Destination Choice Modeling Primer

presented by

César A. Segovia, AICP

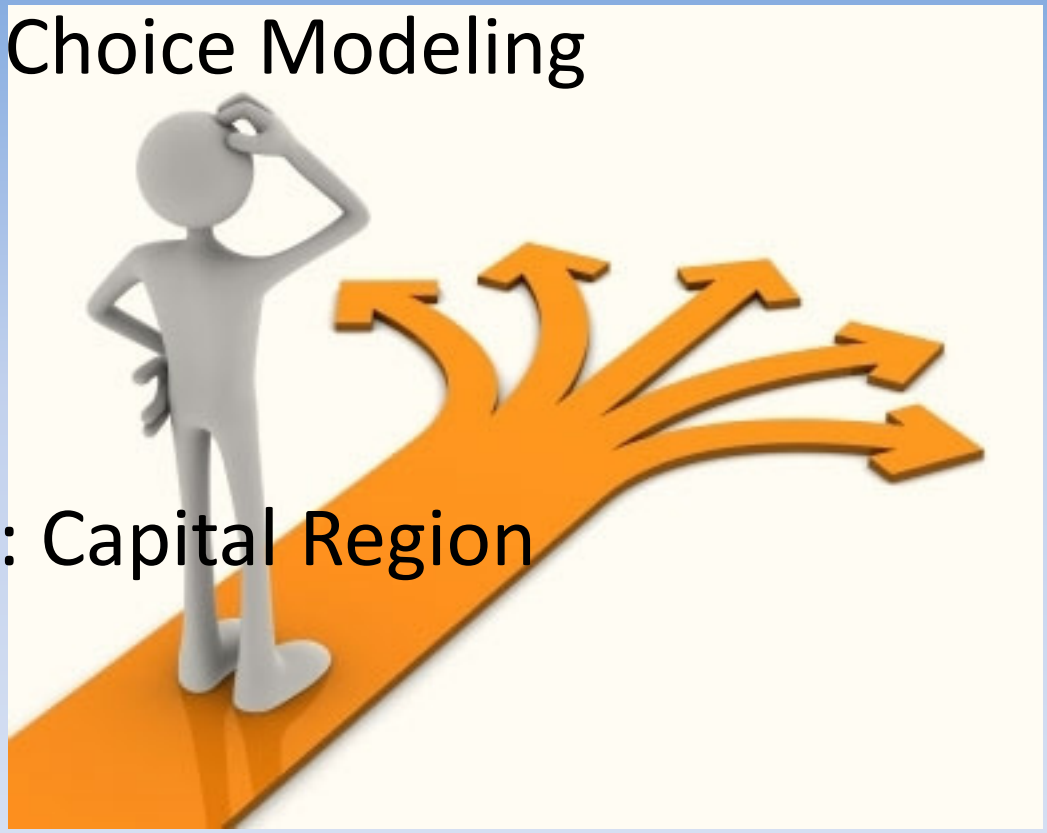
Senior Transportation Planner

AECOM

September 10, 2015

Today's Webinar Content

- An Introduction to Choice Modeling
- Destination Choice
- Florida Application: Capital Region



Destination Choice: A Modeling Primer

Choice Modeling

Discrete Choice Modeling: What's that?

- The key word in there is "CHOICE". It may make you think in terms of "human behavior", which is a very appropriate idea if we are trying to replicate travel patterns of real people.
- A lot of disciplines ranging from Behavioral Economics to Mathematical Programming, including Probabilities Theory and Statistical Sciences, inform transportation planners on how to replicate trip making decision processes.

Discrete Choice Modeling: What's that?

- If you thought you could get away from travel demand modeling (TDM) without knowing *choice modeling*…think again!
- This choice modeling method is a popular one when trying to establish: destination, route, activity, auto ownership, residential/commercial location, to mention only the most relevant ones for travel demand models.
- You may have heard of "mode choice", where the TDM chooses how many people will travel by car (drive alone or HOV), or by some kind of transit, or by a non-motorized mode.
- Although the most universal use is the *travel mode choice*, the others are spreading fast within activity-based models (ABM) and managed lanes modeling.

Discrete Choice Modeling: Definition

- "Discrete choice models can be used to analyze and predict a decision maker's choice of one alternative from a finite set of mutually exclusive and collectively exhaustive alternatives ."[4] The objective is to be able to predict the decision making of a group of individuals (an individual or decision maker can be a person, a household, a firm, etc).

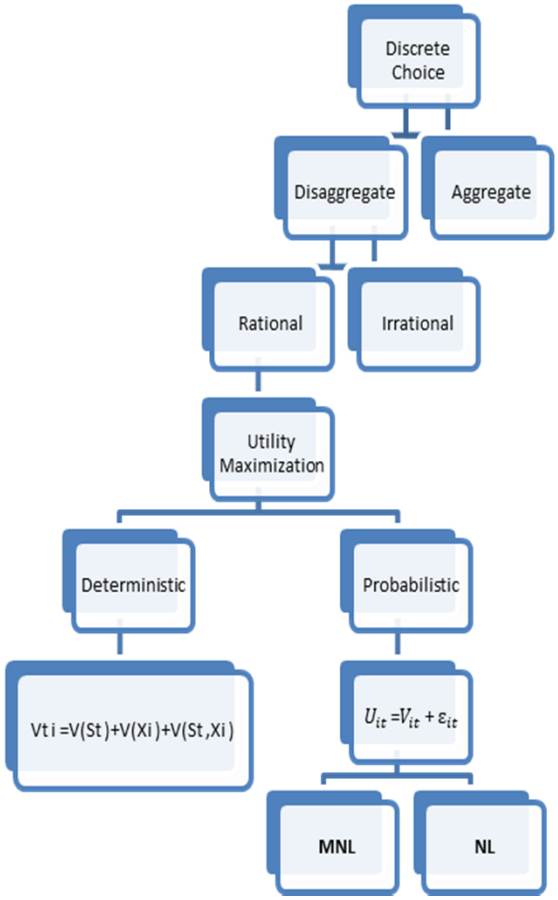
[4] Bhat, C., Koppelman, F.S. (2006) "A self-instructing Course in Mode Choice modeling: Multinomial and Nested Logit Models" Final report, prepared for the U.S. Department of Transportation, Federal Transit Administration, Washington, D.C.

Choice Modeling: What's that?

- This behavior prediction can be made in an aggregate or disaggregate manner. In the aggregate approach, one is modeling a group of individuals which decisions are based on some statistical associations of the group's attributes with alternatives characteristics.
- On the other hand, the disaggregate approach recognizes that this group behavior is based on individuals making decisions independently of each other as a function of their individual attributes and the characteristics of the available alternatives.
- The disaggregate approach provides a better understanding of why an individual make choices rather than the statistical aggregation of individuals making it more sensitive to behavior changes. This creates a more direct cause/effect connection which is more transferable in time and space.

Choice Modeling: What's that?

- Utility is an indicator of value to an individual. Generally, we think about utility as being derived from the attributes of alternatives or sets of alternatives; e.g., the total set of groceries purchased in a week. The utility maximization rule states that an individual will select the alternative from his/her set of available alternatives that maximizes his or her utility.
- If analysts thoroughly understood all aspects of the internal decision making process of choosers as well as their perception of alternatives, they would be able to describe that process and predict choice using deterministic utility models. Experience has shown, however, that analysts do not have such knowledge; they do not fully understand the decision process of each individual or their perceptions of alternatives and they have no realistic possibility of obtaining this information. The data and models used by analysts describe preferences and choice in terms of probabilities of choosing each alternative rather than predicting that an individual will choose a particular mode with certainty.



Choice Modeling: Utilities


- Utilities are represented by mathematical functions. These functions or equations seek to correlate attributes of the individual (such as income, car ownership, etc.) and the alternatives' characteristics (for example, travel cost, travel time, etc.)
- Any attribute that the modeler thinks has an influence on the individual when making a choice should be included. More on this later.
- The absolute value of the function is not important. What the modeler wants is the relative value of the different alternatives being considered from the choice set and how each of those attributes impact the individuals' selection process.
- *The data and models used by analysts describe preferences and choice in terms of probabilities of choosing each alternative rather than predicting that an individual will choose a particular one with certainty*

- Utility associated with the Attributes of the Alternative:

$$V(X_i) = \gamma_1 \times X_{i1} + \gamma_2 \times X_{i2} + \cdots + \gamma_k \times X_{iK} \qquad 3.4$$

where γ_k is the parameter which defines the direction and importance of the effect of attribute k on the utility of an alternative and

X_{ik} is the value of attribute k for alternative i.

- An example for three different travel modes (Drive Alone (DA), Shared Ride (SR) and Transit (TR) alternatives is:

$$V(X_{DA}) = \gamma_1 \times TT_{DA} + \gamma_2 \times TC_{DA}$$
$$V(X_{SR}) = \gamma_1 \times TT_{SR} + \gamma_2 \times TC_{SR}$$
$$V(X_{TR}) = \gamma_1 \times TT_{TR} + \gamma_2 \times TC_{TR} + \gamma_3 \times FREQ_{TR}$$

- The Utility associated to the characteristics of the Decision Maker, in the specific example above would be:

$$V(S_{DA}) = \beta_{DA,0} \times 1 + \beta_{DA,1} \times Inc_t + \beta_{DA,2} \times NCar_t$$
$$V(S_{SR}) = \beta_{SR,0} \times 1 + \beta_{SR,1} \times Inc_t + \beta_{SR,2} \times NCar_t$$
$$V(S_{TR}) = \beta_{TR,0} \times 1 + \beta_{TR,1} \times Inc_t + \beta_{TR,2} \times NCar_t$$

Choice Modeling: Logit Models


- Different surveys and data collection methods are used to build these utility functions.
- The intuition, statistical analysis and judgment are used to specify the utility function mathematical form. Past experience has demonstrated the superiority of certain forms over others, depending on what type of choice one wants to model (toll choice, mode choice, destination choice, etc.)
- The multinomial logit model (MNL) has been widely to use to determine the choice probabilities of each alternative under consideration. The general expression for the probability of choosing alternative "i" (i= 1, 2,…,J) from a set of "J" alternatives is:

$$\Pr(i) = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)}$$

- Following our mode choice example, this can be written as:

$$\Pr(DA) = \frac{\exp(V_{DA})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})}$$

$$\Pr(SR) = \frac{\exp(V_{SR})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})}$$

$$\Pr(TR) = \frac{\exp(V_{TR})}{\exp(V_{DA}) + \exp(V_{SR}) + \exp(V_{TR})}$$

- Utility associated with the Attributes of the Alternative:

$$V(X_i) = \gamma_1 \times X_{i1} + \gamma_2 \times X_{i2} + \cdots + \gamma_k \times X_{iK} \qquad 3.4$$

where γ_k is the parameter which defines the direction and importance of the effect of attribute k on the utility of an alternative and

X_{ik} is the value of attribute k for alternative i.

- An example for three different travel modes (Drive Alone (DA), Shared Ride (SR) and Transit (TR) alternatives is:

$$V(X_{DA}) = \gamma_1 \times TT_{DA} + \gamma_2 \times TC_{DA}$$
$$V(X_{SR}) = \gamma_1 \times TT_{SR} + \gamma_2 \times TC_{SR}$$
$$V(X_{TR}) = \gamma_1 \times TT_{TR} + \gamma_2 \times TC_{TR} + \gamma_3 \times FREQ_{TR}$$

- The Utility associated to the characteristics of the Decision Maker, in the specific example above would be:

$$V(S_{DA}) = \beta_{DA,0} \times 1 + \beta_{DA,1} \times Inc_t + \beta_{DA,2} \times NCar_t$$
$$V(S_{SR}) = \beta_{SR,0} \times 1 + \beta_{SR,1} \times Inc_t + \beta_{SR,2} \times NCar_t$$
$$V(S_{TR}) = \beta_{TR,0} \times 1 + \beta_{TR,1} \times Inc_t + \beta_{TR,2} \times NCar_t$$

Choice Modeling: Utility Estimation

- Data from surveys include information describing the trip maker and his/her household, the chosen travel mode, some trip context (time of day, frequency of travel OD data). Network analysis can provide zone-to-zone in-vehicle travel time for highway and transit, toll and parking cost, other relevant supply-side attribute.
- This presentation will not discuss survey design, and sampling techniques. However, the modeler should make judgment calls on the appropriate variables' data to be collected in enough market segments that are relevant to the choice under analysis: If mode choice is the issue then, how many travel modes should be included; how many trip purposes; how many income level segments, presence of workers in the household, etc..
- One should collect enough data on each market segment to be able to improve the statistical validity of the estimation results.

Data Layout Type I: Trip Format

Trip Number	*Income*	*Alternative 1 Time*	*Alternative 1 Cost*	*Alternative 2 Time*	*Alternative 2 Cost*	*Alternative 3 Time*	*Alternative 3 Cost*	*Alternative Chosen*
1	30000	30	150	40	100	20	200	1
2	30000	25	125	35	100	0	0	2
3	40000	40	125	50	75	30	175	3
4	50000	15	225	20	150	10	250	3

Data Layout Type II: Trip-Alternative Format

Trip Number	*Alternative Number*	*Number of Alternatives*	*Income*	*Time*	*Cost*	*Alternative Chosen*
1	1	3	30000	30	150	1
1	2	3	30000	40	100	0
1	3	3	30000	20	200	0
2	1	2	30000	25	125	0
2	2	2	30000	35	100	1
3	1	3	40000	40	125	0
3	2	3	40000	50	75	0
3	3	3	40000	30	175	1
4	1	3	50000	15	225	0
4	2	3	50000	20	150	0
4	3	3	50000	10	250	1

Choice Modeling: Utility Estimation

- There are a number of statistical packages' software that can be used to estimate the parameters of a MNL formulation, such as: MATLAB, ALOGIT, LIMDEP/NLOGIT, ELM, BIOGEME software.
- The modeler inputs the survey data and the software provides the values of the coefficients (or parameters γ_i in equation 3.4) and depending on the software, a number of statistical tests to measure how strongly this estimated parameters belong to the utility as specified.
- Based on those statistics one can accept or reject the null hypothesis, which is that the estimated parameter is equal to zero.

- http://www.alogit.com/index.htm; http://www.limdep.com/;
- http://elm.newman.me/;
- http://transp-or.epfl.ch/index.php

```
Hague Consulting Group                                          Page  5
ALOGIT Version 3.8F (135)                          17:45:02 on  1 Jun 98

      Data --> SF.DAT; Input--> SFM1.BIN; MNL Model 1; Cost,Time,Income

Convergence achieved after   5 iterations
Analysis is based on   5029 observations

Likelihood with Zero Coefficients =  -7309.6010
Likelihood with Constants only    =  -4283.5050
Initial Likelihood                =  -7309.6010
Final value of Likelihood         =  -3626.1860
   Rho-Squared w.r.t. Zero        =  .5039
   Rho-Squared w.r.t. Constants   =  .1535

ESTIMATES OBTAINED AT ITERATION  5
Likelihood =  -3626.1860

              Travel     Travel     Income     Income     Income     Income
              Time       Cost       SR 2       SR 3+      Transit    Bike
Estimate  -.4920E-02 -.5134E-01 -.2170E-02  .3576E-03 -.5286E-02 -.1281E-01
Std. Error  .239E-03   .310E-02   .155E-02   .254E-02   .183E-02   .532E-02
"T" Ratio  -20.6      -16.6       -1.4        .1        -2.9       -2.4

            Income    Constant   Constant  Constant   Constant   Constant
            Walk      SR 2       SR 3+     Transit    Bike       Walk
 Estimate  -.9686E-02 -2.178     -3.725     -.6709    -2.376     -.2068
Std. Error  .303E-02   .105       .178       .133       .305       .194
"T" Ratio  -3.2       -20.8      -21.0      -5.1       -7.8       -1.1
```

Choice Modeling: Utility Estimation

- Each of these statistical packages can present results in slightly different ways: The LIMDEP software outputs has the "z" statistics (other packages label this as a "t" statistic).
- The test-statistics should be usually |t|>1,or more precisely,

Confidence Level	Critical t-value (two-tailed test)
90%	1.645
95%	1.960
99%	2.576
99.5%	2.810
99.9%	3.290

- The sign of the parameter is also important since they should be consistent with theory, intuition and judgment regarding the expected impact of the corresponding variables.

LIMDEP Estimation Results

```
Discrete choice (multinomial logit) model
Maximum Likelihood Estimates
Dependent variable                Choice
Number of observations             5029
Iterations completed                  6
Log likelihood function        -3626.186
Log L: No coefficients   =  -9010.75837
```

Variable	Coefficient	Standard Error	z=b/s.e.	P[|Z|≥z]
Travel Cost	-0.49204E-02	0.23890E-03	-20.597	0.00000
Travel Time	-0.51341E-01	0.30994E-02	-16.565	0.00000
Income Shared Ride 2	-0.21700E-02	0.15533E-02	-1.397	0.16241
Income Shared Ride 3+	0.35756E-03	0.25377E-02	0.141	0.88795
Income Transit	-0.52864E-02	0.18288E-02	-2.891	0.00385
Income Bike	-0.12808E-01	0.53241E-02	-2.406	0.01614
Income Walk	-0.96863E-02	0.30331E-02	-3.194	0.00141
Shared Ride 2 Constant	-2.1780	0.10464	-20.815	0.00000
Shared Ride 3+Constant	-3.7251	0.17769	-20.964	0.00000
Transit Constant	-0.67095	0.13259	-5.060	0.00000
Bike Constant	-2.3763	0.30450	-7.804	0.00000
Walk Constant	-0.20682	0.19410	-1.066	0.28664

Choice Modeling: Computation Example


Multinomial Utility Expression

$$U_a = \beta_{wat} \times Time_{wat} + \beta_{wt} \times Time_{wt} + \beta_{ivtt} \times Time_{ivtt} + \beta_{pc} \times Cost_{pc} + \beta_{oc} \times Cost_{oc} + k_a$$

Where:

U_a = Utility of mode "a"

β_t = Time coefficient (walk access time, WAT; wait time, Wait; in-vehicle travel time, IVTT)

β_c = Cost coefficient (parking cost, PC; other cost, OC)

$Time_a$ = Time of travel using mode "a" (walk access time, WAT; wait time, Wait; in-vehicle time, IVTT)

$Cost_a$ = Cost of travel using mode "a" (parking cost, PC; other cost, OC)

k_a = Mode specific constants

Impedances	Transit	Drive Alone
Walk Access Time (WAT)	43.2	8
Wait Time (WT	5	0
In vehicle travel time (IVTT)	18.8	15.9
Parking cost (PC)	5	25
Other Cost (OC)	75	114
Coefficients (for HBW)		
Walk access time	0.08	0.08
Wait time	0.14	0.14
In vehicle travel time	0.015	0.015
Parking cost	0.02	0.02
Other cost	0.005	0.005
Constants (for HBW)		
Mode specific constant	1.087	1.587

Transit Utility

$$\begin{cases} U_{tr} = 0.08 \times 43.2 + 0.14 \times 5 + 0.015 \times 18.8 + 0.02 \times 0 + 0.005 \times 75 + 1.087 \\ U_{tr} = 5.90 \\ e^{U_{tr}} = 365.037 \end{cases}$$

Drive Alone Utility

$$\begin{cases} U_{da} = 0.08 \times 8 + 0.14 \times 0 + 0.015 \times 15.9 + 0.02 \times 25 + 0.005 \times 114 + 1.587 \\ U_{da} = 3.54 \\ e^{U_{da}} = 34.312 \end{cases}$$

So the mode share results for this particular O-D pair is:

$$\Pr(da) = \frac{e^{U_{da}}}{e^{U_{da}} + e^{U_{tr}}} = \frac{34.312}{34.312 + 365.037} = 47.65\%$$

$$\Pr(tr) = \frac{e^{U_{da}}}{e^{U_{da}} + e^{U_{tr}}} = \frac{365.037}{34.312 + 365.037} = 52.35\%$$

Choice Modeling: Computation Example

- The calculations as described in the previous example is an application to one pair of zones: Trips produced in one traffic analysis zone (TAZ) going to one other zone. It shows how those person trips would split by travel mode.
- In the context of a full area model, there would be thousands even millions of i-j pairs of TAZs and the results of previous calculation would go in a table for each mode, within a matrix, for each trip purpose.
- The example was about mode choice calculation and every input variable in the MNL utility expression (time, cost, utility coefficients and constants) are being pulled out from corresponding data files, either text, dbf, or matrix formats.
- This presentation is not including Nested Logit models, which in essence, are a series of MNL expressions grouped in different tiered levels (nested levels). The nesting coefficients are also estimated from the available data using available estimation software.

mainme00.app (C:\FSUTMS\Statewide\T... | PANDA_C07_PK.DBF (C:\FSUTMS\State... | ZONEDATA_07C.dbf (C:\FSUTMS\Statew...

MF_1AUTO	MF_2AUTO	HMDU	HMOCC	HMPOP	IND_EMP	COMM_EMP	SERV_EMP	TOT_EMP	SCHOOL	LONGPARK	SHORTPARK	NE
60	12	0	0	0	1	20	25	46	0	0	0	
60	12	0	0	0	0	2	9	11	0	0	0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	
60											0	

TFARE - Transit Fares

OD	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	75	75	0	50	50	50	0
2	0	0	0	0	0	0	0	0	0	0	0
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											

mainme00.app (C:\FSUTMS\Statewide\T... | CONGSKIM_C07_AM.MAT-*1 STOLL (C:\...

*1 STOLL | 2 SDISTANCE | 3 STIME | 4 STERMINALTIME | 5 STIME2 | 6 HTOLL | 7 HDISTANCE | 8 HTIME | 9 HTIME2

	1	2	3	4	5	6	7	8	9	10
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										

mainme00.app (C:\FSUTMS\Statewide\T... | PANDA_C07_PK.DBF (C:\FSUTMS\Statew...

TAZ	HBWP1	HBWP2	HBWP3	HBWP4	HBWP5	HBWP6	HBWP7	HBSHP1	HBSHP2	HBSHP3	HBSHP4
1	0	0	0	0	0	0	0	0	0	0	0
2	0.028	0.037	0.09	0.087	0.448	0.959	1.659	0.033	0.017	0.024	0.015
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	1.371	0.487	1.23	0.933	2.073	5.194	8.518	1.141	0.222	0.29	0.184
6	0.08	0.026	0.063	0.058	0.267	0.626	1.08	0.087	0.027	0.029	0.014
7	0.102	0	0	0	0.144	0.359	0.381	0.112	0	0	0
8	0.034	0	0	0	0.075	0.19	0.225	0.045	0	0	0
9	0.223	0.085	0.221	0.161	34.069	0.829	1.269	0.234	0.09	0.078	0.045
10	10.31	3.644	9.429	7.367	14.508	37.034	58.314	8.264	1.287	1.963	1.316
11	0	0	0	0	46.482	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0
16	0.606	0.215	0.558	0.46	0.849	2.171	3.388	0.468	0.052	0.103	0.072
17	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0
23	0.332	0.121	0.27	0.191	0.604	1.403	2.405	0.316	0.113	0.105	0.053

Destination Choice: A Modeling Primer

Destination Choice

Destination Choice: How it all started

- Trip distribution was always a difficult problem due to the lack of good O-D data and household's survey, which forces calibration/re-calibration using transferred parameters.

- Poor trip distribution results in poor mode choice, which generates poor transit/highway assignment.

- FSUTMS trip distribution model relies on the gravity model (GM), which groups all production and attractions and distribute them based on impedance only.

Destination Choice: How it all started

- Trip distribution reviews performed as early as 2008[1] and even before through research performed by the Lehman Center at Florida International University[2] suggested the use of alternative explanatory variables beyond just impedance, and different model structures beyond gravity model.

- Another study implemented those suggestions in 2012. The report states : "The gravity analogy works for trip distribution in mono-centric urban regions where accessibility to transit plays little to no role in choice of destination. This is no longer the case in some Florida urban areas, where there may be more than one dominant attraction region, multiple and important suburban-to suburban trip flows, and where there is interest in understanding the contribution of transit towards achieving more sustainable urban development patterns."[3]

[1] New FSUTMS Framework, a report prepared by AECOM for the Florida Department of Transportation, Systems Planning Office, August 21, 2008
[2] Refinement Of FSUTMS Trip Distribution Methodology, Final Report, Contract No. BB942. Prepared for the Research Office of the Florida Department of Transportation by the Lehman Center for Transportation Research, Department of Civil & Environmental Engineering, Florida International University, September 2004.
[3] http://www.fsutmsonline.net/images/uploads/reports/Trip_Distribution_Review_and_Recommended_Improvements_.pdf. Accessed on August 28, 2015

Destination Choice: Advantages

- Destination choice model is widely suggested in the literature as a good replacement for gravity model. A better market segmentation using better household characteristics (like income level, car ownership, workers present, etc.) would avoid some of the issues mentioned earlier. It provides a better match between, for instance, work trips made by households of a particular income level and job sites corresponding to that income level.

- The basic assumption is that better defining the market segments, the distribution may be more sensitive to changes in demographics, and accessibility changes. It is known that the GM may respond illogically to changes in levels of service — *improved accessibility to a given destination may cause a disproportionate increase in total trips, and/or an increase of trips using the mode(s) whose accessibility did not change. Both results are undesirable and may lead to erroneous assessments of the impact of transit (or highway) improvements*.

Destination Choice: Advantages

- For example, total person trips between two zones increase because of better travel times (usually when you improve capacity in either transit or highway), one should expect that the additional demand would be allocated to the corresponding mode that was improved.

- Destination choice model is very flexible, which it should be obvious from the utility function expressions: One can include the variables of interest that capture the key market differences. This is a characteristic of choice models in general, and it is only limited to the ability to capture enough data (sample size) for the specific market or sub-market to derive statistically significant parameters.

Destination Choice: Proposed Solution

- The MNL model is the preferred functional form to implement destination choice. This is the reason why this presentation was paying attention on this form of choice model.
- When traveling, users wants to maximize their trip utility. This implies not just minimize time and cost of the trip, but also from all possible destinations, where is the best location to go to: Not a blind shortest path type of distribution, like in GM.
- *The utility of a destination is a function of multi-modal accessibilities and preferences, the attractiveness of the destination zone, person and household attributes, and other unknown, un-included attributes of the trip maker or the destination zone*.

Destination Choice: Model Form and Utility Function

- The probability that trip m produced in zone i chooses destination zone j is given by the utility of zone j and the utility of all other possible destinations.

- "m" actually represents trips done by different "markets", in the case of the Florida prototype it means:

i. Zero Car Households – All Income
ii. Cars less than Workers – Low Income
iii. Cars less than Workers - Medium Income
iv. Cars less than Workers – High Income
v. Cars equal to or greater than Workers - Low Income
vi. Cars equal to or greater than Workers - Medium Income
vii. Cars equal to or greater than Workers – High Income

- Special trip market populations could include the following:

- Seasonal residents
- Visitors
- Air passengers
- Special events

- The mode choice Logsum inclusion accounts for travel impedance by mode. The mode choice utilities were developed using modal preferences by income-stratified impedance coefficients, so that the choice of the destination and the choice of mode are linked to each other..

$$P_{ijm} = \frac{e^{U_{ijm}}}{\sum_j e^{U_{ijm}}}$$

$$U_{ijm} = \theta \times L_{ijm} + \sum_k \beta^k D_{ij}^k + \sum_k \delta_m^k N_m^k D_{ij}^k + \sum_k \gamma_m^k M_i^k IZ_j + Ln(A_{jm})$$

Where:

θ, β^k, δ_m^k and γ_m^k are parameters to be estimated

L_{ijm} is the mode choice logsum of trip market m,

$\sum_k \beta^k D_{ij}^k$ Represents distance

$\sum_k \delta_m^k N_m^k D_{ij}^k$ Represents attributes of the trip market (e.g., income, auto availability)

$\sum_k \gamma_m^k M_i^k IZ_j$ Represents attributes of trip production zone (e.g., residential density) to account for intra-zonal activity

$Ln(A_{jm})$ This is called the "size" variable, which is the equivalent of the attractions in a GM. It could be calculated as a regression equation based on different employment categories, school enrollment, population, etc.

Destination Choice: Model Estimation

- The required dataset to build the utility functions are the same as a household survey data used to construct the mode choice utility expressions, although relevant data fields might be slightly different.
- For instance, in mode choice the data of interest needs to include the transportation mode.
- Developing destination choice functions, the chosen attraction zone has to obviously be included. Of course, depending on the market segmentation determined to be represented with specific parameters/constants, additional variables such as household income, car availability, number of workers in the household, etc. would also be selected.
- As the utility functions needs highway travel time and distance, employment, mode choice logsums per market segment, then these variables need to be added to the household survey from previous travel demand model runs.
- The selection of the choice set data is important for practical reasons: One cannot include all possible TAZs in the region; instead a sample of destinations is chosen for each TAZ (e.g., based on number of total employment, and distance). Several choice sets could be used to improve the robustness of the estimation results.

Trip record data	Trip Id
	Production TAZ (PTAZ)
	Attraction TAZ (choice)
	Trip purpose
	Household income
Destination choice set data	TAZ 1
	Travel time PTAZ to TAZ 1
	Travel distance PTAZ to TAZ 1
	Mode choice logsum PTAZ to TAZ 1
	Employment TAZ 1

	TAZ N
	Travel time PTAZ to TAZ N
	Travel distance PTAZ to TAZ N
	Mode choice logsum PTAZ to TAZ N
	Employment TAZ 3

Source: http://www.fsutmsonline.net/images/uploads/reports/Trip_Distribution_Review_and_Recommended_Improvements_.pdf

Destination Choice: Model Calibration

- The calibration of a destination choice model is no different than the calibration of a gravity model, in that the goal is to reproduce the observed trip length frequencies and district-level flow patterns.

- Typical destination choice calibration targets include:
  - ❑ Average trip lengths
  - ❑ Trip length frequency distribution
  - ❑ Percent of intra-zonal trips
  - ❑ Coincidence ratio
  - ❑ ACS worker flow pattern (county level or smaller geography)
  - ❑ Trip flow pattern (20-district geography)
  - ❑ Implied transit mode shares

- Any of these targets may be stratified by trip purpose, time period and trip market segment, as allowed by the sample size.

Destination Choice: Disadvantages


- Utility functions estimation: As in any other discrete choice models, data collection and processing is needed to estimate the coefficients and constants required by the utility function. This adds time/budget to the overall TDM development.

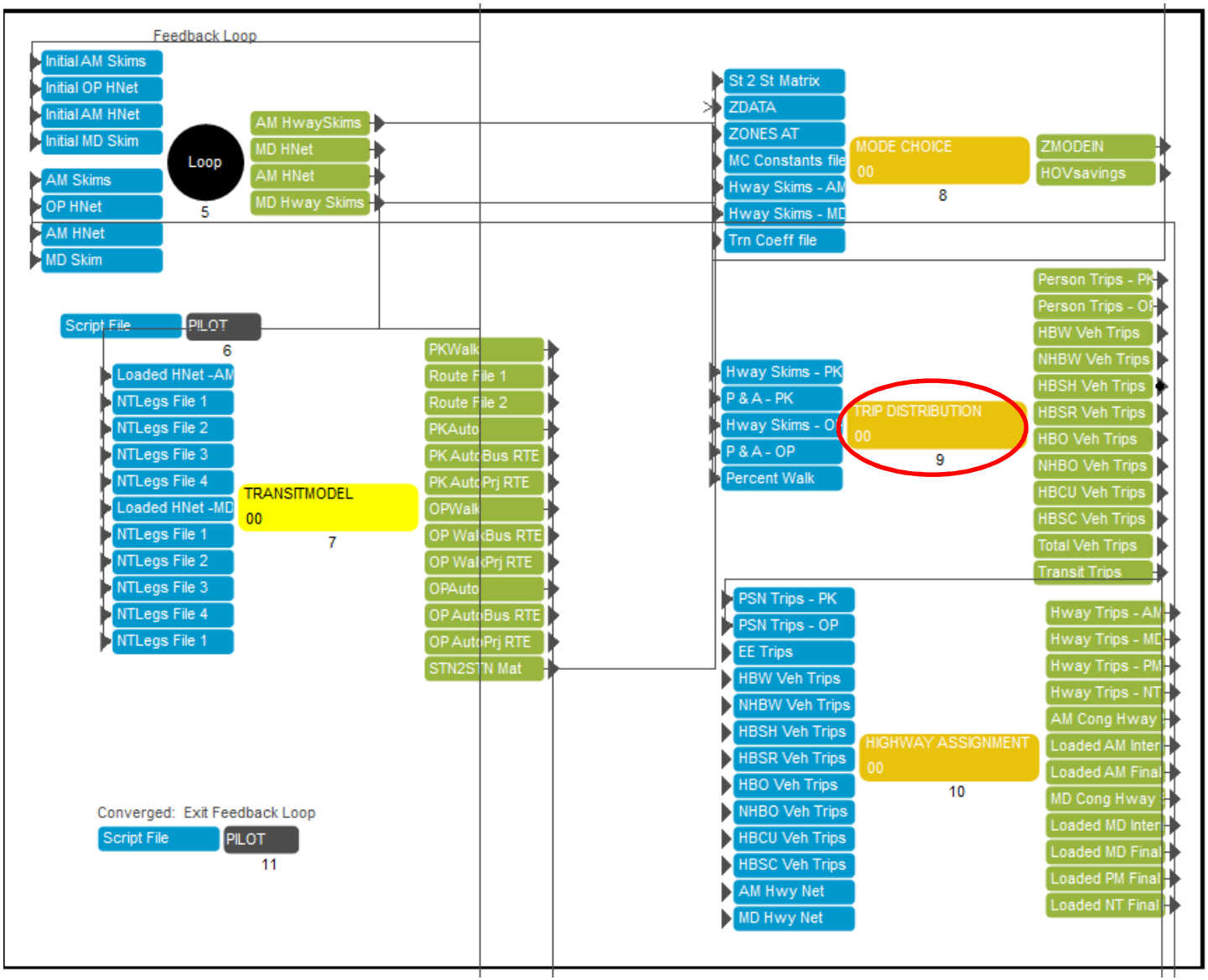
- Sample size should be sufficient in every market segment you want to analyze and sometimes it is not readily available. For example, the demonstration model that follows used the National Household Travel Survey (NHTS) Florida add-on, which did not have enough data in some of the market segments (for the Tallahassee region) to be able to obtain a statistically valid result, so parameters from elsewhere were used.

- Run-time of the models could be significantly higher than traditional methods (less defined market segments and trip purposes, by time of day)

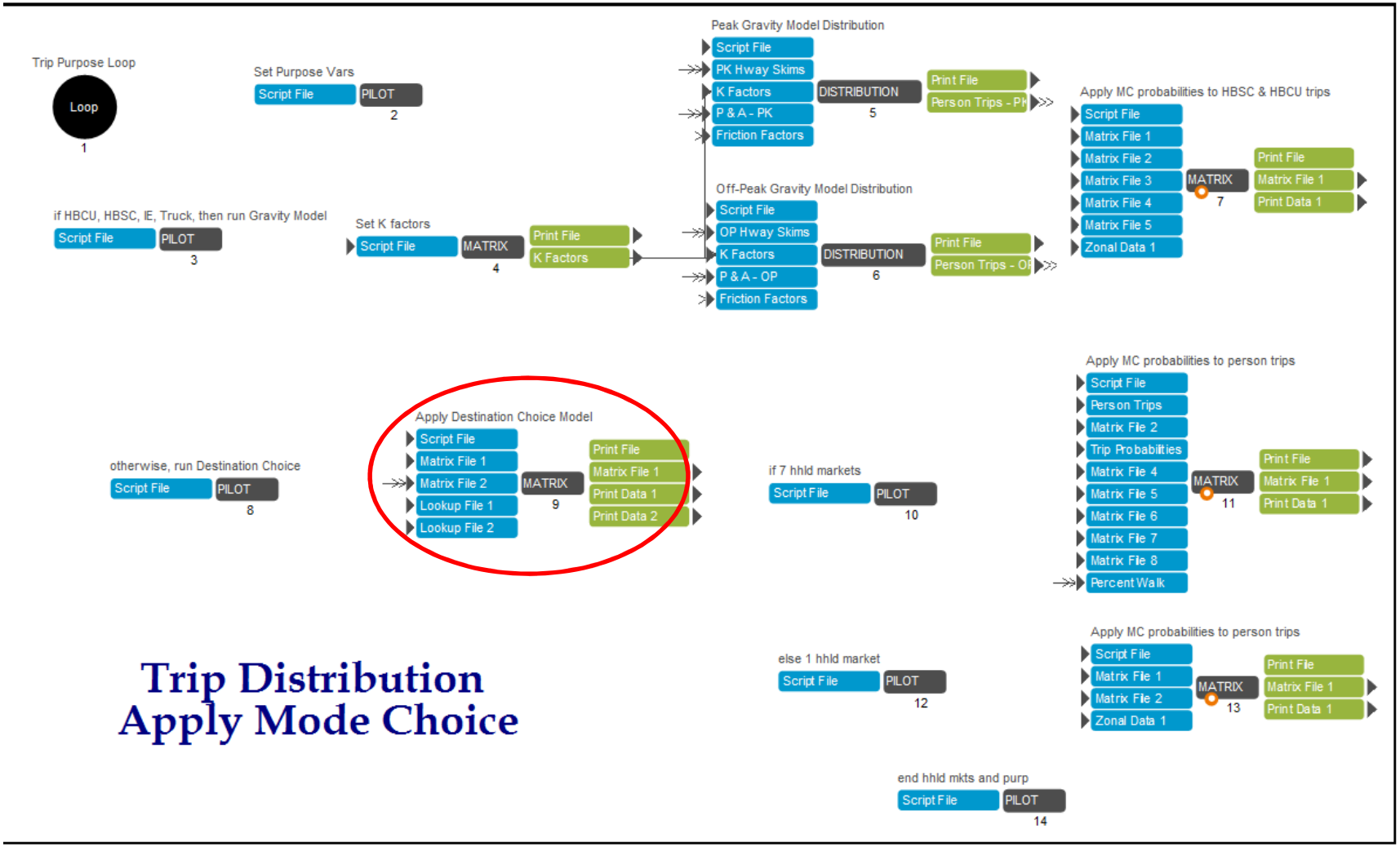
Destination Choice: A Modeling Primer

Florida Application: CRTPA model

Destination Choice: CRTPA demo



Destination Choice: CRTPA demo



Destination Choice: CRTPA Market Segments

- As mentioned before, *FSUTMS uses gravity models to distribute all trips. These trip distribution models are stratified by trip purpose, but not by household attributes (income or auto ownership, for example) or time period*

- *The lack of trip market segmentation besides trip purpose can also lead to incorrect forecasts of trip distribution patterns. Persons of restricted mobility (low income, zero car households, for example) exhibit, on average, shorter trip lengths than other persons. Low income workers are more likely to be employed in retail and service occupations, while high income workers are more often professionals.*

- *Without appropriate trip market segmentation the distribution model is unable to link the "right" workers to the "right" jobs.*

- To accomplish this market segmentation, the trip generation step has to include trip purposes that are including these better definition of mainly income, auto availability and presence of workers in the household. So for the CRTPA, these are the trip purposes:

- Home-Based Work (HBW)
- Home-Based College/University (HBCU)
- Home-Based School (HBSC)
- Home-Based Shop (HBSH)
- Home-Based Social/Recreation (HBSR)
- Home-Based Other (HBO)
- Non-Home-Based Work (NHBW)
- Non-Home Based Other (NHBO)

- All of the Home-based trips (except HBCU and HBSC) were further stratified in seven additional segments:

i. Zero Car Households – All Income

ii. Cars less than Workers – Low Income

iii. Cars less than Workers - Medium Income

iv. Cars less than Workers – High Income

v. Cars equal to or greater than Workers - Low Income

vi. Cars equal to or greater than Workers - Medium Income

vii. Cars equal to or greater than Workers – High Income

Source:
http://www.fsutmsonline.net/images/uploads/reports/Trip_Distribution_Review_and_Recommended_Improvements_.pdf

Destination Choice: CRTPA application

- The utility estimated coefficients/constants for the chosen utility are contained in a *dbf file, (destchoice_parameters.dbf).
- The logsum values were calculated in the previous mode choice step for every i-j pair of zones. In a simpler model, with no transit for example, this could be replaced with a impedance skim matrix.
- Distances are provided processing the network to output a matrix of every i-j distance.
- The size variable is provided in the trip generation step (the panda.dbf)

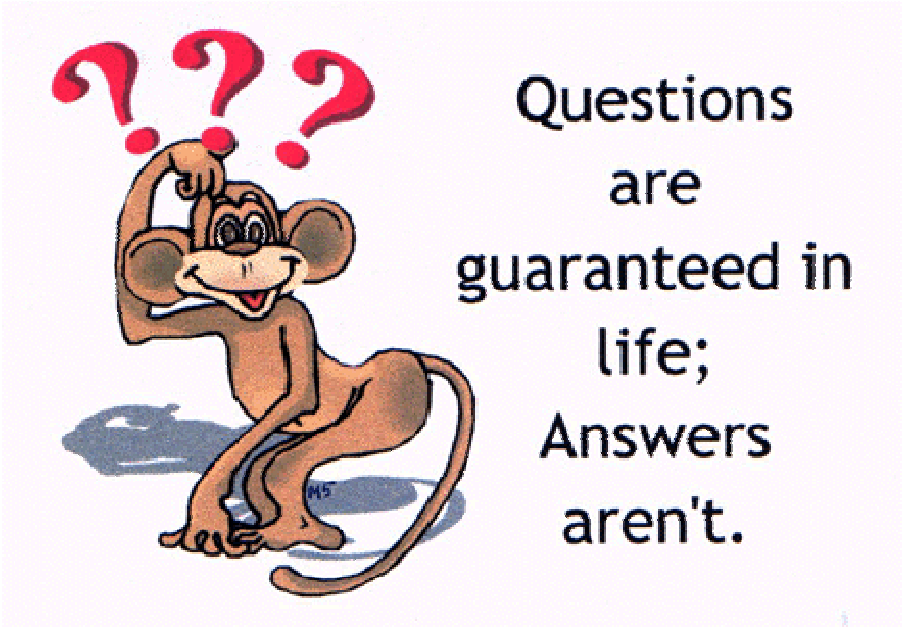
$$P_{ijm} = \frac{e^{U_{ijm}}}{\sum_j e^{U_{ijm}}}$$

$$U_{ijm} = \theta \times L_{ijm} + \sum_k \beta^k D_{ij}^k + \sum_k \delta_m^k N_m^k D_{ij}^k + \sum_k \gamma_m^k M_i^k IZ_j + Ln(A_{jm})$$

```
; COMPUTE DESTINATION CHOICE PROBABILITIES

LOOP _m = 1, @MARKETS@

  JLOOP

 INTRAZN = 0
 IF(J == I) INTRAZN = 1

  ; utility
  IF(ATTS[_m][J] > 0.0)

  MW[100+_m] = CLSUM * MW[_m] +
       CDIST1[_m] * MW[12] + CDIST2[_m] * POW(MW[12],2) + CDIST3[_m] * POW(MW[12],3) +
            CIZ[_m] * INTRAZN + LN(ATTS[_m][J])

   MW[110 + _m] = EXP(MW[100+_m])

  ELSE
   MW[100+_m] = 0.0
   MW[110 + _m] = 0.0
  ENDIF

 ENDJLOOP

  ; destination choice probabilities

  DENOM[_m] = ROWSUM(110+_m)

  IF(DENOM[_m] > 0)
    MW[120+_m] = MW[110+_m]/DENOM[_m]
  ELSE
    MW[120+_m] = 0.0
  ENDIF

  ; person trips

  MW[130+_m] = PRODS[_m][I] * MW[120+_m]
```

Source:
http://www.fsutmsonline.net/images/uploads/reports/Trip_Distribution_Review_and_Recommended_Improvements_.pdf

Questions



Contact info

Gabrielle.Matthews@dot.state.fl.us

Cesar.Segovia@dot.state.fl.us

Cesar.Segovia@aecom.com