RSG
the science of insight

Gen3 Model Development Project

Travel Forecasting Subcommittee Meeting

September 24, 2021

IN PARTNERSHIP WITH

BASELINE MOBILITY

Metropolitan Washington Council of Governments

Discussion Topics

- Model estimation in ActivitySim
- Data for estimation
- Model estimation results
  - Tour mode choice
  - Tour destination choice

Model Estimation in ActivitySim

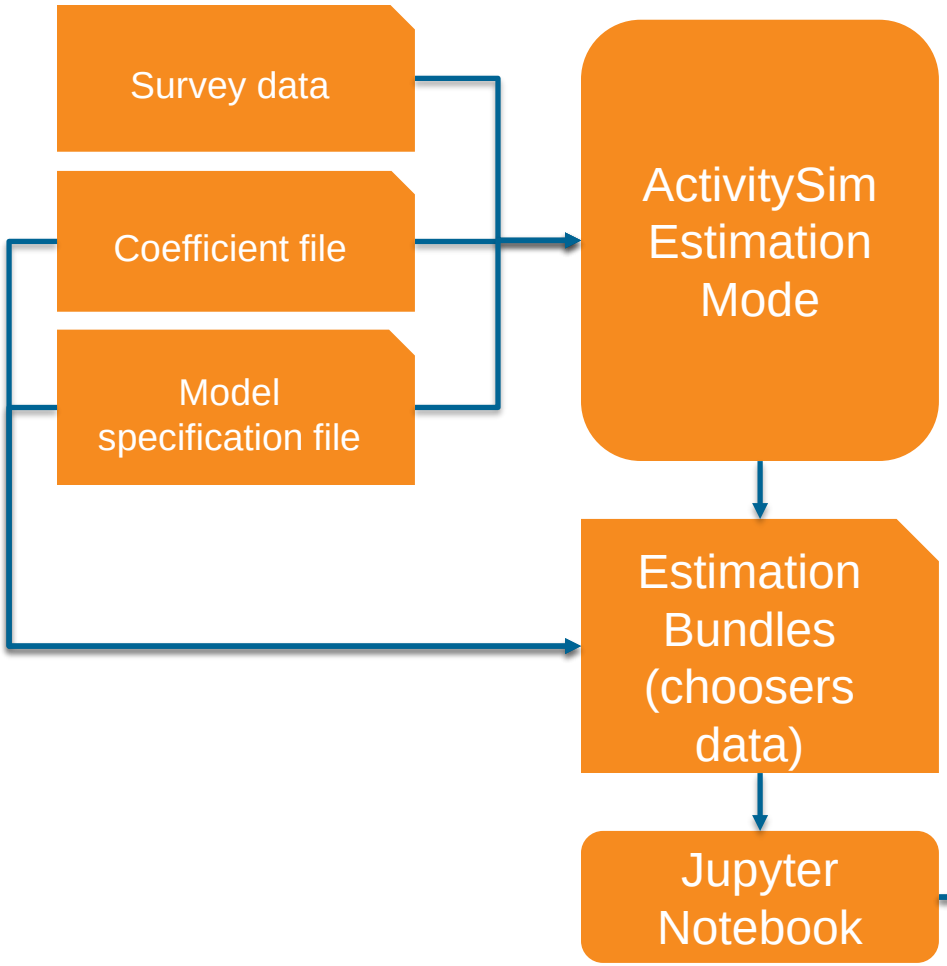
ActivitySim “Estimation Mode”

- ActivitySim is a disaggregate activity-based travel model
  - A synthetic population is run through each model component
- The software builds a choice model that is specific to each household and person, taking into account
  - Attributes of the synthetic population
  - Choice outcomes of previous models in model system
  - Logsums from downstream model components
- It is now possible to run a ***survey population*** through the software
  - With ***same attributes*** as synthetic population
  - Observed choices ***override*** simulated choices
  - Logsums created exactly as per model specification

ActivitySim “Estimation Mode”

- The output of ActivitySim estimation mode is a set of “estimation data bundles” for each model
- A table of data where:
  - Rows are decision makers (households, persons, tours, trips, etc.)
  - Columns are data for each alternative to be used in utility equations
- This data, along with the ActivitySim input coefficient file(s) and model specification file is read by a Jupyter Notebook that re-estimates the model specification in Larch
  - Larch is a logit model estimation package in Python that is built on top of the Python Scipy package

Estimation process



www.python.org

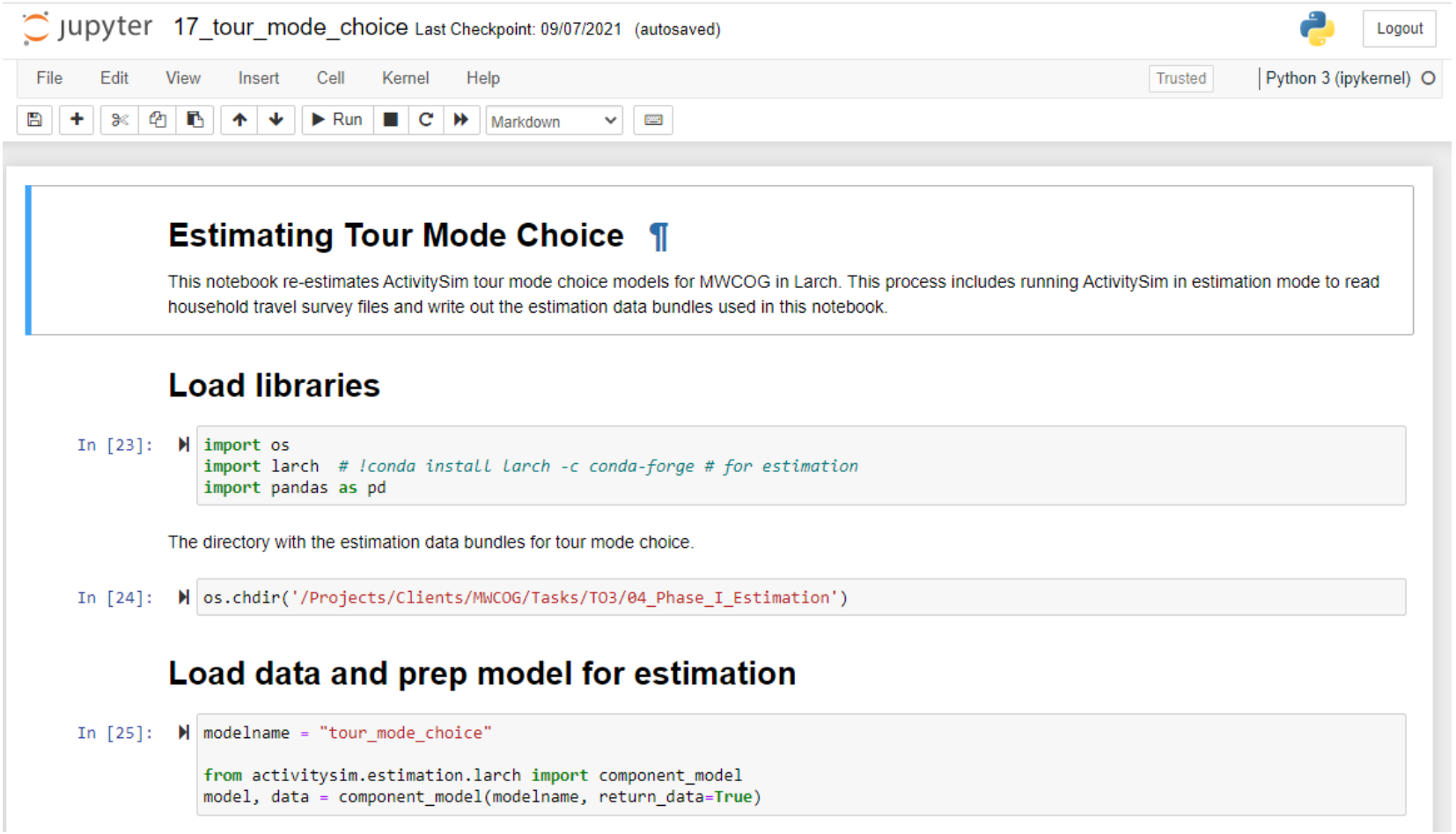
https://larch.newman.me/

Scipy.org

Data preparation

1. Run SPA tool for 2017-2018 COG/TPB Regional Travel Survey (RTS)/2018-2019 Maryland Statewide Household Travel Survey (MTS)
   - Groups trips into tours, determines tour and trip modes and purposes
2. Run Jupyter notebook that re-formats SPA output into ActivitySim input
   - Cleans data, imputes missing variables, fixes data inconsistencies.
     - HH size must match # of persons in the household
     - Missing school and work TAZs are imputed
     - FT workers cannot make school tours (change to PT)
     - Start/end time constraints
     - Valid tour destination
     - Tours that fall outside allowed frequencies are removed

Jupyter Notebook (1): Tour Mode Choice



Jupyter Notebook (2): Tour Mode Choice

In [*]:
```
result_dir='/Projects/Clients/MWCOG/Tasks/TO3/04_Phase_I_Estimation/estimation/tour_mode_choice'
model.maximize_loglike(method="SLSQP", options={"maxiter": 1000})
```

Iteration 009

Currently using SLSQP, Best LL = -inf

	value	initvalue	nullvalue	minimum	maximum	holdfast	note
-999	-999.000000	-999.0	-999.0	-999.0	-999.0	1	
KNR_BM_ASC_nonmandatory	-999.000000	-999.0	0.0	NaN	NaN	1	
KNR_CR_ASC_nonmandatory	-999.000000	-999.0	0.0	NaN	NaN	1	
KNR_MR_ASC_nonmandatory	-999.000000	-999.0	0.0	NaN	NaN	1	
PNR_BM_ASC_nonmandatory	-999.000000	-999.0	0.0	NaN	NaN	1	
...	...	...	...	...	...	...	...
walk_MR_ASC_atwork	-54.048308	0.0	0.0	NaN	NaN	0	
walk_transit_ASC_auto_deficient_atwork	-56.082305	0.0	0.0	NaN	NaN	0	
walk_transit_ASC_auto_sufficient_atwork	-159.896064	0.0	0.0	NaN	NaN	0	
walk_transit_ASC_no_auto_atwork	-42.127637	0.0	0.0	NaN	NaN	0	
walk_transit_CBD_ASC_atwork	-160.708671	0.0	0.0	NaN	NaN	0	

444 rows × 7 columns

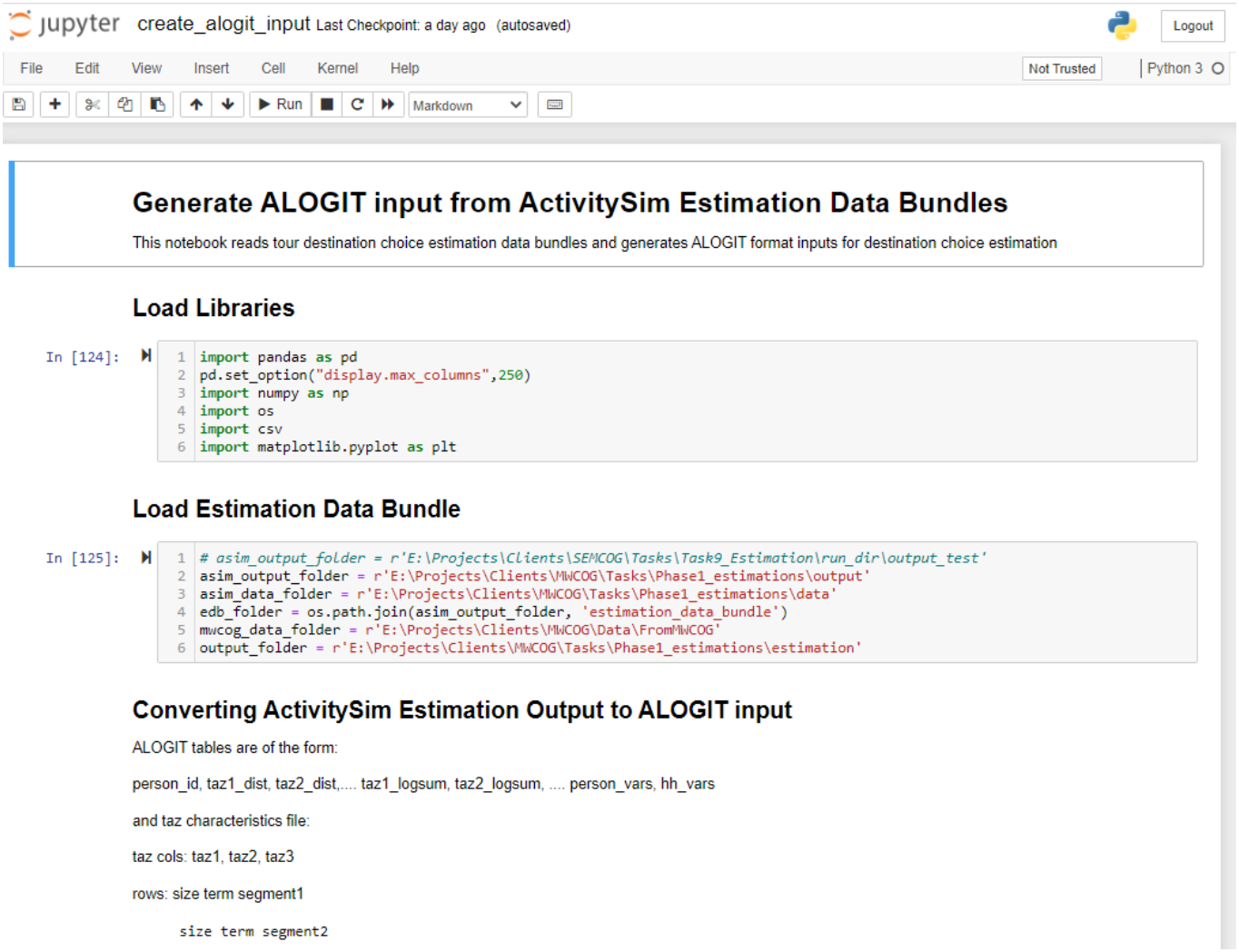
Estimated coefficients

In [*]:
```
model.calculate_parameter_covariance()
result_dir='/Projects/Clients/MWCOG/Tasks/TO3/04_Phase_I_Estimation/estimation/tour_mode_choice'
model.to_xlsx(
        result_dir+"model_estimation_all_005_nested.xlsx",
        data_statistics=True,
    )
```

XLSX Estimation Report

1	<< Back to Table of Contents						
2	**Table 1: Parameters**						
3		**Value**	**Std Err**	**t Stat**	**Signif**	**Like Ratio**	**Null Value**
4	-999	-999.	0.00	NA		NA	-999
5	KNR_BM_ASC_nonmandatory	-5.95	1.86e-09	-BIG	***	NA	0
6	KNR_CR_ASC_nonmandatory	-3.69	4.58e-09	-BIG	***	NA	0
7	KNR_MR_ASC_nonmandatory	12.3	0.938	13.10	***	NA	0
8	PNR_BM_ASC_nonmandatory	6.75	3.67e-05	BIG	***	NA	0
9	PNR_CR_ASC_nonmandatory	-3.73	1.08e-10	-BIG	***	NA	0
10	PNR_MR_ASC_nonmandatory	18.6	1.38	13.55	***	NA	0
11	bike_ASC_auto_deficient_discretionary	-1.48	0.300	-4.94	***	NA	0
12	bike_ASC_auto_sufficient_discretionary	-3.66	0.205	-17.84	***	NA	0
13	bike_ASC_no_auto_discretionary	1.57	0.465	3.38	***	NA	0
14	coef_age010_wlktrn_nonmandatory	-1.49	0.347	-4.30	***	NA	0
15	coef_age1619_da_nonmandatory	-1.23	0.181	-6.80	***	NA	0
16	coef_age16p_sr_nonmandatory	0.790	0.0918	8.61	***	NA	0
17	coef_bike_nonmandatory	-0.0522	0.00612	-8.53	***	NA	0
18	coef_cost_nonmandatory	-0.00106	8.71e-05	-12.11	***	NA	0
19	coef_density_bike_nonmandatory	0.0410	0.0123	3.32	***	NA	0
20	coef_density_knr_nonmandatory	0.00	0.00	NA		NA	0
21	coef_density_pnr_nonmandatory	0.00	0.00	NA		NA	0
22	coef_density_walk_nonmandatory	0.0579	0.00555	10.42	***	NA	0
23	coef_density_wlktrn_nonmandatory	-0.0157	0.00844	-1.86		NA	0
24	coef_distpen_drvtrn_nonmandatory	0.00	0.00	NA		NA	0
25	coef_drvacc_knr_nonmandatory	0.0681	0.134	0.51		NA	0
26	coef_drvacc_pnr_nonmandatory	-0.336	0.128	-2.62	**	NA	0
27	coef_drvratio_knr_nonmandatory	0.00	0.00	NA		NA	0
28	coef_drvratio_pnr_nonmandatory	0.00	0.00	NA		NA	0
29	coef_hhsize1_sr_nonmandatory	-1.13	0.0689	-16.35	***	NA	0
30	coef_hhsize2_sr_nonmandatory	-0.757	0.0492	-15.37	***	NA	0

Jupyter Notebook (3): Tour Destination Choice



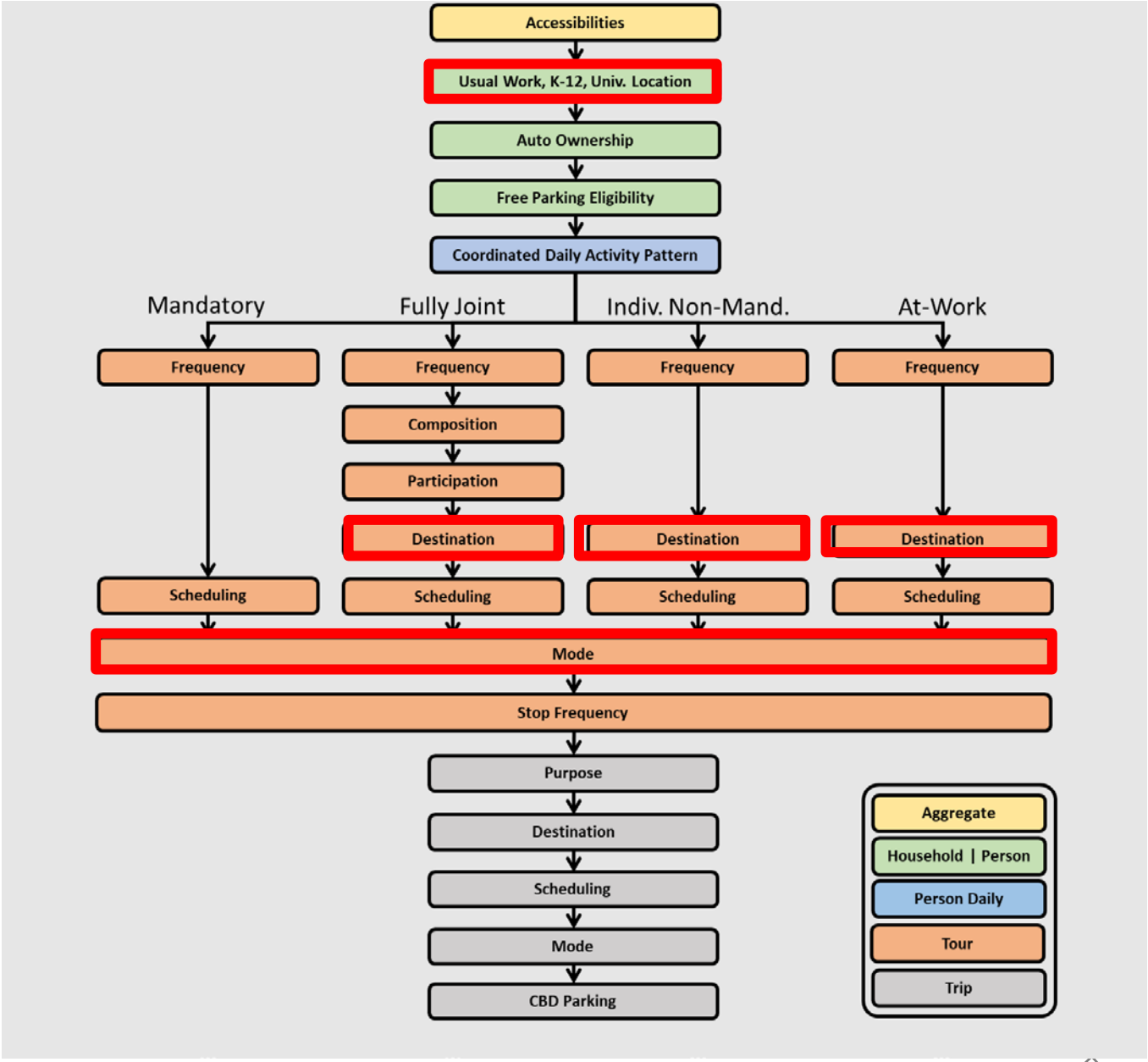
Jupyter Notebook (4): Tour Destination Choice

In [220]:

```
# List of variables to export to ALOGIT
export_vars = [x for x in work_ca_alogit.columns if (x.endswith('_dist') or x.endswith('_logsum'))]
export_vars.insert(0,'person_id')
other_vars = ['pemploy','age','SEX', 'is_student', 'num_children','auto_ownership','income_segment_y','survey_choice_seq
export_vars = export_vars + other_vars

#write out alogit input variables
filename = os.path.join(output_folder,
                                        'destination_choice',
                                        'work_location_choice',
                                        'alogit_input_variables.csv')

with open(filename, 'w', newline='', encoding='utf-8') as f:
    writer = csv.writer(f)
    writer.writerow(['variable'])
    for val in export_vars:
        writer.writerow([val])

#export_vars[-10:]
```

MWCOG Phase I Model Estimation

Gen3 Phase I estimation plan



Tours By Purpose and Mode

Mode	Tour Purpose									Total
	work	univ	school	escort	shop	other maint	eat out	social	other discr	
Drive alone	8,093	252	100	2	2,162	1,351	603	355	2,312	15,230
Shared ride (2 person)	1,482	121	489	1,245	956	793	651	219	1,164	7,120
Shared ride (3+ person)	635	95	803	936	355	411	305	132	684	4,356
Walk	713	54	355	371	418	208	286	111	582	3,098
Bike	293	20	40	12	20	15	17	8	72	497
WALK-All Bus	409	28	40	7	81	57	13	16	65	716
WALK-Metrorail	1,183	36	13	3	38	30	23	11	78	1,415
WALK-Bus+Metrorail	378	22	13	2	11	11	2	3	20	462
WALK-Commuter Rail	22	1	-	-	-	-	-	-	2	25
PNR-All Bus	107	1	-	-	1	-	-	-	1	110
PNR-Metrorail	546	18	1	3	5	5	4	3	36	621
PNR-Bus+Metrorail	54	1	-	-	-	1	1	-	4	61
PNR-Commuter Rail	120	-	-	-	-	-	-	-	-	120
KNR-All Bus	15	1	4	-	-	3	-	-	2	25
KNR-Metrorail	126	-	2	2	2	4	3	1	13	153
KNR-Bus+Metrorail	39	-	3	-	-	1	-	-	1	44
KNR-Commuter Rail	11	-	-	-	-	1	-	-	-	12
SCHOOLBUS	5	15	1,538	2	1	1	-	-	4	1,566
Taxi	83	11	6	1	5	17	5	1	7	136
TNC-Single Payer	143	10	4	1	10	14	19	11	29	241
TNC -Shared	34	1	-	-	1	1	6	1	7	51
Total	14,491	687	3,411	2,587	4,066	2,924	1,938	872	5,083	36,059

Tour Mode Choice Results

Work Coefficients (ASCs not shown)

- Average wage rate was around $34/hr in 2018; cost is too low
- Transfer walk time asserted to be equal to walk access/egress time
- Drive access time asserted to be 2x IVT
- KNR – CR will be turned off

Coefficient	Value	T-Stat	Ratio to IVT
In-vehicle time	-0.0250	-17.16	1.00
Cost	-0.0004	-10.02	$39.68/hr
Walk time	-0.0344	-16.45	1.38
Bike time	-0.0367	-10.58	1.47
Walk to/from transit time	-0.0457	-21.92	1.83
Transit transfer walk time	0.0062	0.25	-0.25
Drive-access time, PNR	0.0615	2.91	-2.46
Drive access time, KNR	-0.0454	-0.83	1.82
Transit first wait time	-0.0372	-5.47	1.49
Transfer wait time	-0.0382	-6.93	1.53
Walk-transit, CBD constant	0.4790	5.54	-19.16
Drive-transit, CBD constant	1.1700	4.35	-46.80
Density, walk mode	0.0580	8.95	-2.32
Density, bike mode	0.0145	1.77	-0.58
age 16-19, drive alone	-0.2050	-0.92	8.20
age 16+, shared ride	-0.4540	-3.24	18.16
household size 1, shared ride	-1.5000	-16.59	60.00
household size 2, shared ride	-0.8950	-16.73	35.80
Walk transit, Metrorail only	1.0100	9.73	-40.40
Walk transit, Bus + Metrorail	0.1490	1.23	-5.96
Walk transit, Commuter rail	1.2000	2.62	-48.00
PNR transit, Metrorail only	-0.8760	-2.66	35.04
PNR transit, Bus + Metrorail	-1.8900	-2.48	75.60
PNR transit, Commuter rail	2.0800	4.79	-83.20
KNR transit, Metrorail only	0.4300	0.57	-17.20
KNR transit, Bus + Metrorail	0.5580	0.61	-22.32
KNR transit, Commuter rail	-8.8800	#VALUE!	355.20

Non-Mandatory Coefficients (ASCs not shown)

- Not enough observations to warrant drive-transit for this purpose
- Transfer walk time asserted to be equal to walk access/egress time
- Transit first wait time asserted to be 2x IVT

Coefficient	Value	T-Stat	Ratio to IVT
In-vehicle time	-0.0213	-6.63	1.00
Cost	-0.0011	-12.11	$ 12.06
Walk time	-0.0488	-32.92	2.29
Bike time	-0.0522	-8.53	2.45
Walk to/from transit time	-0.0371	-9.81	1.74
Transit transfer walk time	0.0236	0.49	-1.11
Transit first wait time	-0.1930	-22.21	9.06
Transfer wait time	-0.0129	-1.08	0.61
Walk-transit, CBD constant	0.2720	1.89	-12.77
Drive-transit, CBD constant	0.0951	0.12	-4.46
Density, walk mode	0.0579	10.42	-2.72
Density, bike mode	0.0410	3.32	-1.92
Age 0-10, walk-transit	-1.4900	-4.30	69.95
Age 16-19, drive alone	-1.2300	-6.80	57.75
Age 16+, shared ride	0.7900	8.61	-37.09
household size 1, shared ride	-1.1300	-16.35	53.05
household size 2, shared ride	-0.7570	-15.37	35.54
Walk transit, Metrorail only	-0.3020	-1.75	14.18
Walk transit, Bus + Metrorail	-0.9790	-3.39	45.96
Walk transit, Commuter rail	2.0100	1.64	-94.37

At-Work Subtour Coefficients (ASCs not shown)

- Not enough observations for walk-Bus+Metro, Comm. Rail, or drive-transit mode estimation
- Bike time asserted equal to walk time
- Transfer walk time asserted to be equal to walk access/egress time
- Transit transfer wait time asserted to be equal to first wait time

Coefficient	Value	T-Stat	Ratio to IVT
In-vehicle time	-0.0361	-2.62	1.00
Cost	-0.0004	-1.62	$ 49.91
Walk time	-0.0709	-15.86	1.96
Bike time	-0.0347	-1.21	0.96
Walk to/from transit time	-0.0414	-2.77	1.15
Transit transfer walk time	0.0835	0.36	-2.31
Transit first wait time	-0.1000	-2.03	2.77
Transfer wait time	0.0171	0.24	-0.47
Walk-transit, CBD constant	0.5800	1.03	-16.07
Density, walk mode	0.0437	2.13	-1.21
Walk transit, Metrorail only	1.9900	3.10	-55.12

Overall comments – Tour Mode Choice

- Mostly reasonable in-vehicle and out-vehicle coefficients, density coefficients, and household/person variables
- Drive-transit estimation suffers from too few observations
- Transit mode-specific constants reflect some underlying biases in survey data and need to be refined based on calibration to on-board data
- School and university mode choice coefficients from MTC donor model will be maintained, calibrated to local conditions

Tour Destination Choice Results

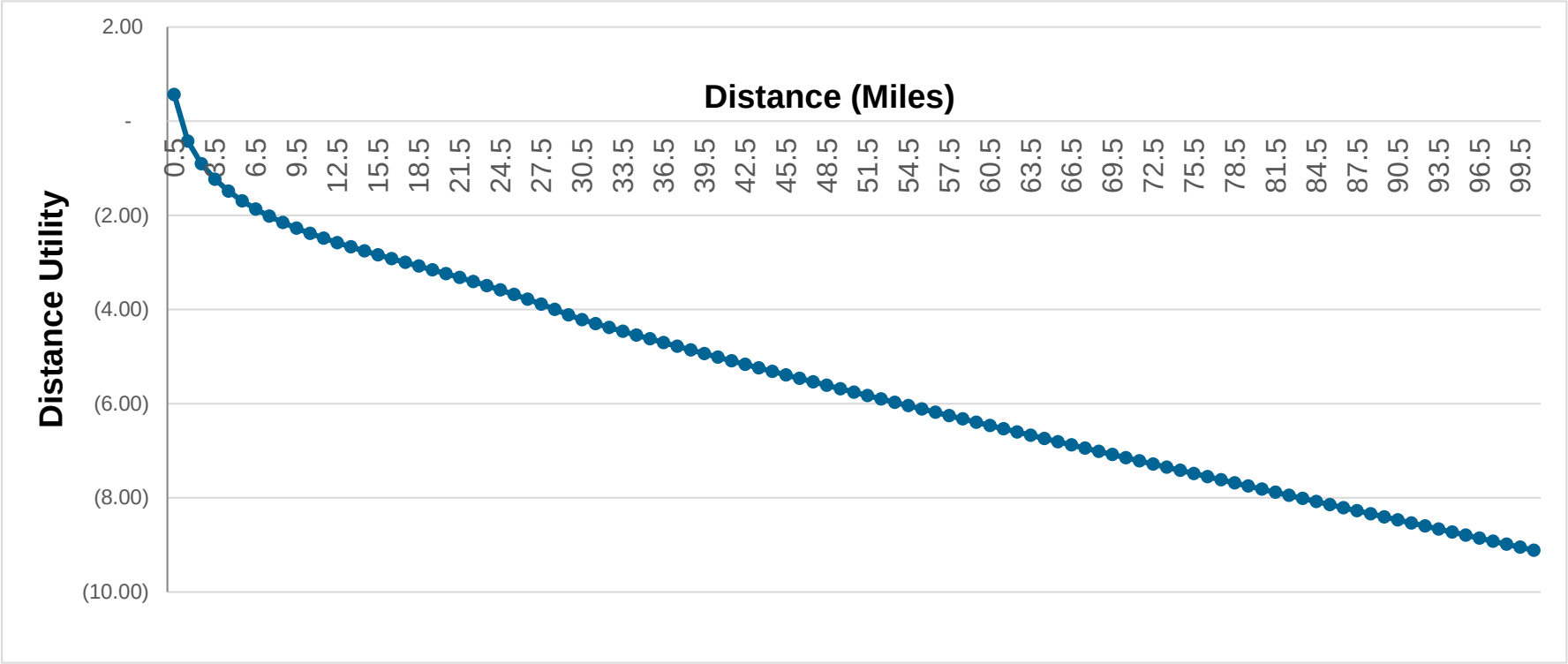
Work Destination Choice Coefficients

17,466 worker records

Coefficient	Value	T-value
Impedance Variables		
Distance	-0.043	-14.73
Distance - Squared (cap at 30)	0.003	8.75
Distance - Cubed (cap at 30)	-7.4E-05	-8.20
Log(1 + Distance)	-0.854	-25.75
LogSum	0.250	
Distance X Demographics		
Young (age<=25)	-0.036	-9.24
Old (age>=65)	-0.014	-4.35
Female	-0.011	-5.38
Part-time worker	-0.016	-7.29
Student	-0.026	-4.71
Zero Auto HH	-0.034	-4.87
Low Income < 50K	-0.019	-5.77
Medium Income 50K-100K	-0.005	-2.31
Very High Income >= 150K	0.006	2.92

Coefficient	Value	T-value
Size Variables		
Office Employment X Low Income	1.000	
Office Employment X Med Income	1.000	
Office Employment X High Income	1.000	
Office Employment X Very High Income	1.000	
Industry Employment X Low Income	1.000	
Industry Employment X Med Income	0.284	-4.63
Industry Employment X High Income	0.351	-5.43
Industry Employment X Very High Income	0.151	-7.03
Retail Employment X Low Income	2.298	6.71
Retail Employment X Med Income	0.573	-4.88
Retail Employment X High Income	0.258	-7.99
Retail Employment X Very High Income	0.134	-9.70
Other Employment X Low Income	2.245	6.28
Other Employment X Med Income	1.305	3.79
Other Employment X High Income	0.807	-4.30
Other Employment X Very High Income	0.807	-4.30

Work Location – Distance Decay



Work Location Size Variables

Employment Category	Income Category: Low	Med	High	V. High
Office	1	1	1	1
Industry	1.000	0.284	0.351	0.151
Retail	2.298	0.573	0.258	0.134
Other	2.245	1.305	0.807	0.807

- Office is the base category
- HH Income groups: <$50K, $50K-$100K, $100K-$150K, >$150K
- Workers from low-income households are more likely to work in zones with retail and other employment
- Propensity to work in non-office employment decreases with increase in household income

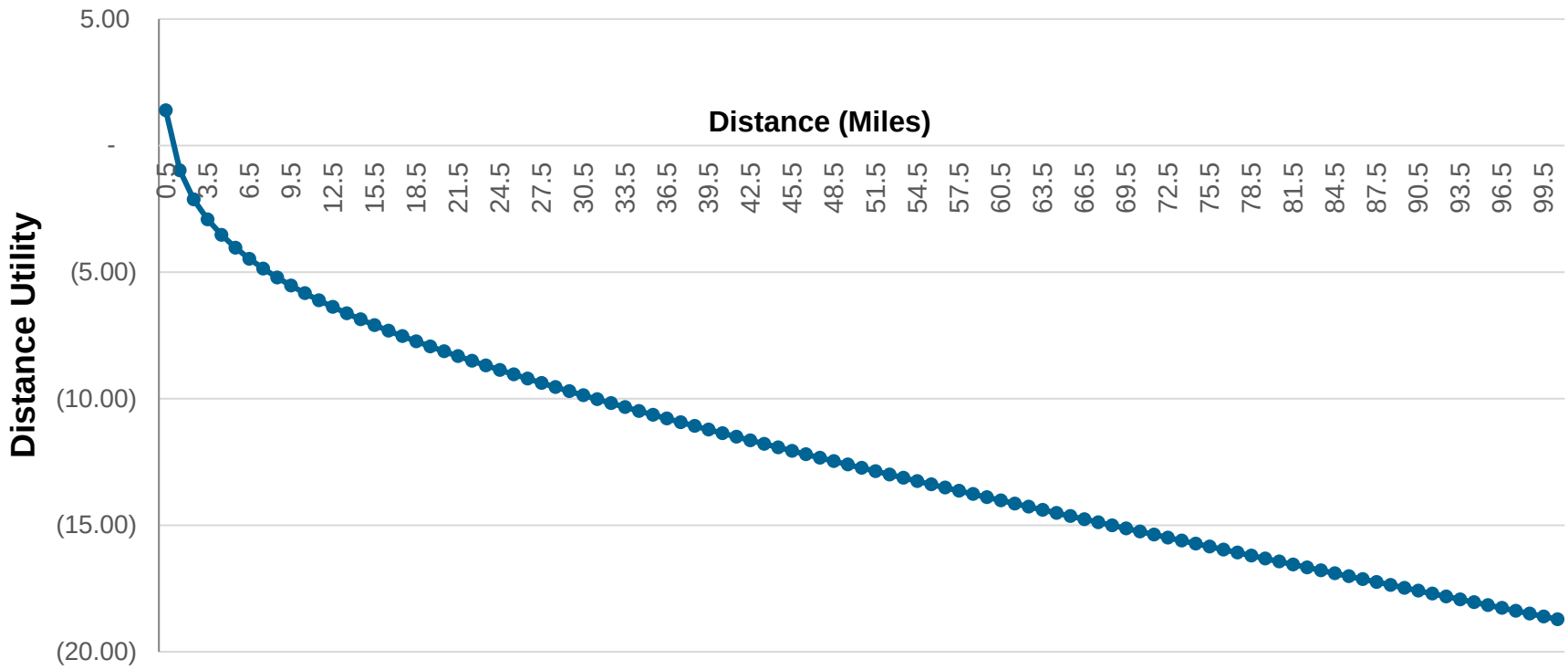
School Destination Choice Coefficients

4,864 K-12 Student records, not enough university records

Coefficient	Value	T-value
Impedance Variables		
Distance	-0.043	-14.73
Log(1 + Distance)	-0.854	-25.75
LogSum	0.300	
Distance X Demographics		
Child aged 6 to 12	-0.036	-9.24
Part-time worker	-0.014	-4.35
Low Income < 50K	-0.011	-5.38
Very High Income >= 150K	-0.016	-7.29

School enrollment is the size variable: K_8 and G9_12

School Location – Distance Decay



Overall comments – Destination Choice

- Reasonable impedance variable coefficients
- Intuitive size variable coefficients
- Constrained logsum coefficients based on simple model specification
- Non-mandatory destination choice results under review

RSG

the science of insight

Joel Freedman

SENIOR DIRECTOR

Joel.Freedman@rsginc.com