A Tour-Based National Model System to Forecast Long Distance Passenger Travel

Mark Bradley, Nazneen Ferdous & Maren Outwater: RSG; Brad Gudzinas: Federal Highway Administration; Stephane Hess & Andrew Daly: University of Leeds; Chandra Bhat: University of Texas; Ram Pendyala: Arizona State University; Jeff LaMondia: Auburn University

This research brief describes work to create a national model system to predict all passenger trips of 50 miles or more made by auto, air, rail and bus in a year period. It is a micro-simulation model, predicting long distance tours and trips made by a prototypical sample of all households residing in the United States. The model is being created and implemented as part of a project for the Federal Highway Administration to create a new framework for modeling long distance passenger travel.

The model design shares several key features with activity-based travel demand models:

- It simulates travel for individual households.
- There is a similar set of "longer term" models predicting household auto ownership and workplace location for all workers.
- It uses a level of spatial detail for trip ends that is very fine compared to the size of the study area (in this case, all Census tracts in the US).
- The models are "vertically integrated", meaning that mode choice, destination choice, and tour generation and scheduling are all mutually inter-related with causal effects in both directions.

The major differences with respect to activity-based travel models are:

- The long distance model predicts only tours to destinations 50 miles or more from home.
- The model deals with multi-destination tours, but does not predict <u>all</u> intermediate stop locations visited during long distance tours. It only predicts trips of 50 miles or more, but does not simulate shorter trips made while staying away from home.
- The model does not schedule long distance tours and trips within a day. Rather, long distance tours are scheduled across the months and weeks of a year. (Further breakdown of the results by day of week and time of day can then be done using a factoring approach.)
- The model jointly predicts destination and mode choice.
- The long distance model scheduling models include count, duration, regression, and multiple discrete continuous extreme value (MDCEV) methods.

Table 1 contains a list of the components of the model system. Each of the models has an analogous counterpart in a tour-based urban regional model system, including population synthesis and longer term models of auto ownership and workplace location. As in a activity-based model, the tour-level models of mode choice, destination choice and scheduling are the "core" models of the system. The trip-level models below the tour level, including stop generation, stop location, trip mode choice, and trip scheduling, are included, but are generally not as important as for urban models, because most long-distance tours have a single destination (not including short trips made during the tour).

Table 1: Long distance model components

Model	What is predicted
Longer term models	
Synthetic population generation	The household and person characteristics of residents of each Census tract (Done using PopGen software)
Workplace location	The Census tract of the usual work location for each worker
Auto ownership	The number of vehicles owned by the household
Tour level ("core") models	
Tour generation, duration & scheduling	The number of long distance tours made for each purpose category and duration category, scheduled across the months of the year
Tour party size & composition	The number of adults and children participating in each tour
Tour primary destination and mode	The Census tract of the tour primary destination, and the main mode used for the tour (auto, air, rail or bus)
Trip-level models	
Intermediate stop generation	The number of intermediate long-distance destinations visited during the tour, and the purpose of each stop
Intermediate stop location	The Census tract of each intermediate stop visited
Trip level mode choice	The mode for each trip, depending mainly on the main tour mode
Trip level departure time	The day of week and time period of day for each trip, depending on the tour period and duration

Table 2: Input data

Input data type	What is included
Land use data	Population, employment by sector, university enrolment, parks and land coverage at Census tract level.
Synthetic population	Household and person records for the entire US synthetic population, located to Census tract.
Rail network	Rail station-to-station matrices of-vehicle time, distance, transfers, frequency and fares, based on Amtrak data.
Air network	Airport-to-airport matrices of distance, in-vehicle time, transfers, frequency, fare, and on-time reliability, from the DB1B ticket and On-time databases.
Auto network	A base national auto network, with connectors added for each Census tract, airport and rail station.
Auto skims	Distance, time and toll for each TAZ-TAZ pair. TAZ's are an intersection of PUMAs and counties, to avoid having too much population or land area in any one TAZ.
Airport and station access	Auto travel distance and time from each Census tract to all airports and stations within a 100 mile radius.
Air and rail skims	Best tract-to-tract air and rail paths are found "on the fly" via all reasonable airport or station pairs.
Mode/destination accessibility logsums	Calculated for each Census tract / income / car ownership / purpose / party size combination, for tours to destinations in various distance bands.

Table 2 lists the input data used by the model system. Most of this data is defined at a Census tract level. The exceptions are the air network, which is at the airport-to-airport level, the rail network at the station-to-station level, and the road network skims, which are at the TAZ-to-TAZ level. Although the road network has connectors for each Census tract, so that the skims could theoretically be tract-to-tract, the size of such skims would be impractically large (there are over 70,000 Census tracts in the US), and the gain in accuracy is not vital for modeling long distance trips. We have defined TAZ's as the intersection of PUMA's and counties. Typically, PUMAs are smaller than counties, but counties can be used as TAZ's in low-density areas where there are multiple counties per PUMA. This leads to a system with roughly 4,500 TAZ's across the US. (Note that using a microsimulation model application framework and sampling of destinations makes it possible to use many more elemental spatial alternatives –Census tracts in this case--than the number of network TAZ's that are used.

Using Census tracts for land use attractions and for rail and air access/egress allows us to take advantage of the micro-simulation software framework to incorporate more spatial detail than would be possible in an aggregate zone-based model structure. Using Census tracts also allows us to use much more accurate mode/destination accessibility logsum to represent the attractiveness of making both short and long trips from each Census tract for various purposes.At the current time, the initial model estimation has been completed, and the final phase of estimation and implementation is beginning. Some interesting results from the model estimation include:

Tour mode and destination choice: This model predicts both the mode and destination for long distance tours simultaneously, with a nested model structure. The original models have been estimated using data from the 2012 California Statewide Household Travel Survey long-distance component, with roughly 30,000 tours. With 4 long distance modes (car, air, rail and bus) and over 5,000 relevant destination alternatives, the models have been estimated using over 20,000 choice alternatives, with separate models for Vacation/Leisure, Visit Friends and Relatives, Employer's Business, and Personal Business, with the latter including a number of different purposes, including school, shopping, and medical. Models for long distance Commuting are being estimated from different survey data.

The nested model has a flexible structure, in that destinations can be nested under modes or vice versa. In estimation tests so far, the best order of nesting appears to vary by tour purpose, with destination nested above mode for employer's business and vacation purposes. However, a cross-nested model structure appears to work better than either nesting order by itself.

Another interesting finding from the modeling is that there is significant heterogeneity in the model scale for destinations of different distances from the tour origin. Because long distance models can cover a very wide range of distances, from 50 miles up to more than 5,000 miles (e.g. New York to Honolulu), this issue has been recognized in previous long distance modeling studies. For the California High Speed Rail model, for example, (Outwater, et al. 2010); the travel purposes were segmented explicitly between trips under 150 miles and trips over 150 miles, and only destinations with the relevant distances were included as alternatives for each segment. While this is a way of addressing the scale issue, it is also rather arbitrary and can lead to discontinuous response sensitivities at the segmentation distances. In this project, we are testing alternative specifications where the error scale for each destination alternative is a non-linear function of the distance to reach that alternative. While this approach does lead to models with better fit, further testing is being done on the policy response of such models. This topic is a critical and interesting one for long distance modeling in general.

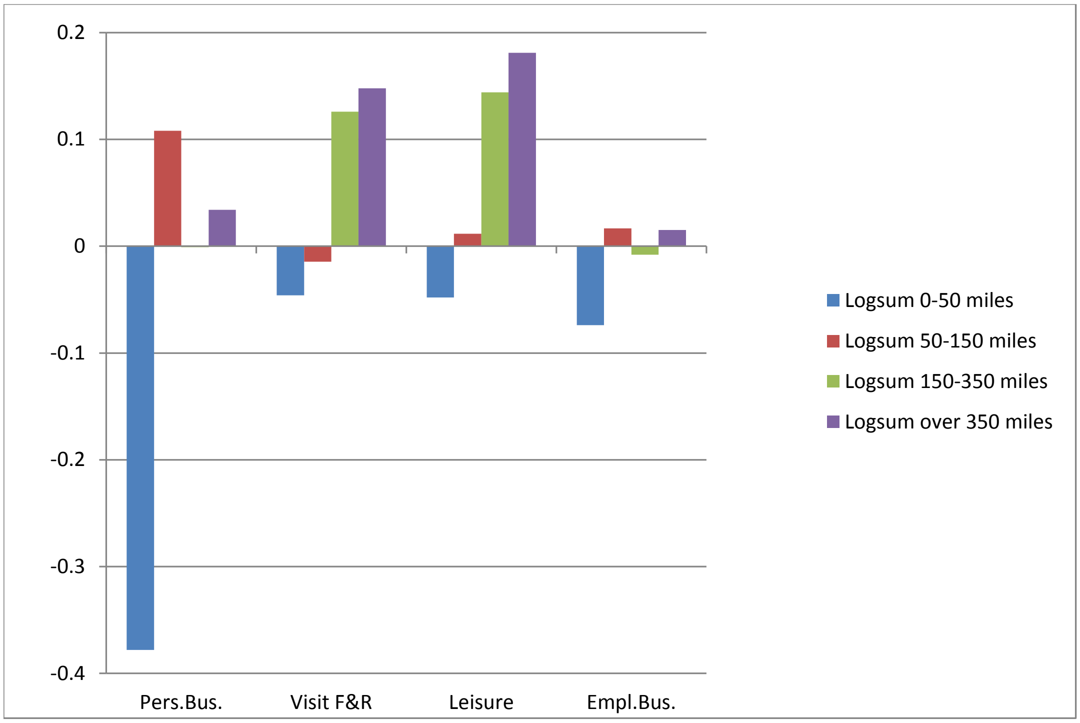
The study team is now in the process of finalizing national-level road, air and bus networks, and the spatial models will be re-estimated on the final data, also using data from other long distance travel surveys such as those from Ohio, Wisconsin and New York (along with the California data that was used originally).

Tour generation and scheduling: Separate long distance tour generation and scheduling models have been estimated for the same tour purpose segments as listed above. For employer's business travel, there does not appear to be completion effects between scheduling tours in different seasons across the year – in other words, making a business tour in the Winter does not make it less likely that the same person or household will schedule another business tour in the Summer, all else equal. That does not appear to be the case for other, non-work-related long distance tours. For those purposes, an MDCEV structure was used to estimate a simultaneous model of long distance tour generation, purpose, duration, and scheduling (season of the year). This model structure can represent "satiation" in the number and duration of long distance tours, and thus represent budget effects and substitution in scheduling tours across different times of year (Bhat 2008).

The tour generation and scheduling models described above were estimated using the 1995 American Travel Survey data set. Although that data set is quite dated, it is, unfortunately, still the only long distance US data set that reports a full year of long distance travel for each household. Two shortcomings of that data set, other than the fact that it is almost 20 years old are that it provides very little spatial or temporal detail – the date of travel is only recorded within a 3-month period, and the location of the destination is only provided at the state level. These shortcomings do not allow us to include detailed seasonality variables or accessibility-variables in models estimated on that data.

To address these shortcomings, additional tour frequency models were estimated using the 2012 California Statewide Travel Survey long distance data, the same dataset used for the mode and destination choice models. One of the more interesting findings from these models is that the accessibility to make tours for the given travel purpose near to versus far away from the tour origin has a significant influence on the frequency of making such tours. Figure 1 shows that for all 4 long distance purposes, the more accessible the relevant destinations within 50 miles, the less likely people are to make long distance tours to destinations 50 or more miles away. This effect is strongest for "personal business", as many of these tours are made by people living in rural areas who do not have many alternatives within a 50 mile range for purposes such as shopping and medical visits. For that purpose, having such destinations within a 50-150 mile range has a positive logsum coefficient (more tours). For "leisure/vacation" and "visit friends and relatives", accessibility to relevant destinations over 150 miles away have the largest positive effects on tour rates. Accessibility logsums have the least effect for the "employer's business" purpose, which is presumably the least discretionary of the purposes.

Figure 1: Mode/Destination Choice Accessibility Logsums Effects in Tour Frequency Models



Summary

The project described here, which is nearing the end of the model estimation stage and close to beginning the application phase, will produce the first national long-distance passenger forecasting model to use a disaggregate, tour-based approach, comparable to the methods used in state-of-the-art regional activity-based models. This brief describes a few of the more innovative features of the model system, including joint tour mode/destination choice models, with cross-nesting and non-linear, distance-related scale effects; joint tour generation/duration/scheduling models using the MDCEV approach, accessibility logsum linkage between the tour mode/destination choice models and the tour generation models, and a practical micro-simulation framework to apply these models to a representative population for the entire US, at the Census tract level.

References

Bhat, C.R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological* 42(3), 274-303.

Outwater, M., K. Tierney, M. Bradley, E. Sall, A. Kuppam, and V. Modugula (2010). California Statewide Model for High-Speed Rail. *Journal of Choice Modelling* 3(1),58-83.