SACSIM/05 Activity-Based Travel Forecasting Model for SACOG **Featuring** *DAYSIM***—the Person Day Activity and Travel Simulator**

Technical Memo Number 5 **Intermediate Stop Location Models**

July 31, 2006—Draft 2

Prepared for

Sacramento Area Council of Governments

Prepared by

John L. Bowman, Ph. D.

Transportation Systems and Decision Sciences

28 Beals Street, Brookline, MA 02446 USA +1-617-232-3478 John_L_Bowman@alum.mit.edu http://JBowman.net

MARK BRADLEY

BRADLEY RESEARCH & CONSULTING 524 Arroyo Ave., Santa Barbara, CA 93109, USA. +1-805-564-3908 mark_bradley@cox.net

Table of Contents

Introduction

This is the fifth in a series of technical memos being produced according to a work program in which Mark A. Bradley and John L. Bowman are developing the activity-based demand model components of a new travel demand forecasting model system for the Sacramento Area Council of Governments (SACOG), depicted in **Figure 1**. For a description of the entire model system, see memo 1 in this series, entitled Model System Design.

The current memo presents the estimation and preliminary validation results for the intermediate stop location 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 2**. In a minor change from technical memo 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 oneby-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 2—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.

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 of the new SACOG model system, 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 parcellevel 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. Errors in the data come from incompletely reported locations in the survey, geocoding based on imprecisely located TIGER shapes, and incomplete, inaccurate or aggregate base year parcel attribute information. Extensive efforts were made by SACOG staff and the consultants to make this data as accurate as possible.

Since over 700,000 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 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**.

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 in **Table 1**. 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.

Table 1: 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:

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

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,

 β_k , $k = 1,2,..., K^v + K^s$ 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 (-lnq). 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:

| walk speed | 3 mph |
|--|-----------------|
| bike speed | 8 mph |
| school bus travel time as multiple of SOV time | 3 |
| perceived auto operating cost | \$0.12 per mile |
| distance under which walk LOS is assumed | .25 mile |
| value of transit in-vehicle time, as multiple of value of auto IVT | 1.0 |
| value of walk time, as multiple of value of auto IVT | 1.5 |
| value of bike time, as multiple of value of auto IVT | 1.5 |
| value of wait time, as multiple of value of auto IVT | 2.0 |
| value of one transit boarding, in terms of auto IVT | 7 minutes |
| value of traveler time, trips with HH income under \$15K/year | \$5/hr |
| value of traveler time, trips with HH income under \$15-50K/year | \$10/hr |
| value of traveler time, trips with HH income under \$50-75K/year | \$15/hr |
| value of traveler time, trips with HH income under \$75-100K/year | \$25/hr |
| value of traveler time, trips with HH income under \$100+K/year | \$37.50/hr |
| value of traveler time, trips with unreported HH income | \$10/hr |

Table 2: Assumptions used in calculation of generalized time

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 distancebased 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 halftour, 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 (prxs), **proximity to tour origin** (prxo), (units of $1/(10 \text{ min})$: $1=10$ 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. Parameters 2-24 are the generalized time parameters, including the square and cubic components. Figures 3 and 4 graph the resulting

effect of generalized time for various trip purposes and tour modes. In all cases, the curve is sshaped, with sensitivity to generalized time gradually diminishing up to a certain point, and then it 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 (123 and 124) capture the tendency to distance-limit stops on transit tours and all escort stops.

Parameter 25 captures a tendency for shorter trips when they are constrained by a short time window.

John L. Bowman, Ph. D., Transportation Systems and Decision Sciences September 9, 2005 *MARK* **A.** BRADLEY, BRADLEY RESEARCH & CONSULTING **page 12**

Parameters 28-54 capture the tendency for trips of various types to occur near the stop origin (28-40) or tour origin (41-54). For example, parameter 41 indicates that trips have a tendency to occur near the tour origin, but parameter 48 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. Parameters 55-58 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 (parameter 59) 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** $[\ln(1+\text{prkg*emb}]/(\text{prkg+emb})).$ 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 3** 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, as estimated by parameters 89-121 in Appendix 1. 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 3: Size variables in the intermediate stop location model

Zonal density variables

The attractiveness of a parcel can also be affected by employment, housing and school enrollment in the surrounding neighborhood. The zonal density variables, in a logarithmic form, capture these neighborhood effects:

ln[1+(medical empl)*100/Million Sq Ft (Msqft)] in zone

ln[1+(service empl)*100/Million Sq Ft (Msqft)] in zone

ln[1+(retail employment)*100/Msqft] in zone

ln[1+(restaurant employment)*100/Msqft] in zone

ln[1+(government+office+education empl)*100/Msqft] in zone

ln[1+(industrial+other empl)*100/Msqft] in zone

ln[1+(total employment)*100/Msqft] in zone

ln[1+(# households)*100/Msqft] in zone

ln[1+(K-12 enrollment)*100/Msqft] in zone

ln[1+(Univ. enrollment)*100/Msqft] in zone

Estimation results. Zonal 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. Parameters 64-88 in Appendix 1 show the strength and statistical significance of all these effects.

| 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 |

Table 4: Density variables in the intermediate stop location model

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.

Appendix 1—Intermediate Stop Model Estimation Results

John L. Bowman, Ph. D., Transportation Systems and Decision Sciences September 9, 2005

John L. Bowman, Ph. D., Transportation Systems and Decision Sciences September 9, 2005 *MARK A. BRADLEY, BRADLEY RESEARCH & CONSULTING page 18*

John L. Bowman, Ph. D., Transportation Systems and Decision Sciences September 9, 2005

Appendix 2—Sampling of alternatives for stop location choice

This appendix describes the choice set sampling procedure used in the intermediate stop location choice model. This and the other destination choice models predict the choice of a particular parcel. This 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 decisionmaker 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 decisionmaker 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 intermediate stops, when the traveler is departing from one known location, stopping at an unknown location, then moving on to another 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 $q(j), j = 1,..., J$. Let $n_j, j = 1,..., J$ be the number of times alternative *j* was drawn.

Add the chosen alternative. Set $\tilde{n}_i = n_i + \delta_{ic}, j = 1, ..., J$, where $\delta_{ic} = 1$ 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: $\tilde{p}(i | \tilde{D}) = \frac{\exp[v_i - \ln(q(i) / \tilde{n}_i)]}{\sum \exp[v_i - \ln(q(j) / \tilde{n}_i)]}$ $j = \ln(q(j)/n_j)$ *j D* $\tilde{p}(i | \tilde{D}) = \frac{\exp[v_i - \ln(q(i) / \tilde{n}_i)]}{\sum \exp[v_i - \ln(q(j) / \tilde{n}_i)]}$ ∈ $=\frac{\exp[v_i - \frac{1}{2}]}{\sum_{i=1}^{n} \exp[v_i - \frac{1}{2}]}$ $\sum_{j \in \tilde{D}} \exp[v_j \tilde{p}(i | \tilde{D}) = \frac{\exp[v_i - \ln(q(i) / \tilde{n}_i)]}{\sum \exp[v_i - \ln(q(j) / \tilde{n}_i)]}$

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 $\tilde{n}_i p$.

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 , $l = 1,..., L$, are the strata, with sampling probabilities $q(r_l)$

 t_k , $k = 1, ..., K$, are the TAZs with conditional sampling probabilities $q(t_k | r_i)$

j, $j = 1, ..., J$, are the parcels with conditional sampling probabilities $q(j | t_k)$

The unconditional parcel sampling probabilities are therefore calculated as $q(j) = q(r_i)q(t_i | r_i)q(j | t_k).$

There are five strata, defined as follows:

$$
r_1 = \{t^{o_s}\}, M_{t^{o_s}}^{p^s} \ge \delta
$$

\n
$$
= \{\}, \text{otherwise}
$$

\n
$$
r_2 = \{t^o\}, M_{t^{o}}^{p^s} \ge \delta
$$

\n
$$
= \{\}, \text{otherwise}
$$

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

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

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

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° is the TAZ of the tour origin,

 \tilde{d}_{o_sko} is impedance measured in direction of travel along the path from t^{o_s} to t_k to t^o ,

 d_{o_so} is impedance measured in direction of travel along the path from t^{o_s} to t^o ,

 $d_{o,ko}$ is $\tilde{d}_{o,ko} - d_{o,o}$, 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, $\theta = (\theta_1, ..., \theta_h, ..., \theta_H)$, 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. θ 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:

$$
q(t_k | r_i) = 1/n_i^t
$$

$$
q(j | t_k) = M_j^{p^s} / \sum_{j \in t_k} M_j^{p^s}
$$

where

 n_i^t is the number of TAZ centroids in r_i , and

 $M_i^{p^s}$ 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_i . 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 $q(j) = q(r_i)q(t_k | r_i)q(j | t_k).$

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

Appendix 3—Application of model on estimation data

This appendix provides statistical results from applying the model on the estimation data. Table A3.1 lists the trip characteristics, which are all 0/1 variables, with the value 1 indicating membership in the category. The second column gives the percentage of the sample in each category.

Tables A3.2 through A3.28 compare the observed and predicted distribution of travel time for various subsets of the trips (see column headings) under the base conditions used for model estimation. The comparison is made by identifying the number of intermediate stops (observed and predicted) falling into each of 5 travel time bands (see row headings in the left hand column), where travel time is the approximate additional travel time required when making the stop instead of proceeding directly from stop origin to tour origin. The estimated standard deviation of the observed choices is also provided, and the number of stars for a prediction indicates the number of standard deviations by which the predicted deviates from the observed.

The results are within 2 standard deviations in most categories. There are, however, a few fairly large categories where the prediction is off by 3 standard deviations. Most notable is an underprediction of medium length stops (incremental travel time between 5 and 10 minutes). The model should be adequate for the initial implementation, but there is room for subsequent improvement by re-estimation with additional variables related to the problem categories.

Below the main table on each page, the predicted average value of ten intermediate stop attributes is also provided for each trip category. These attributes are:

- ttim2 travel detour time (10ths of minutes)
- gtim2 generalized detour time (10ths of minutes)
- empm2 medical employment at stop parcel
- emps2 service employment at stop parcel
- empr2 retail employment at stop parcel
empf2 restaurant employment at stop pa
- restaurant employment at stop parcel
- empo2 government, office and education employment at stop parcel
- hhld2 households at stop parcel
- enrs2 grade school enrollment at stop parcel
- enru2 university enrollment at stop parcel

This section of Table A3.3 is especially informative because it shows how effective the model is at matching trips of specific purposes with parcels that have appropriate levels of employment or enrollment of specific types.

Tables A3.29 through A3.35 examine elasticities calculated from a second application of the estimation data with all travel times increased by 10%. A so-called "range elasticity" is calculated. Range is calculated as the ratio of the predicted incremental travel time to the required incremental travel time for any given intermediate stop. The range elasticity is calculated as the percentage change in range divided by the percentage change in required travel time. It is calculated as arc

elasticity for a 10% across the board increase in travel time for all available modes to all available stop locations.

We expect range to decrease as required travel time increases (expect elasticity<0), but not to the extent that the resulting predicted travel time actually decreases (expect elasticity>-.91). The elasticities are approximately -0.4 for work and university stops, -0.6 for school stops, and -0.8 for other purposes, except for escort stops in households with children, where the elasticity exceeds - 0.9.

Table A3.1: Frequency of trip characteristics in estimation and test application sample

Total number of trip records is 7143.

Table A3.2: for pertype

Table A3.3: for trippurp

Table A3.4: for mode

Table A3.5: for tod

Table A3.6: for tourtype

Table A3.7: for inc6

Table A3.8: for hhsize

INFORMATION 572: number of **stars** in table is 31

Table A3.9: for totveh

Table A3.10: for distsotoc

Table A3.11: for gend

Table A3.12: for comp

WARNING 570: table rows or columns omit5744.0 observations INFORMATION 572: number of **stars** in table is 2

Table A3.13: for mult

WARNING 570: table rows or columns omit3111.0 observations INFORMATION 572: number of **stars** in table is 4

Table A3.14: for sect

WARNING 570: table rows or columns omit5070.0 observations INFORMATION 572: number of **stars** in table is 8

Table A3.15: for wbas

WARNING 570: table rows or columns omit6973.0 observations INFORMATION 572: number of **stars** in table is 2

Table A3.16: for shsh

WARNING 570: table rows or columns omit6598.0 observations INFORMATION 572: number of **stars** in table is 4

Table A3.17: for fkid

WARNING 570: table rows or columns omit5809.0 observations INFORMATION 572: number of **stars** in table is 4

Table A3.18: for essc

WARNING 570: table rows or columns omit6904.0 observations INFORMATION 572: number of **stars** in table is 6

Table A3.19: for bman

WARNING 570: table rows or columns omit5993.0 observations INFORMATION 572: number of **stars** in table is 0

Table A3.20: for wrksch

WARNING 570: table rows or columns omit6729.0 observations INFORMATION 572: number of **stars** in table is 8

Table A3.21: for auto

WARNING 570: table rows or columns omit 355.0 observations INFORMATION 572: number of **stars** in table is 8

Table A3.22: for naut

WARNING 570: table rows or columns omit6788.0 observations INFORMATION 572: number of **stars** in table is 4

Table A3.23: for first

WARNING 570: table rows or columns omit2429.0 observations INFORMATION 572: number of **stars** in table is 10

Table A3.24: for nfirst

WARNING 570: table rows or columns omit4714.0 observations INFORMATION 572: number of **stars** in table is 4

Table A3.25: for last

WARNING 570: table rows or columns omit2506.0 observations INFORMATION 572: number of **stars** in table is 8

Table A3.26: for nlast

WARNING 570: table rows or columns omit4637.0 observations INFORMATION 572: number of **stars** in table is 12

Table A3.27: for bwork

WARNING 570: table rows or columns omit6233.0 observations INFORMATION 572: number of **stars** in table is 0

Table A3.28: for nwrkt

WARNING 570: table rows or columns omit2392.0 observations INFORMATION 572: number of **stars** in table is 14

Note: Range is calculated as the ratio of the predicted incremental travel time to the required incremental travel time for any given intermediate stop. The range elasticity is the calculated as the percentage change in range divided by the percentage change in required travel time. It is calculated as arc elasticity for a 10% across the board increase in travel time for all available modes to all available stop locations. We expect range to decrease as required travel time increases α (elasticity $\langle 0 \rangle$), but not to the extent that the resulting predicted travel time actually decreases (elasticity>-.91).

Table A3.32: Elasticity of range with respect to required travel time—by time of day

Table A3.33: Elasticity of range with respect to required travel time—by tour purpose

Table A3.34: Elasticity of range with respect to required travel time—by XY distance from stop origin to tour origin

John L. Bowman, Ph. D., Transportation Systems and Decision Sciences September 9, 2005 *MARK* **A. B**RADLEY, BRADLEY RESEARCH & CONSULTING **page 57**