represent aggregations of targets.

Table 7.6 Mode Choice Model Validation Targets – HBS and HBO Trip Purposes

Income	Mode	Geo 1 to 4	Geo 5	Geo 6	Geo 7	Geo 8	Total
All Income Levels	SOV	300,090	1,949,069	1,631,525	106,349	72,332	**4,059,366**
	HOV2+3	470,710	2,924,366	2,441,738	164,405	106,515	**6,107,734**
	KNR+PNR	35,133	11,053	6,547	1,692	5,047	**59,473**
	WTrn	95,918	30,177	17,875	4,621	13,780	**162,370**
	Total	**901,850**	**4,914,665**	**4,097,685**	**277,067**	**197,674**	**10,388,942**

Note: Bolded cells represent aggregations of targets.

Table 7.7 Mode Choice Model Validation Targets – NHBW and NHBO Trip Purposes

Purpose	Mode	Geo 1 to 4	Geo 5	Geo 6	Geo 7	Geo 8	Total
NHBW	SOV	104,896	550,588	497,427	n/a	35,897	**1,188,808**
	HOV2+3	27,688	145,333	131,300	n/a	9,475	**313,797**
	KNR+PNR	18,488	2,358	2,829	n/a	5,858	**29,533**
	WTrn	50,475	6,438	7,723	n/a	15,993	**80,629**
	Total	**201,547**	**704,718**	**639,279**		**67,223**	**1,612,767**
NHBO	SOV	63,058	741,151	575,556	n/a	25,694	**1,405,459**
	HOV2+3	79,868	938,731	728,990	n/a	32,544	**1,780,133**
	KNR+PNR	7,201	2,134	1,270	n/a	2,273	**12,878**
	WTrn	19,659	5,826	3,468	n/a	6,207	**35,160**
	Total	**169,786**	**1,687,842**	**1,309,285**		**66,717**	**3,233,631**
NHBW+O	SOV	**167,954**	**1,291,740**	**1,072,983**	90,069	**61,591**	**2,684,337**
	HOV2+3	**107,557**	**1,084,064**	**860,291**	46,416	**42,019**	**2,140,346**
	KNR+PNR	**25,689**	**4,492**	**4,099**	1,299	**8,131**	**43,710**
	WTrn	**70,134**	**12,265**	**11,191**	3,547	**22,199**	**119,336**
	Total	**371,334**	**2,392,560**	**1,948,564**	**141,331**	**133,940**	**4,987,729**

Note: Bolded cells represent aggregations of targets.

Calibration Approach

The validation targets were generated by examining trips across four different dimensions: trip purpose, mode, household income, and geographic area. For the first three of these dimensions, calibrating the mode choice models to match targets is fairly straightforward. Since each trip purpose uses a separate model with distinct coefficients, mode-specific constants (i.e., alternative-specific constants) can be adjusted so that targets by those two dimensions are matched. In the case of household income, alternative-specific income indicator (or dummy) variables are included in the model specifications, so those can be adjusted to match income targets. The geographic area targets, however, require a different approach since geographic area constants are not part of the model.[39]

Several variables' coefficients were adjusted to match geographic targets. These variables include the following:

1. Transit accessibility variables – The transit accessibility variables measure the relative transit accessibility from a zone to all other zones.[40] These variables are correlated to transit network density, employment and households density, and relative accessibility via auto modes. For these reasons, we

[39] These were not included specifically to ensure that geographic variations in modal use were reflected via correct model sensitivities of key land use and accessibility variables. Including geographic constants can ensure the model matches modal variations by geography in the base year, but it does not guarantee the model is properly sensitive to key policy variables.

[40] See Section 4.0, Mode Choice Model Enhancements/Estimation, for more details on these variables.

would expect that increasing the transit accessibility coefficient for a transit mode would result in a larger number of trips using transit modes to or from higher density zones, which tend to be in Washington, D.C. and other urban areas of the region. These variables are specified in the mode choice models at both the production and attraction trip ends. Adjustments were primarily made to the attraction trip end coefficients.

2. No transit access variables – These variable coefficients were adjusted in line with adjustments made to the transit accessibility variables. This has to do with ensuring reasonable choices across zones with no transit access and zones with poor transit access, the latter of which would be expected to see higher transit shares than the former.

3. Cul-de-sacs variables – These variables measure the number of cul-de-sacs in a zone at either the production or attraction end. Typically zones with a larger number of cul-de-sacs will be in suburban and residential areas, where transit usage is typically lower. Coefficients of these variables, therefore, can be adjusted to influence the geographic distribution of transit trips.

4. Level-of-service variables – These include the travel time and cost variables. The relationships between travel time and cost coefficients were not changed in calibration. However, what was considered in model calibration was rescaling all of the level-of-service variables in tandem (e.g., by some factor). Ultimately, such changes were only made to the HBO model, where the original estimated coefficients of the level-of-service variables were lower in magnitude than expected. In this case, all level-of-service variable coefficients were multiplied by a factor of 2.0.

5. Urban core indicator variables – These variables were not part of the original model estimation, but were added in an attempt to better match transit mode targets for trips attracted to Washington, D.C. They were applied at the production and/or attraction end of trips.

In some cases, new variables were added to the model specifications because they were not included in the original estimated models. In these cases (with the exception of the urban core indicators), the reason the variables were not included in the estimated models was generally due to the statistical significance of the estimated coefficients being low, and as a result, dropped from the estimated model specification. For urban core indicators, these were not considered in model estimation.

Coefficients of the variables above were adjusted in a trial-and-error approach in order to steer the model results to match targets more closely. One important point to note is that by changing the coefficients of these variables, the average levels of the utility functions change, which necessitates corresponding changes to the modal constants. For instance, the transit accessibility variables are always negative. When transit accessibility is high, the variable has a higher (less negative) value, and when transit accessibility is low, the variable has a lower (more negative) value. As a result, when the transit accessibility variable coefficient is increased, it has the effect of reducing transit mode shares across the board, more so for lower accessible zones and less so for higher accessible zones. In order to compensate, the transit constants necessarily must increase. Therefore, interpreting changes to the transit constants is not a particularly useful exercise here.

Mode Choice Model Validation Results

The companion Excel file[41] compares the mode choice model results with the validation targets that are derived from the available observed data (transit counts and household survey data) using the procedure described in Section 7.1. The transit targets reflect the observed boarding data, converted to linked trips using transfer rates from the survey data and segmented by trip purpose, geography, and income level using the survey data. The auto targets reflect the difference between the trip distribution outputs (already segmented by trip purpose, income level, and geographic segment) and the transit targets; the auto targets are segmented into SOV and HOV trips using the survey data. The total trips in the targets and the model results by purpose, geographic segment, and income level are therefore equal.

The differences between the model results and the targets are shown both in terms of the nominal difference in the number of trips in the segment and the percentage difference. Of course, in cases where the number of trips in the segment is low, large percentage differences reflect small differences in the number of trips.

Table 7.8 compares the observed targets and modeled results for the HBW trip purpose, aggregated across geographies.[42] Looking at the income groups for this trip purpose, the model is somewhat overestimating park-and-ride for the highest income group and underestimating for the middle group. These differences offset, and the total trip differences are not too large. We recommend reexamining this after examining transit assignment results. Otherwise, the model appears to be performing well at this aggregate level, ignoring geographic areas for the moment.

[41] Entitled ‘MC val summary 6-25-17.xlsx’.

[42] The reader is referred to the Excel file for full results for each geographic level.

Table 7.8 Aggregated Mode Choice Validation Comparisons – HBW Trip Purpose

Income	Mode	Observed	Modeled	Pct Difference	Trips Difference
<$50K	SOV	397,847	397,975	0%	128
	HOV2+3	65,193	66,304	2%	1,111
	KNR	7,572	7,674	1%	101
	PNR	26,938	26,643	-1%	-295
	WTrn	94,220	93,175	-1%	-1,045
	Total	591,771	591,771	0%	0
$50-149.9K	SOV	1,430,440	1,434,764	0%	4,325
	HOV2+3	234,397	233,212	-1%	-1,185
	KNR	15,055	15,048	0%	-7
	PNR	53,557	43,079	-20%	-10,478
	WTrn	187,321	194,666	4%	7,345
	Total	1,920,769	1,920,769	0%	0
>$150K	SOV	729,863	702,044	-4%	-27,819
	HOV2+3	119,598	121,616	2%	2,018
	KNR	12,000	13,025	9%	1,025
	PNR	42,688	56,630	33%	13,941
	WTrn	149,308	160,143	7%	10,835
	Total	1,053,457	1,053,457	0%	0
All Income Values	SOV	2,802,256	2,805,553	0%	3,297
	HOV2+3	459,188	464,118	1%	4,929
	KNR	39,858	38,487	-3%	-1,371
	PNR	141,792	137,668	-3%	-4,124
	WTrn	495,933	493,202	-1%	-2,731
	Total	3,939,028	3,939,028	0%	0

Table 7.9 and Table 7.10 compare the observed targets and modeled results for HBS, HBO, NHBW, and NHBO trip purposes, aggregated across geographies. Overall, results of the model and observed targets are very close on a percentage basis. For both HBS and HBO trip purposes, HOV trips are underpredicted in the model by about 30,000 trips each and SOV trips are overpredicted by an opposite amount. Since this only amounts to a one or two percentage point difference to observed trip totals, we think the results are reasonable.

Table 7.9 Aggregated Mode Choice Validation Comparisons – HBS and HBO Trip Purposes

Trip Purpose	Mode	Observed	Modeled	Pct Difference	Trips Difference
HBS	SOV	1,434,365	1,465,761	2%	31,396
	HOV2+3	1,778,769	1,747,997	-2%	-30,772
	KNR+PNR	10,882	10,755	-1%	-127
	WTrn	29,708	29,211	-2%	-497
	Total	3,253,723	3,253,723	0%	0
HBO	SOV	2,626,814	2,659,542	1%	32,728
	HOV2+3	4,327,152	4,294,663	-1%	-32,489
	KNR+PNR	48,591	48,531	0%	-60
	WTrn	132,662	132,482	0%	-180
	Total	7,135,219	7,135,219	0%	0

Table 7.10 Aggregated Mode Choice Validation Comparisons – NHBW and NHBO Trip Purposes

Trip Purpose	Mode	Observed	Modeled	Pct Difference	Trips Difference
NHBW	SOV	1,236,866	1,236,467	0%	-399
	HOV2+3	333,981	335,044	0%	1,063
	KNR+PNR	30,182	30,240	0%	58
	WTrn	82,403	81,681	-1%	-721
	Total	1,683,433	1,683,433	0%	0
NHBO	SOV	1,445,969	1,438,618	-1%	-7,351
	HOV2+3	1,807,866	1,815,952	0%	8,086
	KNR+PNR	13,528	13,656	1%	128
	WTrn	36,934	36,071	-2%	-862
	Total	3,304,297	3,304,297	0%	0

Overall, the number of linked transit trips (1.05 million) seems consistent with expectations. The transit trips in the targets total 1.06 million.

In terms of geographic segments (see the Excel file with full results), the results look good for geographic segments 1-4 and 5. The model overestimates transit in Segment 6, and underestimates it in Segment 8 (which represents a small percentage of regional trips). If the model substantially overestimates boardings in inner Virginia in the transit assignment, it might be worth revisiting this result.

Final Coefficients

As detailed above, several variables were considered during model validation and changes to model coefficients were performed. A full set of model coefficients values is included in a companion Excel file[43]. In this file, the estimated value of the coefficients is shown side by side with the value in the final calibrated model. As noted above, the change in coefficient values for the mode constants were largely a result of compensating for changes in other coefficient values. Prior to making any adjustments to the model, the overall mode shares generated from the estimated models were actually fairly close to observed numbers. The primary reason that changes were made was to adjust the mode shares across the eight geographic areas. The estimated models (prior to adjustment) generated too few transit trips to and from Washington, D.C. and too many elsewhere. The nature of our coefficient adjustments was to shift the transit mode share results to a more closely resemble the observed geographic split. The final coefficients for each mode choice model are presented in Table 7.11 through Table 7.15.

[43] Entitled, 'MC_calibrated_coefficients 06-29-17.xlsx'.

Table 7.11 Final Mode Choice Coefficients – HBW Trip Purpose

Variable	Mode	Coefficient
ASC	S2	-2.178
ASC	S3	-3.268
ASC	PNR	0.458
ASC	KNR	-0.932
ASC	WTRN	7.939
Cost ($), Income 1	All	-0.318
Cost ($), Income 2 (Constrained)	All	-0.197
Cost ($), Income 3 (Constrained)	All	-0.150
Cost ($), Income 4 (Constrained)	All	-0.080
Avg. Base IVT (min), Income 1 (Constrained)	All	-0.038
Avg. Base IVT (min), Income 2 (Constrained)	All	-0.039
Avg. Base IVT (min), Income 3 (Constrained)	All	-0.038
Avg. Base IVT (min), Income 4 (Constrained)	All	-0.038
Income 1	S2	-0.044
Income 1	S3	0.373
Income 1	PNR	3.900
Income 1	KNR	2.846
Income 1	WTRN	5.181
Income 4	WTRN	-1.178
No Transit Access, Prod	PNR	-6.331
No Transit Access, Prod	KNR	-4.409
No Transit Access, Prod	WTRN	-20.000
No Transit Access, Attr	All Transit	-7.200
Transit Accessibility, Prod	PNR	-0.496
Transit Accessibility, Prod	KNR	-0.152
Transit Accessibility, Prod	WTRN	0.945
Transit Accessibility, Attr	PNR	2.256
Transit Accessibility, Attr	KNR	1.296
Transit Accessibility, Attr	WTRN	1.805
Diversity Index, Prod	WTRN	0.595
Cul-de-sacs, Prod	WTRN	-0.061
Cul-de-sacs, Attr	All Transit	-0.030
Urban Core Indicator, Prod	PNR	0.00
Urban Core Indicator, Prod	KNR	0.00
Urban Core Indicator, Prod	WTRN	0.15
Urban Core Indicator, Attr	PNR	0.50
Urban Core Indicator, Attr	KNR	0.60
Urban Core Indicator, Attr	WTRN	0.20

Table 7.12 Final Mode Choice Coefficients – HBS Trip Purpose

Variable	Mode	Coefficient
ASC	S2	-0.337
ASC	S3	-0.877
ASC	PNR	11.445
ASC	KNR	1.224
ASC	WTRN	6.436
Cost ($), Income 1	All	-0.260
Cost ($), Income 2 (Constrained)	All	-0.161
Cost ($), Income 3 (Constrained)	All	-0.122
Cost ($), Income 4 (Constrained)	All	-0.065
Avg. Base IVT (min), Income 1 (Constrained)	All	-0.021
Avg. Base IVT (min), Income 2 (Constrained)	All	-0.021
Avg. Base IVT (min), Income 3 (Constrained)	All	-0.021
Avg. Base IVT (min), Income 4 (Constrained)	All	-0.021
Income 4	S2	-0.025
Income 4	S3	-0.426
Income 1	PNR	6.040
Income 2	PNR	1.164
Income 4	PNR	-4.780
Income 1	KNR	3.290
Income 4	KNR	-2.179
Income 1	WTRN	4.146
Income 2	WTRN	0.665
Income 4	WTRN	-1.848
No Transit Access, Prod	PNR	-3.600
No Transit Access, Prod	KNR	-3.600
No Transit Access, Prod	WTRN	-20.000
No Transit Access, Attr	All Transit	-7.700
Transit Accessibility, Prod	WTRN	0.410
Transit Accessibility, Attr	PNR	5.714
Transit Accessibility, Attr	KNR	2.246
Transit Accessibility, Attr	WTRN	2.249
Diversity Index, Prod	WTRN	0.509
Cul-de-sacs, Prod	WTRN	-0.059
Cul-de-sacs, Attr	All Transit	-0.037
Cul-de-sacs, Prod	S2	0.001
Cul-de-sacs, Prod	S3	0.003
Urban Core Indicator, Prod	PNR	0.00
Urban Core Indicator, Prod	KNR	0.00
Urban Core Indicator, Prod	WTRN	0.15
Urban Core Indicator, Attr	PNR	0.60
Urban Core Indicator, Attr	KNR	0.60
Urban Core Indicator, Attr	WTRN	0.20

Table 7.13 Final Mode Choice Coefficients – HBO Trip Purpose

Variable	Mode	Coefficient
ASC	S2	-0.016
ASC	S3	-0.373
ASC	PNR	6.291
ASC	KNR	0.726
ASC	WTRN	6.836
Cost ($), Income 1	All	-0.172
Cost ($), Income 2 (Constrained)	All	-0.107
Cost ($), Income 3 (Constrained)	All	-0.081
Cost ($), Income 4 (Constrained)	All	-0.043
Avg. Base IVT (min), Income 1 (Constrained)	All	-0.014
Avg. Base IVT (min), Income 2 (Constrained)	All	-0.014
Avg. Base IVT (min), Income 3 (Constrained)	All	-0.014
Avg. Base IVT (min), Income 4 (Constrained)	All	-0.014
Income 1	S2	-0.191
Income 1	S3	-0.367
Income 1	PNR	3.047
Income 4	PNR	-3.067
Income 1	KNR	1.829
Income 4	KNR	-0.663
Income 1	WTRN	3.817
Income 2	WTRN	0.305
Income 4	WTRN	-2.311
No Transit Access, Prod	PNR, KNR	-7.658
No Transit Access, Prod	WTRN	-20.000
No Transit Access, Attr	All Transit	-7.900
Transit Accessibility, Prod	PNR	-0.605
Transit Accessibility, Prod	KNR	-0.583
Transit Accessibility, Prod	WTRN	0.717
Transit Accessibility, Attr	PNR	4.539
Transit Accessibility, Attr	KNR	2.762
Transit Accessibility, Attr	WTRN	2.168
Diversity Index, Prod	WTRN	0.159
Cul-de-sacs, Prod	WTRN	-0.058
Cul-de-sacs, Attr	All Transit	-0.049
Cul-de-sacs, Prod	S2	0.0001
Cul-de-sacs, Prod	S3	0.0003
Urban Core Indicator, Prod	PNR	0.00
Urban Core Indicator, Prod	KNR	0.00
Urban Core Indicator, Prod	WTRN	0.15
Urban Core Indicator, Attr	PNR	0.60
Urban Core Indicator, Attr	KNR	0.60
Urban Core Indicator, Attr	WTRN	0.20

Table 7.14 Final Mode Choice Coefficients – NHBW Trip Purpose

Variable	Mode	Coefficient
ASC	S2	-1.584
ASC	S3	-2.240
ASC	PNR	2.545
ASC	KNR	0.382
ASC	WTRN	8.647
Cost ($)	All	-0.381
Avg. Base IVT (min, Constrained)	All	-0.065
No Transit Access, Prod	PNR, KNR	-6.758
No Transit Access, Prod	WTRN	-20.000
No Transit Access, Attr	All Transit	-8.200
Transit Accessibility, Prod	PNR	-0.474
Transit Accessibility, Prod	KNR	-0.710
Transit Accessibility, Prod	WTRN	1.400
Transit Accessibility, Attr	PNR	2.191
Transit Accessibility, Attr	KNR	1.5
Transit Accessibility, Attr	WTRN	1.4
Cul-de-sacs, Prod	WTRN	-0.130
Cul-de-sacs, Attr	All Transit	-0.054
Urban Core Indicator, Prod	PNR	0.00
Urban Core Indicator, Prod	KNR	0.00
Urban Core Indicator, Prod	WTRN	0.00
Urban Core Indicator, Attr	PNR	0.00
Urban Core Indicator, Attr	KNR	0.00
Urban Core Indicator, Attr	WTRN	0.00

Table 7.15 Final Mode Choice Coefficients – NHBO Trip Purpose

Variable	Mode	Coefficient
ASC	S2	0.267
ASC	S3	-0.365
ASC	PNR	2.796
ASC	KNR	2.039
ASC	WTRN	9.678
Cost ($)	All	-0.412
Avg. Base IVT (min, Constrained)	All	-0.070
No Transit Access, Prod	PNR, KNR	-1.273
No Transit Access, Prod	WTRN	-20.000
No Transit Access, Attr	All Transit	-11.000
Transit Accessibility, Prod	PNR	-0.2
Transit Accessibility, Prod	KNR	-0.2
Transit Accessibility, Prod	WTRN	1.1
Transit Accessibility, Attr	PNR	1.7
Transit Accessibility, Attr	KNR	1.7
Transit Accessibility, Attr	WTRN	2.0
Cul-de-sacs, Prod + Attr	S2, S3	0.001
Cul-de-sacs, Prod	WTRN	-0.090
Cul-de-sacs, Attr	All Transit	-0.040
Urban Core Indicator, Prod	PNR	0.00
Urban Core Indicator, Prod	KNR	0.00
Urban Core Indicator, Prod	WTRN	0.00
Urban Core Indicator, Attr	PNR	0.00
Urban Core Indicator, Attr	KNR	0.00
Urban Core Indicator, Attr	WTRN	0.00

Next Steps

As noted in Section 7.1, Model Validation Plan, the mode choice model validation cannot be considered complete until we have run the transit assignment to see how the modeled transit boardings compare with the observed. We believe that the mode choice model is sufficiently calibrated to proceed with transit assignment. Transit assignment validation should proceed, as described in Section 7.1. The main validation check for transit assignment is comparison of the assigned transit volumes to boarding counts. This can proceed by first aggregating transit boarding information. For instance, bus routes can be grouped into route groups within specific corridors. Rail stations may be grouped by line or line segment, as has been done in past TPB model validation efforts.

In addition to transit assignment validation, sensitivity tests should be performed. Sensitivity tests are important to ensure the model's sensitivity to key inputs is reasonable and appropriate. Section 7.1 details several such sensitivity tests related to the mode choice model that we recommend as follows:

- Changing transit fares for particular transit user segments (or across the board);
- Changing service frequencies for selected transit services;
- Changing auto operating or parking costs;

- Changing income level distributions; and
- Assuming different land use patterns by changing values for variables such as the diversity index.

7.4 Preliminary Assignment Validation

Prior to the expiration of the work program, preliminary validation was undertaken on both the highway and transit assignments. The results presented in this section are based on a preliminary final model run. It is expected that additional refinement will be achieved during the next steps activities.

Highway Assignment

In terms of daily traffic volumes and daily VMT, the overall assignment results of the updated model are comparable with the observed data and also with the assignment results of existing model (see Figure 7.1, Table 7.16, Table 7.17, and Table 7.18).

Two main issues are observed from the preliminary assignment results:

1. Overestimation of traffic volumes on Interstate road segments with HOV3+/HOT facilities (see Figure 7.2)
2. Underestimation of traffic volumes on toll facilities (MD 200 and Dulles Toll Road) (see Figure 7.3)

The issue of overestimation of traffic volumes on Interstate with HOV3+/HOT lanes seems to be limited to the Interstate road segments with HOV3+/HOT lanes (e.g., I-495 and I-95). This issue may be associated with the "2-stage" assignment process and the "HOV+3 skims replacement" process as implemented in the current model setup. These two processes were implemented with the purpose to improve the loading the traffic volumes on HOV3+ facilities. Some further investigation (such as sensitivity tests) of the updated model with the use of these special assignment processes is needed.

The issue of the underestimation of traffic volumes on MD 200 and Dulles Toll Road is potentially associated with the specified "values of time" for various VOT segments in the updated assignment procedure. Figure 7.4 summarizes the specified time values in the existing and the updated traffic assignment processes. As shown in the figure, the time values as specified in the existing model is in general close to values of VOT3 segment (with highest values of time) in the updated model. As a result, the sensitivity of updated model on toll is higher and hence resulting in lower assigned volumes on toll road segments, as compared with the existing model. Thus the time values as specified in the updated assignment may need to be refined.

Figure 7.1 Comparison of Observed and Estimated Daily Traffic Volumes

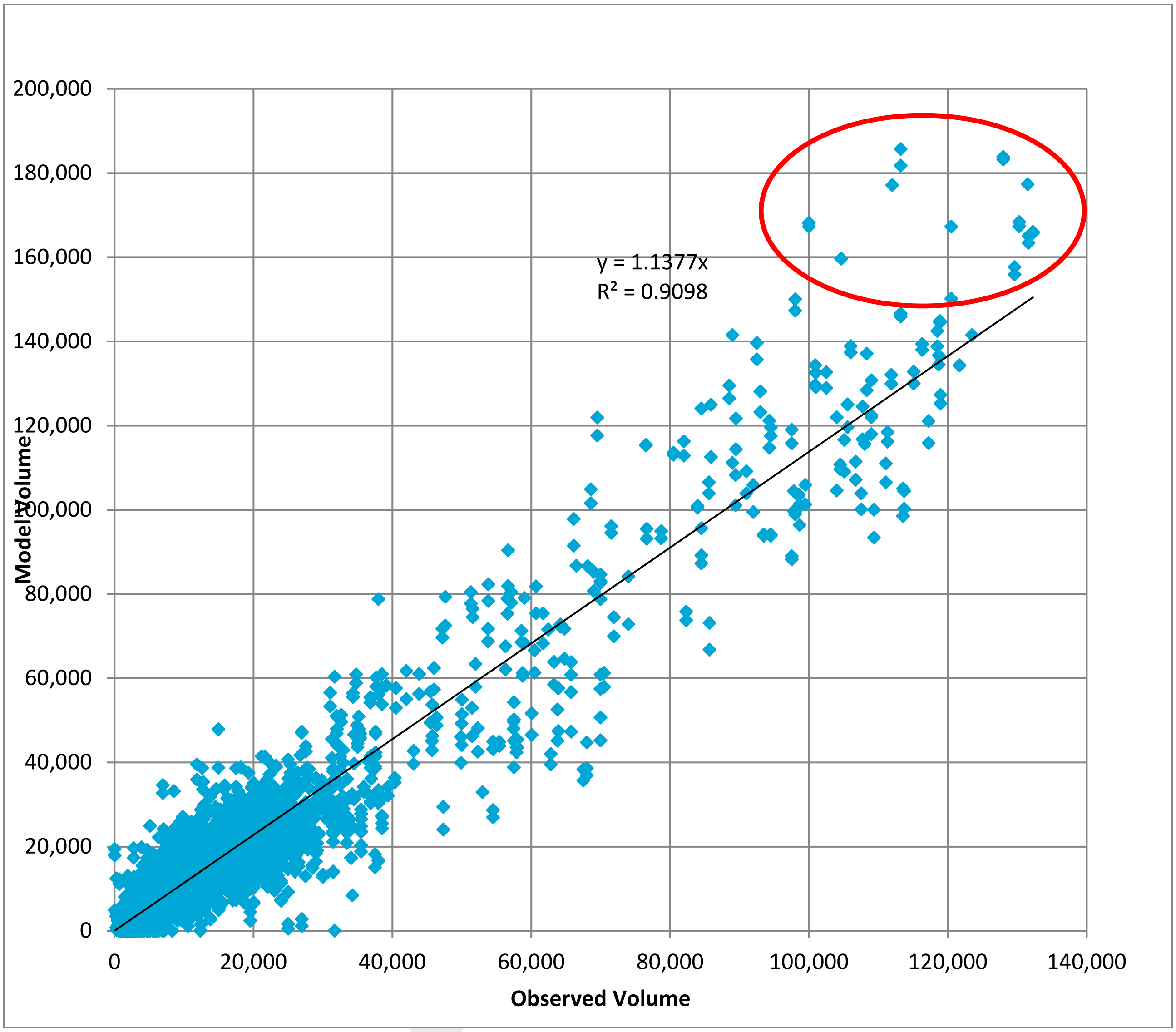


Table 7.16 RMSE by Functional Classification

Functional Class	RMSE	% RMSE	R^2
Interstate/freeway	17913.51	33.56	0.8587
Primary arterial	6763.28	41.16	0.4942
Minor arterial	3913.54	53.84	0.5269
Collector	2966.33	78.85	0.4269

Table 7.17 RMSE by Daily Volume Group

Daily Volume Group	RMSE	% RMSE	R^2
0-5,000	3228.70	119.87	0.1493
5,000-10,000	4477.95	61.13	0.1013
10,000-25,000	6425.62	40.40	0.2251
25,000-50,000	10908.35	32.25	0.4337
50,000-100,000	19686.10	28.44	0.5681
>100,000	28884.37	25.70	0.2678

Table 7.18 VMT by Jurisdiction

Jurisdiction	Observed VMT	Estimated VMT	% Diff.
District of Columbia	7,922,357	9,190,022	16.0%
Montgomery County	19,757,260	23,145,610	17.1%
Prince George's County	23,646,575	25,325,242	7.1%
Arlington County	4,046,638	4,519,765	11.7%
City of Alexandria	2,016,133	2,909,649	44.3%
Fairfax County	26,663,007	28,812,002	8.1%
Loudoun County	6,623,699	7,762,141	17.2%
Prince William County	9,425,332	9,977,823	5.9%
Frederick County	7,798,767	9,625,808	23.4%
Howard County	10,546,027	11,868,866	12.5%
Anne Arundel County	15,493,973	16,332,502	5.4%
Charles County	3,276,575	3,133,258	-4.4%
Carrol County	3,290,959	4,159,363	26.4%
Calvert County	1,987,808	1,795,912	-9.7%
St. Mary's County	2,246,712	2,291,058	2.0%
King George County	871,306	768,006	-11.9%
City of Fredericksburg	929,927	851,800	-8.4%
Stafford County	4,006,798	4,581,962	14.4%
Spotsylvania County	3,442,058	2,275,095	-33.9%
Fauquier County	3,439,861	3,707,556	7.8%
Clarke County	810,485	1,115,934	37.7%
Jefferson County	1,177,470	1,483,599	26.0%
Total	159,419,729	175,632,973	10.2%

Figure 7.2 Daily Volumes of Freeway Segments with HOV Facilities

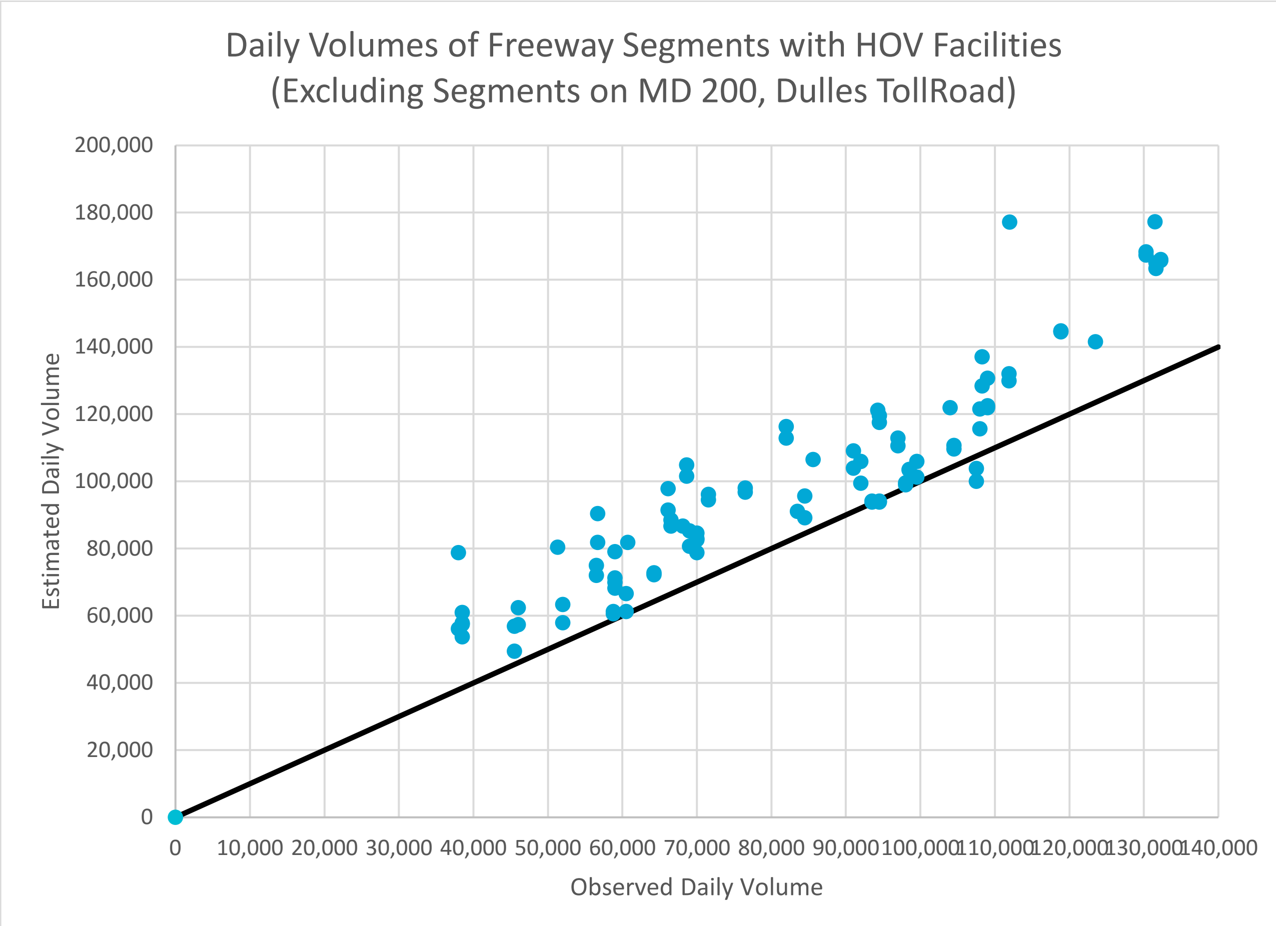


Figure 7.3 Daily Volumes on Toll Facility Segments

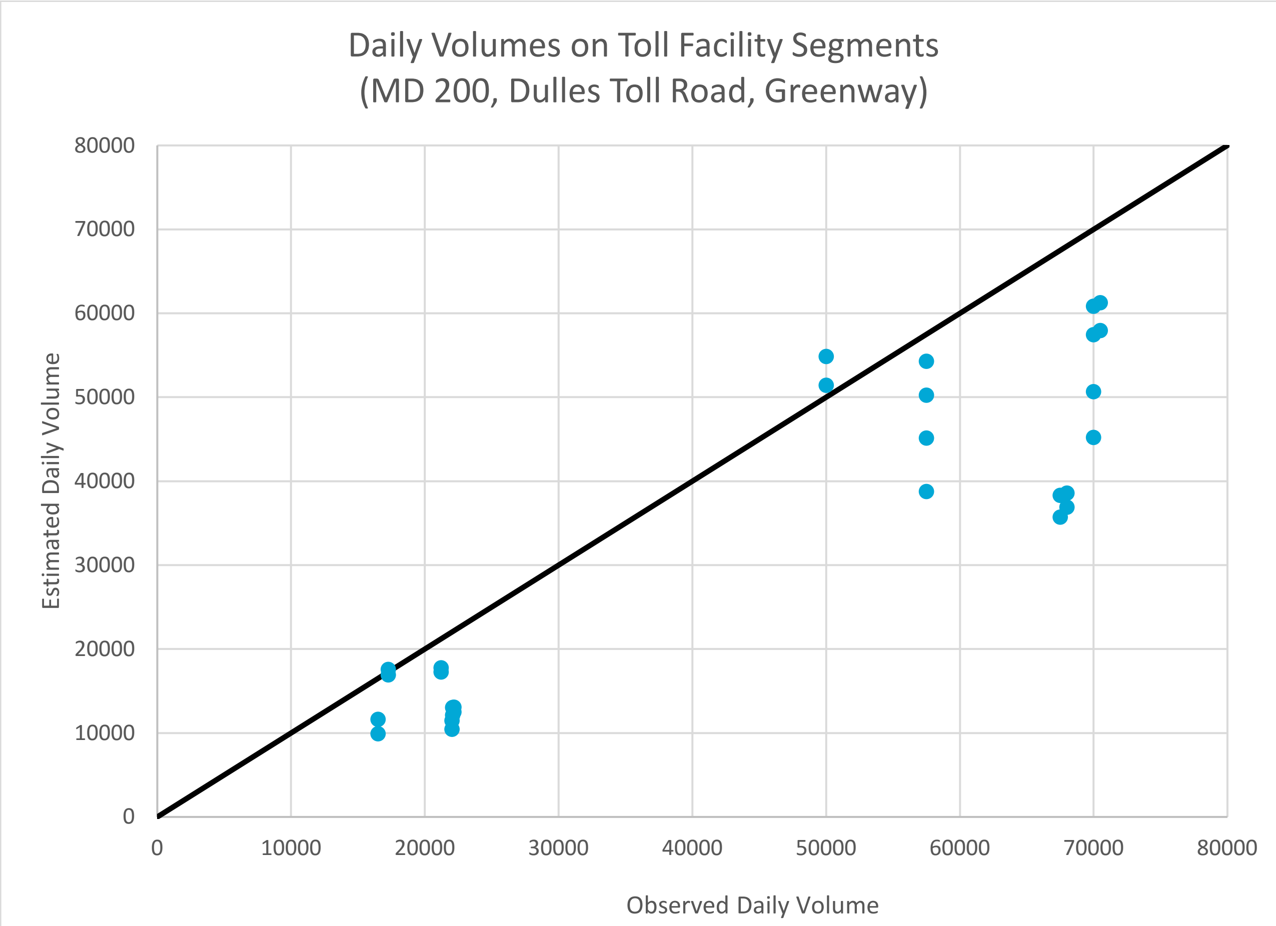


Figure 7.4 Equivalent Toll Minutes by Time Period and Vehicle Type (in minutes per 2007 dollar)

```
; Time Values of Existing Model
; =============================
; AM Peak                 Midday                 PM Peak                  Night
  SVAMEQM = 2.5           SVMDEQM = 3.0          SVPMEQM = 3.0            SVNTEQM = 3.0 ; <--- SOVs
  H2AMEQM = 1.5           H2MDEQM = 4.0          H2PMEQM = 2.0            H2NTEQM = 4.0 ; <--- HOVs-2 Occ
  H3AMEQM = 1.0           H3MDEQM = 4.0          H3PMEQM = 1.0            H3NTEQM = 4.0 ; <--- HOVs-3+Occ

  CVAMEQM = 2.0           CVMDEQM = 2.0          CVPMEQM = 2.0            CVNTEQM = 2.0 ; <--- Comm Veh
  TKAMEQM = 2.0           TKMDEQM = 2.0          TKPMEQM = 2.0            TKNTEQM = 2.0 ; <--- Trucks
  APAMEQM = 2.0           APMDEQM = 2.0          APPMEQM = 2.0            APNTEQM = 2.0 ; <--- Apaxs
;
; Time Values of Updated Model
; =============================
; AM Peak                 Midday                  PM Peak                Night
; VOT1 Segment
  SVAMEQMVT1 = 22.22      SVMDEQMVT1 = 22.22      SVPMEQMVT1 = 22.22     SVNTEQMVT1 = 22.22 ; <--- SOVs
  H2AMEQMVT1 = 22.22      H2MDEQMVT1 = 22.22      H2PMEQMVT1 = 22.22     H2NTEQMVT1 = 22.22 ; <--- HOVs-2 Occ
  H3AMEQMVT1 = 22.22      H3MDEQMVT1 = 22.22      H3PMEQMVT1 = 22.22     H3NTEQMVT1 = 22.22 ; <--- HOVs-3+Occ
;
; VOT2 Segment
  SVAMEQMVT2 =  7.24      SVMDEQMVT2 =  7.24      SVPMEQMVT2 =  7.24     SVNTEQMVT2 =  7.24 ; <--- SOVs
  H2AMEQMVT2 =  7.24      H2MDEQMVT2 =  7.24      H2PMEQMVT2 =  7.24     H2NTEQMVT2 =  7.24 ; <--- HOVs-2 Occ
  H3AMEQMVT2 =  7.24      H3MDEQMVT2 =  7.24      H3PMEQMVT2 =  7.24     H3NTEQMVT2 =  7.24 ; <--- HOVs-3+Occ
;
; VOT3 Segment
  SVAMEQMVT3 =  2.19      SVMDEQMVT3 =  2.19      SVPMEQMVT3 =  2.19     SVNTEQMVT3 =  2.19 ; <--- SOVs
  H2AMEQMVT3 =  2.19      H2MDEQMVT3 =  2.19      H2PMEQMVT3 =  2.19     H2NTEQMVT3 =  2.19 ; <--- HOVs-2 Occ
  H3AMEQMVT3 =  2.19      H3MDEQMVT3 =  2.19      H3PMEQMVT3 =  2.19     H3NTEQMVT3 =  2.19 ; <--- HOVs-3+Occ

  CVAMEQM    =  2.0       CVMDEQM    =  2.0       CVPMEQM    =  2.0      CVNTEQM    =  2.0  ; <--- Comm Veh
  TKAMEQM    =  2.0       TKMDEQM    =  2.0       TKPMEQM    =  2.0      TKNTEQM    =  2.0  ; <--- Trucks
  APAMEQM    =  2.0       APMDEQM    =  2.0       APPMEQM    =  2.0      APNTEQM    =  2.0  ; <--- Apaxs
```

Transit Assignment

The estimated Metrorail system ridership is comparable to the observed Metro ridership.

The estimated numbers (unlinked trips) for other modes are lower than the reported numbers (as provided by MWCOG). This may be related to the total transit linked trips that were used to validate the model (about **1.06M linked trips**), as compared with the **1.4 M unlinked trips** (per data provided by MWCOG).

Some weighting factors (like IVT weights of various modes) in the PT model may need to be further refined with more detailed analysis of the assignment results.

Table 7.19 Transit Ridership by Mode

Transit Submode	Observed	Estimated
Metrorail*	721,804	764,833
Commuter Rail	54,217	37,656
MARC	36,051	30,394
VRE	18,166	7,262
Metrobus	445,623	311,283
Other Bus	202,460	149,431
Total	1,424,103	1,300,859

Note: * Does not count intra-system transfers

8.0 Next Steps and Guidance

The FY17 work program was ambitious. The budget and schedule available to make enhancements served as constraints on some aspects of the work program. The Cambridge Systematics team focused on making the enhancements to the areas of the model that had been agreed to and on delivering a working model set to allow future refinement of both those core features as well as, potentially, additional features. TPB staff will want to familiarize themselves with the delivered model set, perform additional validation tests and further calibration activity, and consider selected further enhancements before moving towards adopting the new model into regular practice.

8.1 Additional Validation Tests and Further Calibration Activity

As described in Section 7.0, Cambridge Systematics produced a validation plan for the updated model set. Many of the described tests were performed, but there are additional suggested validation tests which remain for TPB staff to perform as they familiarize themselves with the delivered model set. Additionally, the results of some tests may lead TPB staff to want to make parameter adjustments to improve the level of validation. For example, based on the preliminary highway assignment, the overall model fit is not yet in our typically acceptable range. Overall VMT is 10 percent high and the R^2 values seem low. These issues may relate to the results being from a preliminary assignment (employing a pre-final version of the mode choice model) or may be related to model components that were not revised in our work (trip generation, trip distribution, truck model, external travel).

Similarly, on the transit assignment it is implied in the preliminary results that there is more work needed. Observed ridership seems to be 9 percent low. Since the validation target is based on the (presumably same) observed ridership, this means that one or more of the following are happening:

- Too few modeled transfers compared to observed due to skim settings or transfer penalties;
- Total number of transit trips are right, but not enough are in places where transferring is more likely/necessary; or
- Some error in the model (e.g., in creating transit trip tables for assignment).

These tests and adjustments are to be expected as part of the work to adopt the delivered enhancements into the agency's production model.

8.2 Future Enhancements

This section presents a few potential future enhancements or known limitations with the delivered model set which could be developed as part of a future work program.

Cube Voyager Issue

During the final month of the project, the Cambridge Systematics team encountered delays due to an undocumented limitation within the Cube Voyager software platform. This limitation required the dividing up of batch controls/script files so as to have a separate instance responsible for driving each iteration of the model. Citilabs indicated that their development team was reviewing the issue. If it is corrected by Citilabs, the implementation batch controls/scripts could then be streamlined to remove the need for separate instances.

Trip Distribution

Addressing the trip distribution models was not part of the FY 17 work program. However, we needed to adapt to the current model (V2.3.66) trip distribution in order to incorporate the new skimming process and resulting skims to replace the old ones. The enhanced model (V2.5) required highway skims segmented by three VOTs and general transit skims (not Metrorail-specific). In order to streamline the adaptation of the existing trip distribution process to support the new transit skimming, mode choice, and assignment routines, the new process uses AM/OP walk and PNR transit skims; and for highway, uses skims for VOT level 2 (average). Future work could potentially review this approach and determine if additional work should be done in "connecting" the existing trip distribution procedures to the enhanced procedures developed as part of the FY 17 work program.

Transit Capacity Constraint

When activated, the transit capacity constraint in the current model (V2.3.66) limits Metrorail trips through the regional core (Downtown DC and Arlington/Alexandria) to not exceed the 2020 level. Excess Metrorail trips are shifted back into the SOV trip table. This is intended to capture the lack of rail car capacity that is currently programmed in the CLRP. The new model works off of generic transit trip tables by mode of access. Since it does not generate Metrorail-specific trip tables, the current transit capacity constraint methodology will not work.

While the new model set does not offer a way to address transit capacity constraint within the mode choice model, the use of Cube Voyager's PT module offers additional options for modeling transit beyond those we have implemented in this work program, including capacity constraint and shadow pricing capabilities. Work could be undertaken in the future to refine the delivered model set to explore setting these parameters as a way to address concerns around transit capacity or transit reliability. Additional functionality could also be developed to set a constraint on Metrorail trips into and through the regional core.

HOT/HOV Two-Step Process

In the current model (V2.3.66), two model runs are conducted: a base run for generating skims for HOV trips, and a final run which has toll present in the skims for SOV trips and the base run HOV skims used for HOV trips. While this current approach mimics the policy of HOT lane pricing being used to maintain free-flow conditions for HOV users, it introduces some inconsistencies into mode choice modeling. Resource and scope of work constraints led the Cambridge Systematics team to maintain this HOT/HOV two-step process in the newly implemented model. However, a future enhancement could be to implement a streamlined one-step HOT/HOV process, which will significantly reduce the run time over the current two-step process.