MN－GEV モデルを用いた地域間交通モード選択行動の分析

Analyzing Inter-regional Travel Mode Choice Behavior with Multi Nested Generalized Extreme Value Model

屋井研究室　04M53017　庄　暁韵（Zhuang XiaoYun）

Keywords: multi-modal route choice, inter-regional travel, multi-nested GEV model, air, high-speed rail

1. Introduction

1.1 Background

Since the construction of High-Speed Rail (HSR) started in Japan from last 1960's, HSR and airlines have always maintained a conflict relationship. HSR were able to compete with airs on destinations which were formerly only served by airlines. They were ascendant in a way that having a more positive impact on the environment and by using less fuel. Once it became clear that HSR could provide comparable trip times in middle- distance travel more efficiently than air, indifference gave way to action and the competition between them became real.

Evaluations of proposal to improve the competitiveness of HSR in the middle-distance trips require the use of forecasting and policy analysis tools, for example, route choice model. In most studies, only the specifications of Level-of-service in trunk mode such as HSR or airlines are considered in the analysis of travellers' mode choice behaviours. However, passengers more often tend to use a combination of modes in a interregional trip. To carry on a study on an inter-regional trip, it is also necessary to consider the preceding/subsequent modes (access/egress modes) in the total trip. Before and after taking a trunk mode such as Air, HSR etc., access and egress modes connected to trunk modes exist, as the components of "Multi-modal Inter-regional Trips". These relationships should be analyzed thoroughly.

1.2 Research Objective

Evaluations of proposal to improve the competitiveness of HSR in the middle-distance trips requires the use of forecasting and policy analysis tools, for example, route choice model. However, passengers more often tend to use a combination of modes in an interregional trip. To conduct a study on an inter-regional trip, it is necessary to consider the preceding and subsequent modes (access/egress modes) in the total trip. To make it easy for policy analysis in multi-model trips, it is necessary to catch the similarities among different combinations. As for MNL and NL models, it is impossible to capture these properties, while it is hard for Probit or Mixed Logit Models to interpreter these similarities directly although they can help to estimate with a correlated structure. Therefore, Multi-Nested Generalized Extreme Value Model, or MN-GEV model appears to be the best solve for the problem.

A recent study by Coldren et al.[1)] (2005) firstly employs weighted nested logit model (MN-GEV model) as a tool evaluating the competition among air-travel itinerary shares of all East West markets in the United States and Canada. And then, Bovy et al.[2)] (2005) also chooses MN-GEV model as the way to estimate the result of intercity train service within the Rotterdam-Dordrecht region in The Netherlands.

This paper aims to give contributions to the evaluation of the policies aiming to more efficient and smooth transfer between travel modes. The paper focuses on the mode choice behavior in the entire trip covering each part of the trip, especially the transfer points. The paper also conducts a stability analysis of an advanced Multi-Nested GEV (MN-GEV) model to catch the bias between the estimation results and true values to show the capability and the limitation of this model with artificial datasets. At last, MN-GEV model is developed in this paper to deal with the analysis of multi-modal travel behaviors, using the empirical data of Japan.

2. Outline of MN-GEV Model

Passengers more often tend to use a combination of modes in an inter-regional trip, which can be also called an inter-modal trip. It could be roughly divided into 3 parts: home-end part (access mode), main part (trunk mode), and activity-end part (egress mode). There are varieties of models that could be introduced into the real case study.

Multi Nested GEV model, also called weighted nested-logit model, is a special case of GNL model. And, it is an extension of usual 2-level nested logit models. A recent study by Coldren and Koppelman[1)] (2005) firstly employs weighted nested logit model (Multi-Nest Generalized Extreme Value Model, or MN-GEV model) as a tool evaluating the competition among air-travel itinerary shares of all East West markets in the United States and Canada. And then, Bovy and Hoogendoorn-Lanser[2)] (2005) also chooses MN-GEV model as the way to analyze intercity travel behavior within the Rotterdam-Dordrecht region in The Netherlands.

For a model with two dimensional MN-GEV model, the mixing distribution can therefore be written as:

$$P_i = \alpha_1 P_{i,d1} + \alpha_2 P_{i,d2} \qquad (1)$$

where P_i : probability of the alternative *i* being chosen.

$P_{i,dn}$: probabilities of the alternative *i* being chosen along dimension *n*

α_d : weight for dimension *d*, which can be fixed, estimated or defined as a function of the logsum parameters. ($\sum \alpha_d = 1$ and $0 \le \alpha_d \le 1$).

Table 1. Categories of exiting models[3)] developed from Generalized Extreme Value (GEV) or not

Based on GEV theory	Multinomial Logit Model (MNL) – McFadden (1974)
	Nested Logit Model(NL) – Ben-Akiva (1973)
	Cross-Nested Logit Model (CNL) – P.Vovsha (1997)
	Generalised-Nested Logit Model (GNL) – Wen & Koppelman (2001)
	C-Logit Model – Cascetta (2001)
	Path-Size Logit Model – Ben-Akiva & Bierlaire (1999)
	Network GEV Model – Ramming & Daly (2006)
Not GEV theory	Probit Model - Daganzo (1976)
	Mixed Logit Model - Train & McFadden (2001)

3. Stability Analysis of the MN-GEV Model

3.1 Analysis Strategy and artificial data settings

A simple stability study of a practical example is to be executed to demonstrate the performance of MN-GEV model through the estimated parameters. Artificial data are introduced into the model building, including parameters to be estimated, explanatory variables and error terms. The objective of this analysis it to test to which extent the estimated results of MN-GEV model will be biased from the setting parameters with a variety settings of data. And also it can be treated as an exercise for the estimation of MN-GEV model.

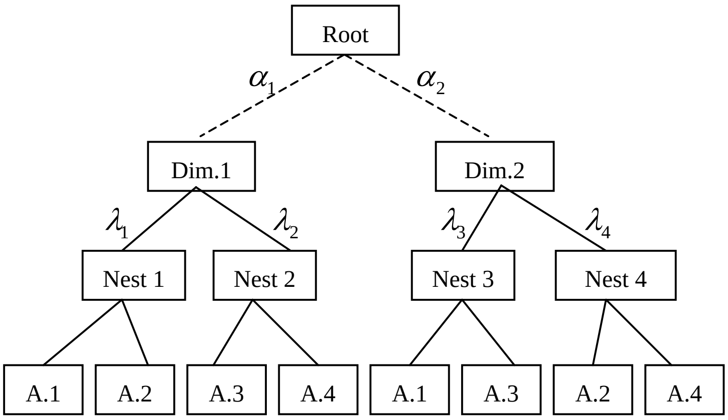


Figure 1. A simple structure for 2-dimension, 2-level MN-GEV model

* Dim. – dimension; A. – alternative.

A dataset with 500 observations will be set. And 20 trials will be done for one setting of α or λ.

The utility function for models in GEV family can be written as:

$$U_{ni} = V_{ni} + \varepsilon_{ni} + \ln G_i \qquad (2)$$

Where, G_i is the first derivative of GEV function; $V_{ni} = a_i + bx_i$, where a_i is the constant and x_i is the explanatory variable. And the GEV function[4)] for MN-GEV model as:

$$G = \sum_{d\in D}[\alpha_d \sum_{k\in d}(\sum_{j\in B_k} Y_j^{1/\lambda_k})^{\lambda_k}] \qquad (3)$$

The artificial data set here include $a_i = 5$ (for any alternative i), x_i (variable), $b = 1$ (the coefficient of these variables) and ε_i (error term within each alternative i respectively). Where x_i is drawn from normal standard distribution and ε_i is standard Gumbel distributed. And the third term can be expressed as:

$$\ln G_i = \ln \sum_{d\in D}[\alpha_d \sum_{k\in d}(\sum_{j\in B_k} Y_j^{1/\lambda_k})^{\lambda_k - 1} \cdot Y_i^{1/\lambda_k - 1}] \qquad (4)$$

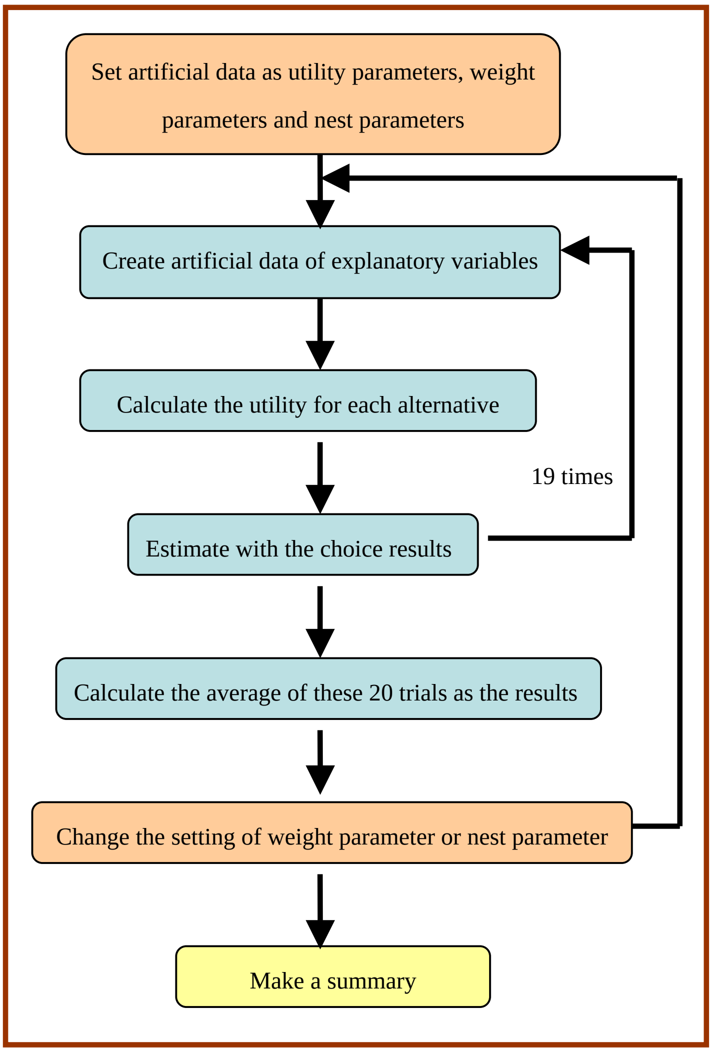


Figure 2. Algorithm for stability analysis

3.2 Analytical Results and Discussion

The estimations of all these sixteen settings are conducted with 20 trials for each. The average is shown as the estimated result of each setting from table 2 through table 5.

- *weight parameter* α

From these tables, almost all the estimated weight parameter α_1 is greater than the setting value of it.

As the exception, in table 4, relatively great logsum parameters are set in the first dimension and small ones in the other. It indicates a high independence of the nests in the first dimension and a low independence of the nests in the second dimension. In this case, the estimated weight parameters turn to be most close to the true value and the standard deviation is the least among all estimations.

Table 2. Estimation results for stability study (1)

True value	$\lambda_1=\lambda_2=\lambda_3=\lambda_4=0.5$			
	$\alpha_1=0.05$	$\alpha_1=0.1$	$\alpha_1=0.3$	$\alpha_1=0.5$
α_1	0.284	0.323	0.47	0.617
λ_1	0.782	0.705	0.639	0.648
λ_2	0.81	0.723	0.634	0.586
λ_3	0.479	0.454	0.497	0.515
λ_4	0.154	0.173	0.22	0.18
a_1	2.112	2.121	2.184	2.109
a_2	4.078	4.068	4.187	4.164
a_3	5.988	5.974	6.12	6.104
b	1.96	1.952	1.982	1.998

Table 3. Estimation results for stability study (2)

True value	$\lambda_1=0.1,\lambda_2=0.9,\lambda_3=0.2,\lambda_4=0.8$			
	$\alpha_1=0.05$	$\alpha_1=0.1$	$\alpha_1=0.3$	$\alpha_1=0.5$
α_1	0.311	0.262	0.382	0.578
λ_1	0.437	0.411	0.214	0.13
λ_2	0.797	0.78	0.732	0.83
λ_3	0.317	0.291	0.228	0.216
λ_4	0.383	0.399	0.353	0.324
a_1	1.875	1.823	1.846	1.871
a_2	4.031	3.944	3.903	3.875
a_3	5.955	5.868	5.817	5.79
b	1.988	1.979	1.966	1.957

While in table 2,3,5, small logsum parameters are set in the first dimension and great ones in the other. In this case, the estimated results most deviate form the true value and the estimated weight parameters in the first dimension are all greater than half.

It might be concluded from these two cases that **the group of nests, which have higher independence, will be lower weighted than the groups of nests with low independence.**

- *logsum parameter* λ

As the estimated logsum parameters, the results are always totally different from the true values. The estimations must be strongly bias from the average and not be stable.

Table 4. Estimation results for stability study (3)

True value	$\lambda_1 = 0.8, \lambda_2 = 0.9, \lambda_3 = 0.1, \lambda_4 = 0.2$			
	$\alpha_1 = 0.05$	$\alpha_1 = 0.1$	$\alpha_1 = 0.3$	$\alpha_1 = 0.5$
α_1	0.054	0.091	0.235	0.311
λ_1	0.376	0.415	0.734	0.757
λ_2	0.503	0.512	0.584	0.788
λ_3	0.15	0.161	0.212	0.366
λ_4	0.169	0.203	0.213	0.201
a_1	2.386	2.786	2.031	2.038
a_2	4.417	4.778	3.986	3.939
a_3	6.358	6.716	5.893	5.85
b	1.985	1.973	1.939	1.928

Table 5 Estimation results for stability study (4)

True value	$\lambda_1 = 0.1, \lambda_2 = 0.2, \lambda_3 = 0.8, \lambda_4 = 0.9$			
	$\alpha_1 = 0.05$	$\alpha_1 = 0.1$	$\alpha_1 = 0.3$	$\alpha_1 = 0.5$
α_1	0.574	0.566	0.703	0.796
λ_1	0.777	0.75	0.662	0.525
λ_2	0.808	0.808	0.713	0.704
λ_3	0.613	0.637	0.585	0.528
λ_4	0.25	0.262	0.215	0.116
a_1	1.922	1.897	2.027	2.175
a_2	3.812	3.786	3.959	4.142
a_3	5.671	5.651	5.84	6.056
b	1.902	1.907	1.932	1.962

- *Constants and parameter of explanatory variables*

From the estimation results, the estimated constants and parameter of explanatory variables turn to be around twice of true values, but all numerically similar.

As the standard deviation of these estimation results are much smaller than the average, the constant and the parameter of explanatory variables could be regarded as stable estimation results

4. Trip Data Used for Empirical Analysis

4.1 Data Source and Target Area

All data used in the estimation are taken from National Corridor Trips Survey of Japan in 2000. This survey was carried by Ministry of Land Infrastructure and Transportation of Japan, which made a detailed summary of yearly national travel description. Tokyo-Osaka corridor is chosen as the target area in this study, not only due to its proper distance around 500km, but also because it is a typical corridor in Japan for inter-regional travelers.

4.2 The data specification

Because in Tokyo-Osaka corridor, HSR and air have a 95% market share according to survey, the analysis here only treats these two transport modes as trunk modes in travels. In order to have a more advantaged dataset and make it easier to carry on the analysis, the unknown and minority alternatives are all eliminated. Finally, the new set of data decreases to 2588 individual trips.

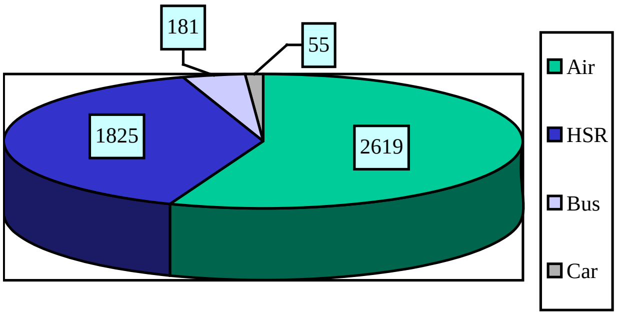


Figure 3. Share of Trunk Modes in Tokyo-Osaka trips

In the original data, there are 7 categories of access/egress mode: rail, bus, car, taxi, chartered bus, others and unknown. Since the first 4 modes dominate among all, observations with another 3 alternatives are eliminated at last and the final dataset is decreased to 2542 observations.

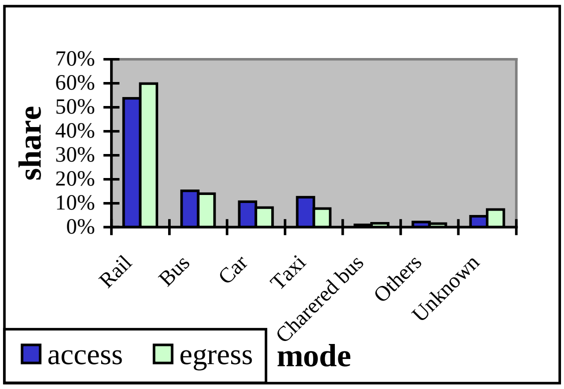


Figure 4. Share of access and egress modes in Air travel

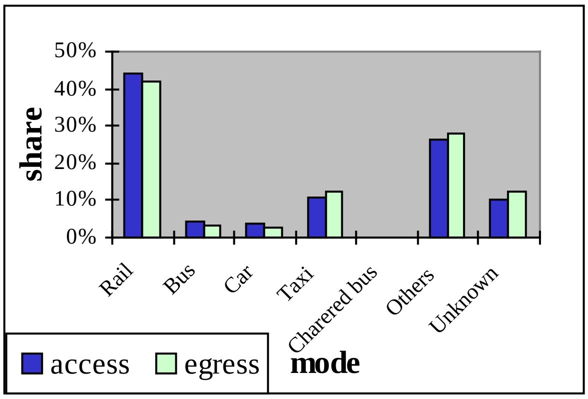


Figure 5. Share of access and egress modes in Hsr travel

In order to execute model analysis to find the preference pattern of travelers in inter-regional areas, all possible Level-of-Service variables should be listed in the dataset, including time and cost for each mode. It's easy to find the necessary data for trunk modes through ticket center. And a software named NITAS (National Integrated Transportation Analysis System) is used to calculate the variables (travel time and cost) of the best route (general lowest cost) for the access or egress modes.

5. Estimation Results

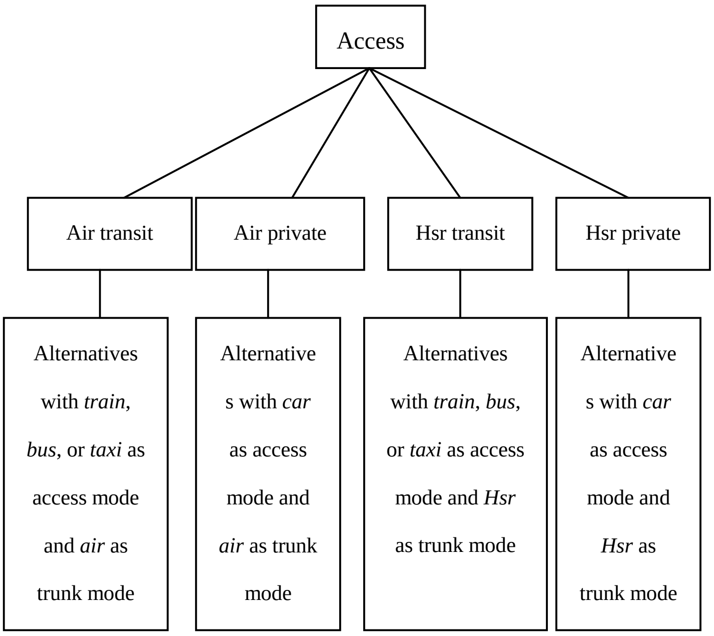
5.1 Explanatory variables and constants

Each alternative is defined as a combination of three components: access, trunk, and egress mode. Since there are 4 access choices, 2 trunk choices and 4 egress choices, totally there will be 32 alternatives. Travel time, cost for access or egress mode and only travel time for trunk mode (because the cost for HSR and air are very close.) were included in the model. And every mode in each part of the total trip has its specified constant. Thus, there are totally 10 alternative specific constants. The utility function is shown below as:

$$U = C_{access} + \beta_1 \cdot AccessTime + \beta_2 \cdot AccessCost + C_{trunk} + \beta_3 \cdot TrunkTime + C_{egress} + \beta_4 \cdot EgressTime + \beta_5 \cdot EgressCost + \varepsilon \quad (5)$$

5.2 Definition of nest structure

A three dimensional nested structure is set for the MN-GEV model estimation in this case. In access and egress dimension, there're 4 nests including AIR-transit (rail, bus, or taxi), AIR-private (car), HSR-transit and HSR-private. In trunk dimension there're 2 nests (AIR or HSR). The nest structure of access part is shown in figure 6. And the nest structure of the other parts are similar to it.



Figrue 6. nest structure of access part

5.3 Estimation results

The software named "Biogeme" is employed in the estimation. With the help from super-computer in Tokyo Institute of Technology, it still cost around 120 hours for one calculation. The parameters of MNL, NL and MN-GEV models are estimated. The estimated results of MNL and NL model are shown in table 6, while the results of MN-GEV model are shown in table 7.

For MN-GEV model estimation, there are varieties of the combinations of estimation results. With different nest parameter settings, the estimated results are totally different. From the estimation result (Table 8.), as expected, MN-GEV model is the best model in the term of goodness of fit. And NL provides the next best fit followed by MNL model.

Table 8. Adjusted rho-test from each model estimation

	MNL	**NL-access**	**NL-egress**	**MN-GEV**
adjusted rho-test	0.2889	0.2954	0.2962	0.3047

Table 6. Estimation results for MNL model and NL models

Model type	*MNL*		*NL-access*		*NL-egress*	
	value	t-test	value	t-test	value	t-test
constants						
Access						
rail	0		0		0	
bus	-1.19	-19.667	-0.567	-9.235	-0.77	-8.482
car	-2.239	-29.788	-2.703	-18.802	-1.564	-9.789
taxi	-1.107	-10.827	-0.567	-7.401	-0.737	-7.589
Trunk						
air	0		0		0	
hsr	0.721	1.543	1.383	3.013	0.0598	0.117
Egress						
rail	0		0		0	
bus	-1.294	-21.135	-0.628	-9.339	-0.802	-8.588
car	-2.64	-31.551	-1.483	-10.591	-2.46	-25.102
taxi	-1.145	-9.981	-0.631	-7.401	-0.669	-7.672
parameters						
Access time	-0.0307	-16.385	-0.0191	-9.852	-0.0239	-10.288
Access cost	-0.000155	-7.778	-0.000086	-6.809	-0.000132	-7.567
Trunk time	-0.0426	-7.63	-0.0442	-8.17	-0.0326	-4.962
Egress time	-0.0318	-15.358	-0.0223	-10.836	-0.0251	-13.215
Egress cost	-0.000233	-9.346	-0.000134	-7.536	-0.0002	-7.94
logsum parameters						
Air transit	-	-	0.477	10.432	0.587	9.785
Air private	-	-	0.907	6.61	0.776	8.995
Hsr transit	-	-	0.531	10.262	0.728	11.04
Hsr private	-	-	0.946	6.94	0.678	6.016
initial log-likelihood	-8809.9	-	-8814.04	-	-8814.65	-
final log-likelihood	-6252.38	-	-6191.38	-	-6184.64	-
likelihood ratio test	5115.05	-	5237.05	-	5250.51	-
adjusted rho-test	0.2889	-	0.2954	-	0.2962	-
number of individuals	2542	-	2542	-	2542	-

Table 7. Estimation results of MN-GEV model

	value	**t-test**		**value**	**t-test**		**value**	**t-test**		**value**	**t-test**
Constants			rail	0	-	**Logsum parameters**			hsr	0.253	0
Access			bus	-0.339	-9.168	**Access** air private	1(fixed)	-	**Alpha**		
rail	0	-	car	-3.218	-32.258	air transit	0.175	1.322	access	0.103	4.137
bus	-0.294	-8.583	taxi	-0.359	-6.735	hsr private	1(fixed)	-	egress	0.828	23.066
car	-1.041	-6.920	**Parameters**			hsr transit	1(fixed)	-	trunk	0.069	4.797
taxi	-0.315	-6.612	access time	-0.0149	-8.895	**Egress** air private	1(fixed)	-	init log-likelihood	-8812.49	
Trunk			access cost	-0.0000879	-7.533	air transit	0.197	7.022	final log-likelihood	-6105.41	
air	0	-	trunk time	-0.0437	-8.129	hsr private	0.696	5.150	likelihood ratio test	5397.21	
hsr	1.558	3.436	egress time	-0.0206	-11.008	hsr transit	0.145	6.435	adjusted rho-test	0.3047	
Egress			egress cost	-0.000151	-9.083	**Trunk** air	1(fixed)	-	number of individuals	2542	

Some conclusions from the estimation results of MN-GEV model:

- Constants: Rail, Hsr, rail modes appear to be the preferred mode in three parts of trips respectively. Car turns out to be the least preferred mode both in access and egress mode maybe because of its inconvenience in the transfer station.
- Utility Parameters: Travel time has more significant affect on the decision making rather than cost. Maybe because we only employed travel data in business purpose, travel time is much higher important for business men.
- Logsum Parameters: Lower logsum parameter means less independence of the nest, or greater correlations in the nest. In another word, it means larger similarity in the nest. We can conclude from the results for logsum parameters that the nests in access dimension are more independent than the nests in egress dimensions. Or the nests in egress dimension has a larger similarity.
- According to the value of alpha, the egress part in the travel is most weighted in all three dimensions.

5.4 Summary

The estimations of MNL, NL and MN-GEV model have been executed in this chapter. With the estimation results of each model, we can conclude that MN-GEV model performs best among three in multi-modal choice behavior, because it allows correlations at each portion of the trip. However, the estimation of MN-GEV model is a huge work that it will take a long time for one estimation.

6. Conclusion and Future Work

MN-GEV model has been used as a tool in the analysis of inter-regional multi-modal route choice behavior. Despite of its computational complex, this model shows its advantage in the estimation compared with MNL and NL models in the situation of an inter-modal travel. In the future, some political analysis could be evaluated by the result of the estimation, or by re-estimating by other variables with this tool.

Reference

1). Gregory M.C, and Frank S.K. (2005): Modeling the competition among air-travel itinerary share: GEV model development, Transportation Research Part A 39(2005) 345-365

2). Piet HL.Bovy, and Sascha H-L. (2005): Modelling route choice behviour in multi-modal ransport networks, Transportation (2005) 32: 341-368

3). Michael Scott Ramming. (2002): Network Knowledge and Route Choice, Doctor of Philosophy in Transportation, MIT

4). Chieh-Hua Wen, and Frank S.K. (2001): The Generalized Nested Logit Model, Transportation Research Part B 35 (2001) 627-6