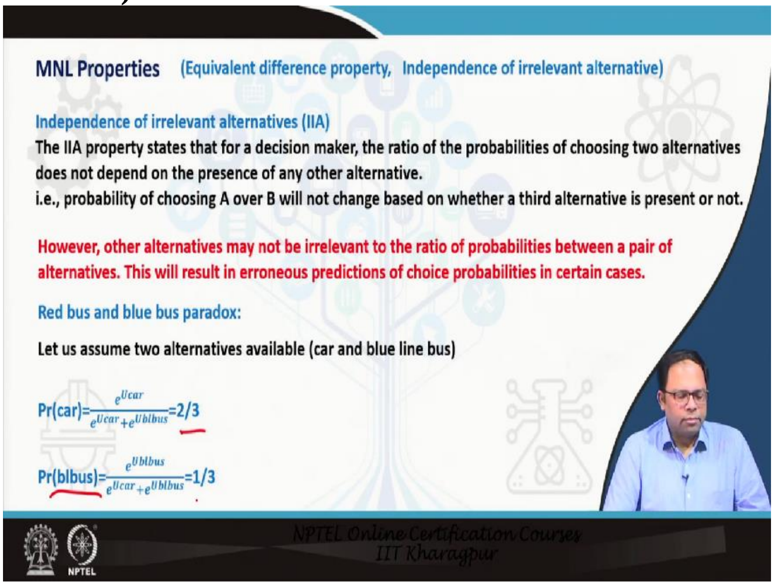
Urban Landuse and Transportation Planning
Prof. Debapratim Pandit
Department of Architecture and Regional Planning
Indian Institute of Technology, Kharagpur

Lecture - 40
Nested Logit Model

The different concepts covered in this lecture are independence of irrelevant alternatives (IIA) property of the MNL model, the nested logit model, the logsum parameter, and the complex nested logit structures that is possible in a nested logit model.

(Refer Slide Time: 00:49)



Independence of irrelevant alternatives (IIA):

The multinomial logit model has two intrinsic properties which are equivalent difference property and the independence of irrelevant alternatives property. As discussed in discrete choice theory (refer lecture 26), both these properties provide flexibility to apply MNL model in choice scenarios. However, the IIA property does not consider the relationship among groups of alternative.

The IIA property states that, for a decision maker, the ratio of probabilities of choosing two alternatives does not depend on the presence of any other alternative. It means that the probability of choosing alternative ‘A’ over ‘B’ will not change based on whether a third alternative is present or not. So, this property helps in estimating a model where different individual faces a different choice set, and therefore probability of one alternative is predicted in reference to another without considering other available alternatives.

On the other hand, this also creates certain problems. For example, other alternatives may not be irrelevant to the ratio probabilities between a pair of alternatives. So, this results in some erroneous prediction of choice probabilities in some cases. In order to explain this, let us consider a classic example of red bus and blue bus paradox.

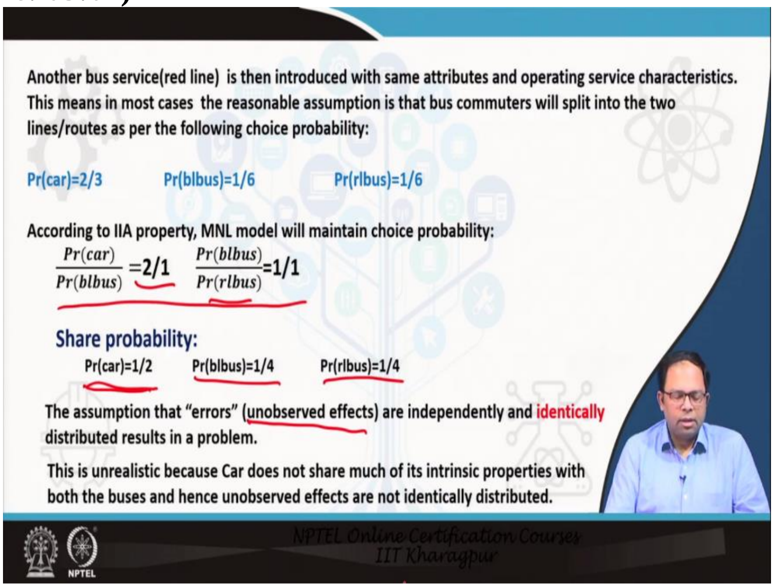
Suppose, there are two mode choices available in a city. One is a car, and other is a blue line bus. Also, assume that the probability of choosing car and blue bus will be 2/3 and 1/3 respectively. Mathematically, it can be written as:

$$P_{Car} = \frac{e^{U_{car}}}{e^{U_{car}} + e^{U_{Bus}}} = \frac{2}{3}$$

$$P_{Bus} = \frac{e^{U_{Bus}}}{e^{U_{car}} + e^{U_{Bus}}} = \frac{1}{3}$$

So, the ratio of the choice probabilities of car and bus will be 2:1.

(Refer Slide Time: 03:04)



Next, a red bus line service is introduced in the city which has the same attributes and operating service characteristics. The only difference between both the bus services is colour. In most cases, the reasonable assumption is that bus commuters will split into two lines/routes as per the following choice probabilities:

$$P_{car} = \frac{2}{3};\ P_{blue\ bus} = \frac{1}{6};\ P_{red\ bus} = \frac{1}{6}$$

In here, the choice probabilities of blue bus and red bus becomes half of the choice probability of bus, because both bus line service have same attributes and provides same service. So, people will either choose blue bus or red bus in 50:50 ratio. Also, the red line service introduction does not have any effect on the commuter's choice for car.

However, based on IIA property, the MNL model will maintain the ratio of choice probabilities of car and bus as 2:1. Also, the probability of blue bus line and red bus line will be same because both have same utility. So, the ratio of choice probabilities of blue bus service and red bus service will be 1:1. Then the share probabilities for car, blue bus, and red bus will be:

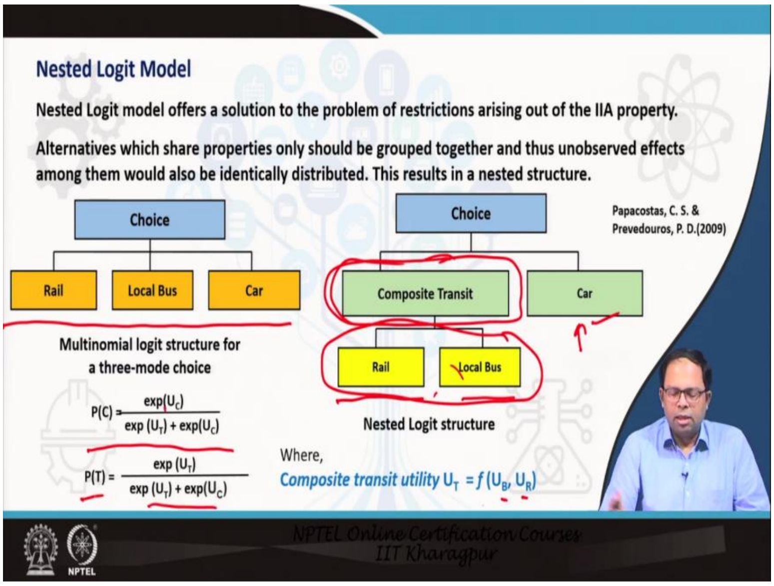
$$P_{car} = \frac{1}{2};\ P_{blue\ bus} = \frac{1}{4};\ P_{red\ bus} = \frac{1}{4}$$

So, the probability of car is reduced from 2/3 to 1/2. This happened because it is assumed that the errors (unobserved effects) are independently and identically distributed. This is unrealistic because car does not share much of its intrinsic properties with both buses and hence unobserved effects are not identically distributed.

Similar to RP-SP model, when there are two different modes, or two separate variances or two different error types, a scale parameter is introduced to equalise the variance. It means that the utility of both the modes needs to be brought at the same platform. Since, bus and car have different variance, then a nested structure can be formed where both the bus alternatives are put within a nest and assigned a single variance term, and car is put just below the root with a variance term. It is done to estimate the model parameters.

So, there are two things which are to be remembered. One is the scale parameter, and other is to put identical alternatives in a group in order to use a same parameter for this group.

(Refer Slide Time: 06:31)



Nested logit model:

Nested logit model offers a solution to the problem of restrictions of IIA property. As discussed in previous section, alternatives which share properties should be grouped together and thus unobserved effects among them would also be identically distributed. This results in a nested structure.

Consider there are three mode choices which are rail, bus, and car. The basic multinomial logit structure for these three mode choices can be represented as the left figure. In nested logit model structure represented in the right figure, the car is kept at the upper nest. While the bus and rail are put in a group which is termed as composite transit. Now, the rail and bus are part of this composite transit which is the part of the upper nest. So, it is similar to RP-SP model where dummy nests were created each with a single SP alternative only (or single-alternative nest).

It is important to mention that the utility equation for composite transit alternative should only include variables/characteristics which are common to both bus and rail. For example, consider people value in-vehicle travel time and out-vehicle travel time similar for bus and rail, then these two attributes are common variables to bus and rail. On the other hand, the variables which are alternative specific i.e. specific to bus or rail, should be included in the second level of the nest (lower nest). So, alternative specific parameters will be estimated within the nested logit model. In this way, both the mode specific parameters and common parameters will be estimated.

In a nested logit model, the MNL model assumes the following form:

$$P(C) = \frac{e^{(U_C)}}{e^{(U_C)} + e^{(U_T)}}$$

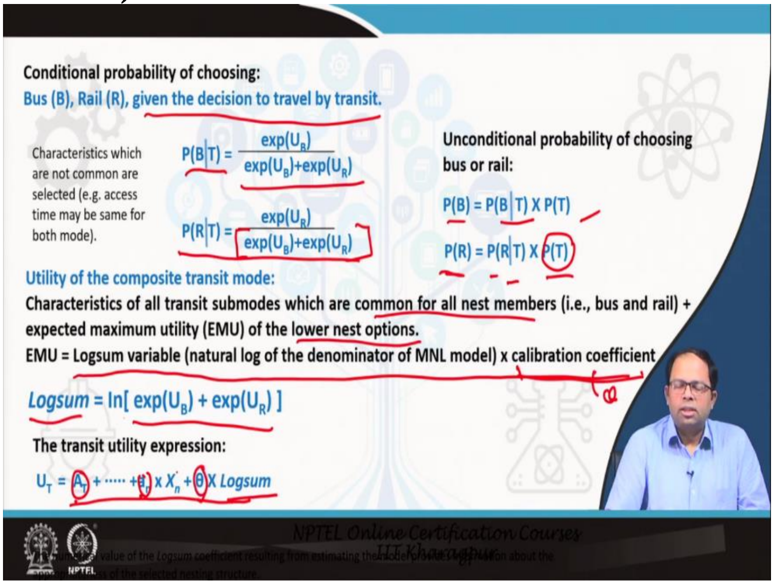
$$P(T) = \frac{e^{(U_T)}}{e^{(U_C)} + e^{(U_T)}}$$

Where, P(C) and P(T) are choice probability of car and transit respectively, U_C is utility of car, and U_T is composite transit utility. It is important to mention that this the first level (upper nest) multinomial model (specifically binary logit model).

In order to develop the composite transit utility equation, a dataset is created out of overall dataset using only the transit variables. These variables are used to form the composite utility equation. Therefore, this composite transit utility is a function of utility of rail (U_R) and utility of bus (U_B). Mathematically, it can be expressed as:

$$U_T = f(U_B, U_R)$$

(Refer Slide Time: 10:20)



Conditional probability of choosing:

Once the probability of choosing transit is determined, the next step is to determine the probability of bus or rail given the decision to travel by transit. The conditional probabilities of choosing bus and rail is given as follows:

$$P(B|T) = \frac{e^{(U_B)}}{e^{(U_B)} + e^{(U_R)}}$$

$$P(R|T) = \frac{e^{(U_R)}}{e^{(U_B)} + e^{(U_R)}}$$

Where, $P(B|T)$ and $P(R|T)$ are the conditional choice probability of bus and rail, U_B is utility of bus, and U_T is utility of rail. This is the second level (lower nest) MNL model. At this level, the utility equations are developed considering the mode specific variables only. It means that the utility equation for bus (U_B) will include only bus specific variables, and utility equation of rail (U_R) will incorporate rail specific variables only as discussed earlier. Based on the data of mode specific variables, the parameters are estimated for this model.

Once, all the parameter values are estimated and conditional probabilities are known, the unconditional choice probabilities for bus and rail can be calculated. The unconditional choice probability of bus and rail are as follows:

$$P(B) = P(B|T) * P(T)$$

$$P(R) = P(R|T) * P(T)$$

Now, the question is how to estimate the probability of transit ($P(T)$). In order to estimate the probability of transit($P(T)$), the composite transit utility equation (U_T) needs to be developed which is a function of utility of bus and utility of rail. The composite transit utility equation is equal to the characteristics of all transit sub modes which are common for all nest members (i.e. bus and rail) + the expected maximum utility (EMU) for the lower nest options. The EMU is given by the product of the natural log of the denominator of the MNL model (logsum) and the calibration coefficient. So, the logsum incorporates the lower nest of the structure, and the calibration coefficient is the scale coefficient.

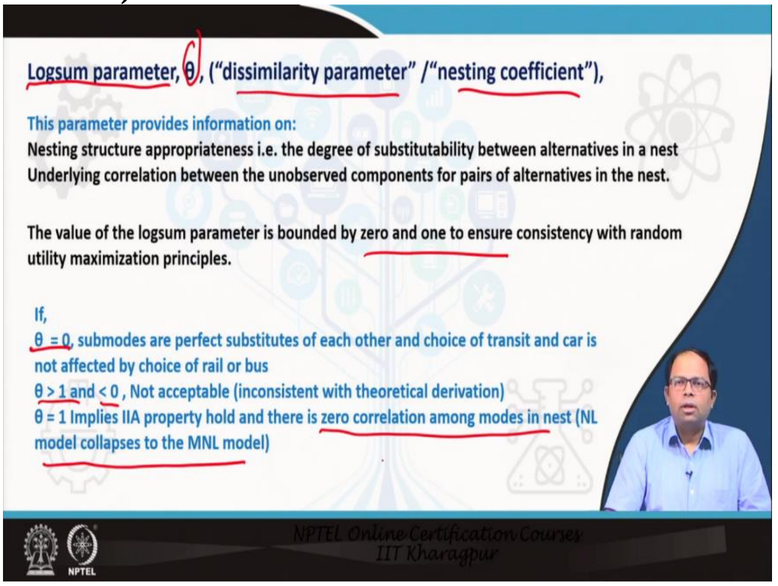
The utility of composite transit U_T can be mathematically expressed as:

$$U_T = A_T + \ldots \ldots . + a_n * X_n + \theta * Logsum$$

$$Logsum = ln[e^{U_B} + e^{U_R}]$$

Where, A_T is the bias, X_n are common characteristics of the transit mode, and θ is the calibration coefficient. It is important to understand that if the logsum is not included in the composite transit utility equation, then the uncommon characteristics of the different sub modes would be lost while estimating the model.

(Refer Slide Time: 15:00)



Logsum parameter (θ):

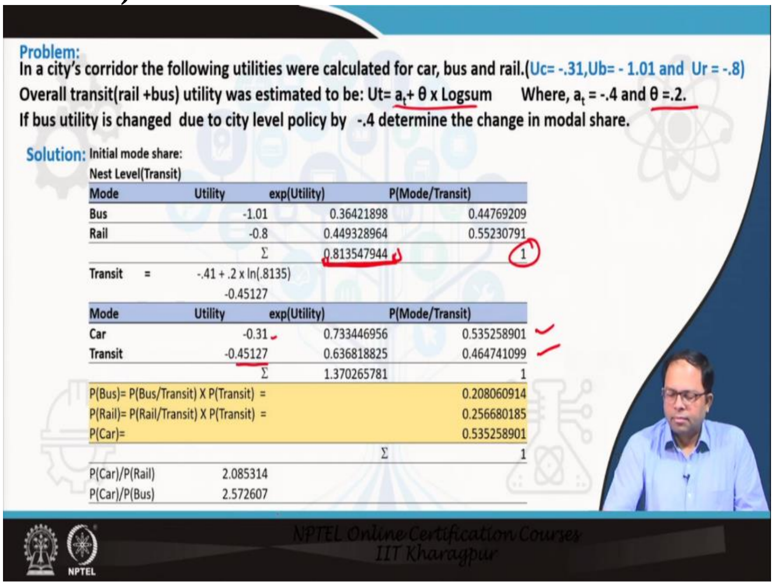
The logsum parameter or 'θ' is known as dissimilarity parameter or nesting coefficient also. It provides a lot of information on the effect of the logsum variable. In addition to this, it gives information on the nesting structure appropriateness, which means that the degree of suitability between alternatives in a nest can be evaluated. The underlying correlation between the unobserved components for pair of alternatives in the nest can be also understood using this parameter.

The logsum parameter is the coefficient of logsum variable. For a particular logsum variable, the value of the logsum parameter **is bounded by 0 and 1** to ensure the consistency with random utility maximization principles. Since, the logsum parameter is within the range 0 to 1, all the beta coefficients are also bounded by 0 and 1. If this parameter is equal to **zero**, then all sub modes are perfect substitutes of each other. For example, the choice of transit and car is not affected by choice of rail or bus. So, one can independently estimate the composite transit utility without considering the utility of the sub nest (bus and rail). In other words, **there will be no effect of the change within the sub nest based on upper level choices**. If the value of **θ>1 and <0**, then it is **inconsistent with the theoretical derivations**. Whereas, if the value of this parameter is equal to **one**, then it implies that the IIA property holds and there is zero correlation among modes in the nest. So, the **nested logit model collapses to**

multinomial logit model. Based on the value of the parameter, one can decide which model to use i.e. nested logit model or MNL model.

So, this is how scale parameter or the logsum parameter is can be interpreted. Based on the value of the logsum parameter; one can decide to opt for a MNL model or a nested logit model. For example, in the red bus blue bus paradox, introducing a mode which is almost similar to another mode, probably will not affect the other mode, contrary to the IIA property, where it was affecting the other modes. So, if the theta parameter or the logsum parameter is equal to 1 that means it implies that it will affect the other modes in the same exact way in the same proportion. And so, there is no need for developing an MNL model.

(Refer Slide Time: 18:48)



Consider a case where, In a city's corridor, the utility of car was found to be (-0.31); utility of bus was found to be (-1.01); and utility of rail was found to be (-0.8). The overall utility of transit i.e. rail and bus was estimated to be $U_t = a_t + \theta \times Logsum$. '**θ**' and **Logsum** are same as discussed in the previous parts of this lecture, and '**a$_t$**' is the bias term. The value of **a$_t$** is to be taken as (-0.4), and that of **θ** is to be taken as (0.2). Given that no other variables are there to impact the utilities, if bus utility changes by (-0.4), due to a city level policy, the change in model share is to be determined.

Since the problem has been designed in a nested logit framework, the change in utility of bus affects the utility of modes in the nest more as compared to the utilities of mode out of the nest. If the nests would not have existed, the impact on other modes would have been more profound. Following is the table of initial mode share calculations:

Initial mode share:			
Nest Level(Transit)			
Mode	**Utility**	e^{Utility}	**P(Mode/Transit)**
Bus	-1.01	0.36421898	0.44769209
Rail	-0.8	0.449328964	0.55230791
	Σ	0.813547944	1
Transit =	-.41 + .2 x ln(.8135)		
	-0.45127		
Mode	**Utility**	e^{Utility}	**P(Mode/Transit)**
Car	-0.31	0.733446956	0.535258901
Transit	-0.45127	0.636818825	0.464741099
	Σ	1.370265781	1
P(Bus)= P(Bus/Transit) X P(Transit) =			0.208060914
P(Rail)= P(Rail/Transit) X P(Transit) =			0.256680185
P(Car)=			0.535258901
		Σ	1
P(Car)/P(Rail)	2.085314		
P(Car)/P(Bus)	2.572607		

With the values of the utility of **bus** and **rail**, their respective exponentials are calculated as **0.36421898** and **0.449328964**, respectively. The sum of these two exponentials are calculated to be **0.813547944**. Hence, the probability of bus and rail can be estimated as $\boldsymbol{P_{Bus} = \frac{0.36421898}{0.813547944} = 0.44769209}$, and the probability of rail can be estimated as $\boldsymbol{P_{Rail} = \frac{0.449328964}{0.813547944} = 0.55230791}$. But these probabilities are not the true probability, rather the conditional probabilities, given that transit is chosen. It can be also represented as $\boldsymbol{P_{(Bus|\,Transit)} = 0.44769209}$, and $\boldsymbol{P_{(Rail|\,Transit)} = 0.55230791}$. Since, both rail and bus are inside the nest of transit, the probability of choosing transit also needs to be estimated. The utility of **transit** can be used to calculate the probability of transit, given the utility of **car**. The utility of transit can be estimated as $\boldsymbol{U_{transit} = -0.31 + 0.2 \times \ln(0.36421898 + 0.449328964) = -0.45127}$. This can be further used to calculate the probability of transit using the $\boldsymbol{U_{car} = (-0.31)}$ as $\boldsymbol{P_{Transit} = \frac{e^{-0.45127}}{e^{-0.45127} + e^{-0.31}} = 0.464741099}$. Similarly $\boldsymbol{P_{Car} = 0.535258901}$ was found.

After calculating these, the actual probabilities can be calculated as follows:

$\boldsymbol{P_{Transit} = 0.464741099}$

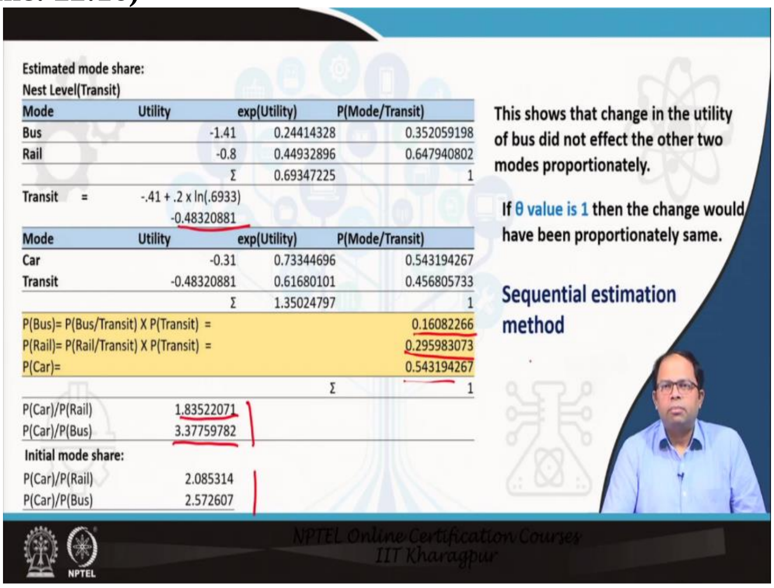
$$P_{Car} = \ 0.535258901$$

$$P_{Bus} = \ P_{(Bus|\ Transit)} \times P_{Transit} = \ 0.44769209 \times \ 0.464741099 = 0.208060914$$

$$P_{Rail} = \ P_{(Rail|\ Transit)} \times P_{Transit} = \ 0.55230791 \times \ 0.464741099 = 0.256680185$$

If the probability of selecting a car is estimated relative to probability of selecting bus is estimated, it can be done as $\frac{P_{Car}}{P_{Bus}} = \frac{0.535258901}{0.208060914} = 2.572607$. Similarly, probability of car with respect to that of trains can be estimated as $\frac{P_{Car}}{P_{Rail}} = \frac{0.535258901}{0.256680185} = 2.085314$.

(Refer Slide Time: 22:16)



Now, as the utility of bus changes by (-0.4), the new utility of the mode becomes $U_{Bus} = -1.41$. As a result, the exponential values, and hence conditional probabilities, i.e. the probabilities within the transit nest changes to $P_{Bus} = \frac{0.24414328}{0.69347225} = 0.352059198$ and $P_{Rail} = \frac{0.44932896}{0.69347225} = 0.647940802$. Consequently, the logsum parameter changes, and so does the utility of the transit nest itself. The new utility of the nest becomes, $U_{transit} = -0.31 + 0.2 \times \ln(0.24414328 + 0.449328964) = -0.48320881$. This can be further used to calculate the probability of transit using the $U_{car} = (-0.31)$ as $P_{Transit} = \frac{e^{-0.48320881}}{e^{-0.48320881} + e^{-0.31}} = 0.456805733$. Similarly $P_{Car} = 0.543194267$ was found. After calculating these, the actual probabilities can be calculated as follows:

$$P_{Transit} = \ 0.456805733$$

$$P_{Car} = \ 0.543194267$$

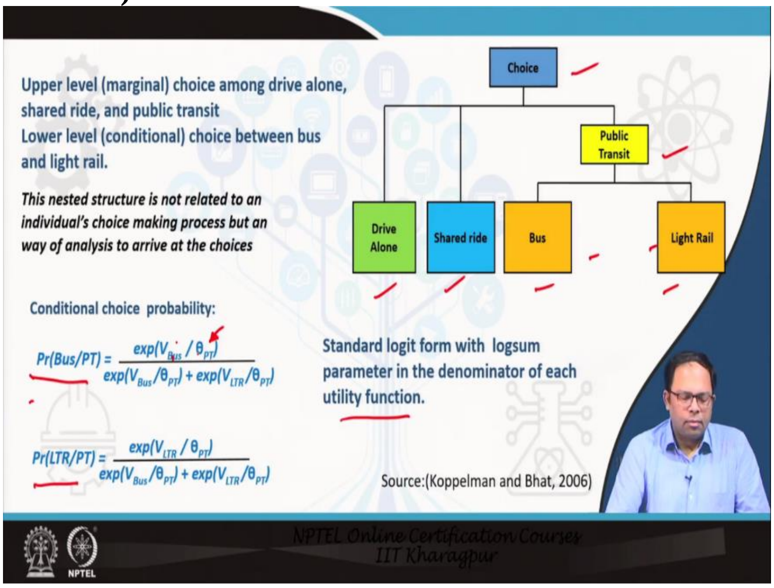
$$P_{Bus} = \ P_{(Bus|\ Transit)} \times P_{Transit} = \ 0.352059198 \times \ 0.464741099 = 0.16082266$$

$$P_{Rail} = P_{(Rail|\,Transit)} \times P_{Transit} = 0.647940802 \times 0.464741099 = 0.295983073$$

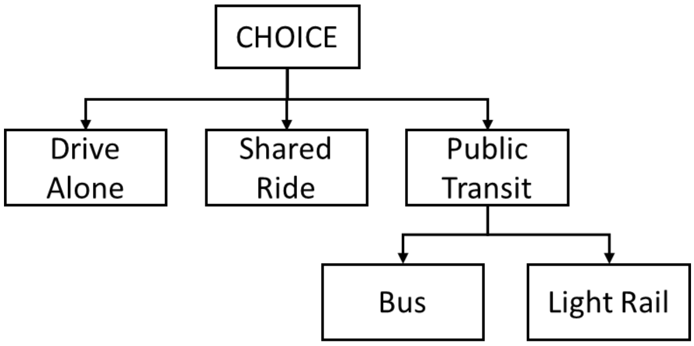
Comparison of choice probabilities		
Initial	**After (-0.4) change in U_{Bus}**	**Difference**
$P_{Car} = 0.535258901$	$P_{Car} = 0.543194267$	0.0079354
$P_{Transit} = 0.464741099$	$P_{Transit} = 0.456805733$	-0.0079354
$P_{Bus} = 0.208060914$	$P_{Bus} = 0.16082266$	-0.047238
$P_{Rail} = 0.256680185$	$P_{Rail} = 0.295983073$	0.039303

So, due to the reduction in utility of bus, a reduction in the probability of bus can be seen. Most of the utility was found to be gained by the rail, and very little effect can be seen on car. This shows that change in the utility of bus did not affect the other two modes proportionately. However, if the **scale parameter** would have been **1**, the change would have been proportional in all the three modes. So, a nested logit model becomes a multinomial logistic regression if the **scale parameter is 1**. This estimation can be also called as a **sequential estimation method** which means sequentially first, the probabilities between bus and rail are estimated; then transit probability is estimated; then obtaining the probabilities of all the alternatives. In this case, the utility equation for transit has been given. There may be a case where the utility equation for transit utility also needs to be estimated, using many independent variables.

(Refer Slide Time: 25:07)



Although sequential approach has been explained in the example above, most of the softwares employ simultaneous approach while estimating a nested logit model, in which some special considerations needs to be taken. For example, as the figure shows below, there are two levels of choice; upper level has **Drive alone**, **Shared ride**, and **Public transit;** within Public transit, there are **Bus** and **Light rail**.



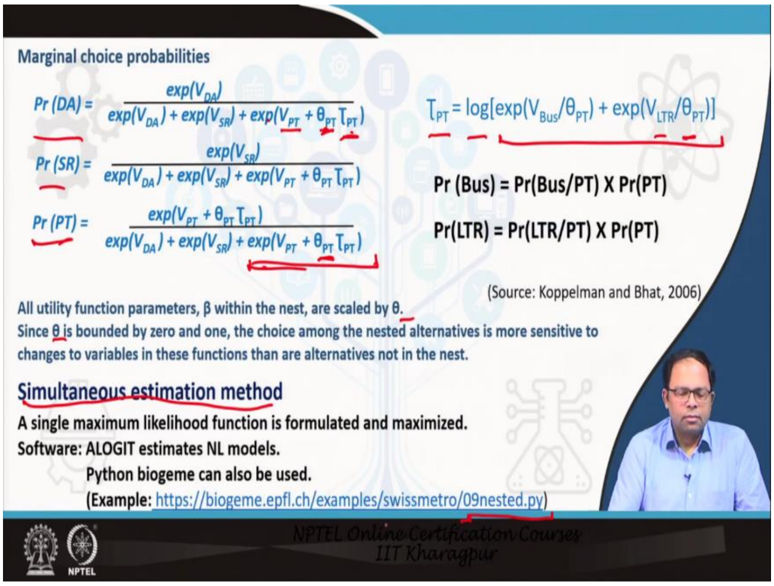
The upper level choice within drive alone, shared ride, and public transit are called **marginal choice**, and the lower level ones within bus and light rail are called **conditional choice**. The term is called so because it is conditional to the selection of the nest i.e. public transit, in this case.

From the well-structured flow of choices, it may seem to be very realistic. That means people, while deciding mode choice, choose between the upper nest, and then choose between the lower nests. But this assumption of human thought process to be sequential is not always true. Most of the time, the alternatives evaluated all at the same time. The nested model is just to make things analytically clear, and facilitate in estimation. So, in order to reflect the nested structure, in the simultaneous approach, the formula for conditional choice probabilities are modified as follows:

$$P_{Bus|PT} = \frac{e^{\frac{V_{Bus}}{\theta_{PT}}}}{e^{\frac{V_{Bus}}{\theta_{PT}}} + e^{\frac{V_{Light\ rail}}{\theta_{PT}}}} \qquad P_{Light\ rail|PT} = \frac{e^{\frac{V_{Light\ rail}}{\theta_{PT}}}}{e^{\frac{V_{Bus}}{\theta_{PT}}} + e^{\frac{V_{Light\ rail}}{\theta_{PT}}}}$$

The **scale parameter**, or the **parameter of the logsum variable** is introduced in the formula of conditional probability, as a denominator of the utility equation, in order to reflect the nested structure. This introduction does not change the formula that was already there, rather it is a re-arrangement of terms in order to facilitate simultaneous estimation. This can be used as a standard formula as when there is no nest, the value of **θ** becomes 1, and the formula reverts to the original probability formula as found in MNL.

(Refer Slide Time: 28:13)



Along the same line, the formula of the marginal probabilities are as follows:

$$P_{DA} = \frac{e^{V_{DA}}}{e^{V_{DA}} + e^{V_{SR}} + e^{(V_{PT}+\theta_{PT}.T_{PT})}}$$

$$P_{SR} = \frac{e^{V_{SR}}}{e^{V_{DA}} + e^{V_{SR}} + e^{(V_{PT}+\theta_{PT}.\mathrm{T}_{PT})}} \qquad \mathrm{T}_{PT} = \log[e^{\frac{V_{Bus}}{\theta_{PT}}} + e^{\frac{V_{Light\ rail}}{\theta_{PT}}}]$$

$$P_{PT} = \frac{e^{(V_{PT}+\theta_{PT}.T_{PT})}}{e^{V_{DA}} + e^{V_{SR}} + e^{(V_{PT}+\theta_{PT}.\mathrm{T}_{PT})}}$$

The marginal probabilities for DA becomes exponential of $\mathbf{V_{DA}}$, divided by the sum of exponential of $\mathbf{V_{DA}}$, exponential of $\mathbf{V_{SA}}$, and exponential of composite utility of public transit. Composite public transit utility is the exponential of sum of the common characteristics of the public transit modes ($\mathbf{V_{PT}}$), and the logsum parameter ($\boldsymbol{\theta_{PT}}$) multiplied by the log of summation of the exponentials of all the alternatives in the nest ($\mathbf{T_{PT}}$). Similarly, the probabilities of SR and PT can be written, as shown above. The choice probabilities of the alternatives in the public transit nest can be calculated using the formula given below:

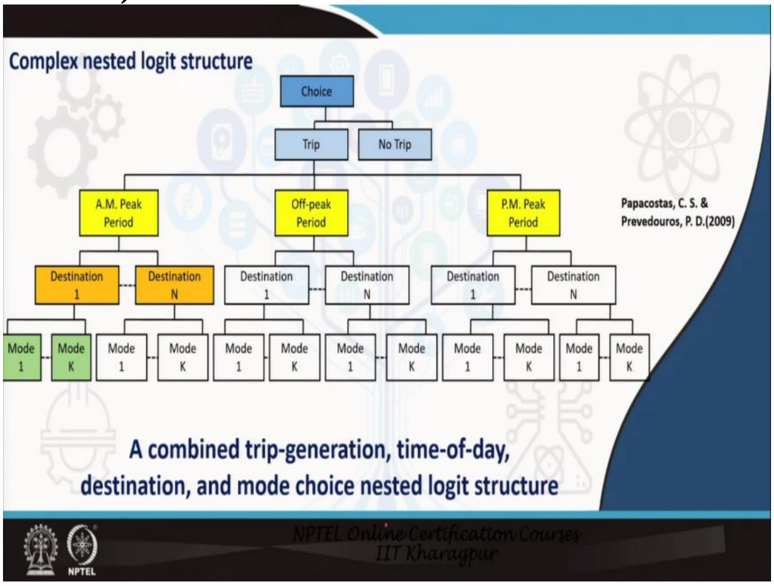
$$P_{Bus} = P_{Bus|PT} \times P_{PT}$$

$$P_{Light\ rail} = P_{Light\ rail|PT} \times P_{PT}$$

The parameters **θ**, and all **β**s are estimated through maximum likelihood estimation for which usually softwares are used. Using the value of the scale parameter (**θ**), the probabilities can be found. Since the value of **θ** is bound between 0 and 1, the choice among the nested alternatives is more sensitive to changes to variables in these particular function, than the alternatives not in the nest.

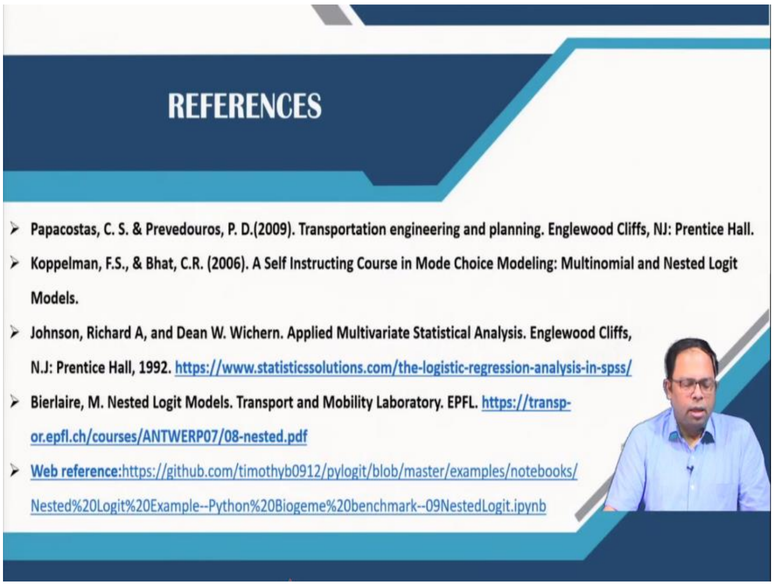
In the simultaneous estimation approach of solving the nested logit problem, a single log likelihood function is formulated and maximized to obtain the values of the parameters. ALOGIT can be used to estimate NL models. Python Biogeme is also an alternative, available freely. The example of such an estimation using Biogeme can be seen in https://biogeme.epfl.ch/examples/swissmetro/09nested.py, where using a data set from Swiss metro, a nested logit model is estimated.

(Refer Slide Time: 32:18)

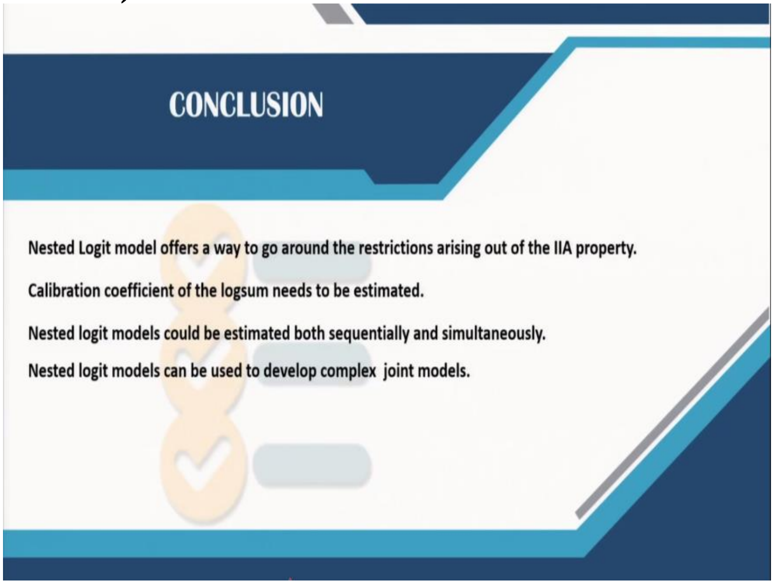


Since joint modelling has been talked about a lot in recent times, nesting structures are very important in landuse transportation modelling, in general. Lately, combined trip generation, time of the day, destination and mode choice models are being discussed a lot, which means there could be a combination of destination and mode choice models, or even time of the day and destination choice models. The figure shows a combined trip-generation, time-of-day, destination choice, and mode choice nested logit structure. As per the structure, a person may either choose to make a trip or not make a trip; if a trip is made, it can be either made during morning peak, afternoon off-peak, or evening peak periods i.e., in either of the time periods. There can be '***n***' destinations to choose from and for each of the destination, there can be '***k***' different modes to choose from. So these kind of complicated nesting structures can be developed for joint models. Apart from mode choice; location choice, destination choice, a combined destination-mode choice, etc. can be developed using the principles discussed .

(Refer Slide Time: 34:04)



(Refer Slide Time: 34:11)



So, these are the references that you can consult. And finally, to conclude, nested logit model offers a way to go around the restrictions arising out of the IIA property. The calibration coefficient of the logsum needs to be estimated. The nested logit models can be estimated both sequentially and simultaneously. Nested logit models can be used to develop complex joint models which actually gives the strength to the discrete choice approach.