LOCATION CHOICE MODEL ESTIMATION

SOUNDCAST: ACTIVITY-BASED TRAVEL FORECASTING MODEL

Puget Sound Regional Council 2015

John L. Bowman, Ph. D. , Mark Bradley, and Suzanne Childress

INTRODUCTION

This document presents the estimation results for the usual work location, usual school location, the tour destinations, and intermediate stop location for all purposes.

The unifying aspect of all the models covered in this document is that they model location choice. The dependent variable is the parcel, and the reader is referred to that memo for a discussion of issues related to modeling at the parcel level of detail.

All these models have a single anchor point, the tour origin, from which impedance is measured. That is, impedance is measured from the tour origin, to the destination, and back to the origin, without direct consideration of the impedance for stops on the way to and from the tour destination. For the usual location models and most tours, the anchor is the person's home; for work-based tours, it is the work location. This simplifies considerably the measurement of impedance, and as a result the model's impedance variables and the sampling of alternatives are much simpler than in the intermediate stop model.

Intermediate stops include all stops on the way to and from the primary destination of a tour, but do not include the primary destination itself. This model occurs within the DaySim portion of the model system, occurring at model step 4.2, as shown in bold in **Figure 1**. The exact number and purpose of stops for each tour (model 4.1) is modeled at the tour level. Then, within each tour, the stops are modeled one-by-one, first for stops before the tour destination, and then for stops after the tour destination. Stops before the tour destination are modeled in reverse temporal sequence. First the stop's location (4.2), and then its trip mode (4.3), and finally the 10-minute time period of the arrival at the tour destination (4.4) are modeled. These results also determine the time period in which the trip from the stop location begins, since the trip mode and travel level of service are known. This continues, constructing the trip chain from the tour primary destination to the tour origin in reverse chronological sequence until the model predicts no more stops (at which point, the "final" trip between the "last" stop and the tour origin is modeled). The reason for modeling in reverse chronological sequence for the first half tour is the hypothesis that people aim to arrive at the primary destination at a particular time, and adjust their tour departure time so as to enable completion of the desired intermediate stops. After the trip chain for the first half-tour is modeled, the trip chain for the second half-tour back to the tour origin is similarly modeled, but this time in regular chronological order.

Figure 1—DaySim models (numbered) within the program looping structure

Begin

{Read run controls, model coefficients, TAZ data, LOS matrices,
population controls, and Parcel data into memory}

{Draw a synthetic household sample if specified}

{Pre-calculate destination sampling probabilities}

{Pre-calculate (or read in) TAZ aggregate accessibility arrays}

{Open other input and output files}

{Main loop on households}

{Loop on persons in HH}

{Apply model **1.1 Work Location for workers**}

{Apply model **1.2 School Location for students**}

{Apply model **1.1 Work Location for students**}

{End loop on persons in HH}

{Apply model **1.3 Household Auto Availability** }

{Loop on all persons within HH}

{Apply model **2.1 Activity Pattern** (0/1+ tours and 0/1+ stops)
and model **2.2 Exact Number of Tours** for 7 purposes}

{Count total home-based tours and assign purposes}

{Initialize tour and stop counters and time window for the person-day before looping on tours}

{If there are tours, loop on home-based tours within person in tour priority sequence,
with tour priority determined by purpose and person type}

{Increment number of home-based tours simulated for tour purpose (including current)}

{Apply model **3.1 Tour destination**}

{If work tour, apply model **3.2 Number and purpose of work-based subtours**}

{Loop on predicted work-based sub tours and insert then tour array after current tour}

{Apply model **3.3 Tour mode**}

{Apply model **3.4 Tour primary destination arrival and departure times**}

{Loop on tour halves (before and after primary activity)}

{Apply model **4.1Half tour stop frequency and purpose**}

{Loop on trips within home-based half tour (in reverse temporal order for 1st tour half)}

{Increment number of stops simulated for stop purpose (including current)}

{Apply model **4.2 Intermediate stop location**}

{Apply model **4.3 Trip mode**}

{Apply model **4.4 Intermediate stop departure time**}

{Update the remaining time window}

{End loop on trips within half tour}

{End loop on tour halves}

{End loop on tours within person}

{Write output records for person-day and all tours and trips}

{End loop on persons within household}

{End loop on Households}

{Close files}

{Create usual work location flow validation statistics}

End.

AVAILABILITY RESTRICTIONS AND ALTERNATIVE SAMPLING

Modeling the choice of a particular parcel makes the universal choice set very large, and presents challenges to appropriately limit the number of alternatives considered when simulating choices.

The reduction of the universal choice set involves two conceptually different methods: availability constraints and sampling of alternatives. The first method removes from the universal choice set those alternatives that the decision-maker would not even consider in making the decision, because they don't accommodate the desired activity or because they are too far away. Each parcel is assigned purpose-specific sizes; for a given purpose, if a parcel has zero size, then it is deemed unavailable. A parcel is also deemed unavailable if reaching it requires more than 125% of the maximum travel time observed in the survey sample for similar tours.

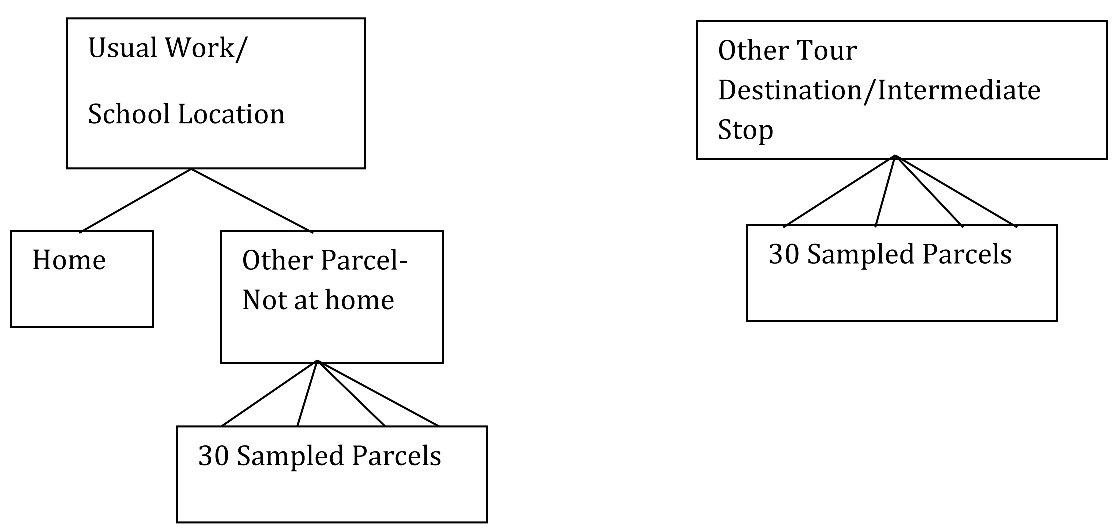
The second method involves taking the remaining alternatives, that would all be reasonable alternatives for the decision maker to consider, and drawing a sample of them to actually use in simulating the choice. This is simply a procedural technique to reduce the computational burden of the model. The employed sampling technique is called importance sampling with replacement. The available alternatives are sampled in a way that allows the probability of being drawn into the sample to be calculated for each drawn alternative. Statistical procedures are then used during model estimation and application to allow the sample to represent the entire set of available alternatives without biasing the results. **Appendix 1** describes the sampling procedure in detail.

MODEL STRUCTURE AND ESTIMATION DATA

The model structure of **Figure 2** imposes an assumed hierarchy of choice among the models, determining what is known and unknown at each level. For the usual location models, auto ownership is assumed to be unknown, based on the assumption that auto ownership is conditioned by work and school locations of household members, rather than the other way around. For the tour destinations, auto ownership levels are treated as given, and affect location choice. For university and grade school students who also work, the usual school location is known when usual work location is modeled; for other workers who also go to school, the work location is known when usual school location is modeled. For the tour destination models, all usual locations are known.

For the work location model, the home location is treated as a special location, because it occurs with greater frequency than any given non-home location, and size and impedance are not meaningful attributes. As a result, both of these models take the nested logit form, with all non-home locations nested together under the conditioning choice between home and non-home, as shown in **Figure 2**.

Figure 2: Structure of the location choice models



The usual work location model was estimated using all survey person records of employed persons, with the reported usual work location as the dependent variable. Similarly, the school location model uses all survey person records of students, with the reported usual school location as dependent variable. Some persons are both employed and student, so they provide observed outcomes for both models. In the estimation data, all workers have a usual work location and all students have a usual school location (counter to our expectation that some workers would not have a usual location), so the model does not have an alternative called "no usual location".

Because a large majority of work tours go to the usual work location, the work tour destination model has this as a special alternative. Therefore, the model is nested, with all locations other than the usual location nested together under the conditioning binary choice between usual and non-usual, as shown in **Figure 2**. In addition, because in the survey sample there were almost no work-based work tours, or work tours by persons with usual work location at home, these alternatives are excluded from consideration.

Since most work tours go to the usual location, there are relatively few data records to provide good parameter estimates of the factors affecting choice among the "non-usual" alternatives. Therefore, the work destination choice model was estimated with a combined data set including all work tour records and also all person records of persons with a non-home usual location. The standard method of combining data from multiple sources was used. This includes the estimation of separate scale of the two data sets and, since ALOGIT was used for estimation, the specification of dummy nodes to accommodate the scale differences. For most utility variables, it was assumed that the effect is the same in the two data sets, but some distinct parameters were estimated for work tours, such as the attractiveness of the usual location, and the effects of distance and street connectivity.

Nearly all school tours go to the usual school location. Therefore, there is no school tour destination choice model. When students with a usual location have a school tour, it is always assigned to the usual location. School tours are excluded from the day pattern choice set of students having home as the usual school location.

Since there are no modeled usual locations for activities other than work and school, the destination choice model of all remaining purposes is simply a multinomial logit model.

UTILITY FUNCTION

The utility function of each regular location alternative includes a regular utility component and a size function component. Equation 1 shows the form of the utility function, with size function included:

$$V_{in} = \sum_{k=1}^{K^v} \beta_k x_{ink} z_{nk} + \mu' \ln \sum_{k=K^v+1}^{K^v+K^s} \exp(\beta_k) x_{ink} z_{nk} \qquad (1)$$

where:

V_{in} is the systematic utility of parcel alternative *i* for tour *n*,

K^v is the number of utility parameters,

K^s is the number of size parameters,

β_k, $k = 1, 2, \ldots, K^v + K^s$ are the utility and size parameters,

x_{ink} is an attribute of parcel alternative *i* for tour *n*,

z_{nk} is a characteristic of tour *n*,

μ' is a scale parameter measuring correlation among elemental activity opportunities within parcels (1—no correlation, 0+--high correlation)

Table 1 provides an overview of the variables (alternative attributes and person/tour characteristics) used in the utility and size functions to explain choice in the models. The left-hand column lists the alternative attributes for the binary choice (special vs. regular alternative) as well as for the conditional MNL choice among regular parcel alternatives. To the right is a column for each of the four models, and in each model's column are the characteristics associated with each of the applicable attributes.

Table 1—Utility function variables in the tour location choice models

Attributes	Usual work location	Work tour destination	Usual school location	Non-work tour destination
Binary choice	**Home vs other**	**Usual vs other**	**Home vs other**	**not applicable**
Constants	by person type*	By person type* tour type	By person type* HH size	
Disaggregte logsum among regular locations	Yes	Yes	yes	
Conditional MNL choice among regular locations				
Disaggregate mode choice logsum to destination	Yes	Yes	Yes	Yes
Piecewise linear driving distance function	For fulltime workers		For children under age 16	By Purpose Priority Pattern type
Natural log of driving distance	For other then fulltime workers by person type* income	By person type* tour type	For persons age 16+ by person type*	By tour type income person type* time available
Distance from usual work location		Yes	for not student aged	
Distance from usual school location	for student aged	for student aged		Yes
Aggregate mode-dest logsum at destination	By person type	By person type	By person type	By purpose
Parking and employment mix	For daily parking in parcel and in TAZ	for daily parking in parcel and TAZ		For hourly parking in parcel and TAZ by car availability

Ratio of neighborhood nodes with 3 or 4 entering links	Yes	By car availability		By car availability
employment, enrollment and households by category:	by person type income	By person type Income	by person type	by purpose (and by 'kids in household' for escort tours)
--Zonal density	--yes	--yes	--yes	--yes
--Parcel size	--yes	--yes	--yes	--yes
Person type categories in the models	full-time worker part-time worker not full- or part-time	full-time worker part-time worker not full- or part-time	child under 5 child 5 to 15 child 16+ university student not student aged	full-time worker part-time worker retired adult other adult university student child 16+ child 5 to 15 child under 5

MODEL ESTIMATION RESULTS FOR TOUR LOCATION MODELS

Tables 2 through **5** show the estimated parameters for all four of the models. Within each table, the parameters appear in the same order as the variables listed in **Table 1**.

In the binary choice between the special alternative and all other possible locations, an alternative specific constant captures the basic tendency to choose one or the other, and dummy variables capture significant differences in this effect among various population segments. The logsum variable from the regular alternatives captures the effect of level of service on this basic choice. In all three cases the parameter is larger than zero, but quite small; that is, the tendency to choose home as the usual location, or to choose the usual location for the work tour, is barely effected by level of service. In the case of the work tour choice, at parameter values close to zero the likelihood function is very flat, so it is difficult to accurately estimate its exact size. Therefore, it is constrained to a specific small value.

Two important variables in all four models are the disaggregate mode choice logsum and network distance. The logsum represents the expected maximum utility from the tour mode choice, and captures the effect of transportation system level of service on the location choice. Distance effects, independent of the level of service, are also present to varying degrees depending on the type of tour being modeled. Since the logsum variable and distance are highly correlated it was difficult in estimation to separately identify the magnitude of their parameters. Therefore, the logsum parameters are constrained to the value one, representing the simple assumption of a multinomial logit form for the joint choice of mode and destination. In nearly all cases, sensitivity to distance declines as distance increases; in some cases this is captured through a logarithmic form of distance. In other cases, where there is plenty of data to support a larger number of estimated parameters, a piecewise linear form is used to more accurately capture this nonlinear effect.

In most cases the models include an aggregate mode-destination logsum variable at the destination. A positive effect is interpreted as the location's attractiveness for making subtours and intermediate stops on tours to this location. A mix of parking and employment, at both the zone and parcel level, as well as street connectivity in the neighborhood, attract workers and tours for non-work purposes. Also, as in the case of intermediate stops, parcel size variables and TAZ-level density variables affect location choice.

Shadow Pricing in Application

In application, the work location and school location choice models do not constrain the choice set to alternatives that have not yet been chosen. In other words, more workers can select a parcel than there are jobs in that parcel. To resolve this inconsistency between workers and jobs, Daysim has a shadow pricing procedure that works on each iteration of a model run in each zone. If a zone is over-selected a negative term will be added to the utility, so that in the next iteration it is less attractive. We have found that workers and jobs are usually fairly well balanced after two to three iterations of running this procedure.

Table 3—Usual Work Location Estimation Results

Alternative Attribute	Person Type	Est.	T-stat
Sampling adjustment factor for estimation		0.98	96.3
Home location	constant	-3.149	-0.2
Home location	PT worker	3.497	0.8
Home location	child or univ. stud.	-5.000	constr*
Home location	female	-0.164	-0.2
Nest parameter		0.25	1.5
Mode choice logsum	FT worker	0.002	1.4
Mode choice logsum	PT worker	0.004	constr
Mode choice logsum	not FT/PT worker	0.004	constr
One-way drive dist--0-3.5 mi	FT worker	-3.325	-8.4
One-way drive dist--3.5-10 mi	FT worker	-1.232	-10.6
One-way drive dist--10+ mi	FT worker	-0.754	-25.2
Nat log (1 + one-way drive dist)	PT worker	-3.581	-22.1
Nat log (1 + one-way drive dist)	not FT/PT worker	-4.421	-10.7
Nat log (1 + one-way drive dist)	female	-0.744	-7.5
Nat log (1 + one-way drive dist from school)	child or univ. stud.	-2.407	-4.4
Aggr. mode-dest logsum at dest	FT worker	-0.131	-4.2
Aggr. mode-dest logsum at dest	PT worker	-0.126	-2.1
Aggr. mode-dest logsum at dest	not FT/PT worker	0.099	0.7
Total Employment in Buffer	FT Worker, Income>75K	0.285	11.3
Size function scale		0.900	constr
size: service empl. in parcel buffer		0.310	
size: education empl. in parcel		0.679	
size: restaurant empl. in parcel		-0.419	
size: retail empl. in parcel		0.644	

Alternative Attribute	Person Type	Est.	T-stat
size: medical empl. in parcel		0.000	
size: industrial empl. in parcel		0.000	
District Pair Constants			
Home Kitsap- Work CBD		1.261	5.0
Home East – Work East		0.654	5.4
Home Kitsap – Work Not Kitsap		-1.425	-7.2
Home East – Work CBD		0.429	2.7
Summary statistics			
Number observed choices		3385	
Log likelihood w coeffs=0		-16636	
Final Log likelihood		-16065	
Rho squared		0.03	

*constr means the coefficient value was constrained during the estimation process

Table 4—Work Tour Destination Estimation Results

Alternative Attribute	Person/Tour Characteristics		Est.	T-stat
Sampling adjustment factor for estimation			1.0000	
Usual location	constant		30.0	const
Usual location	PT worker		-4.83	5.9
Usual location	child or univ. stud.		1.80	0.63
Usual location	pattern has 2+ work tours	primary tour	-4.79	-3.87
Usual location	pattern has intermediate work stop(s)		-12.44	-24.0
Usual location		secondary tour	-8.78	-7.60
Dest choice logsum (in usual location vs other choice)			.1884	40.1
Variables in the Not Usual Alternative Set				
Mode choice logsum	FT worker	tour dest.	0.69	24.65
Mode choice logsum	PT worker		0.78	16.32
Mode choice logsum	not FT/PT worker		1.52	constr*
Log distance from origin	Retired Adult		-3.71	constr
Log distance from usual work			-1.85	constr
Log distance from school	Student		2.58	2.91
Mix of daily parking & (empl+stud) in parcel buffer): ln(1+prkgdens*(empldens+studdens)/ (prkgdens+empldens+studdens)), (dens in units/Msqft)	HH has Cars>=Drivers		-0.69	-6.18
dens of education empl in parcel buffer			0.11	constr
dens of education empl in parcel buffer			0.11	constr
dens of households in parcel buffer			-0.12	constr
dens of industrial employment in parcel buffer			-0.12	constr
Size function scale			.2247	constr

size:education empl. in parcel	1.66	2.48
size: restaurant empl. in parcel	-3.97	-4.58
size: office empl. in parcel	-4.35	-6.48
size: retail empl. in parcel	-2.55	-3.99
size: service empl in parcel	-2.70	-3.53
size: medical empl. in parcel	-1.69	-2.77
size: industrial empl. in parcel	-5.05	-7.80
	0.0	
size: university students in parcel		constr

*constr means the coefficient value was constrained during the estimation process

Likelihood with Zero Coefficients = -25818.2662

Final value of Likelihood = -5073.6890

"Rho-Squared" w.r.t. Zero = .8035

Table 5—School Location Estimation Results

Alternative Attribute	Person Characteristic	Est.	T-stat
Sampling adjustment factor for estimation		0.696	31.5
School at Home Alternative constant		-7.11	-6.49
School at home	Adult	-5.00	constr*
Mode choice logsum	child age <5		
Mode choice logsum	child age 5-15		
Mode choice logsum	driving age stud.	3.2	18.14
Mode choice logsum	univ. stud.	0.8	8.01
Mode choice logsum	adult not univ. stud.	0.9	9.17
One-way drive dist--0-1 mi (10s of mi)	child age <5	-0.9	-0.16
One-way drive dist--1-5 mi (10s of mi)	child age <5	-5.0	-6.80
One-way drive dist--5+ mi (10s of mi)	child age <5	-2.2	-6.97
One-way drive dist--0-1 mi (10s of mi)	child age 5-15	-13.4	-6.25
One-way drive dist--1-5 mi (10s of mi)	child age 5-15	-8.3	-26.55
One-way drive dist--5+ mi (10s of mi)	child age 5-15	-1.5	-13.92
Aggr. mode-dest logsum at dest	child age <5	-0.9	-4.26
Aggr. mode-dest logsum at dest	driving age stud.	0.51	3.39
Total Employment in Buffer	child age <5	.79	5.8
K-8 Students in Buffer	child age 5-15	.001	0.3
Size function scale		0.3	constr
size: education empl. in parcel	child age <5	-5	constr
size: total empl in parcel	child age <5	-2	constr

size: # households in parcel*10	child age <5	-5	constr
size: K-8 enrollment in parcel	child age <5	-3	
size: education empl. in parcel	child age 5-15	0	constr
size: 9-12 enrollment in parcel	child age 5-15	-5	constr
size: K-8 enrollment in parcel	child age 5-15	-0.22	-0.36
size: education empl. in parcel	driving age stud.	0	constr
size: 9-12 enrollment in parcel	driving age stud.	3.1	4.1
size: education empl. in parcel	adult or univ. stud.	0	constr
size: University enrollment in parcel	adult or univ. stud.	7.7	12.5
Nest		0.95	constr
Summary statistics			
Number observed choices	1770		
Log likelihood w coeffs=0	-6786.5		
Final Log likelihood	-3934.7		
Rho squared	.4202		

Table 6. Non-work/Non-school location

Alternative Attribute	Person/Tour Characteristics	Est.	T-stat
Sampling adjustment factor for estimation		1	
ln(1- travel time o to d) / available minutes		9.34	8.79
Mode choice logsum	escort	0.64	16.16
Mode choice logsum	personal business	0.16	4.5
Mode choice logsum	shopping	0.21	4.03
Mode choice logsum	meal	0.11	1.56
Mode Choice logsum	social/rec	0.08	1.02
One-way drive dist--0-1 mi (10s of mi)	escort	-8.98	-5.53
One-way drive dist--1-3.5 mi (10s of mi)	escort	-4.24	-12.47
One-way drive dist--3.5-10 mi (10s of mi)	escort	-1.89	-10.01
One-way drive dist--0-1 mi (10s of mi)	personal business	-15.59	-9.53
One-way drive dist--1-3.5 mi (10s of mi)	personal business	-4.62	-12.47
One-way drive dist--3.5-10 mi (10s of mi)	personal business	-3.11	-20.47
One-way drive dist--10+ mi (10s of mi)	personal	-0.92	-12.39

	business			
One-way drive dist--0-1 mi (10s of mi)	shopping		-22.29	-13.48
One-way drive dist--1-3.5 mi (10s of mi)	shopping		-8.06	-20.34
One-way drive dist--3.5-10 mi (10s of mi)	shopping		-3.81	-20.21
One-way drive dist--10+ mi (10s of mi)	shopping		-1.32	-10.93
One-way drive dist--0-1 mi (10s of mi)	meal		-22.37	-11.06
One-way drive dist--1-3.5 mi (10s of mi)	meal		-6.36	-12.29
One-way drive dist--3.5-10 mi (10s of mi)	meal		-3.61	-15.32
One-way drive dist--10+ mi (10s of mi)	meal		-0.89797	-7.51
One-way drive dist--0-1 mi (10s of mi)	social/recreation		-23.6239	-10.12
One-way drive dist--1-3.5 mi (10s of mi)	social/recreation		-3.50887	-5.86
One-way drive dist--3.5-10 mi (10s of mi)	social/recreation		-2.99452	-12.8
One-way drive dist--10+ mi (10s of mi)	social/recreation		-0.94385	-7.93
Nat log (1 + one-way drive dist (10s of mi))	work based tour		-2.88	-6.98
Nat log (1 + one-way drive dist (10s of mi))		HH inc<$15K	0.47	1.42
Nat log (1 + one-way drive dist (10s of mi))		HH inc unreported	0.27	2.29
Nat log (1 + one-way drive dist (10s of mi))		nonworker age 65+	0.06	0.57
Nat log (1 + one-way drive dist (10s of mi))		univ. stud.		
Nat log (1 + one-way drive dist (10s of mi))		child age 5-15	-0.18	-0.91
Nat log (1 + one-way drive dist (10s of mi))		child age <5	-0.4	-2.49
Nat log (1 + one-way drive dist from school (10s of mi))	home based tour		-0.77	-4.47
Nat log (1 + one-way drive dist from work (10s of mi))	home based tour		-0.35	-3.55
Aggr. mode-dest logsum at dest	escort		0.1648	2
Aggr. mode-dest logsum at dest	personal business		0.0206	0.4
Aggr. mode-dest logsum at dest	shopping		0.1892	3.1
Street connectivity: (# 3 & 4 link nodes)/(# 1,3,4-link nodes) within a qtr mile		HH has no car	0.729	0.7
Street connectivity: (# 3 & 4 link nodes)/(# 1,3,4-link nodes) within a qtr mile		1+ cars per driver	0.2101	1.8
dens of educ empl	personal business		-0.921	-12.39
dens of office empl	personal business		0.045	2.03
dens of service empl	personal business		0.007	0.32
dens of households	personal business		-0.002	-0.07
dens of univ enroll	personal business		-0.003	-0.27
dens of educ empl	shopping		-0.2	-9.73
dens of retail empl	shopping		0.45	17.2
dens of office empl	social/recreation		-0.10325	-2.63
dens of service empl	social/recreation		0.120238	2.43
dens of households	social/recreation		-0.26272	-4.6

dens of uni students	social/recreation	6.21E-02	3.25
Size function scale		0.478	61.23
size: education empl. in parcel	escort, HH w/o kids	0	-6.26
size: restaurant empl. in parcel	escort, HH w/o kids	-2.98369	-6.8
size: gov empl. in parcel	escort, HH w/o kids	-2.19	-4.23
size: office empl. in parcel	escort, HH w/o kids	-1.42	-3.93
size: retail empl. in parcel	escort, HH w/o kids	-1.37	-2.49
size: service empl. in parcel	escort, HH w/o kids	-0.8	-9.49
size: medical empl. in parcel	escort, HH w/o kids	-4.81	-23.06
size: industrial empl. in parcel	escort, HH w/o kids	-12.33	2.77
size: # households in parcel	escort, HH w/o kids	0.81	-6.26
size: education empl. in parcel	escort, HH w kids	-1.6	-5.6
size: gov empl. in parcel	escort, HH w kids	-2.31	-4
size: office empl. in parcel	escort, HH w kids	-1.32	-8
size: retail empl. in parcel	escort, HH w kids	-1.49	-8.8
size: service empl. in parcel	escort, HH w kids	-1.1	-9.4
size: medical empl. in parcel	escort, HH w kids	-3.57	-5.3
size: industrial empl. in parcel	escort, HH w kids	-13.75	-7.8
size: # households in parcel	escort, HH w kids	0	constr
size: education empl. in parcel	personal business	-3.17	0.231337
size: restaurant empl. in parcel	personal business	-5.89	0.444016
size: office empl. in parcel	personal business	-2.27	0.135934
size: retail empl. in parcel	personal business	-3.81	0.257259
size: service empl. in parcel	personal business	-3.76	0.21229
size: medical empl. in parcel	personal business	0	0
size: industrial empl. in parcel	personal business	-6.31	0.268066

size: # households in parcel	personal business	-12.85	0.340744
size: K-12 enrollment in parcel	personal business	-46.2	60684.06
size: restaurant empl. in parcel	shopping	-11.58	-8.7
size: office empl. in parcel	shopping	-7.2	-24.35
size: retail empl. in parcel	shopping	0	constr
size: service empl. in parcel	shopping	-7.82	-15.26
size: restaurant empl. in parcel	meal	0	constr
size: office empl. in parcel	meal	-25.04	475.7423
size: total empl. in parcel	meal	-13.31	0.59464
size: # households in parcel	meal	-15.52	1.184757
size: education empl. in parcel	social/recreation	-1.44	constr
size: restaurant empl. in parcel	social/recreation	0	constr
size: retail empl. in parcel	social/recreation	-4.44	-8.79
size: service empl. in parcel	social/recreation	-3.99	-7.08
size: medical empl. in parcel	social/recreation	-2.12	-4.66
size: # households in parcel	social/recreation	-9.74	-20.41
size: University enrollment in parcel	social/recreation	-5.66	-3.22
size: K-12 enrollment in parcel	social/recreation	-3.38	-5.32
Summary statistics			
Number observed choices		12339	
Log likelihood w coeffs=0		-39456.7	
Final Log likelihood		-22828.3	
Rho squared		0.4214	

BASIC FEATURES OF THE INTERMEDIATE STOP MODEL

What is known and not known when location is modeled. At the time that a particular stop's location is modeled, information about the tour (origin, destination, time period arriving and departing the primary destination, and tour mode are known, and can be used to explain the location choice. The number of stops in each half-tour and their purposes are known. Additionally, details about any stops nearer to the primary destination are also known, including the location, trip mode, and the 10-minute time period of departure toward the tour destination (or arrival from the tour destination on the second half-tour).

However, at the time a stop's destination is modeled, several things are NOT known. These include the trip mode for the trip between this stop and the stop nearer to the tour destination, and the departure and arrival times of that trip, which will be modeled immediately after this stop's location. The arrival time from the stop nearer to the tour origin (or departure time to that stop on second half-tour) is also not known because it will be modeled along with stop location and trip mode for the next stop further from the tour origin.

As a result of this modeling approach, two known locations serve as anchor points for calculating travel impedance. These are the stop location immediately toward the tour destination (the tour destination itself for the first stop in a half-tour), which we call the **stop origin,** and the **tour origin**.

Parcel as dependent variable. The dependent variable used in this model, as in all location choice models is the parcel rather than the TAZ. The parcel is used in order to capture as well as possible the effect on activity and travel choices of parcel-level land use and transportation system attributes that may be affected by public policy.

The parcel is such a small unit of geography that, in the parcel data, it is difficult to accurately associate attributes with parcels and to associate survey locations with the correct parcel. This is an important issue because errors in the data could introduce more noise, or even bias, than is in the zonal data.

Since over 1 million parcels comprise the universal set of location choice alternatives, it is necessary to estimate and apply the stop location model with a sample of alternatives. For estimation, a sample of 30-100 parcels was used to represent the choice set for each observed choice. A randomly drawn subset of all parcels is used, with appropriate weighting, to represent the entire set of available parcels. The procedure uses importance sampling with replacement, in three stages: stratum, TAZ and parcel. Each stratum represents a particular band of impedance levels, and strata are sampled in proportion to their observed frequency of choice in the survey sample for a given type of intermediate stop. Strata include the tour origin TAZ, the stop origin TAZ, and three concentric ellipses surrounding those two points, with the size of the ellipses depending on stop characteristics. Since the stratum sampling procedure accounts for the effect of impedance, TAZ are drawn randomly within stratum. Then, within TAZ, parcels are drawn in proportion to their attracting size for the intermediate stop type. Details of the sampling procedure are provided in **Appendix 2 and 3**.

When the sample of parcels is drawn for estimation or application, infeasible destinations are excluded. Excluded parcels include those that lack the employment, school enrollment or households needed to accommodate the stop's activity purpose, as well as those that are too far away in light of the available time, tour mode and stop purpose. The distance constraints are shown below. Constraints for large time windows come from empirical analysis of the household survey data, allowing for stops with distance from 20-50% greater than the greatest observed distance, depending on the survey sample size for the category. Because of small sample size, the constraints for stops with short time windows are based on judgment.

Stop location parcel availability constraints by tour mode, stop purpose and avilable time window. Maximum XY distance in miles from stop origin through parcel and on to tour origin

Trip category	**Available time window less than 1 hour**	**Available time window greater than 1 hour**
Walk and bike tour modes	4 miles	35 miles
Motorized tour modes (by stop purpose)		
work	30	105
school	30	70
escort	40	120
personal business	30	80
shop	30	100
meal	20	70
social/recreation	20	150

Survey estimation data. The intermediate stop location model is estimated using all valid HH survey trip records with a destination other than the tour origin or destination. To be considered valid, the record must have identifiable tour mode, and valid parcels for the tour origin, tour destination, stop origin and stop destination.

Utility function. The model is a multinomial logit (MNL). Each alternative's utility function consists of the sum of several utility terms and one size function. Each utility term consists of an estimated coefficient multiplied by an alternative attribute and a trip characteristic. The trip characteristic is a dummy (0/1) variable that says to which subset of trips the coefficient applies. The alternative attribute is either a scalar value or a dummy variable that is nonzero only for the applicable subset of alternatives. Each utility term measures one aspect of a parcel's attractiveness for a given trip.

Size function. The size function also measures attractiveness of a parcel for a given trip. However, in this case the attractiveness depends on the parcel's size, that is, its capacity for accommodating the stop's activity purpose. The size function consists of several utility-like terms that are combined in the utility function in a form that corresponds with utility theory for aggregate alternatives. Although parcels are quite small, they must still be considered as aggregate alternatives because they have widely differing capacities for accommodating activities. For example, one residential parcel might include a large apartment building and another might have a single-family dwelling; the apartment building has a much larger capacity for accommodating activities that occur in homes. A size function is used instead of a single size variable because the defined activity purposes and size attributes do not have a simple one-to-one correspondence. Rather, several attributes can indicate capacity for accommodating a given purpose. For example, personal business could be conducted at many types of places, such as restaurants, stores or office buildings. The estimated coefficients give different weights to different size variables for a given purpose, and a scale parameter captures correlation among elemental activity opportunities within parcel. Equation 1 shows the form of the utility function, with size function included:

$$V_{in} = \sum_{k=1}^{K^v} \beta_k x_{ink} z_{nk} + \mu' \ln \sum_{k=K^v+1}^{K^v+K^s} \exp(\beta_k) x_{ink} z_{nk} \qquad (2)$$

where:

V_{in} is the systematic utility of parcel alternative *i* for trip *n*,

K^v is the number of utility parameters,

K^s is the number of size parameters,

$\beta_k,\ k = 1, 2, \ldots, K^v + K^s$ are the utility and size parameters,

x_{ink} is an attribute of parcel alternative *i* for trip *n*,

z_{nk} is a characteristic of trip *n*,

μ' is a scale parameter measuring correlation among elemental activity opportunities within parcels (1—no correlation, 0+--high correlation)

TRIP CHARACTERISTIC VARIABLES

The following trip characteristics are used in the utility function, interacting with attributes so that the effect of attributes depends on the characteristics of the trip. They are all 0/1 indicator variables, with 1 corresponding to the identified trip type. In many cases, the variable z_{nk} above represents the interaction of two or more of the characteristics from this list. For example, in one case z_{nk} equals one only for shopping stops with auto tour mode.

STOP PURPOSE

Work or school

University

Grade school

Escort

Personal business

Shop

Meal

Social-recreation

TOUR MODE

Auto

Non-auto

Auto drive alone

Auto shared ride 2

Auto shared ride 3+

Transit auto access

Transit walk access

Bike

Walk

Tour and trip characteristics

Multiple stops on half-tour

Secondary tour

Work-based tour

School tour

Work tour

Nonwork tour

Shop tour

Stop before work or school

First stop from tour destination

Not first stop from tour destination

Not last stop from tour destination

Person type and household characteristics

Female adult

HH with children

HH without children

HH income under $50K

HH income over $75K

HH income unreported

(used in estimation only)

The most important characteristics are the tour mode and the stop purpose. The tour mode restricts the modes available for the stop, and this affects the availability and impedance of stop locations. The availability and attractiveness of stop locations depend heavily on the stop purpose. Tour characteristics also affect willingness to travel for the stop, and the tendency to stop near the stop or tour origin. The above characteristics tend to overshadow the effect of personal and household characteristics in this model.

Alternative attributes and estimation results

The following alternative attributes are used in the utility function.

Alternative sampling adjustment term. This term is technically not a utility term, but rather it weights the alternative by the number of alternatives it represents as a result of the alternative sampling procedure.

Impedance variables

The impedance variables calculated for the intermediate stop model are based on the notion that the perceived impedance of an intermediate stop is a function of the time and cost along the path from the last prior known stop location to the intermediate stop location, and on to the first subsequent known stop location. It is assumed that the traveler forms their tour from the primary tour destination back toward the tour origin. For the first half-tour, this is in reverse chronological order. The reason for this is the hypothesis that people aim to arrive at the primary destination at a particular time, and choose their intermediate stop attributes so as to enable completion of the desired intermediate stops and still arrive at the primary destination on time. These assumptions affect the assumption of what is known when the intermediate stop choice is modeled. The known time and space anchors, used for measuring impedance, are the location and departure time from the stop nearer to the primary destination (or arrival time for second half-tour), and the location of the tour origin. Additionally, assumptions are made about the trip mode for each leg of the journey to and from the intermediate stop location, based on the known tour mode, the half-tour, and the proximity and connectivity of the stop location to the stop origin and tour origin.

Generalized time (100 minute units). The main impedance variable is generalized time. It combines all travel cost and time components according to the following assumptions:

Assumptions used in calculation of generalized time

PathImpedance_PathChoiceScaleFactor = 1.5
PathImpedance_AutoOperatingCostPerMile = 0.10
PathImpedance_TransitInVehicleTimeWeight = 1.0
PathImpedance_TransitFirstWaitTimeWeight = 2.0
PathImpedance_TransitTransferWaitTimeWeight = 2.0
PathImpedance_TransitNumberBoardingsWeight = 8.0
PathImpedance_TransitDriveAccessTimeWeight = 2.0
PathImpedance_TransitWalkAccessTimeWeight = 2.0
PathImpedance_WalkTimeWeight = 2.5
PathImpedance_BikeTimeWeight = 2.5
PathImpedance_WalkMinutesPerMile = 20.0
PathImpedance_TransitWalkAccessDistanceLimit = 1.0
PathImpedance_TransitWalkAccessDirectLimit = 1.0
PathImpedance_TransitSingleBoardingLimit = 1.1
PathImpedance_AutoTolledPathConstant = 0.0
PathImpedance_AvailablePathUpperTimeLimit = 200.0
PathImpedance_TransitLightRailPathConstant = 0.0
PathImpedance_TransitCommuterRailPathConstant = 0.0
PathImpedance_TransitFerryPathConstant = 0.0
PathImpedance_TransitUsePathTypeSpecificTime = true
PathImpedance_TransitLightRailTimeAdditiveWeight = -0.10
PathImpedance_TransitCommuterRailTimeAdditiveWeight = -0.30
PathImpedance_TransitFerryTimeAdditiveWeight = -0.40
PathImpedance_BikeUseTypeSpecificDistanceFractions = false
PathImpedance_TransitUseFareDiscountFractions = true
PathImpedance_TransitFareDiscountFractionChildUnder5 = 0.8
PathImpedance_TransitFareDiscountFractionChild5To15 = 0.5
PathImpedance_TransitFareDiscountFractionHighSchoolStudent = 0.5
PathImpedance_TransitFareDiscountFractionUniverityStudent = 0.5
PathImpedance_TransitFareDiscountFractionAge65Up = 0.5
Coefficients_BaseCostCoefficientPerDollar = -0.15
Coefficients_BaseCostCoefficientIncomeLevel = 30000
Coefficients_CostCoefficientIncomePower_Work = 0.6
Coefficients_CostCoefficientIncomePower_Other = 0.3
Coefficients_MeanTimeCoefficient_Work = -0.03
Coefficients_MeanTimeCoefficient_Other = -0.015

Coefficients_StdDeviationTimeCoefficient_Work = 0.8
Coefficients_StdDeviationTimeCoefficient_Other = 1.0
Coefficients_HOV2CostDivisor_Work = 1.741
Coefficients_HOV2CostDivisor_Other = 1.741
Coefficients_HOV3CostDivisor_Work = 2.408
Coefficients_HOV3CostDivisor_Other = 2.158

Generalized time is used, instead of various separately estimated time and cost coefficients, because the intermediate stop data is not robust enough to support good estimates of the relative values. Higher values of time were considered, and increasing them improved the model fit substantially, indicating that travelers are perhaps more time-sensitive for intermediate stops than for other travel. However, the lower values were retained because of FTA expectations. Higher values of walk, bike and wait time were also considered because of FTA expectations, but in this case the lower values were retained because of better model fit.

Generalized time is calculated by first calculating generalized time for the entire journey from the stop origin, through the stop location, and on to the tour origin, using the above assumptions and information about the known details of the tour and stop. It is then reduced by a distance-based factor to approximate the generalized time for only the detour to the stop location. Thus it might more appropriately be called generalized detour time.

Generalized detour time is further modified by discounting it according to the distance between the stop origin and the tour origin. The discount increases linearly from zero to 30% for distances between 0 and 30 miles, and remains at 30% for distances over 30 miles. This enables a single estimated coefficient to capture distance-based discounting. The discounting is based on the hypothesis that people are more willing to make longer detours for intermediate stops on long tours than they are on short tours. The hypothesis was tested by estimating the model with various discounting assumptions. Model fit improved with discounting and the best fit was with the assumptions of 30% and 30 miles.

Further mention of generalized time refers to discounted generalized detour time as described here.

Generalized time squared and **generalized time cubed**. These components allow for a nonlinear effect of generalized time.

Detour distance (miles) cubed. For transit tours, the generalized time variables and coefficients, as defined, and the imposed availability restrictions, don't adequately account for the tendency to avoid long intermediate stops, and the result is excessively large estimated sensitivity to generalized time. Therefore, for these stops, distance cubed is included as a variable. With it, the model fit improves and the elasticities come down to reasonable levels, as subsequently discussed.

Travel time as a fraction of the available time window. This variable captures the tendency to choose nearby activity locations if there are tight time constraints on the stop. If the stop occurs on the first half-tour (on the way to the primary tour activity) then the available time window begins at the beginning of the tour origin activity, and ends at the end of the activity immediately after the modeled stop or upon arrival at the primary destination. If it occurs on the second half-tour, then the available window begins at the beginning of the preceding activity or upon departure from the primary destination, and ends at the end of the subsequent tour origin activity. A similar variable was attempted that divided the available time window among all remaining stops on the half-tour, but it did not fit as well.

Proximity to stop origin , **proximity to tour origin** , (units of 1/(10 min): 1=10 min, .1=100min). Prxs is inverse travel time between stop destination and stop origin. It captures the tendency to stop near the stop origin. Analogously, prxo captures the tendency to stop near the tour origin.

Estimation results for the impedance variables. Appendix 1 provides the estimation results for all the coefficients in the intermediate stop model. increases again. Sensitivity is lower for work and school purposes, and higher for shopping, meals, and escorting (HH with kids). Sensitivity is higher for auto tour modes and lower for transit tour modes.

The distance cubed parameters capture the tendency to distance-limit stops on transit tours and all escort stops.

There is a time constraint variable that captures a tendency for shorter trips when they are constrained by a short time window.

Several parameters capture the tendency for trips of various types to occur near the stop origin or tour origin. For example, one parameter indicates that trips have a tendency to occur near the tour origin, but another nearly nullifies the effect if it is not the last stop before returning to the tour origin (on the second half-tour) or the first stop after departing from the tour origin (on the first half-tour).

CONNECTIVITY VARIABLES

These variables measure aspects of network connectivity in the vicinity of a parcel that impact its accessibility by non-auto modes.

Walk and transit both unavailable for 1 leg (0/1 indicator variable). A value of 1 indicates that the stop location is accessible by neither walk nor transit from either the stop origin or the tour origin, but it is accessible by walk and/or transit from the other.

Walk and transit both unavailable for both legs (0/1 indicator variable). A value of 1 indicates that the stop location is accessible by neither walk nor transit from the stop origin, and is similarly inaccessible from the tour origin.

Four-link density (n4lq). Number of road network nodes with 4 links within a quarter mile of parcel. A large value of this measure indicates a high degree of road connectivity.

Three-link density (n3lq). Number of road network nodes with 3 links within a quarter mile of parcel. A large value of this measure indicates a large number of nodes that are not fully connected.

Four-link ratio (n4sq, range 0-1). (# 4-link nodes)/(# 1,3, and 4-link nodes) within a quarter mile. A large proportion of nodes with 4 entering links indicates a highly connected grid-type street network.

Dead-end ratio (n1sq, range 0-1). (# 1-link nodes)/(# 1,3, and 4-link nodes) within a quarter mile. A large proportion of dead-end nodes indicates a lack of connectivity of the street network.

Estimation results. Several parameters show the expected strong tendency to avoid parcels that aren't connected by walk or transit to the stop and tour origins when the tour mode is transit with walk access. The effect is neither strong nor significant for transit with auto access because it is possible to make the stop during the auto portion of the tour. Although several variations of the other connectivity variables were tried, in an attempt to capture the effect of walkability on location choice for walk and bike tours, only the dead-end ratio captured the expected effect; it was retained even though the result is not statistically significant.

PARKING VARIABLES

The parking variables capture the coincidence of attractions and available parking.

Mix of hourly parking & employment in zone [ln(1+ prkgdens*empldens/ (prkgdens+ empldens)). A large value of this interaction variable indicates that the zone is very attractive for short-term activities and has parking available to match the attractiveness. A small value indicates that the zone lacks either attractions or parking or both. In the formula, dividing by the sum of parking and employment densities removes simple density effects that are accounted for by the density variables.

Mix of hourly parking & employment in parcel [ln(1+prkg*empl/(prkg+empl)). This is like the zonal variable except it is an absolute measure, instead of density, and measures parking and employment on the parcel itself.

Estimation results. These effects are statistically very significant in the model. Availability of parking within the zone draws auto trips to parcels in zones with many attractions (parameter 60) although the effect is not quite as strong for auto drive alone mode. Availability of parking on the parcel itself draws auto trips to parcels with many attractions (parameter 62).

PARCEL SIZE VARIABLES

These are the variables that are included in the size function described above:

Medical employment in parcel

Service employment in parcel

Retail employment in parcel

Restaurant employment in parcel

Industrial and other employment in parcel

Government, office and school employment in parcel

Total employment in parcel

Number of households in parcel

K-12 enrollment in parcel

University enrollment in parcel

Estimation results. In the size function, one size variable serves as the 'base', setting the scale of the function, and parameters are estimated for all the other variables in the function, measuring their effect relative to the base. In the model, the size function differs by stop type. **Table 2** below shows the base size variable for each stop type, along with the other variables. It also identifies the effect of the other variables in the size function relative to the base variable. For most stop types, only one size variable has a significant effect. This is a very good result, indicating that the stop types and size variables have been defined narrowly enough so that relative parcel size in the various categories clearly impacts modeled location choice.

Table : Size variables in the intermediate stop location model

Stop type	Base size variable	Other variables in size function
Escort (HH with kids)	K-12 enrollment	total employment households
Escort (HH with no kids)	total employment	K-12 enrollment university enrollment households
Meal	restaurant employment	total employment households
Personal business	medical employment	service employment restaurant employment industrial and other employment gov., office and educ. employ. retail employment university enrollment households
Grade school	K-12 enrollment	total employment households
University	university enrollment	total employment
Shopping	retail employment	service employment medical employment total employment
Social-recreation	service employment	retail employment medical employment total employment households
Work	total employment	none

DENSITY VARIABLES USING BUFFER

The attractiveness of a parcel can also be affected by employment, housing and school enrollment in the surrounding neighborhood. The density variables, calculated within a distance-weighted buffer around the parcel in a logarithmic form, capture these neighborhood effects:

Medical Jobs
Service Jobs
Retail Employment
Restaurant Employment
Education Employment
Industrial Employment
Total Employment
Total Households
K-12 Enrollment
University Enrollment

Estimation results. Density effects are estimated only for non-mandatory purposes, under the hypothesis that work and school stops are determined strictly by the need to visit a particular location, regardless of its surroundings. For the other trip types, the following table summarizes the estimation results, identifying the zonal density variables that attract stops to a parcel, and those with a negative effect. Only the most statistically significant effects are shown below.

Table : Density variables in the intermediate stop location model

Stop type	Zonal density that attracts stops at parcel	Zonal density that repels stops at parcel
Escort (HH with kids)	K-12 enrollment	medical employment households
Escort (HH no kids)	gov., office and educ. employ.	households
Meal	medical employment	restaurant employment households
Personal business	gov., office and educ. employ. medical employment	service employment households university enrollment
Shopping	restaurant employment gov., office and educ. employ. service employment	industrial and other employment retail employment university enrollment
Social-recreation		households

MIXED USE VARIABLES

Several variables were tried in the specification measuring the mix of housing and employment in the zone, in an effort to capture the attractiveness of parcels in mixed-use neighborhoods for intermediate stops. However, the variables failed to capture the expected effect and were dropped from the model. It may be because the size and impedance variables would already capture the tendency of mixed use developments to reduce trip lengths for intermediate stops.

Intermediate Stop location Model Estimation Results

Variable Description	Coefficient	T-stat
sampling adjustment factor	1	const
ln(travel time/available Window)	2.31	7.5
detour generalized time	-14.43	-46.4
detour generalized time squared	10.83	15.4
detour generalized time cubed	-3.09	-9.3
distance cubed (100s of miles up to .5 cubed)	-2.45	-1.3
1 / (Max(1, travel minutes from stop origin) / 10 (this is called proximity to stop)	0.1	13.5
1 / (Max(1, travel minutes from tour origin) / 10) (this is called proximity to origin)	0.14	12.3
household has income less than 25 K * detour generalized time	0.38	2.5
household has 100K plus income * detour generalized time	-0.42	-2.5
Household has missing income * detour generalized time	0.14	0.7
Number of Children * detour generalized time	-0.7	-4
non-work tour * detour generalized time	0.01	0.1
not first tour * proximity to the destination	0.08	10.2
work-based tour * proximity to origin	-0.18	-3.4
school purpose * proximity to origin	-0.03	-1.3
work or school purpose * first half tour * proximity to origin	0.11	6.7
hov mode * proximity to stop alternative	0.01	1.5
hov mode * proximity to origin	-0.03	-2.2
auto mode * intersection density at stop alternative	-0.01	-24.2
sov mode * parking, employment, commercial mix in parcel	-0.67	-10.6
auto mode * parking, employment, commercial mix in pacel	0.34	2.9
non-auto mode * detour generalized time	1.44	2.1
non-auto mode * detour generalized time squared	-0.02	0
non-auto mode * detour generalized time cubed	-0.33	-0.5
non-auto mode * proximity to origin	0.07	2.6
bike mode * proximity to origin	-0.22	-3.1
walk mode * proximity to origin	-0.37	-4.9

transit mode* walk and transit unavailable on one leg	-1.01	-7.3
transit* walk and transit unavailable on both legs	-2.11	-11.8
work or school destination purpose * generalized time	0.76	3.7
personal Destination Purpose *University Student * Log(destination Parcel Students University Buffer1)	-0.02	-0.6
work Destination Purpose * Log(destination Parcel Employment Total Buffer1 + 1)	0.41	15.6
work Destination Purpose * Log(destination Parcel Students K12 Buffer1 + 1)	-0.04	-2.2
escort stop HH with kids * generalized time	-1.41	-5.6
escort stop HH with kids * proximity to stop alternative	-0.01	-0.9
escort stop HH no kids * proximity to origin	0.05	1.8
escort stop HH with kids* proximity to origin	0.07	3.4
escort stop HH with kids * Log(destination Parcel Employment Industrial Buffer 1 + destination parcel Employment Agriculture Construction Buffer1 +1)	-0.12	-6.0
escort stop HH with kids * Log(destination parcel Students K12 Buffer1 +1)	0.1	6.2
escort stop HH no kids * Log(destination parcel Employment Total Buffer1 + 1)	0.25	7.7
escort stop HH no kids * Log(destination parcel Students K12 Buffer1 +1)	-0.02	-0.9
personal Or Medical Destination Purpose * Log(destination parcel Employment Medical Buffer1 + 1)	-0.09	-5.1
personal Or Medical Destination Purpose * Log(destination parcel Employment Food Buffer1 + 1)	0.23	8.9
personal Or Medical Destination Purpose * Log(destination parcel Employment Retail Buffer1 + 1)	0.1	4.3
shopping Destination Purpose * generalized time	-2.57	-12.2
shopping Stop On Shop Tour * proximity to destination	-0.07	-4.5
shopping Destination Purpose * Log(destination parcel Employment Retail Buffer1 + 1)	0.39	23.7
meal Destination Purpose * generalized time	-1.72	-6.9
meal Destination Purpose * Log(destination parcel Employment Food Buffer1 + 1)	0.17	7.2
social Or Recreation Destination Purpose * generalized time	-0.28	-1.6
social Or Recreation Destination Purpose * Log(destination parcel Employment Food Buffer1 + 1)	0.12	3.5
social Or Recreation Destination Purpose * Log(destination parcel Employment Industrial Buffer 1 + destination parcel Employment Agriculture Construction Buffer1 +1)	-0.3	-8.0
social Or Recreation Destination Purpose * Log(destination parcel Employment Service Buffer1 + 1)	-0.08	-1.6
social Or Recreation Destination Purpose * Log(destination parcel Employment Total Buffer1 + 1)	0.39	6.5
social Or Recreation Destination Purpose * Log(destination parcel HouseholdsBuffer1 + 1)	-0.3	-6.7
size function coefficient (subsequent coefficients are for size variables)	0.07	86.3

work Destination Purpose * (destination parcel Employment Government + destination parcel Employment Office + destination parcel Employment Education)	0	
work Destination Purpose * destination parcel Employment Total	-13.07	-14.2
child under 15 on school stop * (destination parcel Employment Government + destination parcel Employment Office + destination parcel Employment Education)	-32.95	-6.5
driving age child on school stop * (destination parcel Employment Government + destination parcel Employment Office + destination parcel Employment Education)	-29.75	-3.6
not child driving age or younger, on school stop * (destination parcel Employment Government + destination parcel Employment Office + destination parcel Employment Education)	-61.72	-11.3
not child driving age or younger, on school stop * destination parcel Students University	0	
escort stop HH with kids * (destination parcel StudentsK8 + destination parcel Students High School)	0	
escort stop HH with kids * destination parcel Employment Total	-31.81	-29.5
escort stop HH with kids * destination parcel Households	-136.08	-39.2
escort stop HH no kids * destination parcel Employment Total	0	
escort stop HH no kids * (destination parcel StudentsK8 + destination parcel Students High School)	23.31	9.1
escort stop HH no kids * destination parcel Households	-86.33	-22.2
personal Or MedicalDestination Purpose * destination parcel Employment Food	-11.66	-3.8
personal Or MedicalDestination Purpose * (destination parcel Employment Industrial + destination parcel Employment Agriculture Construction)	-10.04	-4.8
personal Or MedicalDestination Purpose * destination parcel Employment Medical	33.33	21.7
personal Or MedicalDestination Purpose * (destination parcel Employment Government + destination parcel Employment Office + destination parcel Employment Education)	31.33	22.5
personal Or Medical Destination Purpose * destination parcel Employment Retail	5.44	2.9
personal Or MedicalDestination Purpose * destination parcel Employment Service	0	
personal Or Medical Destination Purpose * destination parcel Households	-78.84	-27
shopping Destination Purpose * destination parcel Employment Retail	0	
shopping Destination Purpose * destination parcel Employment Service	-50.66	-22.3
shopping Destination Purpose * destination parcel Employment Total	-55.78	-38.2
shopping Destination Purpose * destination parcel Households	-40	constr
mea Destination Purpose * destination parcel Employment Food	0	
meal Destination Purpose * destination parcel Employment Total	-88.73	-22
meal Destination Purpose * destination parcel Households	-80	constr

social Or Recreation Destination Purpose * destination parcel Employment Food	-15.68	-8.2
social Or Recreation Destination Purpose * destination parcel Employment Medical	-30	constr
social Or Recreation Destination Purpose * destination parcel Employment Service	0	constr
social Or Recreation Destination Purpose * destination parcel Employment Total	-34.72	-31.1
social Or Recreation Destination Purpose * destination parcel OpenSpaceType2Buffer1 > 0	-80	constr
social Or Recreation Destination Purpose * destination parcel Households	-80.3	-74.5

Analysis is based on 18344 observations

Likelihood with Zero Coefficients = -70121.5469
Likelihood with Constants only = -37359.8780
Final value of Likelihood = -50351.3072
"Rho-Squared" w.r.t. Zero = .28

Appendix 1—Sampling of Alternatives for Destination Choice

This appendix describes choice set sampling procedures used in the destination choice models. Modeling the choice of a particular parcel makes the universal choice set very large, and presents challenges to appropriately limit the number of alternatives considered when simulating choices.

The reduction of the universal choice set involves two conceptually different methods. The first method involves attempting to remove from the universal choice set those alternatives that the decision-maker would not even consider in making the decision; they would appropriately be assigned a probability of zero. Examples of these include parcels that cannot be reached in the available time, and parcels that don't accommodate the desired type of activity. There is a behavioral basis for removing these parcels from the choice set, because there is no chance that they will even be considered.

The second method involves taking the remaining alternatives, that would all be reasonable alternatives for the decision-maker to consider, and drawing a sample of them to actually use in simulating the choice. This is simply a procedural technique to reduce the computational burden of the model.

The procedures described in this paper employ both methods. The first method includes two aspects. First, each parcel is assigned purpose-specific sizes. For a given purpose, if a parcel has zero size, then it will be unavailable. Second, the approximate time required to reach a parcel is compared to an estimate of the available time. If the parcel can't be reached in time, then it is eliminated from consideration.

The second method uses a technique called importance sampling with replacement. The available alternatives are sampled in a way that allows the probability of being drawn into the sample to be calculated for each drawn alternative. Statistical procedures are then used during model estimation and application to allow the sample to represent the entire set of available alternatives without biasing the results.

The following material describes importance sampling with replacement, and then describes its implementation for usual locations and tour destinations, cases where the traveler is departing from a known location, visiting an unknown destination, and then returning to the original known location.

Importance sampling with replacement for MNL models—estimation procedure

(per Moshe Ben-Akiva, MIT course 1.205, Fall 1993)

The following procedure yields consistent MNL estimates:

Draw R times from the full choice set *C* with replacement and selection probabilities $q(j), j = 1, \ldots, J$. Let $n_j, j = 1, \ldots, J$ be the number of times alternative *j* was drawn.

Add the chosen alternative. Set $\tilde{n}_j = n_j + \delta_{jc}, j = 1, \ldots, J$, where $\delta_{jc} = 1$ for *j = c* and 0 otherwise and *c* denotes the chosen alternative.

Create the set $\tilde{D}$ as $\tilde{D} = \{ j \in C \mid \tilde{n}_j > 0 \}$

Estimate the following MNL: $\tilde{p}(i \mid \tilde{D}) = \dfrac{\exp[v_i - \ln(q(i)/\tilde{n}_i)]}{\sum_{j \in \tilde{D}} \exp[v_j - \ln(q(j)/\tilde{n}_j)]}$

Notes:

a. This procedure has **not** been proven to yield consistent estimates for nested logit models.

b. The correction factor expands the exponentiated utility of each sampled alternative by the inverse of the sampling probability, giving it the weight of all the unsampled alternatives it represents.

c. The correction factor is not part of the true model. It is removed for model application with a full choice set. However, it is retained when simulating choices with a similarly generated sample of alternatives.

d. In model application with a similarly generated sample of alternatives, it is not necessary to remove duplicates of sampled alternatives; instead, each occurrence of each alternative can simply be assigned $\tilde{n}_j = 1$. Statistically, the effect is identical; in one case there are $\tilde{n}_j$ identical alternatives with probability *p*, and in the other there is one alternative with probability $\tilde{n}_j p$.

TOUR DESTINATION SAMPLING

The procedure uses 2-stage importance sampling with replacement. For each parcel to be drawn, first a TAZ is drawn, and then a parcel within the TAZ. To formalize, define the following notation:
$t_k,\ k = 1,\ldots,K$, are the TAZs with sampling probabilities $q(t_k)$
$j,\ j = 1,\ldots,J$, are the parcels with conditional sampling probabilities $q(j \mid t_k)$

The unconditional parcel sampling probabilities are therefore calculated as $q(j) = q(t_k)q(j \mid t_k)$.

TAZ are sampled according to size and impedance based importance weights, and parcels are sampled according to size-based importance within TAZ, as follows:

$$
\begin{aligned}
q(t_k) &\equiv W_{t_k h} / \sum W_{thg} \\
&= M^p_{kh} \exp(-\alpha_h d_k) \Big/ \sum\nolimits_{t_{\tilde{k}} \mid d_{\tilde{k}} < d_g} M^p_{\tilde{k}h} \exp(-\alpha_h d_{\tilde{k}}) \ \text{ if } d_k < d_g \\
&= 0 \text{ otherwise}
\end{aligned}
$$

$$
q(j \mid t_k) = M^p_{jh} \big/ M^p_{kh}
$$

where

h is the importance weighting scheme

d_g is an impedance threshold beyond which locations are unavailable

$W_{t_k h} \equiv M_{kh}^{p} \exp(-\alpha_h d_k)$ is the importance weight for t_k, given h

$\sum W_{thg} \equiv \sum_{t_k | d_k < d_g} M_{kh}^{p} \exp(-\alpha_h d_k)$ is the sum of importance weights, given h and d_g.

$M_{kh}^{p} = \sum_{j \in t_k} M_{jh}^{p}$

M_{jh}^{p} is the size of parcel j for tour purpose p, given h

α_h is a mixing parameter that sets the relative influence of impedance and size

d_k is the impedance measured along the path from t^o to t_k and back,

t^o is the TAZ of the tour origin.

The importance weighting scheme, h, and the impedance threshold, d_g are selected at the time of the draw, and depend on known characteristics of the tour. h has a corresponding vector of parameters, θ_h, chosen from a small set of such vectors, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_h, \ldots, \theta_H)$, with $\theta_h = (\alpha_h, M_h)$. M_h are the parameters of a particular size function that generates the size of all TAZ. $\boldsymbol{\theta}$ will have been empirically derived to represent the full range of characteristics of all possible tour stop situations.

THE TOUR DESTINATION SAMPLING PROCEDURE:

To draw a sample of tour destinations for a given choice situation, the draw proceeds as follows:

Select the impedance threshold g and the importance weighting scheme, h, with its corresponding vector of weighting parameters, θ_h.

Look up the importance weight of all available TAZ in the region, $\sum W_{th}$, using the weight formula determined by θ_h.

For each needed destination alternative, draw a random number, y, between 0 and 1, and pass sequentially through TAZ in order of decreasing importance weight, $W_{t_k h}$, selecting the TAZ at the point where the cumulative importance weight exceeds $y^* \sum W_{thg}$. Retain its ID and its sampling probability, $q(t_k)$.

For each drawn TAZ, draw a random number between 0 and 1, and pass sequentially through its parcels in order of decreasing sampling probability, selecting the parcel at the point where the cumulative sampling probability exceeds the drawn random number. For each drawn parcel calculate and retain its unconditional sampling probability, $q(j) = q(t_k) q(j | t_k)$.

For estimation only, add the chosen parcel to the choice set (again, if it was already drawn randomly) and count the number of occurrences of each parcel. Retain only one copy of each distinct parcel ID, j, along with its unconditional sampling probability $q(j)$ and the number of times it was drawn, $\tilde{n}_j$

Appendix 2—Tour Destination Sampling Parameters

This appendix presents the details of the weighting schemes prescribed in appendix 3. The reason for weighting in the sampling of alternatives is to improve the statistical efficiency of the choice models. A choice model estimated and applied with a sample of alternatives is most efficient when the alternatives appear in the sample in proportion to their actual choice probabilities. If the sample is inefficient, the estimation or prediction is still statistically consistent, but less efficient (precise) than it might be. However, complex schemes designed for maximum statistical efficiency can cause severe computational inefficiency. Therefore, the choice of schemes constitutes a trade-off between statistical efficiency and computational efficiency.

Each scheme is defined by the attraction (size) variables used for sampling, and by the relative importance of travel impedance and activity attractiveness. Tours that have a similar spatial distribution, relative to tour origin, and that are attracted to the same kind of locations, share a weighting scheme. **Table A2.1** shows the groupings that have been chosen for sampling schemes, based on simple unweighted data analysis of the survey sample. The primary variable determining scheme is purpose, because attraction variables differ substantially by purpose. After that, the factors that affect the spatial distribution are primarily person type (especially full-time vs other persons for work tours), and tour priority (other things being equal, tours with longer distances are assigned higher priority in the sample).

Table A2.1—Groupings for tour sampling schemes

	Purpose	Person Type	Tour Priority
1	Work	Full-time worker	Usual location, Primary tour
2	Work	Full-time worker	Secondary tours Work-based tours
		Not full-time worker	Usual location and all tours
3	School	Full-time worker, Part-time worker, Non-worker 65+ Non-worker 18-64, University student	Usual location and all tours
4		Driving age student, Child age 5-15, Child under age 5	Usual location and all tours
5	Escort	All	All
6	Personal business	All	Primary tour
7		All	Secondary tours,

			Work-based tours
8	Shopping	All	Primary tour
9		All	Secondary tours, Work-based tours
10	Meal	All	Primary tour, Secondary tours
11		All	Work-based tours
12	Social/recreation	All	Primary tour, Work-based tours
13		All	Secondary tours

The following tables provide details from the sample data analysis upon which the grouping decisions were made.

INTERMEDIATE STOP SAMPLING

INTERMEDIATE STOP LOCATION SAMPLING

A key feature of intermediate stops that makes them different from tour destinations is that travel impedance is a function of three locations instead of two: the intermediate stop location, as well as locations before it and after it in the half tour. Accounting for different locations before and after the stop expands the number of relevant impedances geometrically, and makes it infeasible to use impedance-based weights for sampling at the TAZ level. Thus the intermediate stop sampling is done differently than tour destination sampling.

We model choices emanating from the tour destination, in reverse temporal sequence before the tour destination, and in regular temporal sequence after the tour destination. Therefore, the two known locations surrounding the modeled stop are the stop immediately nearer to the tour destination (subsequently called stop origin for convenience, even though on the first half tour it is actually the stop destination), and the tour origin.

The procedure uses importance sampling with replacement, in three stages: stratum, TAZ and parcel. The stratum sampling stage handles the effect of impedance in a way that is simple enough to make it feasible. Each stratum represents a particular band of impedance levels, and strata are sampled in proportion to their observed frequency of choice in the survey sample for a given type of intermediate stop. The first two strata represent special TAZ that are particularly attractive for intermediate stops. The first stratum is the TAZ of the stop origin, and the second stratum is TAZ of the tour origin. The reason for giving these TAZ their own stratum is the fact that a disproportionate number of stops occur in them, perhaps due to familiarity effects. The third through fifth strata consist of the remaining TAZ in three bands of increasing distance, where distance is measured from the stop origin, through the potential stop location, and on back to the tour origin. TAZ are excluded from the strata if they have zero attracting size for the stop purpose or if they cannot be reached given the time constraints.

Since the stratum sampling procedure accounts for the effect of impedance, TAZ are drawn randomly within stratum. Then, within TAZ, parcels are drawn in proportion to their attracting size for the intermediate stop type.

To formalize, define the following notation:
$r_l,\ l=1,\ldots,L$, are the strata, with sampling probabilities $q(r_l)$
$t_k,\ k=1,\ldots,K$, are the TAZs with conditional sampling probabilities $q(t_k \mid r_l)$
$j,\ j=1,\ldots,J$, are the parcels with conditional sampling probabilities $q(j \mid t_k)$

The unconditional parcel sampling probabilities are therefore calculated as
$q(j)=q(r_l)q(t_k \mid r_l)q(j \mid t_k)$.

There are five strata, defined as follows:

$$r_1=\{t^{o_s}\}, M_{t^{o_s}}^{p^s} \geq \delta$$
$$=\{\}, \text{otherwise}$$

$$r_2=\{t^{o}\}, M_{t^{o}}^{p^s} \geq \delta$$
$$=\{\}, \text{otherwise}$$

$$r_3=\{t_k \mid d_{o_s ko} < d_3, \tilde{d}_{o_s ko} < d_{\max}, M_{t_k}^{p^s} \geq \delta, t_k \notin r_1, t_k \notin r_2\}$$

$$r_4=\{t_k \mid d_3 < d_{o_s ko} < d_4, \tilde{d}_{o_s ko} < d_{\max}, M_{t_k}^{p^s} \geq \delta, t_k \notin r_1, t_k \notin r_2\}$$

$$r_5=\{t_k \mid d_4 < d_{o_s ko} < d_5, \tilde{d}_{o_s ko} < d_{\max}, M_{t_k}^{p^s} \geq \delta, t_k \notin r_1, t_k \notin r_2\}$$

where:

t^{o_s} is the TAZ of the stop origin,
p^s is the stop purpose,
$M_t^{p^s}$ is the attracting size of TAZ *t* for the stop purpose,
δ is a small size, below which attracting size is considered equal to zero,
t^o is the TAZ of the tour origin,
$\tilde{d}_{o_s ko}$ is impedance measured in direction of travel along the path from t^{o_s} to t_k to t^o,
$d_{o_s o}$ is impedance measured in direction of travel along the path from t^{o_s} to t^o,
$d_{o_s ko}$ is $\tilde{d}_{o_s ko} - d_{o_s o}$, the incremental impedance caused by the stop at t_k,
d_3, d_4, d_5 are impedance thresholds separating available stop locations into groups, and
$d_{\max}$ is the impedance beyond which stop locations are considered infeasible.

Strata impedance thresholds and sampling probabilities are selected at the time of the draw. This vector of parameters is chosen from a small set of such vectors, $\boldsymbol{\theta} = (\theta_1, ..., \theta_h, ..., \theta_H)$, with $\theta_h = (q_h(r_1), q_h(r_2), q_h(r_3), q_h(r_4), q_h(r_5), d_{3h}, d_{4h}, d_{5h}, d_{\max h})$. The selection of *h* depends on the values x^s, which are known characteristics of the tour and stop. $\boldsymbol{\theta}$ are empirically derived to represent the full range of characteristics of all possible intermediate stop situations. TAZ are sampled randomly within strata, and parcels are sampled according to purpose-specific size-based importance within TAZ, as follows:

$$q(t_k \mid r_l) = 1 / n_l^t$$
$$q(j \mid t_k) = M_j^{p^s} \Big/ \sum_{\tilde{j} \in t_k} M_{\tilde{j}}^{p^s}$$

where

n_l^t is the number of TAZ centroids in r_l, and

$M_j^{p^s}$ is the attracting size of parcel j for the stop purpose

THE INTERMEDIATE STOP SAMPLING PROCEDURE:

To draw a sample of stop locations for a give intermediate stop location choice situation, the draw proceeds as follows:

Set strata sampling probabilities. Select the strata impedance thresholds and sampling probabilities, θ_h.

Retrieve the TAZ sampling probabilities. For strata 3 through 5, retrieve the number of available TAZ in the stratum from a matrix, $n_l^t[]$, containing these values precalculated for all possible combinations of stop origin TAZ, tour origin TAZ, impedance band, stop purpose, and maximum impedance. The inverse is the TAZ sampling probability within stratum.

Sample the strata. Sample the strata *C* times, according to their sampling probabilities, retaining the number of times each stratum is drawn, C_l.

Sample TAZ within strata. Draw from all TAZ randomly with replacement, keeping the first C_l for each stratum, until each stratum has reached its quota, C_l. Retain the TAZ ID and stratum of each drawn TAZ.

Sample parcels within TAZ. For each drawn TAZ, draw a random number between 0 and 1, and pass sequentially through its parcels in order of decreasing sampling probability, selecting the parcel at the point where the cumulative sampling probability exceeds the drawn random number. For each drawn parcel calculate and retain its unconditional sampling probability $q(j) = q(r_l) q(t_k \mid r_l) q(j \mid t_k)$.

Adjust sample (for estimation only). For estimation only, add the chosen parcel to the choice set (regardless of whether it was drawn randomly) and count the number of occurrences of each parcel. Retain only one copy of each distinct parcel ID, j, along with its unconditional sampling probability $q(j)$ and the number of times it was drawn, $\tilde{n}_j$.