

GEN3, PHASE 2, MODEL ESTIMATION

Results, Findings and Lessons Learned by COG Staff

Feng Xie, Ray Ngo, Nazneen Ferdous
TPB Transportation Engineers

Travel Forecasting Subcommittee
January 27, 2023

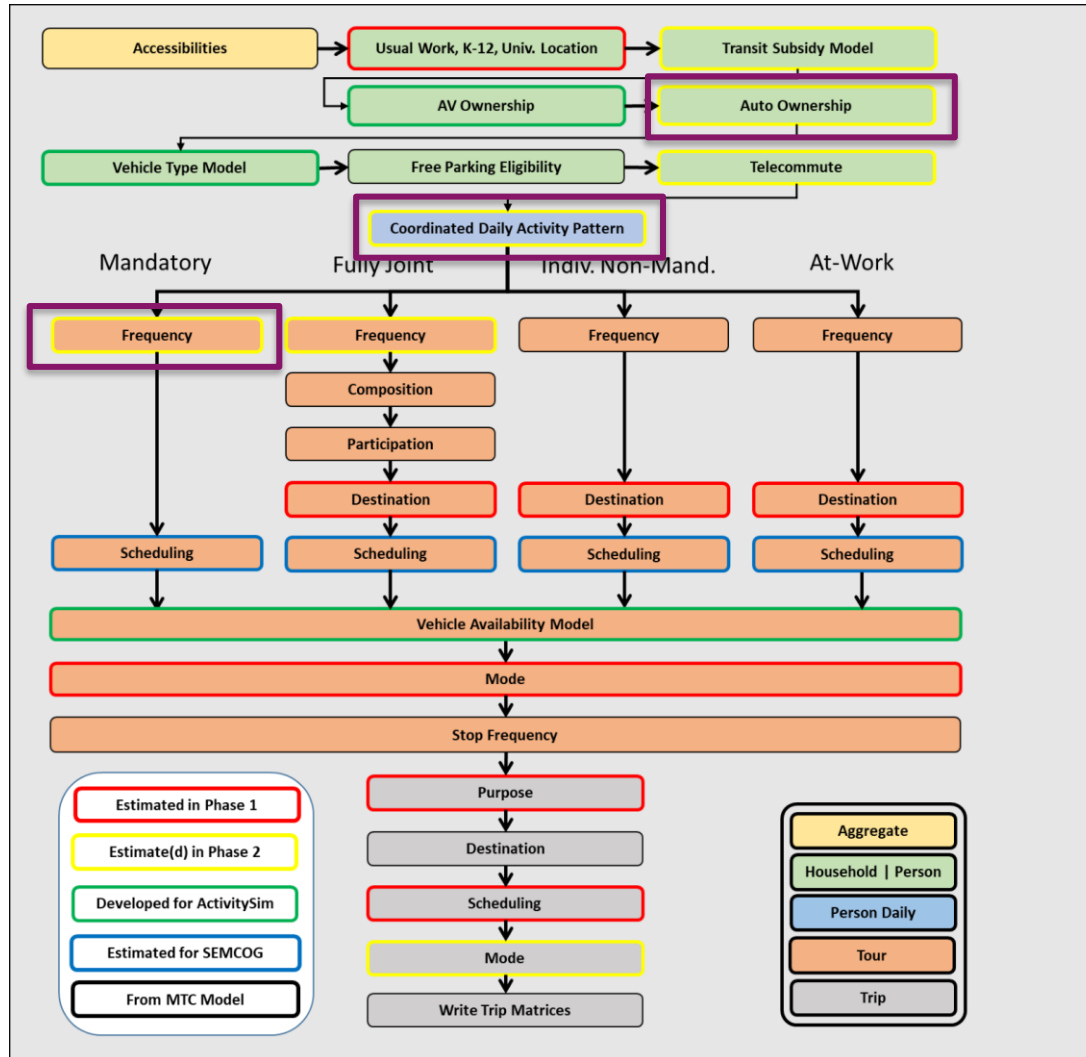


Background

- The Gen3 Travel Model consists of many sub-models.
- Key sub-models have been estimated by RSG and COG staff in both Phase 1 and Phase 2 developments of the Gen3 Model.
- In Phase 2, COG staff estimated the following sub-models:
 - Auto Ownership, by Ray
 - Coordinated Daily Activity Pattern (CDAP), by Nazneen
 - Mandatory Tour Frequency (MTF), by Feng
- We will, in turn, discuss the main results/findings and general experiences from our estimation work conducted by COG staff.

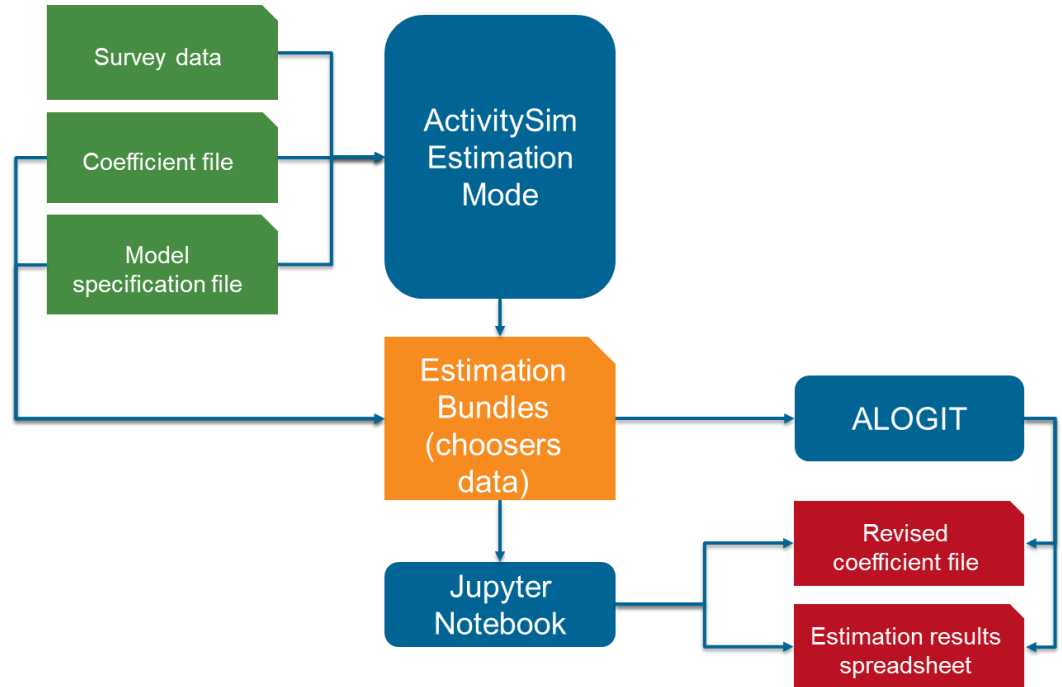


Gen3 Model Structure



Overview of the Estimation Process

- The 2017-18 RTS and 2018-19 MTS survey data were pre-processed by RSG's Survey Processing Application (SPA).
- ActivitySim was run in Estimation Mode to generate Estimation Data Bundles (EDBs) that are needed for model estimation.



- The Phase 2 model estimation work was conducted using a Jupyter Notebook that calls up Larch, an open-source logit model estimation tool built on top of the Python SciPy library.
- The final model coefficient/specification files were generated by the Jupyter Notebook in an iterative process.

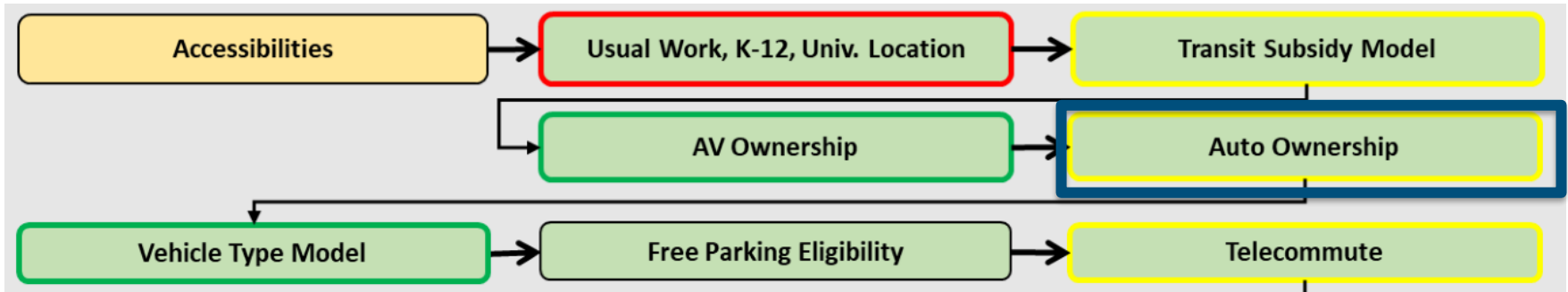


Auto Ownership Model

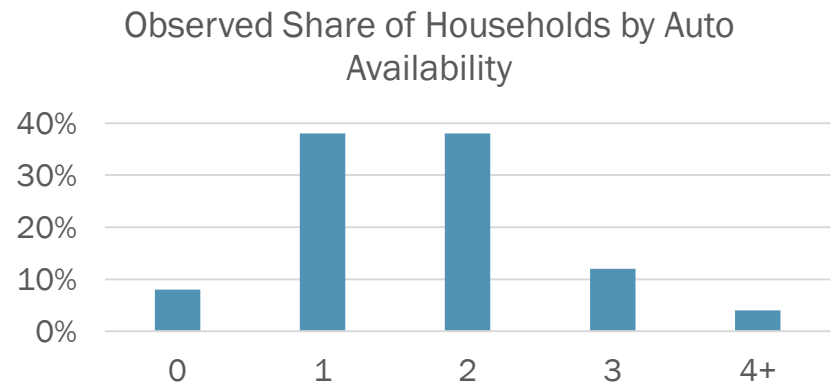
Ray Ngo



Overview



- The auto ownership model estimates the number of autos owned by each household in the synthetic population
- Model type: Multinomial logit
- Choice set: 0, 1, 2, 3, 4+
autos per household

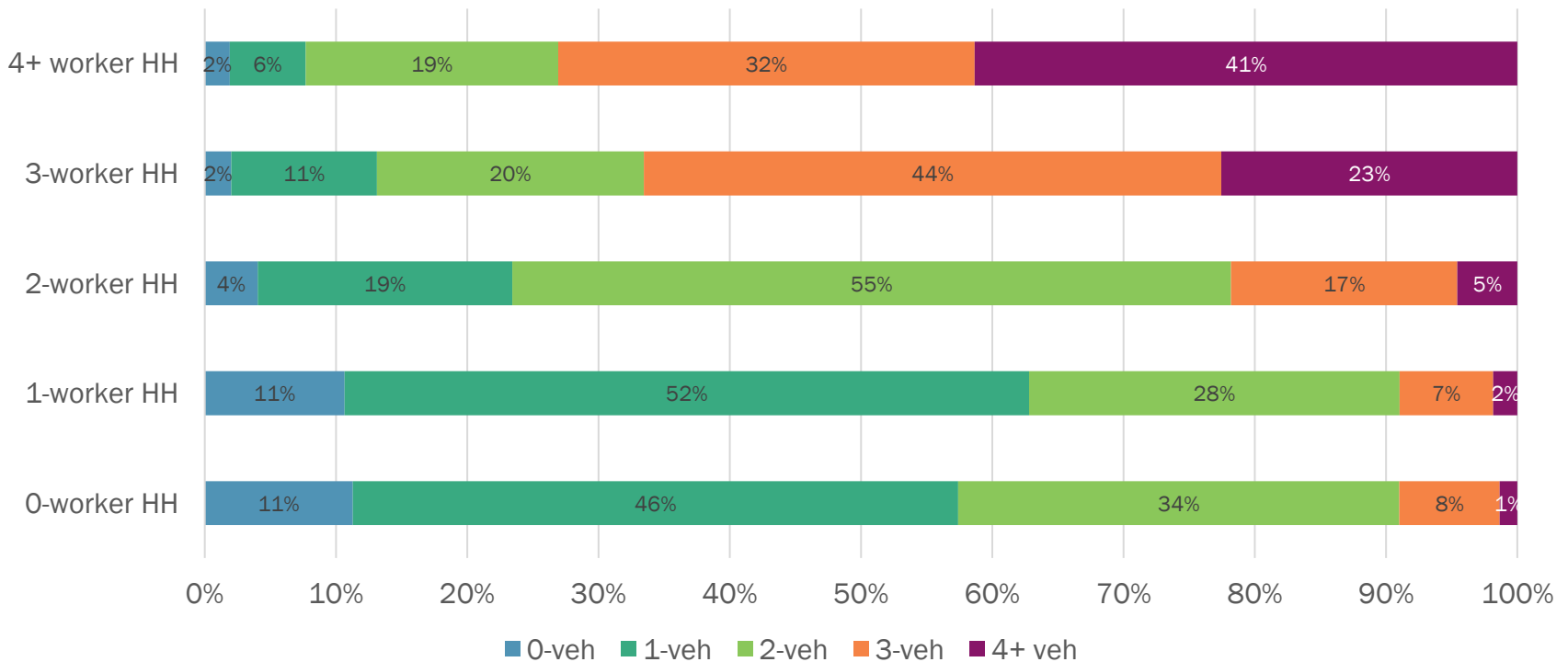


Explanatory Variables

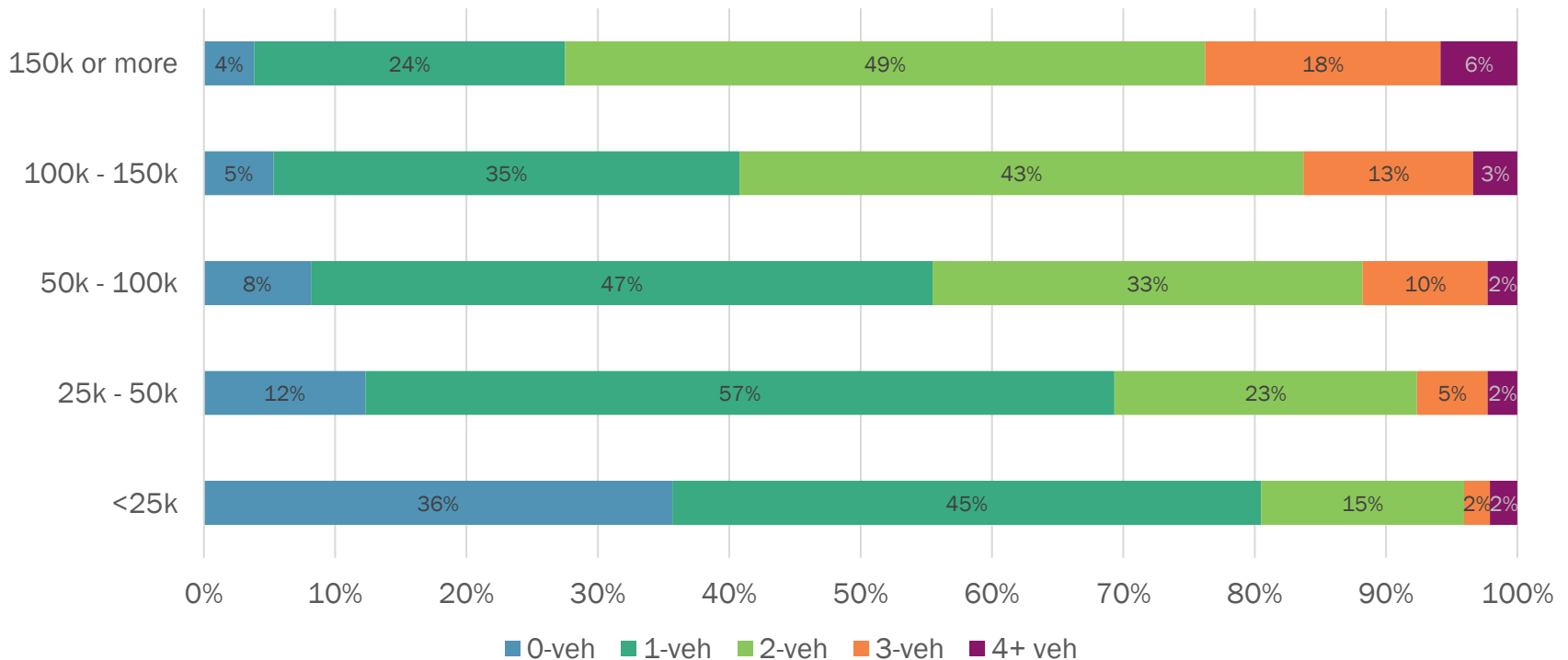
- Household composition variables
 - Number of driving-age household members in household
 - Ratio of number of workers to driving-age household members
 - Presence of young children (age 0-15)
 - Ratio of driving-age children (age 16-17) to driving-age household members
- Household income group: Low, Medium-Low, Medium, Medium-High, High
- Zonal density index, which is a function of TAZ household density and TAZ employment density, simultaneously measures zonal density and jobs/housing balance in a zone
- Zonal accessibility indices from residential zones to potential destinations
- Auto travel time savings per worker to work (compared to walk or transit)



Observed Share of Households by Auto Availability, Stratified by Number of Workers



Observed Share of Households by Auto Availability, Stratified by Household Income Group



Auto Ownership Model Estimation Results

Variable	Coefficient and T-Stat by Choice (Base: 0)							
	1 car		2 cars		3 cars		4+ cars	
	Coef.	T-stat	Coef.	T-stat	Coef.	T-stat	Coef.	T-stat
Household Constant	1.17	9.37	-0.71	-4.26	-2.88	-11.24	-4.57	-13.53
2 drivers in HH	0.33	4.19	3.34	35.04	3.82	27.56	3.94	16.24
3 drivers in HH	0.00	NA	3.49	29.39	5.79	36.14	6.37	24.45
4+ drivers in HH	0.00	NA	3.88	19.01	6.57	27.82	8.37	26.68
Workers/drivers in HH	0.56	3.01	0.90	4.52	1.32	6.21	1.87	7.65
People aged 16-17/drivers in HH	0.00	NA	-2.42	-6.63	-4.18	-8.90	-5.50	-8.72
People aged 18-24/drivers in HH	-0.45	-2.54	-1.34	-5.69	-1.63	-5.60	-1.63	-5.60
People aged 25-34/drivers in HH	-0.52	-6.15	-0.99	-9.70	-1.44	-10.43	-1.99	-8.51
Presence of children aged 0-15	0.24	3.45	0.38	6.52	0.00	NA	0.00	NA
Household income <25k	ref	ref	ref	ref	ref	ref	ref	ref
Household income 25k-50k	1.41	12.13	1.51	9.36	1.88	7.63	1.88	7.63
Household income 50k-100k	1.98	18.52	2.62	18.03	3.08	13.79	3.08	13.79
Household income 100-150k	2.26	18.48	3.39	21.58	3.92	16.93	3.92	16.93
Household income >150k	2.31	17.51	3.65	22.27	4.39	18.61	4.39	18.61
Density index	-0.01	-1.67	-0.01	-1.67	-0.01	-1.67	-0.01	-1.67
Auto time savings / workers (over walk or transit)	0.84	5.71	0.96	5.82	0.96	5.82	0.96	5.82
Retail accessibility by non-motorized	0.00	NA	-0.09	-6.06	-0.15	-7.62	-0.23	-8.12
Retail access. (2/3 pk, 1/3 op.) by transit, if 0 workers	-0.28	-10.82	-0.49	-14.93	-0.68	-14.02	-0.68	-14.02
Retail access. (2/3 pk, 1/3 op.) by transit, if 1+ workers	-0.50	-19.62	-0.82	-28.69	-1.00	-30.41	-1.09	-24.11



Estimation Result Example

	Coefficient and T-Stat by Choice (Base: 0)							
Variable	1 car		2 cars		3 cars		4+ cars	
	Coef.	T-stat	Coef.	T-stat	Coef.	T-stat	Coef.	T-stat
Workers/drivers in HH	0.56	3.01	0.90	4.52	1.32	6.21	1.87	7.65



Main Findings

- The auto ownership model was estimated in multiple rounds by changing utilities specification and coefficients.
- Not all tested variables are included in the results.
- The auto ownership model estimation results are reasonable, for example:
 - All else being equal, the higher the number of drivers in a household, the more likely the household owns more cars.
 - Households tend to own one or two cars when they have young children.
 - The low-medium-income households are more likely to own at least one car than zero car. Other income groups share the same pattern, but to a higher degree with higher income.
 - The probability of households owning two or more cars is lower if they have better accessibility to retail and service establishments by non-motorized modes.



Main Findings (Continued)

- The auto ownership model estimation task not only provided a great hands-on experience on model estimation, but also helped staff learn more about the Gen3 Model.
- The model settings, coefficients, and utilities specification from the final estimation round will be used directly in the Gen3 Model.

source\configs\activitysim\configs



auto_ownership.csv



auto_ownership.yaml



auto_ownership_coefficients.csv

- Larch is a very helpful estimation tool for ABM models using ActivitySim, but users may need some guidance.

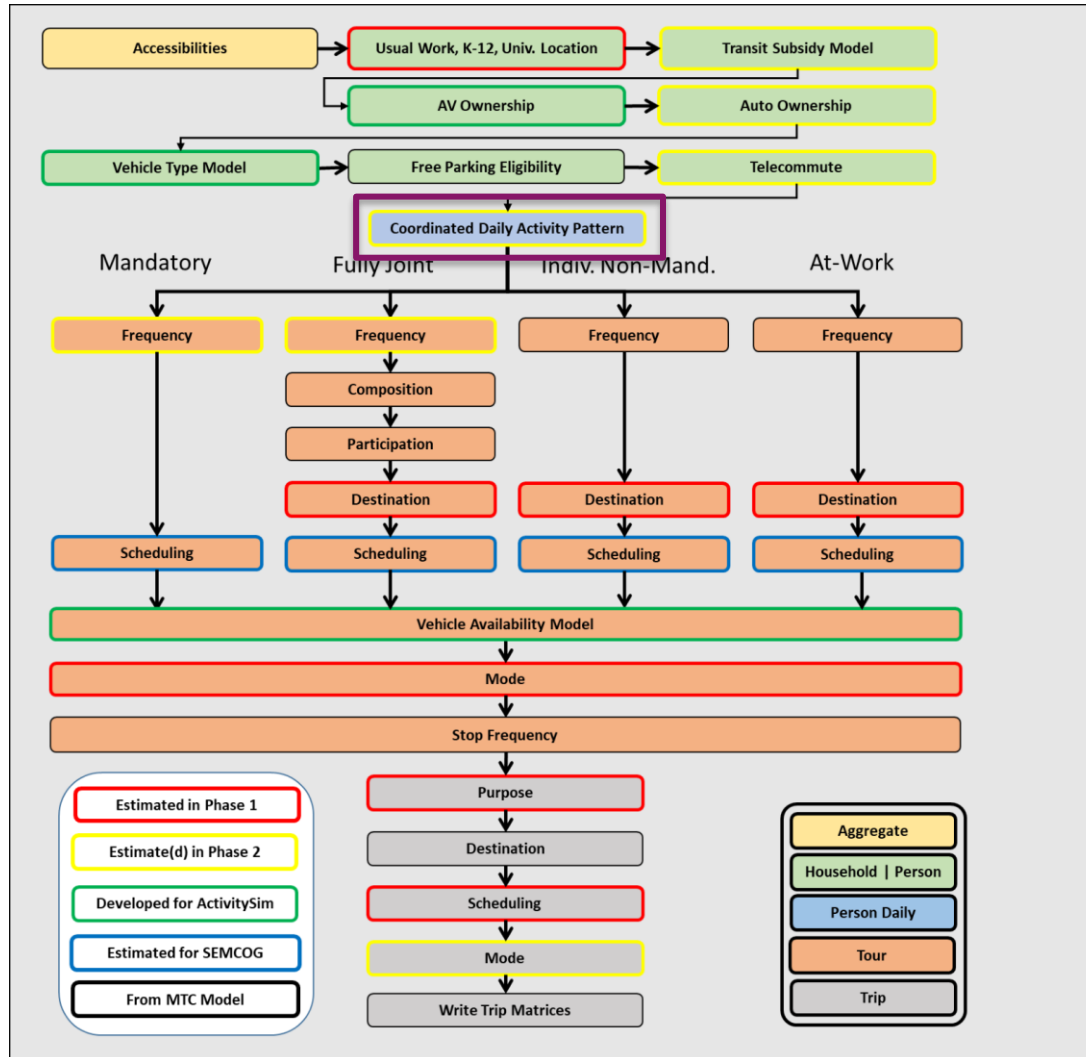


Coordinated Daily Activity Pattern (CDAP) Model

Nazneen Ferdous



Gen3 Model Structure



Overview

- The Coordinated Daily Activity Pattern (CDAP) model predicts the daily activity pattern (DAP) for each person in the household simultaneously. A person can have only one pattern per day.
 - Mandatory Pattern (M): Work, university, or school tour.
 - Non-Mandatory Pattern (N): Shopping, eating out, discretionary, social tours, etc.
 - At-Home Pattern (H): The person stays home all day.

Person Type	Mandatory Pattern (M)	Non-Mandatory Pattern (N)	At-Home Pattern (H)
1=Full time worker	√	√	√
2=Part time worker	√	√	√
3=University student	√	√	√
4=Non-worker under 65	×	√	√
5=Retiree 65+	×	√	√
6=Driving age school child 16-17	√	√	√
7=Pre-driving age school child 6-15	√	√	√
8=Preschool child under 6	√	√	√



Model Structure

- The CDAP model has a multinomial logit (MNL) structure.
 - Alternatives for one-person household: M, N, H.
 - Alternatives for two-person household: MM, MN, MH, NM, NN, NH, HM, HN, HH.
- The utility formation includes individual components as well as pair-wise components
 - An example utility function for a two-person household where both members choose an at-home pattern:

$$U_{HH} = U_H^1 + U_H^2 + U_H^{12}$$

where,

U_{HH} is the utility for both person 1 and person 2 to choose at-home pattern

U_H^1 and U_H^2 are the utilities for person 1 and person 2 to choose at-home patterns, respectively

U_H^{12} is the utility/interaction term for both person 1 and person 2 to choose at-home patterns



Observed Frequency of DAP by Person Type

Person Type (ptype)	% ptype	Absolute Frequency				Relative Frequency		
		M	N	H	Total	M	N	H
Full-time worker (FW)	46.4%	12036	1848	2839	16723	72.0%	11.1%	17.0%
Part-time worker (PW)	8.4%	773	1250	1005	3028	25.5%	41.3%	33.2%
University student (US)	4.2%	733	319	473	1525	48.1%	20.9%	31.0%
Non-worker under 65 (NW)	9.4%	-	2052	1350	3402	-	60.3%	39.7%
Retired/Retiree 65+ (RT)	15.0%	-	3374	2021	5395	-	62.5%	37.5%
Driving age student (SD)	2.0%	531	57	145	733	72.4%	7.8%	19.8%
Pre-driving age student 6-15 (SP)	8.6%	2210	358	518	3086	71.6%	11.6%	16.8%
Preschool child under 6 (PS)	5.9%	533	1024	564	2121	25.1%	48.3%	26.6%
Total	100.0%	16816	10282	8915	36013	46.7%	28.6%	24.8%



CDAP Model Estimation Results

Variable	Coefficient (T-Stat)		
	Mandatory (M)	Non-Mandatory (N)	Stay at-Home (H)
Constant			
FW (Full-time worker)	1.896 (35.93)	-0.238 (-4.24)	
PW (Part-time worker)	0.341 (2.99)	0.230 (2.27)	
US (University student)	1.154 (13.72)	0.237 (2.47)	
NW (Non-worker under 65)		-1.666 (-3.80)	
RT (Retired/Retiree 65+)		-0.868 (-2.56)	
SD (Driving age student)	1.277 (8.73)	-0.840 (-4.15)	
SP (Pre-driving age student)	1.441 (12.56)	-0.150 (-1.20)	
PS (Pre-school child under 6)	-0.438 (-2.98)	0.484 (5.05)	
Age			
PS – Age 0 to 1	-1.123 (-6.24)		
PS – Age 4 to 5	1.356 (10.65)		
SP – Age 6 to 9			-0.165 (-1.31)
SP – Age 13 to 15	0.161 (1.20)	-0.321 (-1.78)	
PW – Age 40 or older		0.226 (2.31)	
RT – Age over 80			0.459 (5.90)
Gender			
PW – Female	0.415 (3.55)		
RT – Female			0.287 (4.54)
Auto Sufficiency			
NW – More cars than workers		0.511 (5.93)	
RT – More cars than workers		0.749 (7.37)	
FW – Fewer cars than workers			0.160 (2.74)
RT – Fewer cars than workers			0.416 (1.37)
SP – Fewer cars than workers			0.273 (2.01)
Household Income			
FW – Income < \$20,000			0.318 (1.87)
FW – Income > \$100,000			-0.124 (-2.60)
PW – Income < \$20,000			0.275 (1.33)
NW – Income > \$100,000			-0.159 (-2.18)
RT – Income \$50–100,000		0.096 (1.67)	
PS – Income < \$20,000			0.559 (2.36)
PS – Income > \$100,000			-0.462 (-4.83)



CDAP Model Estimation Results

Variable	Coefficient (T-Stat)		
	Mandatory (M)	Non-Mandatory (N)	Stay at-Home (H)
Accessibility			
FW – Peak accessibility to employment by transit	0.032 (5.21)		
NW – Off-peak accessibility to retail by auto		0.187 (4.45)	
RT – Off-peak accessibility to retail by auto		0.110 (3.50)	
Work and School Locations			
FW – Usual work location is home		0.228 (2.99)	
PW – Usual work location is home		0.209 (2.16)	
FW – No usual work location			3.192 (33.28)
PW – No usual work location			1.127 (6.08)
US – No usual school location			1.839 (13.34)
SD – No usual school location			2.517 (6.87)
SP – No usual school location			1.364 (9.17)
Telecommute Frequency			
Telecommutes 1 day per week		0.336 (4.37)	0.306 (4.26)
Telecommutes 2–3 days per week		0.886 (12.89)	0.733 (10.93)
Telecommutes 4 days per week		0.886 (12.89)	0.733 (10.93)
2-Way Interactions			
Full-time worker x Full-time worker	0.200 (5.27)	0.229 (1.81)	0.671 (9.45)
Full-time worker x Part-time worker		0.276 (2.16)	0.736 (8.28)
Full-time worker x University student		0.507 (3.11)	0.487 (4.81)
Full-time worker x Non-worker under 65		0.196 (2.32)	0.484 (6.38)
Full-time worker x Retired		0.219 (1.89)	0.433 (4.46)
Full-time worker x Driving age student	0.445 (5.04)	0.591 (2.06)	0.476 (3.09)
Full-time worker x Pre-driving age student	0.182 (4.93)	0.421 (4.18)	0.413 (5.95)
Full-time worker x Pre-school child under 6	0.187 (2.76)	0.265 (3.86)	0.326 (4.53)
Part-time worker x Part-time worker		0.308 (1.75)	1.036 (6.41)
Part-time worker x University student			0.745 (4.45)
Part-time worker x Non-worker under 65			0.780 (5.90)
Part-time worker x Retired		0.340 (2.86)	0.994 (8.36)
Part-time worker x Driving age student		0.946 (3.26)	
Part-time worker x Pre-driving age student		0.449 (3.87)	0.505 (4.95)
Part-time worker x Pre-school child under 6		0.902 (9.42)	1.071 (9.64)

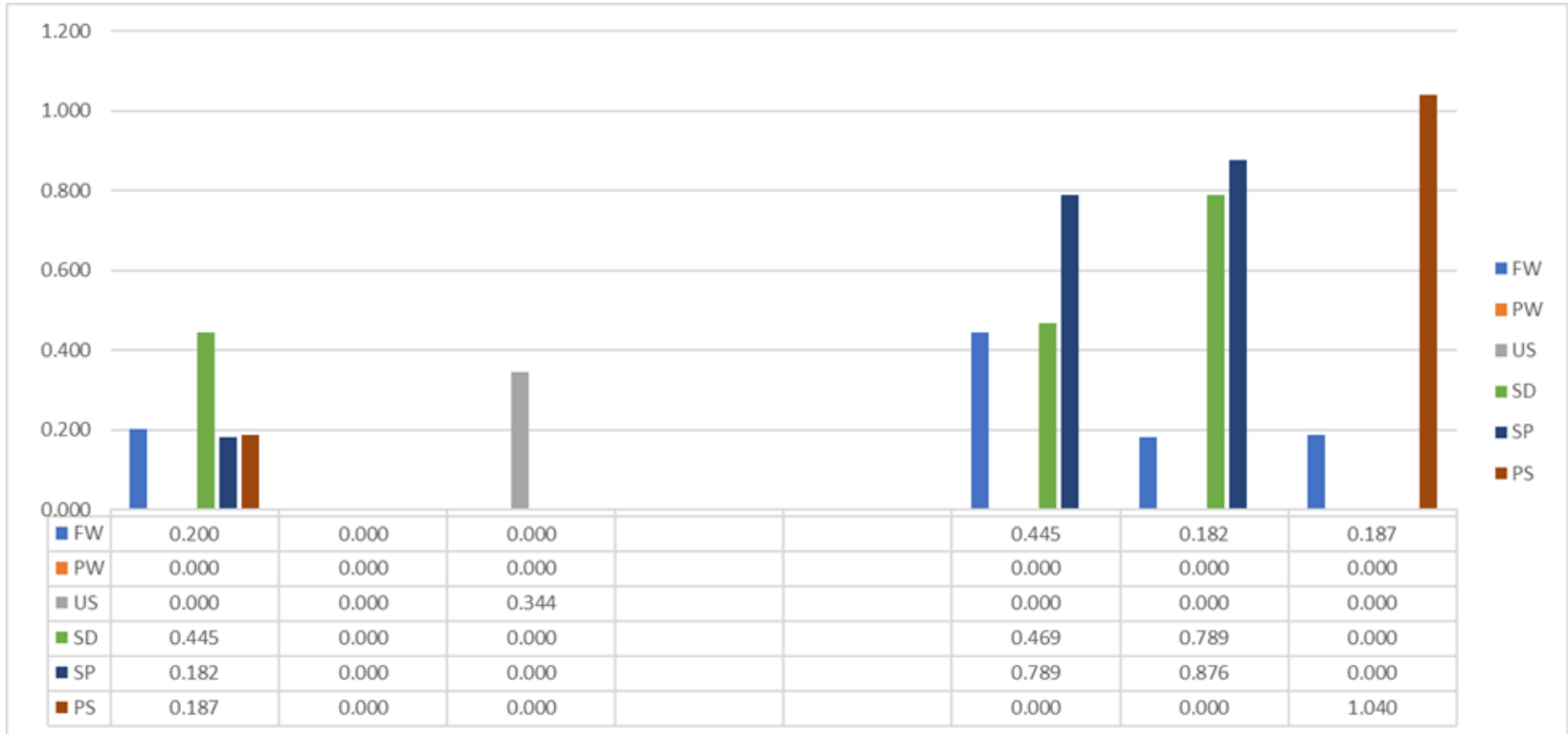


CDAP Model Estimation Results

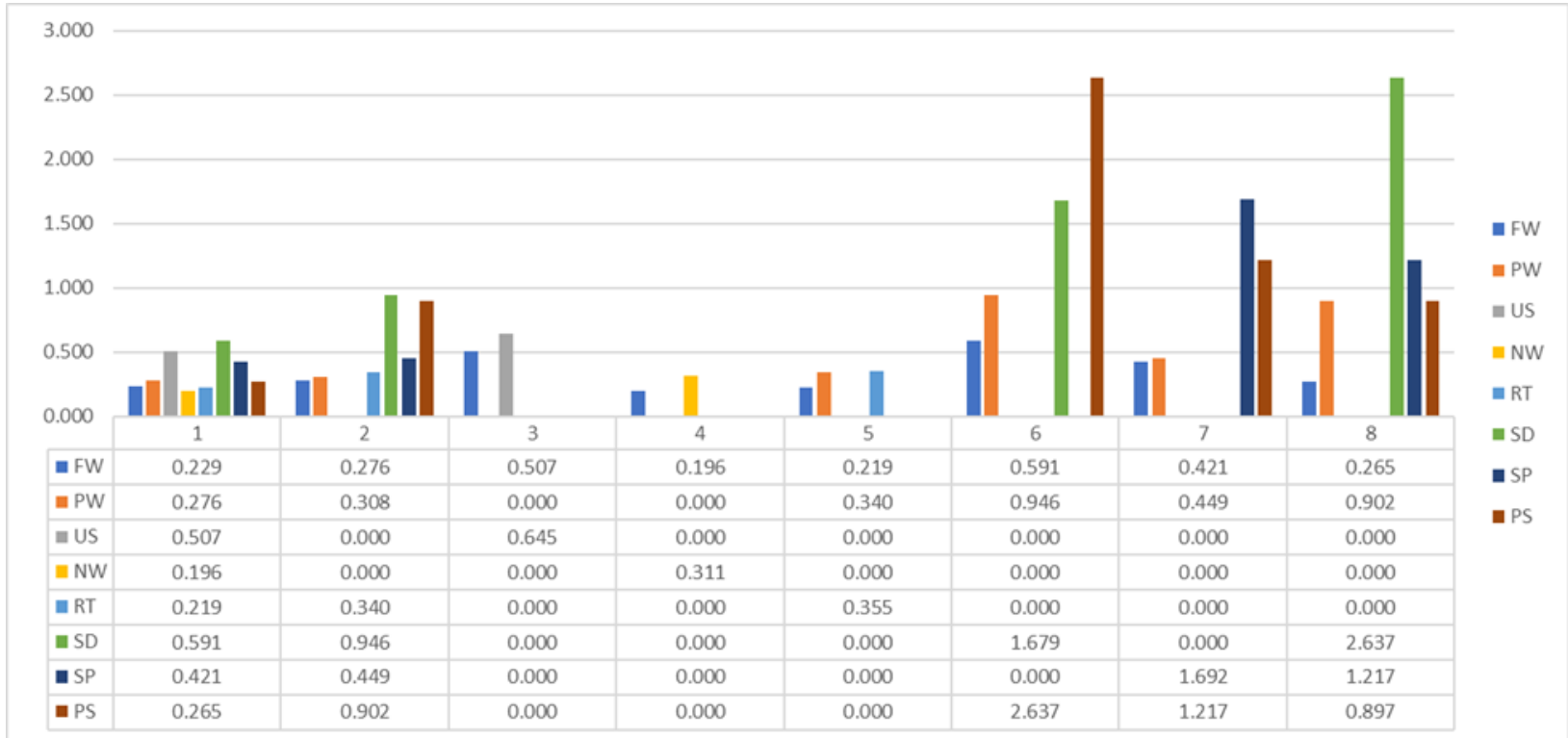
Variable	Coefficient (T-Stat)		
	Mandatory (M)	Non-Mandatory (N)	Stay at-Home (H)
University student x University student	0.344 (2.33)	0.645 (2.73)	0.973 (5.87)
University student x Non-worker under 65			0.510 (2.73)
Non-worker under 65 x Non-worker under 65		0.311 (2.46)	0.890 (7.03)
Non-worker under 65 x Retired			0.878 (8.68)
Retired x Retired		0.355 (4.65)	0.940 (10.77)
Driving age student x Driving age student	0.469 (1.43)	1.679 (2.19)	1.699 (3.91)
Driving age student x Pre-driving age student	0.789 (1.19)		
Driving age student x Pre-school child under 6		2.637 (2.09)	2.050 (1.52)
Pre-driving age student x Pre-driving age student	0.876 (11.15)	1.692 (14.29)	1.564 (14.89)
Pre-driving age student x Pre-school child under 6		1.217 (5.46)	0.646 (2.64)
Pre-school child under 6 x Pre-school child under 6	1.040 (6.80)	0.897 (8.40)	1.531 (12.94)
3-Way Interactions			
FW x FW x PW	-0.431 (-2.28)		
FW x FW x NW	-0.431 (-2.28)		
Estimation Statistics			
Final log-likelihood	-24886.29		
Rho ² (0)	0.29		



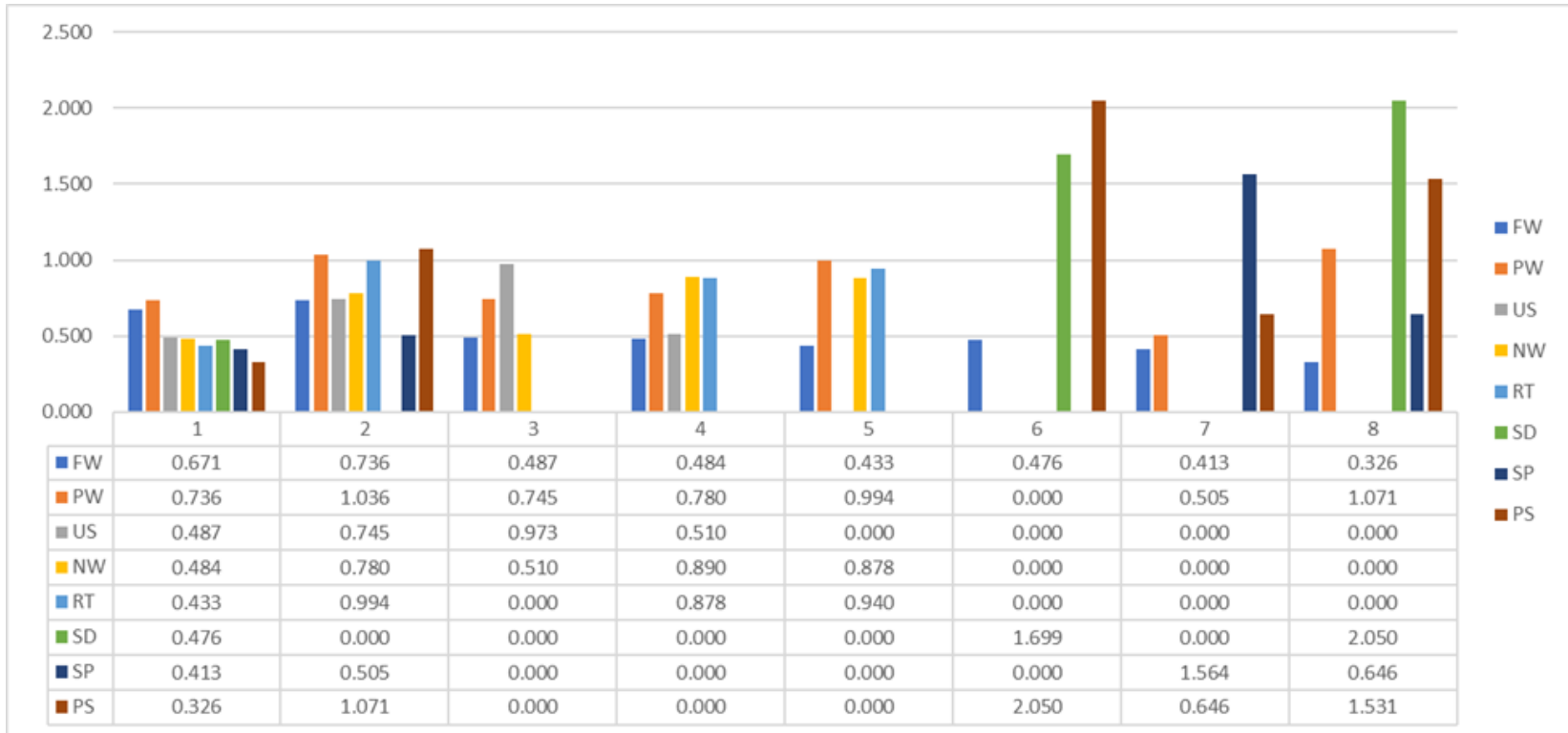
2-Person Interactions – Mandatory Patterns



2-Person Interactions – Non-Mandatory Patterns



2-Person Interactions – Stay At-Home Patterns



Mandatory Tour Frequency (MTF) Model

Feng Xie



Observed Frequencies

- MTF predicts the number and purpose of mandatory tours for simulated persons who choose the mandatory pattern in the CDAP Model.
- Observed frequencies of mandatory tours by person type and alternative from the RTS/MTS data (shown below) informed model specification.
- Constraints associated with specific person types in terms of mandatory tour making evolved over time.

Person Type	Number of Mandatory Tours				
	1 Work Tour	2+ Work Tours	1 School Tour	2+ School Tours	Work & School Tours
1=Full time worker	11,658	378	0	0	0
	96.86%	3.14%	0.00%	0.00%	0.00%
2=Part time worker	757	16	0	0	0
	97.93%	2.07%	0.00%	0.00%	0.00%
3=University student	90	1	543	15	84
	12.28%	0.14%	74.08%	2.05%	11.46%
4=Non-worker under 65	0	0	0	0	0
	N.A.	N.A.	N.A.	N.A.	N.A.
5=Retiree 65+	0	0	0	0	0
	N.A.	N.A.	N.A.	N.A.	N.A.
6=Driving age school child 16-17	13	0	472	26	20
	2.45%	0.00%	88.89%	4.90%	3.77%
7=Pre-driving age school child 6-15	0	0	2,118	92	0
	0.00%	0.00%	95.84%	4.16%	0.00%
8=Preschool child under 6	0	0	526	7	0
	0.00%	0.00%	98.69%	1.31%	0.00%



Main Results and Findings

- Screenshot of the final estimation results table (partial)

Variable	Relevant person types	Coefficient & T-Stat by Choice Alternative					Notes
		1 Work	2+ Work	1 School	2+ School	Work & School	
Constant	1=Full-time worker	0 (reference)	-2.23 (-14.12)				All alternative specific constants have the expected negative signs; The constants could be further adjusted in the model calibration step
	2=Part time worker	0 (reference)	-3.05 (-10.22)				
	3=University student	-0.0351 (-0.17)	-4.27 (-4.18)	0 (reference)	-3.64 (-7.19)	-0.309 (-1.19)	
	4=Non-worker U65						
	5=Retiree 65+						
	6=School child 16-17			0 (reference)	-2.73 (-7.75)	-0.58 (-1.98)	
	7=School child 6-15			0 (reference)	-3.18 (-10.66)		
Person is female (dummy)	1=Full-time worker	0 (reference)	-0.124 (-1.18)	0 (NA)		-0.477 (-2.06)	Similar to the ARC model, females have a significantly lower tendency to make "2+ Work" tours and "Work & School" tours, as well as a higher tendency to make "2+ School" tours. The coefficient for "1 School" is turned off because 1) workers or retirees do not make any school tours in the MWCOC's regional travel survey (RTS), and 2) "1 School" should be the reference alternative for University Students.
	2=Part time worker	0 (reference)	-0.124 (-1.18)	0 (NA)		-0.477 (-2.06)	
	3=University student	-0.477 (-2.24)	-0.124 (-1.18)	0 (NA)	0.0407 (0.23)	-0.477 (-2.06)	
	4=Non-worker U65	0 (reference)	-0.124 (-1.18)	0 (NA)			
	5=Retiree 65+	0 (reference)	-0.124 (-1.18)	0 (NA)			
	6=School child 16-17	-0.477 (-2.24)		0 (reference)	0.0407 (0.23)	-0.477 (-2.06)	
	7=School child 6-15						
Person is 25 year old or younger (dummy)	1=Full-time worker	0 (reference)	-0.0546 (-0.26)	0 (NA)		0 (NA)	The threshold age is changed from 35 to 25, which slightly improves the estimation results. Coefficients for "1 School" and "Work & School" are turned off as they were unreasonably signed.
	2=Part time worker	0 (reference)	-0.0546 (-0.26)	0 (NA)		0 (NA)	
	3=University student	0 (NA)	-0.0546 (-0.26)	0 (reference)	0.838 (1.58)	0 (NA)	
	4=Non-worker U65	0 (reference)	-0.0546 (-0.26)	0 (NA)			
Workplace within walking distance of 3 miles (dummy)	1=Full-time worker		0.306 (2.1)				The coefficient is statistically significant and reasonably signed.
	2=Part time worker		0.306 (2.1)				
	3=University student		0.306 (2.1)				
	4=Non-worker U65		0.306 (2.1)				
	5=Retiree 65+		0.306 (2.1)				
School within walking distance (dummy)	3=University student				0.464 (1.84)		The coefficient is statistically significant and reasonably signed.
	6=School child 16-17				0.464 (1.84)		
	7=School child 6-15				0.464 (1.84)		



Main Results and Findings

- Overall, the estimation results look reasonable.
 - All the alternative-specific constants (“asc”) have the expected negative signs, indicating a higher tendency of choosing the most typical reference alternatives (those with zero coefficients) compared to other alternatives.
 - The estimated coefficients for most of the explanatory variables, such as gender, proximity to workplace by walking, proximity to workplace or school by driving, and car availability in household, are reasonable and consistent with the original estimation results, which were developed by the Atlanta Regional Commission (ARC) for its Core Activity/Tour-Based Models during 2004-06.
- A couple of variables with unreasonable estimated coefficients, such as workplace or school within walking distance and residence in urban areas, are dropped altogether.
- For variables with mixed estimation results varying by person type and alternative, the subset of variables with unreasonable results are dropped by setting their corresponding coefficients to zero.



Main Results and Findings

- Rho-squared values shown at the bottom of the table below describe the overall goodness of fit of the model.
 - Rho-squared statistic with respect to Zero (0.7517) is close to its counterpart in the original ARC model (0.8072), indicating that the MTF model estimated for MWCOG and that for ARC have a comparable explanatory power relative to the reference zero coefficients model (or the equally likely model).
 - Rho-squared statistic with respect to Constants (0.0587) is much smaller than its counterpart in the ARC model (0.6561), suggesting that alternative specific constants by person type explain most of the variation in the RTS/MTS data.

Statistics	Value
Initial likelihood (zero coefficients)	-12100.2875
Likelihood with constants only	-3192.5920
Final likelihood	-3005.0276
Rho-squared w.r.t. Zero	0.7517
Rho-squared w.r.t. Constants	0.0587



Lessons Learned

- The estimation was executed in an iterative process: COG staff examined preliminary estimation results, adjusted model specifications/coefficients accordingly, and reran the estimation. The final estimation results were obtained after seven such iterations.
- The estimation initially failed as the Jupyter Notebook reported a likelihood of negative infinity. The Larch documentation did not provide any instruction on how to troubleshoot this issue. With guidance from RSG, COG staff pinpointed the variable that caused the issue by starting with the constants-only model and adding variables back incrementally.
- In order to improve the estimation results, COG staff tested different threshold values for two variables, Person is 25 years old or younger (dummy) and Household income of 100K or higher (dummy). Threshold values that generated the best results were adopted.



Concluding Remarks

- RSG reviewed and approved the estimation work conducted by COG staff and incorporated the final estimated files into the Gen3, Phase 2, Model implementation.
- The Phase 2 model estimation work that was conducted by both RSG and COG staff is documented in a technical report, which will become available on the COG [website](#) once it is finalized.
- The open-source Larch package provides a powerful logit model estimation tool that enables COG staff to conduct model estimation for the Gen3 Model without proprietary software such as ALOGIT.
- Users of the Larch package could benefit from improved documentation: Many of the estimation techniques/workarounds/caveats were learned from the consultant and our trial-and-error experiences, not from the documentation.
- Technical issues were found in the current Larch package. For instances,
 - It failed to estimate the Non-Mandatory Tour Frequency (NMTF) Model.
 - It failed to estimate the person type interaction terms for the CDAP Model.



Acknowledgements

- We would like to express our sincere thanks to Joel Freedman, Andrew Rohne, and David Hensle (RSG) for their guidance and assistance on this estimation task.



Feng Xie

Manager of Model Development, COG/TPB

(202) 962-3259

fxie@mwkog.org

Ray Ngo

Senior Transportation Engineer, COG/TPB

(202) 962-3231

rngo@mwkog.org

Nazneen Ferdous

Transportation Engineer, COG/TPB

(202) 962-3325

nferdous@mwkog.org

mwkog.org/TPB

Metropolitan Washington Council of Governments

777 North Capitol Street NE, Suite 300

Washington, DC 20002



National Capital Region
Transportation Planning Board