

MEMO

TO: Shengyi Gao, SACOG

FROM: Mark Bradley, Nagendra Dhakar, RSG

CC: Yanmei Ou, SACOG

DATE: April 29, 2022

SUBJECT: Report on the re-estimation and calibration of the SACSIM DAYSIM models

Contents

TASK 1: RECOMMENDATION OF THE BEST DAYSIM VERSION	2
TASK 2: PREPARATION OF 2018 HOUSEHOLD TRAVEL SURVEY DATA FOR MODEL ESTIMATION	2
TASK 3: RE-ESTIMATION OF ALL DAYSIM MODEL COMPONENTS .	8
TASK 4: ASSISTANCE WITH CALIBRATION AND VALIDATION OF THE RESULTING MODELS	13
TASK 5: ADDING TNC DEADHEADING TRIPS	17
APPENDIX A DETAILED DESCRIPTION OF THE PURPOSE IMPUTATION PROCESS	20
APPENDIX B: CHECKS FOR TESTING COMPATIBILITY BETWEEN ACTIVITY PURPOSE AND PARCEL LAND USE.....	28
APPENDIX C: COEFFICIENTS ESTIMATED FOR THE PAID RIDESHARE (TNC) UTILITY IN THE SACOG MODE CHOICE MODELS	29

Task 1: Recommendation of the best DAYSIM version

After discussion with SACOG, it was decided to use the latest version of the DAYSIM code (from the Develop branch of the RSG DAYSIM Github repository) along with the latest 2016 base year data inputs from SACOG, including parcel data, zonal data, and networks. The latest SACSIM files were transferred from SACOG to RSG and tested by RSG. Using this version of the model, network skim matrices were produced from a converged base year run to be used in model estimation.

Task 2: Preparation of 2018 household travel survey data for model estimation

The purpose of this task was to perform any further processing of the data from the 2018 SACOG household travel survey to make it suitable for re-estimation of the various choice models in DAYSIM. The main tasks were (a) to perform further logic checks and trip purpose imputation procedures on the survey data, (b) to compare the land use on the geocoded parcel in the survey data to the activity purpose in the data and, if appropriate, choose another nearby parcel with land use that better matches that activity purpose, and (c) to re-run the RSG code to form tours and day-patterns and write out a set of new data files in the correct format for input to DAYSIM for model estimation.

Relevant previous work with the 2018 survey data

As part of the California multi-MPO (MTC, SANDAG, SCAG and SACOG) work program for developing and testing survey methods, RSG has already performed two important work tasks using the 2018 SACOG household travel survey data. (The initial design and pilot for the SACOG survey was also done under the multi-MPO work program, although the main survey data collection was funded by SACOG.)

In Task 7.1 RSG developed new methods for correcting biases in the diary-based survey data relative to the smartphone-based data, both at the day level (too many days in the diary-based data with 0 trips) and at the trip level (lower trip rates in the diary-based data, particularly for shorter discretionary trips). RSG also added additional targets based on ACS commute mode shares, to correct any participation self-selection bias for users of modes such as bike and transit. The methods developed and work done in the Task 7.1 research remain current, so the revised survey weights developed as part of that work were used for this project.

In Task 7.2, RSG estimated several example choice models using the 2018 SACOG survey data to demonstrate the value of using multi-day survey data in model estimation. This work included a limited subset of the full set of model components in DAYSIM, including mode choice models that included the TNC mode, as well as examples of destination choice and intermediate stop generation models. The purpose of the research was to demonstrate the value of estimating models with multiple weekdays of survey data from respondents who used the smartphone-based method.



The reports from both Tasks 7.1 and 7.2 have been provided to SACOG.

Application of new activity purpose imputation methods

After Tasks 7.1 and 7.2 were completed in 2019, MTC, SANDAG, SCAG and SFCTA funded additional work to improve RSG’s methods for data QA/QC, particularly for the smartphone-based data, which is relatively new in terms of its use in modeling, and also provides much more opportunity to double-check the respondents’ answers against the locations, times and speeds captured in the smartphone GPS trace data. The main result of that work that was applied in this project are algorithms to check reported trip purposes against the known locations in the data (such as home, work and school), to flag any inconsistencies, and to impute “corrected” answers where appropriate. These imputation methods can also be used to provide trip purposes where none were reported, although we did not use trips with missing data for model estimation in this project.

Table 1 shows the results of the purpose imputation processing, with technical detail on the various logic tests referred to in the table provided in Appendix A. In total, the tests assign an imputed destination purpose that is different from the reported purpose for about 9% of the rMove (smartphone-based) trips and about 6% of the rSurvey (diary-based) trips.

The logic also attempts to impute more specific purpose data for the cases where people reported purpose as “other” or did not report purpose at all. The latter case of missing trip surveys are not included in the model estimation, but cases where purpose is “other” but all other trip data is complete are used in estimation, so the imputation of a more detailed purpose based on the imputation rules is useful. In most cases, the purpose is imputed from the purpose of another trip that the person made to within 50 meters of that same location within the data collection period.

Table 1: Results of activity purpose imputation logic

	rMove #	rMove %	rSurvey #	rSurvey %
OK-Use as is	134336	71.6%	4810	76.3%
Location is anchor location & purpose is not consistent				
Test 1-change mode ok	78	0.0%	0	0.0%
Test 2A-change to location type	2354	1.3%	55	0.9%
Test 2B-for last trip of day-change to home	1252	0.7%	15	0.2%
Test 2C-2A modified for missing-change to location type	2011	1.1%	54	0.9%
Test 2D-2B modified for missing-change to home	823	0.4%	13	0.2%
Test 3A-prev trip at same location- longer dwell - keep as is	787	0.4%	14	0.2%
Test 3B-next trip at same location- longer dwell - keep as is	436	0.2%	5	0.1%
Test 5-prev trip is opposite mismatch - swap around	34	0.0%	0	0.0%
Test 6-next trip is opposite mismatch - swap around	24	0.0%	0	0.0%
Test 7-change to location type- all purposes in day shifted down 1	370	0.2%	6	0.1%

Test 8-last destination of day to Home loc with stay >3hr- change to home	3434	1.8%	0	0.0%
Test 9-tests distance to usual location in case H/W/S have same loc	2111	1.1%	125	2.0%
Test 10-change work-related to work if at work loc with no adj work trip	486	0.3%	25	0.4%
Test 11-change to loc type if does not create consec trip with same purpo	2894	1.5%	83	1.3%
x/O No test passed - examine further	5633	3.0%	98	1.6%
Location is not anchor location & purpose is not consistent				
Test 1-imputed Escort purpose	193	0.1%	2	0.0%
Test 2B-within 300 m of usual location- no adj conflicts	2	0.0%	0	0.0%
Test 2C-within 500 m of usual location- no adj conflicts	18	0.0%	0	0.0%
Test 3-Is overnight stay away from home- out of region	54	0.0%	0	0.0%
Test 3-Is overnight stay away from home- within region	38	0.0%	0	0.0%
Purpose = other or missing....				
Imputed based on same persons stops within 50 m of same location	19414	10.4%	416	6.6%
Imputed based on same persons stops within 100 m of same location	811	0.4%	6	0.1%
Imputed based on same persons stops within 200 m of same location	867	0.5%	16	0.3%
Imputed based on other persons stops within 50 m of same location	5630	3.0%	418	6.6%
Imputed based on other persons stops within 100 m of same location	1143	0.6%	56	0.9%
Imputed based on other persons stops within 200 m of same location	862	0.5%	39	0.6%
No other visits to location - impute other (no evidence)	1454	0.8%	52	0.8%
Total	187549	100.0%	6308	100.0%



Checking consistency of geocoded parcel IDs against parcel level land use

Because parcels are so small and survey data locations are not always precise, geocoded parcels may not be the actual ones where destination activities took place. It is best for modeling if survey data purpose is consistent with the parcel data, in terms of employment by type, school enrollment by type, and households for the assigned parcel.

Table 2A: Results of parcel land use checks by activity purpose (part 1)

Geocoded parcel in survey data....	home	usual work	other work	College	k-12	meal
matches nearest parcel that passes test 1	12.6%	46.1%	41.5%	46.0%	38.0%	32.4%
does not match nearest, but passes test 1	79.9%	43.5%	37.3%		1.3%	50.8%
matches nearest parcel that passes test 2	.1%	1.0%	2.4%	20.7%	6.2%	2.8%
does not match nearest but passes test 2	.2%	4.3%	11.3%	20.4%	2.5%	3.3%
another parcel within 1/8 mi passes test 1	5.8%	3.9%	4.8%	1.4%	9.3%	7.1%
another parcel within 1/8 mi passes test 2	.0%	.5%	1.4%	1.0%	1.9%	.7%
another parcel within 1/4 mi passes test 1	.8%	.1%	.4%	1.8%	9.6%	1.1%
another parcel within 1/4 mi passes test 2	.0%	.1%	.4%	1.1%	1.7%	.5%
another parcel within 1/2 mi passes test 1	.3%	.1%		.8%	13.9%	.6%
another parcel within 1/2 mi passes test 2	.0%	.0%	.1%	2.8%	2.3%	.4%
no parcel within 1/2 mile passes either test	.1%	.2%	.5%	4.0%	13.2%	.3%
total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 2B: Results of parcel land use checks by activity purpose (part 2)

Geocoded parcel in survey data....	shop	per bus	social rec	Escort	other	school-no usual location	total
matches nearest that passes test 1	45.6%	29.5%	17.3%	22.7%	27.6%	32.1%	26.1%
does not match nearest, but passes test 1	38.9%	35.4%	56.4%	67.5%	62.3%	61.1%	57.8%
matches nearest that passes test 2	2.8%	9.5%	4.0%				2.2%
does not match nearest but passes test 2	3.6%	20.4%	7.3%				3.9%
another parcel within 1/8 mi passes test 1	4.4%	3.7%	11.8%	8.7%	7.6%	6.2%	6.6%
another parcel within 1/8 mi passes test 2	1.1%	1.0%	.1%				.4%
another parcel within 1/4 mi passes test 1	1.0%	.1%	1.8%	.9%	1.3%	.4%	1.2%
another parcel within 1/4 mi passes test 2	1.0%	.1%	.1%				.3%
another parcel within 1/2 mi passes test 1	.7%	.0%	.5%	.1%	.4%	.1%	.7%
another parcel within 1/2 mi passes test 2	.6%	.1%	.0%				.2%
no parcel within 1/2 mile passes either test	.3%	.0%	.7%	.1%	.8%	.1%	.6%
total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Tables 2A and 2B show the results of comparisons where each parcel in the survey data was compared to the nearest parcels in the parcel database (in terms of straight-line distance between the geocoded trip end coordinates and the parcel centroid coordinates) that passed

two different “tests” of whether the parcel has land use data that are consistent with the activity purpose. The tests are shown in Appendix B, with Test 1 being stricter than Test 2. The conditions in the rows of Tables 2A and 2B are checked in the same sequence as the rows in the table. First, if the parcel originally geocoded in the data passes Test 1 or Test 2, that parcel is retained for modeling. That is the first four rows in the tables, that include (in the Total column across all purposes) of trip end locations that 84% that pass Test 1 and another 6% that do not pass Test 1 but pass Test 2.

For the remaining 10% that do not pass either test, there are tests for the next three rows shaded in green—if there is a different parcel in the parcel database within 1/8 of a mile of the survey location that passes Test 1, a different parcel with 1/8 that passes Test 2, or a different parcel between 1/8 mile and 1/4 mile that passes Test 1. A further 8% of trip ends all in these rows, with the highest percentages for K-12 school and social/recreation purposes. For those cases, the alternative parcel is used for modeling.

This leaves about 2% of trip ends in the last four rows of the table for which no alternative parcel that passes a test within a reasonable distance from the survey location is found. In those cases, the original survey parcel is retained for modeling. The cells shaded in yellow show that the purpose for which the most cases fall in the last four rows is K-12 education. Many of those trip ends are for children under 5, so the actual location may be a pre-school or day-care location which is not included as K-12 enrollment in the parcel database. For most other purposes, 1% or fewer of the cases are in the last 4 rows of the table.

Preparing the data for input to DAYSIM estimation mode

Custom RSG code is run to prepare the survey data for input to DAYSIM to estimate models. This code performs the following:

1. Splits the mtelulti-day data into separate household-days.
2. Forms tours from the trips in each travel day
3. Classifies person-days by day patterns based on the tours and other criteria.
4. Selects person days to use in model estimation. (This is described in the following paragraph.)
5. For the selected person days, writes out household, person, household-day, person-day, tour and trip files in the format expected by DAYSIM.

Table 3 shows the results of the selection of person-days to include in model estimation before the extra data imputation and processing done in Task 2, while Table 4 shows the corresponding results when done after the Task 2 data processing. The first three columns correspond to reasons why some person days are not used in model estimation—(a) there is incomplete data (missing trip surveys and/or day survey) for the person, (b) there is incomplete data for that day for another person in the household, or (c) the data is from a Friday, Saturday



or Sunday, which are designated as “weekend days”. In other words, only weekday data (Mon-Thu) from travel days with no trip data for the entire household are used.

The first two rows in each table show the types of travel days patterns that are acceptable for all of the models in DAYSIM—either (a) the person stays home for the entire day, with no trips, or (b) the person makes trips and starts and ends the travel day at home and stays within the SACOG region. The last three rows are people who start and/or end the travel day away from home, leave the SACOG region during the day, or both.

The main difference between Tables 3 and 4 are that the improved purpose imputation yields a higher percentage of rMove person-days that start and end at home and stay in the region. From the orange-shaded cells in Table 3, more than half of the rMove Weekday person days that start/or end away from home (26.7%) now begin or end, increasing the percentage in the row above from 52.8% to 67.0% (the green-shaded cell in Table 4. There is very little change for the rSurvey data, since the purpose imputation improvements apply mainly for the rMove smartphone-based data.

Table 3: Classification of person-days BEFORE extra data processing

	rMove-person-day incomplete	rMove-hhold-day incomplete	rMove-all complete-Weekend*	rMove-all complete-Weekday	rSurvey-all complete-Weekday
No trips made-stayed home	26.9%	21.3%	15.7%	12.2%	23.1%
Start and end at home & stay in region	25.2%	42.8%	48.9%	52.8%	65.8%
Start and/or end away from home	33.4%	25.9%	22.7%	26.7%	6.3%
Leave the region	4.4%	4.5%	5.5%	3.8%	2.8%
Start/end away from home & leave region	10.1%	5.6%	7.3%	4.6%	2.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

* FRI, SAT and SUN are designated as weekends, while MON, TUE, WED and THU are weekdays

Table 4: Classification of person-days AFTER extra data processing

	rMove-person-day incomplete	rMove-hhold-day incomplete	rMove-all complete-Weekend*	rMove-all complete-Weekday	rSurvey-all complete-Weekday
No trips made-stayed home	26.9%	21.3%	15.7%	12.2%	23.1%
Start and end at home & stay in region	42.2%	55.3%	62.8%	67.0%	66.3%
Start and/or end away from home	16.3%	13.3%	8.8%	12.5%	5.8%
Leave the region	5.0%	4.5%	5.2%	4.2%	2.7%
Start/end away from home & leave region	9.6%	5.6%	7.6%	4.2%	2.2%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

Both tables also show that the rSurvey data has a much higher percent of person days with no trips compared to the rMove data (23.1% vs 12.2%, highlighted in the yellow cells in Table 4).

This difference was addressed in model estimation by including dummy variables for the diary-based data in the main tour and stop generation day pattern model, to correct for any bias in trip rates due to the survey method after accounting for the other variables in the model (person type, income, age, etc.)

Task 3: Re-estimation of all DAYSIM model components

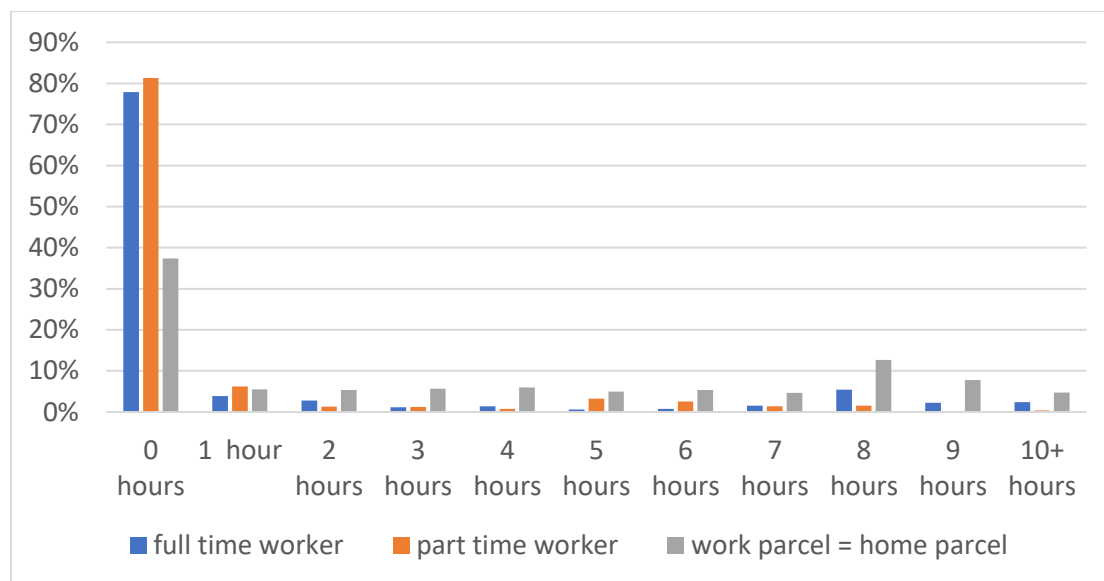
Task 3 required estimating new coefficients for all of the DAYSIM model components, as well as estimating and implementing a new model for people who have out-of-home workplaces but choose to work at home for at least a threshold period of time in the day.

The new Work at Home model

The Work Location model that predicts each worker’s usual work parcel already predicts workers who always work from home (the usual work parcel is the home parcel). Prior to this project, however, DAYSIM did not have a way to predict workers who have an out-of-work place but still choose to do paid work at home during the simulated weekday. The model could predict that a worker would not make a work tour to the usual workplace, but working at home was treated no differently than other reasons for not going to work, (a sick day, on vacation, scheduled day off, etc.).

The 2018 SACOG survey asked each worker at the end of each day how many hours of paid work the person had done at home during that day, regardless of whether that person had (also) commuted to work on that day or not. Figure 1 shows the unweighted distribution of responses to that question. About 80% of full-time and part-time workers with out-of-home workplaces did no paid work at home, while only 35% of those whose workplace is the home parcel did no paid work there on that day.

Figure 1: Hours doing paid work at home on survey day by worker type





Another 7% or so of full-time and part-time workers worked at home only or 1 or 2 hours. For those who worked at home more than two hours, the most common answers were 8 or 9 hours for full-time workers and those whose usual workplace is at home, but a much flatter distribution for part-time workers.

Many of the workers with out-of-home workplaces who reported doing paid work at home also commuted to the usual workplace on the same day. Table 5 shows that for those who worked at home for 0-2 hours, over 50% made commute tours on that day—similar to the overall percentage of 56.7%. For those who worked at home for 3 or more hours, only about one-third made commute tours to the usual workplace, with the values all quite similar ranging from 26.8% to 37.7%. Based on this result, a threshold of 2.5 hours was used for the binary “Work at Home” choice model, with those working 3 or more hours at home choosing the “work at home” alternative.

Table 5: Fraction of workers commuting to usual workplace by hours worked at home (unweighted)

Hours worked at home during day	Commuted to usual workplace that day
0	62.2%
1	57.7%
2	47.4%
3	35.8%
4	32.0%
5	33.6%
6	26.9%
7	30.0%
8	37.7%
9	33.3%
10	26.8%
Total	56.7%

An important feature of the work-at-home model is that it only predicts whether workers with out-of-home usual workplaces choose to work at home for 3+ hours on the simulated day. It does not directly predict whether that person (also) makes a commute tour or other type of work tour on that day. That is done indirectly in DAYSIM by using the prediction from the Work at Home model as an input to the Individual Person Day Pattern Model that predicts whether a person makes tours and/or extra stops for various travel purposes. Note that this is the same way that “work from home” (usual work parcel = home parcel) is handled in the Individual Person Day Pattern Model but using separate variables for “work at home” versus “work from home”. This two-stage method of predicting teleworking provides greater flexibility in applying the models and is consistent with how the survey data was collected.

The Work at Home model was estimated as a binary logit model using 9,221 worker-weekday observations (excluding workers whose usual work parcel is the home parcel). In 1,180 (12.8%) of the cases, the person chose to do paid work at home for more than 2.5 hours. The estimated coefficients are shown in Table 6. Part-time workers were somewhat less likely to work at home relative to full-time workers, while those in households with income below \$50,000 (roughly 35,000 in year 2000 \$), with income above \$150,000, with no vehicles, and with at least one non-working adult and at least one child under age 18 were somewhat more likely to work at home.

Table 6: Estimation results for the Work at Home model

	#.of Cases	% of Cases
Worked at home < 2.5 hours	8,041	87.2%
Worked at home >=2.5 hours	1,180	12.8%
Total worker days w/out-of-home workplace	9,221	100.0%
variable	Coefficient	T-statistic
constant	-1.968	-21.2
part time worker	-0.264	-2.7
income missing	0.295	1.9
income < \$50,000	0.349	2.8
income > \$150,000	0.300	2.7
non-working adult and kids in HH	0.399	3.3
no vehicles in HH	0.412	2.2
fraction medical jobs	-0.415	-2.7
fraction "other" jobs	1.396	2.1
fraction industrial jobs x income > \$50k	0.431	2.3
fraction industrial jobs x income < \$50k	-0.402	-1.7
fraction office jobs x income > \$50k	0.295	1.6
fraction office jobs x income < \$50k	-0.630	-2.7
fraction government jobs x income > \$50k	-0.311	-1.4
fraction government jobs x income < \$50k	-0.650	-2.3
fraction retail / food jobs x income > \$50k	-0.385	-1.8
fraction retail / food jobs x income < \$50k	-0.418	-2.2

The rest of the variables in the model are based on the distribution of employment at the worker's usual work parcel interacted with the worker's household income (above or below \$50,000). With the fraction of service jobs used as the base category, those whose work parcels have high fractions of medical, government, or retail/food service jobs were somewhat less likely to work at home. For most employment types, those with incomes above \$50,000 were more likely to work from home. For industrial and office jobs, the coefficients have different signs depending on income level.



These results are generally consistent with findings elsewhere about what types of workers were most likely to work from home prior to the COVID-19 pandemic. These patterns may change post-pandemic, and the model provides the flexibility to adjust the coefficients to represent different future telework scenarios.

Note that it would probably give a better model if we could use the worker's actual specific industry type/occupation and the worker's own personal income, but neither of those were asked in the survey and neither are available in the synthetic population used for model application. Those questions are asked in the American Community Survey (ACS), so it may be worthwhile to ask them in future travel surveys and possibly include them in the population synthesis process.

The code for the new Work at Home choice model was programmed as a “pre-model” to the Individual Person Day Pattern Model, in the same way as the Automated Vehicle (AV) choice model had been programmed as a “pre-model” to the Auto Ownership model. This means that the coefficients to the model are input via new entries in the properties configuration file rather than as a separate coefficient (.F12) file.

In model application the outcome from the Work at Home model is written to the Person-Day file in the “workathome” field. A value of 0 indicates that the person does not work at home, while any value greater than 0 indicates that the person does work at home for more than the threshold number of hours.

Estimation results for the pre-existing models

Nearly all of the pre-existing choice models in DAYSIM were re-estimated using the internal estimation capabilities. The only exception was the Pay to Park at Workplace model, for which the 2018 SACOG survey did not contain suitable data for estimation.

For model estimation with the survey data, a scaling parameter was implemented via the configuration properties file to scale the 2017 incomes in the survey back to 2000 \$, since the income-related variables in the models were based on year 2000 survey data with 2000 \$. (In model application, if the incomes in the synthetic population are already scaled to year 2000 \$, then this configuration scaling parameter should be set at 1.0 (the default value).

Rather than tabulate the model coefficients in this document, they are tabulated in an accompanying Excel workbook “SACOG DAYSIM model variable descriptions-new 2022.xlsx”. There is a worksheet for each choice model containing at least three sets of coefficients:

1. The previously estimated coefficients, based on the 2000 SACOG travel survey, along with the t-statistics related to the coefficients.
2. The calibrated coefficients of the previously estimated model, along with the difference between the calibrated coefficients and the original coefficients. Cells where the difference is non-0—meaning that the coefficient was changed in calibration—are highlighted in yellow.

3. The newly estimated model coefficients, along with the corresponding t-statistics.

Both of the t-statistics columns use conditional shading with darker red for the most negative t-statistics and darker green for the most positive t-statistics. A quick comparison of the colors in the two columns generally indicates similar results from the previous models and the new models in terms of the signs and relative significance of the estimates, although there are a few cases where coefficients change sign.

Below the coefficients are additional rows showing the number of observations, the final log-likelihood, the rho-squared relative to a model with 0 coefficients, and, for models that include constants, the rho-squared relative to a model with alternative-specific constants only. In most cases, the number of observations in the new data is at least twice as high as the number of observations used to estimate the previous models. This is due largely to the use of multi-day survey data with multiple weekdays of travel data from many households.

Some results to note in the new models are:

- The mode choice models now include the TNC mode, for which new variables were added. Details please refer to Appendix B.
- A key variable in the mode choice models is the “timeutil” variable which is the logsum from the PathTypeModel across all relevant time and cost variables (using tour-specific and person-specific VOT) and all available path types for the mode. One would expect the coefficient for this variable to be in the range of 0.5 to 1.5 with high t-statistics. In the new mode choice models, this is the case. Compared to the previous models, the timeutil estimates are generally closer to 1.0 with higher t-statistics. This is also generally true for the “costutil” variable that relates only to parking cost, although that variable is much less important to the overall model sensitivity than the “timeutil” variable.
- The WorkUsualLocation, SchoolUsualLocation and WorkTourDestination sheets show two new models, as these models are estimated in two stages—stage 1 omitting the “special alternative” (usual work parcel=home parcel, usual school parcel=home parcel, and destination parcel=usual work parcel, respectively), and stage 2 constraining the estimates from stage 1 and adding the special alternative in a nested structure.
- The IndividualPersonDayPattern model has several new variables using the outcome from the new Work at Home model. These new variables are highlighted in yellow on the worksheet. The most important variable is 134-W-WATHOME which has a coefficient of -1.21 and a t-statistic of -16.6. This indicates that those who work at home for more than 2.5 hours in the day are much less likely to make any work tours. This can be compared to variable 124-W-WAHOME which applies to people whose usual work parcel is the home parcel. That coefficient is more negative (-2.66) but with a similar t-statistic (-16.1). Other key new variables are 834-T-WATHOME (0.272 with t-statistic of 2.3) and 934-I-WATHOME (-0.191 with t-statistic of -1.6). These two variables indicate that those who work at home may make more home-based tours (somewhat offsetting the negative



coefficients on most of the purpose-specific variables) but with fewer intermediate stops per tour.

- Two other key new variables in the IndividualPersonDayPattern model are 832-T-NODIAR (estimate -0.80 with t-statistic -6.8) and I-NODIAR (estimate -0.472 with t-statistic -4.3). These variables apply to the data from the rSurvey diary-based method, as opposed to the rMove smartphone-based method. They help adjust the models for the bias towards fewer tours and trips in the diary-based data, so that when the models are applied without these bias variables, they will predict higher, less biased trip rates. Note that these variables are also applied to specific purposes, such as 732 D-NODIAR (estimate -0.172, t-statistic -3.4) to add further purpose-specific bias adjustments that tend to be most significant for the non-mandatory purposes like shopping and social/recreation. (Also note that the purpose-specific NODIAR variables were also estimated in the previous models, but had a different definition. Since the 2000 survey used only diary-based methods, the NODIAR variables were applied for respondents who said they did not use the diary form provided with the survey and were replying only from memory. That variable was not collected in the 2018 survey, so these variables were redefined to capture the difference between smartphone-based and diary-based data.)

In addition to the model spreadsheet, the coefficient (.F12) files for the new models have been supplied to SACOG, as well as the updated DAYSIM code. RSG also gave a 2-hour seminar to SACOG summarizing most of the material in the Task 2 and Task 3 sections of this document.

Task 4: Assistance with calibration and validation of the resulting models

To prepare the model for calibration, we updated DAYSIM configuration files to add/modify settings. Those changes are described in the next section, followed by the details of the model calibration.

DAYSIM Configurations

We made two sets of changes to the DAYSIM configuration files used in the SACSIM model.

- Changed format from XML to Properties
- Added new properties and modified a few existing ones

The SacSim model uses two DAYSIM configuration files which reside under the scenario folder “2016_new_DAYSIM\run_2016_fixDelCurv_newDAYSIM”. The files were in XML format. We changed the file format to Properties for easier edits and also to align with the latest DAYSIM. The two properties file are still under the same scenario folder with the below names:

- config.sacsim.trips.properties

- config.sacsim.wksoloc. properties

We also updated the main Cube run script (run_2016_fixDelCurv_newDAYSIM\RUN_sacsim19_02_01_2016_baseline_TrnLINfile_cluster Order.s) to reflect the configuration file name changes. We also made a few other setup changes to better organize the model setup (e.g., DAYSIM executables and coefficient files were placed in separate folders - "DAYSIM\software" and "DAYSIM\coefficients")

Table 7 presents the list of properties that were either added or modified during this project. The changes are organized in the three categories

- New work-at-home model settings – new settings for the new telecommute model that predicts work-at-home choice for workers that have an out-of-home workplace
- Ride share settings – settings for enabling ride sharing (TNC) and producing outputs in format required to process these trips
- Other settings – other settings that were added or modified

TABLE 7: DAYSIM CONFIGURATION CHANGES

SETTING	OLD	NEW
New Work-At-Home (Telecommute) Model settings		
UseDiaryVsSmartphoneBiasVariables		TRUE
UseProxyBiasVariables		TRUE
UseWorkAtHomeModelAndVariables		TRUE
WorkAtHome_DurationThreshold		2.5
WorkAtHome_AlternativeSpecificConstant		-1.8067
WorkAtHome_PartTimeWorkerCoefficient		-0.264
WorkAtHome_Income0to50Coefficient		0.349
WorkAtHome_IncomeOver150Coefficient		0.3
WorkAtHome_NonWorkerAndKidsInHHCoefficient		0.399
WorkAtHome_NoVehiclesInHHCoefficient		0.412
WorkAtHome_FractionMedicalJobsCoefficient		-0.415
WorkAtHome_FractionEducationJobsCoefficient		0



SETTING	OLD	NEW
WorkAtHome_FractionServiceJobsCoefficient		0
WorkAtHome_FractionOtherJobsCoefficient		1.396
WorkAtHome_FractionGovernmentJobsLowIncomeCoefficient		-0.65
WorkAtHome_FractionIndustrialJobsLowIncomeCoefficient		-0.402
WorkAtHome_FractionRetailFoodJobsLowIncomeCoefficient		-0.418
WorkAtHome_FractionOfficeJobsLowIncomeCoefficient		-0.63
WorkAtHome_FractionGovernmentJobsHigherIncomeCoefficient		-0.311
WorkAtHome_FractionIndustrialJobsHigherIncomeCoefficient		0.431
WorkAtHome_FractionRetailFoodJobsHigherIncomeCoefficient		-0.385
WorkAtHome_FractionOfficeJobsHigherIncomeCoefficient		0.295
Ride Share Settings		
PaidRideShareModelsAvailable	False	TRUE
PaidRideshare_UseEstimatedInsteadOfAssertedCoefficients		True
PaidRideshare_OutputNumberOfPassengersOnTripRecord		True
PaidRideshare_1PassengerShareForWorkTours		0.7
PaidRideshare_2PassengerShareForWorkTours		0.2
PaidRideshare_AverageNumberFor3plusPassengerWorkTours		4
PaidRideshare_1PassengerShareForSchoolTours		0.5
PaidRideshare_2PassengerShareForSchoolTours		0.3
PaidRideshare_AverageNumberFor3plusPassengerSchoolTours		4
PaidRideshare_1PassengerShareForEscortTours		0.1
PaidRideshare_2PassengerShareForEscortTours		0.4

SETTING	OLD	NEW
PaidRideshare_AverageNumberFor3plusPassengerEscortTours		3.3
PaidRideshare_1PassengerShareForOtherTours		0.45
PaidRideshare_2PassengerShareForOtherTours		0.35
PaidRideshare_AverageNumberFor3plusPassengerOtherTours		4.25
Other Settings		
RawPersonPath	.\2016_raw_person.txt	.\2016_prec_pp.csv
DestinationScale	1	0
HouseholdIncomeAdjustmentFactorTo2000Dollars		1
UseWorkShadowPricingForWorkAtHomeAlternative		false
CountAllIntermediateStopsOnPersonDayRecord		true
PathImpedance_BikeUseTypeSpecificDistanceFractions	False	True

Other configuration settings that were changed included:

- The coefficient file names for all the models, using the names of the newly estimated model coefficient files, typically ending with “_2021v2.F12” or “_2021v3.F12”.
- The file names for the synthetic population file, after SACOG adjusted the person type (pptyp) variable in the person file to use new rules to assign person type to be consistent with past practice.

Initial Base Year Model Calibration

As part of the calibration, RSG adjusted the R-based calibration scripts to separate out results for workers with usual work parcel=home parcel and workers who work at home more than 2.5 hours in the day (using the ‘wkathome’ field in the person day output file, so that the effect of the new work at home model could be assessed. The predictions from these models closely matched the percentages in the survey data.

RSG also created new versions of the survey data files that included only complete person-days beginning and ending at home and staying within the SACOG region, as these provide the best basis for comparison to days simulated by DAYSIM and are best to use as inputs to the R calibration scripts. (Survey data files with a larger number of observations had been used in model estimation, since some person-days were useful in estimating the tour-level and trip-level models and the longer-term models, even though they were not appropriate for estimating the



day pattern models and were excluded in the estimation of those models.) These newer versions of the survey data files also have slightly adjusted weights so that the total weighted population for each person type in the survey data matches the total population for the same person type in the synthetic population.

RSG calibrated various DAYSIM model components and assisted SACOG with further calibration in-house. Table 8 provides the list of DAYSIM models that RSG calibrated. The table also provides details of the observed data source and the summaries used in the model calibration. Note that after delivery of the initially calibrated model, RSG assisted SACOG in conducting in-house calibration to further refine model results.

Table 8: Calibrated DAYSIM Models

DAYSIM MODEL	OBSERVED SOURCE	CALIBRATION SUMMARIES
Auto ownership	2016 calibrated model	Number of households by household size and autos owned
Work location	2018 household travel survey	Work-from-home share, work-at-home (telecommute) share, and home to work distance (miles)
School location	2018 household travel survey	School-at-home share, home to school distance (miles)
Individual Day Pattern	2018 household travel survey	Tours by purpose, tours by purpose and person type, stops by purpose
Tour Mode	2018 household travel survey	Tour mode share by purpose (work, school, escort, other, and work-based)

Task 5: Adding TNC deadheading trips

To assign TNC trips to existing auto classes for assignment, we used the DORP (Driver Type) variable on the trip file. Table 9 shows the values in the DORP field and how to assign them (there is no driver for AVs) to vehicle trips. The TNC trips are assigned to the mode shown in the table and to the highest VOT bin.

Even though there is a driver in the TNCs which increases their occupancy for classifying the auto submode (e.g. can use HOV lanes, etc.), the driver's trip is not explicitly modeled by DAYSIM, so we do not count the driver when converting from passenger-trips to vehicle-trips. The other passengers' trips are simulated by DAYSIM, at least in aggregate, so we use the same type of person-trip to vehicle-trip conversion we use for each mode for regular car trips, but not including the driver.

TABLE 9: DORP FIELD ASSIGNMENT TO VEHICLE CLASSES

DORP	DESCRIPTION	ASSIGNED CLASS	VEHICLE TRIPS
11	TNC non-AV, 1 passenger	assign to HOV 2	1 vehicle-trip per trip
12	TNC non-AV, 2 passengers	assign to HOV 3+	0.5 vehicle-trips per trip
13	TNC non-AV, 3+ passengers	assign to HOV 3+	0.25 vehicle trips per trip
21	TNC AV, 1 passenger	assign to SOV	1 vehicle-trip per trip
22	TNC AV, 2 passengers	assign to HOV 2	0.5 vehicle-trips per trip
23	TNC AV, 3+ passengers	assign to HOV 3+	0.25 vehicle trips per trip

In addition to actual passenger trips, TNC trips generate deadheading trips impacting flows on the roadway network. To account for the deadheading impact, we referenced a Fehr and Peers research report¹ that estimates that 40% of TNC VMT is deadheading, compared to 60% transferring passengers. So, for each TNC passenger trip where you assign X vehicle trips to the network, you generate a deadhead trip with $(X * 40/60)$ vehicle trips.

Since we cannot easily predict the origin or destination of the deadhead trip to assign it to the OD of any actual deadhead trips, we just use the same origin and destination and time period as the TNC passenger trip for the deadhead trip but factored by 40/60 (see above) and assigned as an SOV trip in the medium VOT bin. (For some of the deadhead time, the driver be hurrying to pick up the passenger, but a lot of the time they are driving around waiting for a request to respond to.)

The above TNC related changes are made in the main Cube run script (run_2016_fixDelCurv_newDAYSIM\RUN_sacsim19_02_01_2016_baseline_TrnLINfile_cluster Order.s). Figures 2 and 3 show the portions of the script where these changes are made.

¹ Source: <https://www.fehrandpeers.com/what-are-tncs-share-of-vmt/>



FIGURE 2: TNC TRIPS TO ASSIGNMENT VEHICLE TABLES

```
1939 ; - - - - -
1940 ; Set assignment user-classes
1941
1942 if (ri.mode=3) ; SOV
1943     autoclass=(ri.votclass-1)*3 + 1
1944     trips = ri.trexpfac
1945 elseif (ri.mode=4) ; HOV 2
1946     autoclass=(ri.votclass-1)*3 + 2
1947     trips = ri.trexpfac * 0.5
1948 elseif (ri.mode=5) ; HOV 3+
1949     autoclass=(ri.votclass-1)*3 + 3
1950     trips = ri.trexpfac * 0.3
1951 elseif (ri.mode=9) ; TNC - SOV, HOV2, and HOV3+
1952     ;add TNC trips
1953     if (ri.dorp=11)
1954         ; 11 - TNC non-AV, 1 passenger - assign to HOV 2 - 1 vehicle-trip per trip
1955         autoclass=11
1956         trips = ri.trexpfac * 1
1957     elseif (ri.dorp=21)
1958         ; 21 - TNC AV, 1 passenger - assign to SOV - 1 vehicle-trip per trip
1959         autoclass=10
1960         trips = ri.trexpfac * 1
1961     elseif (ri.dorp=12)
1962         ; 12 - TNC non-AV, 2 passengers - assign to HOV 3+ - 0.5 vehicle-trips per trip
1963         autoclass=12
1964         trips = ri.trexpfac * 0.5
1965     elseif (ri.dorp=22)
1966         ; 22 - TNC AV, 2 passengers - assign to HOV 2 - 0.5 vehicle-trips per trip
1967         autoclass=11
1968         trips = ri.trexpfac * 0.5
1969     elseif (ri.dorp=13 || ri.dorp=23)
1970         ; 13 - TNC non-AV, 3+ passengers - assign to HOV 3+ - 0.25 vehicle trips per trip.
1971         ; 23 - TNC AV, 3+ passengers - assign to HOV 3+ - 0.25 vehicle trips per trip
1972         autoclass=12
1973         trips = ri.trexpfac * 0.25 ;HOV 3+
1974     else
1975         autoclass=0
1976         trips = 0
1977     endif
```

FIGURE 3: TNC TRIPS AND DEADHEADING TO ASSIGNMENT VEHICLE TABLES

```
2073 ;step 28
2074     RUN PGM=MATRIX MSG='step 28 autotrips matrix'
2075
2076 mati[1]=autotrips.@per@.dbf, fields=otaz,dtaz,autoclass,@per@, pattern=ijm:v
2077 mato[1]=autotrips.@per@.mat, mo=1-9, dec=9*4
2078
2079 zones=@network._zones@
2080
2081 array Trips=9
2082
2083 @tolls.code1@ mw[01] = mi.1.01
2084 @tolls.code1@ mw[02] = mi.1.02
2085 @tolls.code1@ mw[03] = mi.1.03
2086
2087 @tolls.code2@ mw[04] = mi.1.04 + (40/60) * (mi.1.10 + mi.1.11 + mi.1.12) ;add TNC deadheading to SOV medium vot
2088 @tolls.code2@ mw[05] = mi.1.05
2089 @tolls.code2@ mw[06] = mi.1.06
2090
2091 @tolls.code3@ mw[07] = mi.1.07 + mi.1.10 ;add TNC SOV veh trips - high vot
2092 @tolls.code3@ mw[08] = mi.1.08 + mi.1.11 ;add TNC HOV2 trips - high vot
2093 @tolls.code3@ mw[09] = mi.1.09 + mi.1.12 ;add TNC HOV3+ trips - high vot
```

Appendix A detailed description of the purpose imputation process

The tests described below are based on the (mis)match between the trip destination location type (home, primary work, school, other) and the stated destination purpose category (home, primary work, school, other-non-missing, and missing, based on the variable **d_purpose_category**). The location type variable is based on the destination location being within 100 meters of the primary home, work, or school location, respectively. The relevant types of (mis)match are:

No mismatch: The destination location type (home, school, work, other) matches the stated destination purpose category.

Loc=Home / Purp=Not Home: The destination location type is home, but the stated destination purpose category is not home and is not missing.

Loc=Work / Purp=Not Work: The destination location type is primary work, but the stated destination purpose category is not primary work and is not missing.

Loc=School / P=Not School: The destination location type is school, but the stated destination purpose category is not school and is not missing.

Purp=Home / Loc=Not Home: The destination location type is not home, but the stated destination purpose category is home.

Purp=Work / Loc=Not Work: The destination location type is not primary work, but the stated destination purpose category is primary work.

Purp=School / Loc=Not School: The destination location type is not school, but the stated destination purpose category is school.

Purp=Missing: The stated destination purpose is missing (either an incomplete trips survey, or a stated purpose of “other” that did not have a text response that could be recoded to a more detailed purpose).

A series of sequential tests described below were run on the trips, with different tests applying to the different “mismatch” categories described above. In many cases, the tests also depend on other trips made during the person-day. For some rules, the description refers to “iterations”. Most of these tests were run though in sequence with five iterations across the entire set of tests. This is because changing the purpose in response to one test can have cascading effects for other tests that compare against the previous and next trips, so running several iterations allows any cascading effects to be tested.

Code 1: OK. Use as is: By default, the imputed purpose for the “OK” trips is the same as the stated purpose, although it may be changed by one of the subsequent tests (Test 7).

Codes 2–19 below apply to the Loc=X / Purp=Not X trips (where X=home, work or school), although in some cases they may affect other trip types if there is a “swap” of purposes between adjacent trip. In the descriptions below, the letter X is used to designate Home (H), Work (W) or School (S), when the same test can apply to all three of the location types. (In such cases, the



tests were applied separately to all of the X=Home trips, then all of the X=Work trips, and then all of the X=School trips.)

Code 2: Test 1: “change mode” purpose is ok: If the stated purpose is “change mode”, but the location is within 100 meters of home, work or school, check for reasonableness. A valid change mode stop satisfies one of the following two tests:

- The location type for the previous destination (or the origin if it is the first trip of the day) is X, the mode is Walk or Bike, and the mode for the next trip is Transit, or
- The location type for the next destination is X, the mode is Transit, and the mode for the next trip is Walk or Bike.

If it passes this test, set the imputed purpose to “Change mode”. About 75% of the change mode cases passed one of the tests, mainly the second one where the person gets off a transit vehicle at a stop very near location X and then walks or bikes to X. A large majority of these cases were trips that were split in the transit unlinking, so the stop location is based on GTF data.

Code 3: Test 2A-change purpose to match location type: The criteria are:

- There is no location type/purpose mismatch for either the preceding trip or the next trip, and neither the previous destination nor the next destination is at location X.

If it passes this test, the imputed purpose is set to X (Home, Work or School). There are many such cases in all of the data sets, mainly for X=Home, which are mostly cases of people reporting what they did at home rather than stating “Home” as a purpose. (Changes are planned for future releases of rMove to prevent this issue.)

Code 4: Test 2B- for last trip of day-change to home: This is a variation of Test 2A, applied only to X=Home trips that are the last trip of the day. The criteria are:

- It is the last destination of the person-day
- There is no location type/purpose mismatch for the preceding trip and the previous trip is not at location X (Home).

If it passes this test, the imputed purpose is set to X (Home). There are almost may such cases data—almost as many as for test 2A, which (again) seem to be mostly cases of people reporting what they did when they got home rather than stating “Home” as a purpose.

Code 5: Test 2C- Test 2A modified for missing purpose-change to match location type:

This is a variation of Test 2A that does not count adjacent trips with missing destination purpose as mismatches. The criteria are:

- For both the preceding and next trip there is no location type/purpose mismatch or else the purpose is missing, and neither the previous nor the next trip is at location X.

If it passes this test, the imputed purpose is set to X. Compared to Test 2A, fewer cases meet these criteria.

Code 6: Test 2D-Test 2B modified for missing-change to home: This is a variation of Test 2B, applied only to X=Home trips that are the last trip of the day, that does not count a preceding trip with missing destination purpose as a mismatch. The criteria are:

- It is the last destination of the person-day
- There is no location type/purpose mismatch for the preceding destination, or the purpose for the preceding trip is missing, and the previous trip is not at location X (Home).

If it passes this test, the imputed purpose is set to X (Home). Compared to Test 2B, fewer cases meet these criteria.

Code 7: Test 3A-previous trip at same location with longer dwell time - keep purpose as is: This test is to avoid assigning the same purpose X to two adjacent trips, and to account for the fact that there are several activity purposes that that could conceivably be performed very near (within 100 m) of the home or work location (e.g. shop, meal, social/recreation, errand, escort, work-related), The criteria are:

- It is not the first destination of the day, the destination location of the previous trip is also X with no purpose/location type mismatch, the dwell time at the destination is not greater than 60 min, the dwell time at the previous destination is at least 90 min, and the dwell time at the previous destination is at least 3 times as long as the dwell time at the current destination.

If it passes these criteria, the imputed purpose is set as the stated purpose. There are very few cases that pass.

Code 8: Test 3B-next trip at same location with longer dwell time - keep purpose as is:

This test is analogous to Test 3A but comparing to the next destination instead of the previous one.

The criteria are:

- It is not the last destination of the day, the destination location of the next trip is also X with no purpose/location type mismatch, the dwell time at the destination is not greater than 60 min, the dwell time at the next destination is at least 90 min, and the dwell time at the next destination is at least 3 times as long as the dwell time at the current destination.

If it passes these criteria, the imputed purpose is set as the stated purpose. As with test 3A, there are few cases that pass.

Code 9: Test 4-previous and next trips are opposite mismatch - swap around: This is the potential case where a respondent mixes up trip ends, so reports a Home-Shop-Home actual tour as a Shop-Home-Shop tour. The criteria are:

- It is not the first or last destination of the day, both the previous destination and the next destination have the opposite mismatch. Also, if there are destinations in the day that are two earlier or two later than the current one, then those do not have a purpose/location mismatch.

If it passes this test, then the imputed purposes for the previous and next destinations are set to the stated purpose of the current destination, and the imputed purpose for the current destination is set to the stated purpose for the previous destination. There are few cases that pass.

Code 10: Test 5: previous trip is opposite mismatch - swap around: This is a simpler version of the previous test where only two consecutive destinations have “swapped” purposes.

The criteria are:

- It is not the first trip of the day, the previous destination has the opposite mismatch (stated purpose is X but location type is not X), if it is at least the third destination of the day, there is not a purpose/location mismatch for the destination before the previous destination, and it is either the last destination of the day or the next destination has no purpose/location mismatch. (In other words, the surrounding destinations in the day, if there are any, have no purpose/location type mismatch.)

If these criteria are met, the imputed purpose for the previous destination is set to the stated purpose of the current destination, and the imputed purpose for the current destination is set to the stated purpose for the previous destination. For the Bay Area, there are about 400 such cases, mostly involving the home location.



Code 11: Test 6: next trip is opposite mismatch - swap around: This is the same as Test 5 but comparing to the next destination. The criteria are:

- It is not the last destination of the day, the next destination has the opposite mismatch (stated purpose is X but location type is not X), if there are at least two more destinations in the day, then there is not a purpose/location mismatch for the destination after the next destination, and it is either the first destination of the day or the previous destination has no purpose/location mismatch. (In other words, the surrounding destinations in the day, if there are any, have no purpose/location type mismatch.)

If these criteria are met, the imputed purpose for the next destination is set to the stated purpose of the current destination, and the imputed purpose for the current destination is set to the stated purpose for the next destination. For the Bay Area, there are about 100 such cases, all involving the home or work location.

Code 12: Test 7-purpose to change to match location type- all subsequent purposes in day shifted down one trip: This test is based on an example provide by SFCTA, where inserting a destination purpose and shifting all the purposes for subsequent destinations in the day to the next trip cause no mismatches. The criteria are:

- After the current X/O trip of the day, shifting the destination purposes for all subsequent destinations in the day to the next destination. If this does not cause any purpose/location type mismatches for any of the subsequent trips (a “Missing” purpose is not treated as a mismatch), then the test is passed.

If the test is passed set the imputed purpose for the current destination to X and set the imputed purpose for each subsequent destination to the stated purpose for the preceding destination. This test was revised a number of times to make sure that a last trip of the day at the home location would be retained, and that certain types of purpose shifts were given lower priority—meaning that they were only changed on later iterations of these rules if other rules had not fixed the problem first. There are about 600 trips with this imputation type in the Bay Area data. Note that this is one of the only tests that can change the imputed purpose for an “OK” trip that had no purpose/location type mismatch. Without having POI land use data to relate to the geo-locations, it is impossible to know if this recoding is correct, but in any event, it does not create a mismatch in type and purpose.

Code 13: Test 8-last destination of day at/near Home location with stay >3hr- change purpose to home: After the preceding tests, there were still many cases remaining where the last destination of the day was an overnight dwell at/near the home location, but the stated purpose was not home. The criteria are:

- The location type is Home, it is the last destination of the person-day, and the dwell time as at least 3 hours.

If it passes the test, change the imputed purpose to Home. There are about 400 such cases in the Bay Area data.

Code 14: Test 9-tests distance to usual location for purpose in case home/work/school have same usual location: There are some cases in the data where the primary home, work, and/or school locations can be the same, with the most common being people who primarily work from home (or within 100 m of home). In those cases, the coded destination location type is ambiguous. The criteria for this test are:

- The stated purpose is home and the distance to home is less than 200 m, OR the stated purpose is primary work and the distance to the primary workplace is less than 200 m,

OR the stated purpose is school and the distance to the school location is less than 200 m.

If it passes this test, there is assumed to be no purpose/location type mismatch, so set the imputed purpose to the stated purpose. There are very few such cases.

Code 15: Test 10-change work-related to work if at primary work location with no adj work purpose: To avoid creating a work-work trip, the criteria are:

- The location type is primary workplace, the stated purpose is “work-related” and neither the preceding or next destination (if they exist) have the stated purpose of primary work.

If it passes, the imputed purpose is set to primary work. There are very few such cases.

Code 16: Test 11-change purpose to match location type if it doesn’t create consecutive trip w same purpose: The last test for the X/O trips accepts the usual location type as the best guess at the purpose as long as it won’t create a Home-Home, Work-Work or School-School trip. The criteria are:

- Neither the preceding destination nor the next destination in the day, if they exist, have the stated purpose of X.

If this test is passed, set the imputed purpose to X. There are about 500 such trips in the Bay Area data.

Code 19: Tests 2–16 not passed – keep stated purpose – examine further: If none of the tests above are passed, then the imputed purpose is kept as the stated purpose. As tables 1-3 show there are only about 0.2% of all trips in this category.

The “examine further” text is added to the label to indicate that these would be good cases for agencies to examine visually, one-by-one, to see if any particular imputed purpose would make sense.

Codes 20--37 below apply to the Purpose=X, Location=Not X trips (X=Home, Work or School) trips only. In the descriptions below, the letter X is used to designate Home (H), Work (W) or School (S), when the test was applied to all of the location types. (In such cases, the tests were applied separately to all three types, they were applied first to all trips where X=Home, then all trips where X=Work, and finally all trips where X=School.

Code 20: Test 21-imputed Escort purpose: This test checks for likely “escort” trips dropping off or picking someone up from another person’s home, work, or school. In these cases, it is likely to be someone else’s home, work, or school location, not the respondent’s. The criteria are:

- The destination is not the last one of the day, the mode for the trip is the same as the mode for the next trip, the number of travelers for the trip is different from the number of travelers for the next trip, and the duration of stay at the location is 30 minutes or less.

If these criteria are met, the imputed purpose is “pick up/drop off passenger”. There are about 250 such cases in the Bay Area data.

Code 21: Test 2A-within 200 m of usual location- no adjacent trip with same purpose: Given that the GPS locations are not always precise, and that respondents may delete very short spurious trips near their destination rather than merging them, the distance limit of 100 meters to impute location type may be too strict in some cases. This test relaxes that limit some when the location is very near the corresponding usual location, but not within the 100 m limit. The criteria are:



- The stated destination purpose is home, primary work or school, and the corresponding home, primary work or school location, respectively, is within 200 m of the destination coordinates. Also, the destination location type for the preceding trip and the next trip are different from that of the current trip (so as not to create a home-home, work-work or school-school trip).

If it passes the test, the imputed purpose is set to the stated purpose. For the Bay Area, about 1,100 cases meet these criteria, with more than half for the work location. (Some work and school locations may cover an area larger than 100 m in themselves, and some people may make quick stops close to work that they delete.)

Code 22: Test 2B-within 300 m of usual location- no adjacent trip with same purpose:

This is the same as Test 2A above, but with the allowable distance extended to be in the range between 200 and 300 meters. Less than 100 cases meet the criteria, in the Bay Area, all for X=Home.

Code 23: Test 2C-within 500 m of usual location- no adjacent trip with same purpose:

This is the same as Test 2B above, but with the allowable distance extended to be in the range between 300 and 500 meters. As with 2B, less than 100 cases meet the criteria for the Bay Area, all for X=Home. (Note that for tests 2B and 2C, there are no cases for X=Work or X=School. The prior recategorization of some work and school trips work-related and school-related trips has pre-empted this check in some cases. This means that for very large school or work campuses, some valid work or school cases may be categorized as work-related or school-related trips.

Code 24: Test 3-Is an overnight stay away from home- out of region: Many overnight stays are more than 500 m from the respondent's home location. The criteria are:

- The destination is the last of the respondent's day, with a dwell time of at least 3 hours, and a distance of more than 500 m from the home location. Also, the location is outside the MPO region.

If it passes this test, assign a new code for the imputed purpose "overnight at a location outside the region", and a new code for the imputed purpose category "other person's home". About 250 trips pass this test in the Bay Area.

Code 25: Test 3-Is an overnight stay away from home- within the region: This is the same test as above, but with the overnight location within the region.

If it passes this test, assign a new code for the imputed purpose "overnight at a non-home location within the region". About 450 trips pass this test in the Bay Area.

Codes 31–36: Imputation based on stated destinations purposes of other trips in the survey data: If a destination purpose for a trip of type Purpose=X, Location=Not X could not be imputed based on multiple iterations of tests 1-3 above, or if the destination purpose is missing or other (Purpose=Missing), then a procedure was used based on other trips in the data. This procedure was applied only after multiple iterations of the rules above, and works as follows:

First, look at all the other destinations that the same respondent made during the survey period. Count the number of destinations visited that meet the following criteria:

- The stated destination purpose is not missing, is not "change mode", and is different from the stated purpose for the current destination (since that is a mismatch with the location type, and the distance from the current destination is less than 50 meters.

If there are multiple trips in the dataset that meet these criteria, then one is chosen at random using a Monte Carlo method. (The random number used is written to the trip file.) The

imputation type is set to 31 and the imputed purpose to the stated purpose of the randomly selected trip. About 2,400 trip purposes were imputed in this way for the Bay Area, including about 1,400 for missing purpose/incomplete trips. The random selection has the property that the distribution of imputed purposes at the location will tend towards the distribution of observed eligible purposes at that location.

If none of the respondent's destinations meet the criteria above, then the search radius is expanded from 50 m to 100 m, and then the same procedure is followed, but with the imputation type code now=32. There are about 800 such cases in the Bay Area data, with 500 for missing/incomplete trips.

If none of the respondent's destinations meet the criteria above, then the search radius is expanded once again from 100 m to 200 m, and then the same procedure is followed, but with the imputation type code now=33. There are about 1100 such cases in the Bay Area data, with 700 for missing/incomplete trips.

If no qualifying trips can be found from the respondent's own trips, then a similar search is done across trips made by other respondents in the survey sample. The criteria are very similar to those described above for the respondent's trips, with some variations:

- The stated destination purpose is not missing and is not "change mode", and the distance from the current destination is less than 50 meters. Trips with purpose = work are not allowed unless the person has reported work trips of their own during the survey period. Similarly, trips with purpose = school are not allowed unless the person has reported school trips of their own during the survey period.

If there are multiple trips in the dataset that meet these criteria, then one is chosen at random using a Monte Carlo method. The imputation type is set to 34 and the imputed purpose is set to the stated purpose of the randomly selected trip. Over 3,500 trip purposes were imputed in this way for the Bay Area, including about 2,400 for missing purpose/incomplete trips.

If none of the other respondents' destinations meet the criteria above, then the search radius is expanded from 50 m to 100 m, and then the same procedure is followed, but with the imputation type code now=35. There are about 600 such cases in the Bay Area data, with 450 for missing/incomplete trips.

Finally, if none of the other respondents' destinations meet the criteria above, then the search radius is expanded once again from 100 m to 200 m, and then the same procedure is followed, but with the imputation type code now=36. There are about 550 such cases in the Bay Area data, with 425 for missing/incomplete trips.

Code 37: Test 5-Work purpose not at primary work location, changed to work-related:

This test is more or less redundant now that some work purposes have been recategorized to work-related based on location prior to purpose imputation. There are only a handful of trips with this code.

Code 38: Test 6-School purpose not at school location, changed to school-related:

This test is also redundant now that some school trips have been recategorized as school-related based on location prior to purpose imputation. There are no cases in any of the regions that have this imputation code.



Code 39: No other visits to location - impute Other (no evidence): The final imputation code is for the Purpose=X, Location=Not X trips that do not pass any of the tests above. There are about 1,000 such trips in the Bay Area data, about 200 with imputed purpose=Home and 800 for which the purpose was originally missing. (Note that for the trips with Home purpose, it would be possible to impute a social visit purpose on the assumption that they are other people's homes.)

Imputing origin purposes

Trip origin purpose is imputed based on the imputed destination purpose of the preceding trip where possible.

Code 1: Copied from preceding destination in same day: For the second and subsequent destinations visited in each person-day, the imputed origin purpose is set equal to the imputed destination purpose for the preceding trip. This accounts for about 83% of trips in all regions.

Code 2: Copied from last destination in preceding day: If the preceding day has valid trips, the imputed origin purpose for the first trip of the day is set to the imputed destination purpose of the last destination from the preceding day. This is a further 14% or so of all trips.

Codes 3-6: First trip- Imputed based on origin location type: If it is the first day the person participated in the survey, or if there were no valid trips on the preceding day, the imputed origin purpose for the first trip is based on the origin location type (**o_location_type**) for that trip. The codes are 3 for Home, 4 for Work, 5 for School and 6 for Other. Of the 3% or so of trips that have this code, the large majority have origin at home, and some have origin location type Other (e.g. starting out of region or from someone else's home). Only a handful start the survey period at work or school.

Appendix B: Checks for Testing Compatibility between Activity Purpose and Parcel Land Use

Activity Purpose	Test 1 on parcel land use	Test 2 on parcel land use
Home	households > 0	households + any student enrollment > 0
Work to usual location	total employment > 0	total employment + households > 0
Work to other location	total employment > 0	total employment + households > 0
School- college	university enrollment > 0	univ. enrollment + education empl. > 0
School- k-12	K-12 enrollment > 0	K-12 enrollment + education empl. > 0
Meal	food service employment > 0	total employment + households > 0
Shopping	retail employment > 0	total employment > 0
Personal business	retail + service employment > 0	total employment > 0
Social/ recreation	households + service employment > 0	households + total employment + any enrollment > 0
Escort	households + total employment + any enrollment > 0	None
Other	households + total employment + any enrollment > 0	None
School-no usual location	households + total employment + any enrollment > 0	None



Appendix C: Coefficients Estimated for the Paid Rideshare (TNC) Utility in the SACOG Mode Choice Models

The coefficient values listed below are lines extracted from the F12 coefficient files for mode choice models that were estimated using DaySim with the new SACOG household travel data, and have been tested in application mode with SACOG inputs as well. In order to use these coefficients, it is necessary to (a) insert the

line *PaidRideshare_UseEstimatedInsteadOfAssertedCoefficients = true* in the configuration file, and (b) insert the lines below into the relevant F12 coefficient files that are used in running the model. These will work when running the latest DaySim code with the PaidRideshare mode set to be available, as these utility terms are included in the mode choice model code., using coefficient numbers that had previously been unused. (Note that the EscortTourModeModel does not have paid rideshare as an available mode, so that model is not affected.)

A brief description of the variables is below. In several cases, coefficients were constrained to 0 when the estimation results were highly insignificant.

tnc-const: this is the mode-specific constant, coefficient 80 in each model. The coefficient is negative for all models, particularly for work-based subtrips.

tnc-nocars: This is a dummy variable for households that own 0 vehicles. This is coefficient 81 in all models except for work-based subtrips, and has a positive, fairly significant value in all those models .

tnc-carltd: This is a dummy variable for households that own 1 or more vehicles, but fewer vehicles than driving age adults—a car competition variable. It is coefficient 82 in all models except work-based subtrips, and is constrained to 0 for school tours. It has positive values for the work tour, other tour, and trip models, about one third to one half the size of the 'nocars' effect.

tnc-und35: A dummy variable for people age under 35, as coefficient 83 in the work tour, work-based subtrip, and trip models. It is marginally significant in the work model, and 0 in the other models.

tnc-a65up: A dummy variable people age 65 or older, as coefficient 84 in the work tour, and trip mode models and coefficient 83 in the Other Tour model. It is negative and significant in the Other Home Based Tour model, but 0 in the other models.

tnc-ag1835: A dummy variable for people age 18-35, coefficient 83 in the school tour model. It is positive and fairly significant.

tnc-ageu18: A dummy variable for people under age 18, coefficient 84 in the school tour mode. It was constrained to a value of -5.0, as nobody under 18 used TNC for school tours in the data.

tnc-odens: A linear variable applied to the origin parcel/micrzone buffer 2 households plus total employment plus university enrollment, with no upper cap on the value. This is coefficient 85 in all the models except Other Tours, where it is number 84. It is generally significant, with values in the range of 0.25 E-04 to 0.6 E-04. For work-based subtrips, the value could not be estimated, but was constrained at 0.25 E-04. This variable is a proxy for TNC availability and waiting time at the trip origin.

tnc-ddens: A similar linear variable as tnc-odens, but measured around the destination parcel/microzone. This is variable 86 or 85, depending on the model. It was constrained to 0 in all models after finding that it was less significant than the origin effect, and that collinearity issues in the data were causing lower estimates for the origin effect.

tnc-incu25: A dummy variable for households with income under \$25,000. This is coefficient 87 for the work mode, school mode and trip models, and coefficient 86 for the other models. It is positive for the work tour, other tour, and trip models, and negative for the school tour model. The general finding of high TNC use for low income households was not expected, but may reflect weekday behavior more than weekend TNC use.

tnc-inc100: A dummy variable for households with income \$100,000 or higher. This is coefficient 88 for the work tour, school mode, and trip models, and coefficient 87 in the other models. It is negative for the other tour and trip models, and constrained to 0 elsewhere. If users from other regions are transferring these models and do not think the income effects seem realistic, they can be set to 0.

tnc-mealtr: A dummy variable for meal tours. This is coefficient 88 in the other tour model, and is positive. (A similar variable for social/recreation tours was near 0.)

tnc-shoptr: A dummy variable for shopping tours. This is coefficient 89 in the other tour model, and is highly negative.

Tnc-wktour and tnc-sctour: Dummy variables for work and school tours, coefficients 89 and 90 in the trip model. Both are negative, meaning that mixed mode tours with other modes being used on TNC tours are more common on work and school tours.

Tnc-estour and tnc-shtour: Dummy variables for escort and shopping tours, coefficients 91 and 92 in the trip model. The escort one is constrained at -5, as there are no escort tours or trips by TNC.

Work Tour Mode

80	tnc-const	F	-4.29883879601	1.04448823110
81	tnc-nocars	F	.894698843386	.434622595389
82	tnc-carltd	F	.241969613772	.374680308198
83	tnc-und35	F	.302655257650	.301435816222
84	tnc-a65up	T	.000000000000	.000000000000
85	tnc-odens	F	.237922093646E-04	.101746969378E-04
86	tnc-ddens	T	.000000000000	.000000000000
87	tnc-incu25	F	.932053361216	.520851501671
88	tnc-inc100	T	.000000000000	.000000000000

School Tour Mode

80	tnc-const	F	-3.88777487608	1.56359381801
81	tnc-nocars	F	1.87788342066	1.62710273709
82	tnc-carltd	T	.000000000000	.000000000000
83	tnc-ag1835	F	2.68728051983	1.49645948563
84	tnc-ageu18	T	-5.000000000000	.000000000000
85	tnc-odens	F	.590612008710E-04	.325975807193E-04
86	tnc-ddens	T	.000000000000	.000000000000
87	tnc-incu25	F	-1.27913838926	1.19648215172
88	tnc-inc100	T	.000000000000	.000000000000

Other Home Based Tour Mode

80	tnc-const	F	-3.26437612658	.497253021515
81	tnc-nocars	F	1.33014773057	.638839684283
82	tnc-carltd	F	.463723794664	.503277262975
83	tnc-a65up	F	-1.05928546050	.548617219712
84	tnc-odens	F	.257032004578E-04	.140625094865E-04
85	tnc-ddens	T	.000000000000	.000000000000
86	tnc-incu25	F	.503316885635	.609959998031
87	tnc-inc100	F	-.676731461830	.490027161763
88	tnc-mealtr	F	.398185638837	.509555333970
89	tnc-shoptr	F	-2.36922984640	.811557804984

Work Based Subtour Mode

80	tnc-const	F	-6.79024470461	1.53416323085
81	tnc-sovtow	T	.000000000000	.000000000000
82	tnc-wbktow	T	.000000000000	.000000000000
83	tnc-und35	T	.000000000000	.000000000000



84	tnc-tnctow	T	.000000000000	.000000000000
85	tnc-odens	T	.250000000000E-04	.000000000000
86	tnc-incu25	T	.000000000000	.000000000000
87	tnc-inc100	T	.000000000000	.000000000000

Trip Mode

80	tnc-const	F	-2.45757631033	.451663400559
81	tnc-nocars	F	.839278813889	.488833364231
82	tnc-carltd	F	.405220976676	.454395896716
83	tnc-und35	T	.000000000000	.000000000000
84	tnc-a65up	T	.000000000000	.000000000000
85	tnc-odens	F	.421161491812E-05	.977438141473E-05
86	tnc-ddens	T	.000000000000	.000000000000
87	tnc-incu25	F	1.38276043387	.520139265889
88	tnc-inc100	F	-.522922485742	.425318172503
89	tnc-wktour	F	-2.17244385304	.396171146712
90	tnc-sctour	F	-1.89148010081	.673767076917
91	tnc-estour	T	-5.000000000000	.000000000000
92	tnc-shtour	T	.000000000000	.000000000000