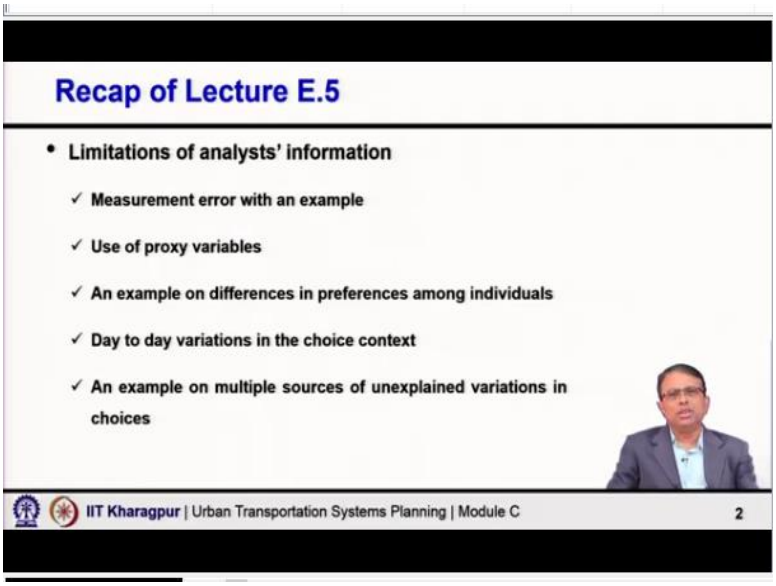
Urban Transportation Systems Planning
Prof. Bhargab Maitra
Department of Civil Engineering
Indian Institute of Technology-Kharagpur

Lecture-36
Disaggregate Mode Choice Models-V

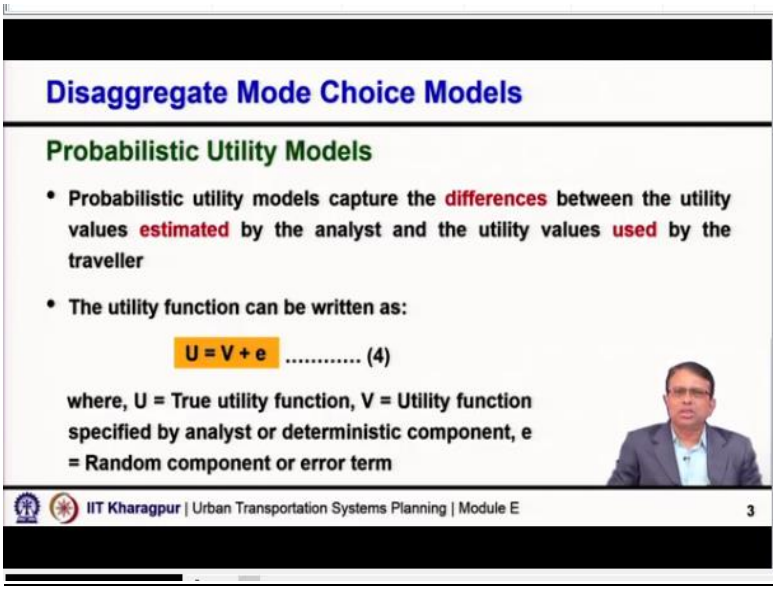
Welcome to module E lecture 6, in this lecture also we shall continue our discussion about desegregate mode choice models.

(Refer Slide Time: 00:23)



In lecture 5 we discussed about various limitations of analysts' information related to measurement error, related to use a proxy variable, difference among individuals, day to day variation in the choice context and some of these we explained with examples and also we gave you an example to tell that how multiple such limitations can impact the choice decisions.

(Refer Slide Time: 01:01)



With this background today we shall discuss about probabilistic utility model. All these limitations lead to a big question that if such limitations are there and if we cannot overcome this decision but also we want to improve our prediction or model output to match closely with the real behaviour then what can be done? The answer is then, think of a different modeling approach which can accommodate all such limitations of analyst, but still can produce reasonable or closer results to the real behaviour.

So, that led to the development of probabilistic utility model. In probabilistic utility model we try to then in a way captured the difference between the utility values as estimated by the analyst using his or her utility equation whatever is used. And the utility value used by the traveller, because in each case multiple examples we gave earlier in each case you have seen that there is an error term.

If I would have considered that in my utility equation then I would have probably able to predict the choice correctly. So, one is the actual utility which is perceived by the user or the traveller and which is actually inflict influencing or impacting his or her choice decision the other is what is the utility that as a modeler I am taking in my model. So, because we know that whatever utility we considered earlier many cases the modeller cannot really take the complete or the true utility.

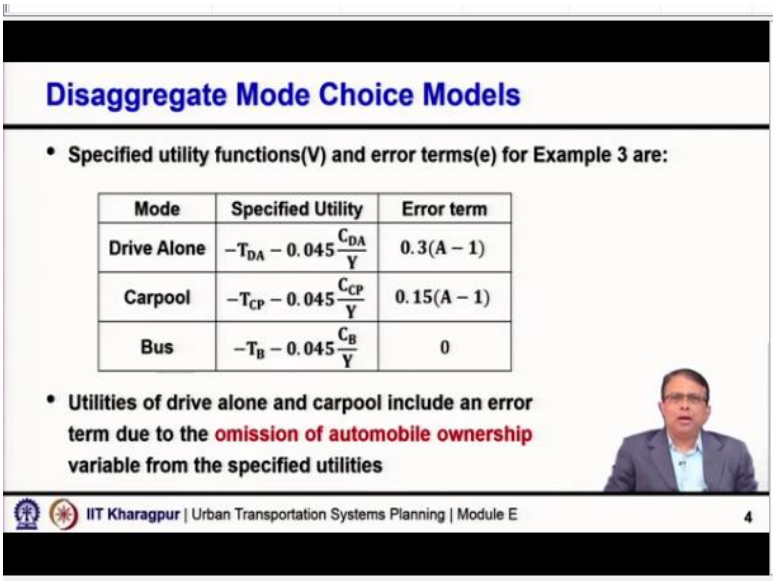
So, we now write the utility function as shown in equation 4, that means the true utility is actually a portion of the utility which as a modeller we are considering, so that we can say utility functions specified by an analyst and this is obviously the deterministic component

because here we are taking attributes and coefficients, all are there and also the attributes as measurable and we are including them in the model.

And this portion we can calculate and for each mode or each alternative we can get a value of this. So, this is called the deterministic component of the utility and in each case there could be an error, so the true utility includes this utilities function specified by the analyst which is V plus an error term e. Now this error term is a random component and because e is random, the true utility is also a random component or random variable.

So, you have now utility function which is a random variable which includes a deterministic component which the modeller can quantify and get a value and there is an error term which is again a random error or random component. So, the whole utility now become a no more a deterministic part, but it becomes altogether a random variable.

(Refer Slide Time: 05:22)



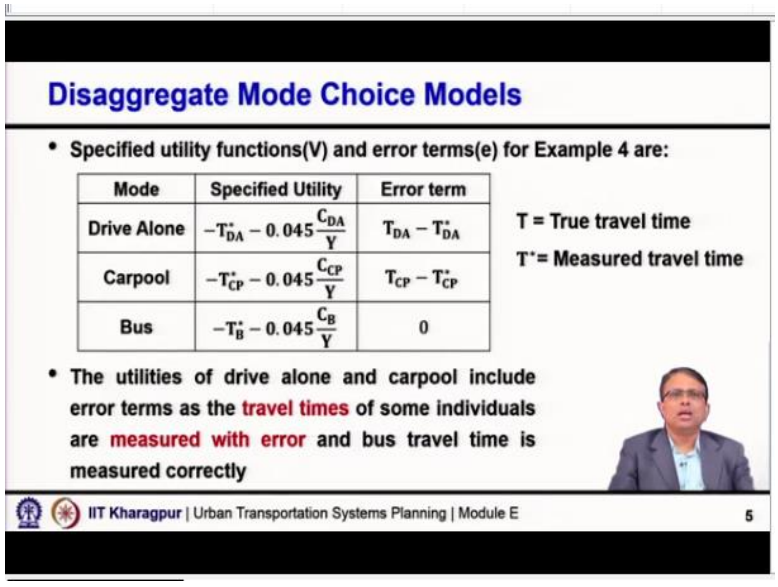
Now, if I connect this statement to various previous examples, what I explained in the context of the limitations of analysts information. So, please try to connect this lecture with the previous lecture and the examples which were given. For example, in example 3 we explained you how due to the omission of a variable the model of utility and true utility could be different, you remember that for drive alone and carpool we showed you that in one case we considered the car ownership and in another case, we did not consider the car ownership, that effect was not considered.

So, what was actually specified without considering the car ownership that I have stated here for driver alone, for carpool and for bus, what are the stated or specified utility by the modeller and then as we said that choice also was influenced by the car ownership. So, the error terms in the drive alone utility, in the carpool utility, what was there that is also shown here.

Now, for bus there was no such error term because bus utility does not impact on the car ownership. So, this error term for the bus was 0. So, what was considered by the modeller what is indicated here in column 2 and the corresponding error terms are considered in column 3. So, because these error terms components that was not considered by the modeller, so, the actual true utility or whatever was the distribution or what was the choice of mode and what was predicted by the modeller using the specified utility equation, the outcomes did not match correctly.

So, here this is the error term. So, I can consider in the context of example 3 my error terms are like this as shown in the last column for driver alone 0.3 into A + 1, for carpool 0.15 into A – 1. So, these are the 2 error terms which are shown here with respect to this example. And this was in the context of omission of irrelevant variable.

(Refer Slide Time: 08:26)



Now, similarly, if I consider example 4 we said that not everybody considers that utility associated with travel time same, he said that there are 2 groups of individual some consider a particular fellow and this was not related to person to person variation but this was related to

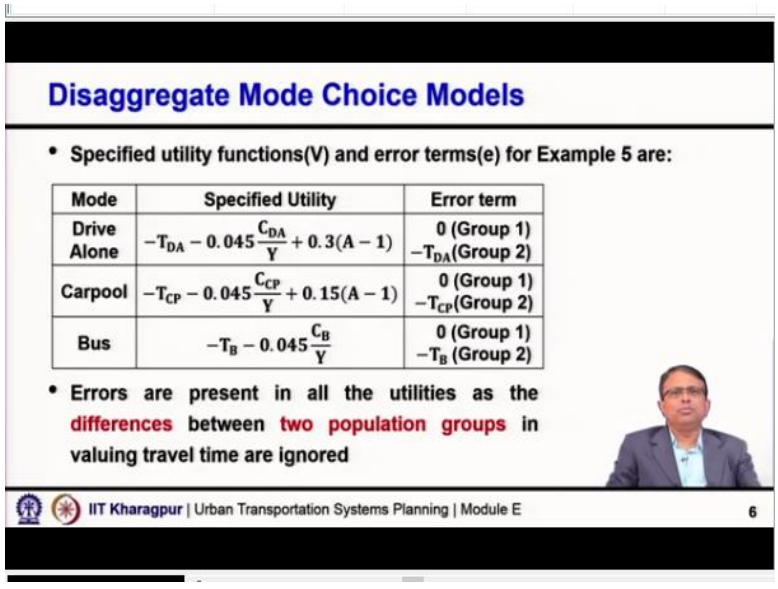
the second example 4. Let us omit this portion and let us say, we consider now another example that is example 4.

Recollect or remember this example where we said that the true utility and the model or specified utility could be different because of the measurement error. And we give you an example that we consider travel time maybe from a zone centroid to zone centroid, but when you are considering drive alone or carpool the actual residential location within the zone may not be of equal distance from the zone centroid.

So, 2 different individual or different groups of individuals who are residing in different band to them the travel time may be different. So, for driver alone what was considered by the modeller that is indicated here as specified utility and the error term indicates what was the actual travel time and what was actually valued taken as the value of travel time and this error is due to measurements, but again this is an error.

So, we have shown you here again in column 3 that what is the error which made the model output using specified utility and deterministic model make the outcome different from the actual observed distribution of mode. So, in the context of this measurement error and the example what we took the error term for drive alone and carpool are as shown in this table in column 3. For bus there was no error.

(Refer Slide Time: 11:14)



Now we also took another example which is example 5 and there we said that person to person they are maybe different weightage to the travel time. The travel time the disutility

due to travel time may not be same to every individual. So, there in that example we considered say 2 group of individual, you can make considered them as 2 income group. So, 2 income groups they do not value or they do not consider the disutility associated with travel time as same.

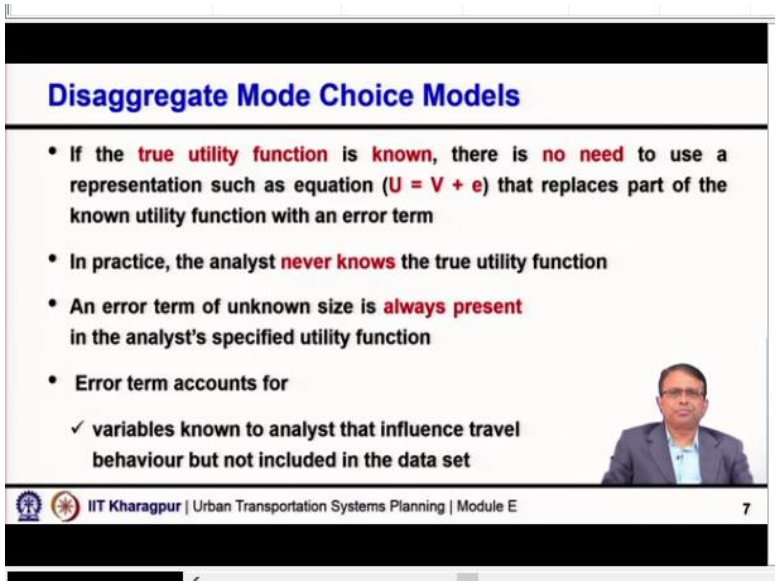
One considered the value double than the other group, but what we did? We did the utility; we did consider the utility as shown here under specified utility. So, this utility considered the disutility associated with travel time same for every user group because we did not consider any user group separately, for everybody we just took one equation, one value. But we have seen in that example that the group 2 value or put much higher double weightage to the travel time.

So, in that case if we can say whatever we considered that was actually appropriate for group 1. So, the error was 0 for group 1 in the case of drive alone, carpool and bus. But because the travel time was the disutility was associated with travel time was a double for the group 2. So, for drive alone group 2 the error term is again minus TDA because they were actually it was - 2 TDA. So, we considered only minus TDA.

So, minus TDA was the error, again for carpool also minus TCP and for bus minus TB. So, in this example the specified utility and actual utility what was the error in the specified utility that is shown here relating to example 5. So, now we actually again discussed this previous 3 examples to tell you every time what was actually specified by the modeler and what was actually the true utility they were different.

Because we did not know that true utility as a modeller only have knowledge about the specified utility. So, what are the error terms in the case of these 3 example problems? One case in this case what I say person to person or different user groups have different value. In the previous one measurement error and in the first case it was due to the omission of a relevant variable.

(Refer Slide Time: 14:40)



Now if the true utility function is known that means if this error terms are also known and exactly as I said in all the 3 examples if we knew this component or if he would have considered them in the specified utility. Then there were no need to use representation such as V + e = u. That means true utility is specified utility plus error term because in each case just imagine that we included probably also error term in our specified utility function.

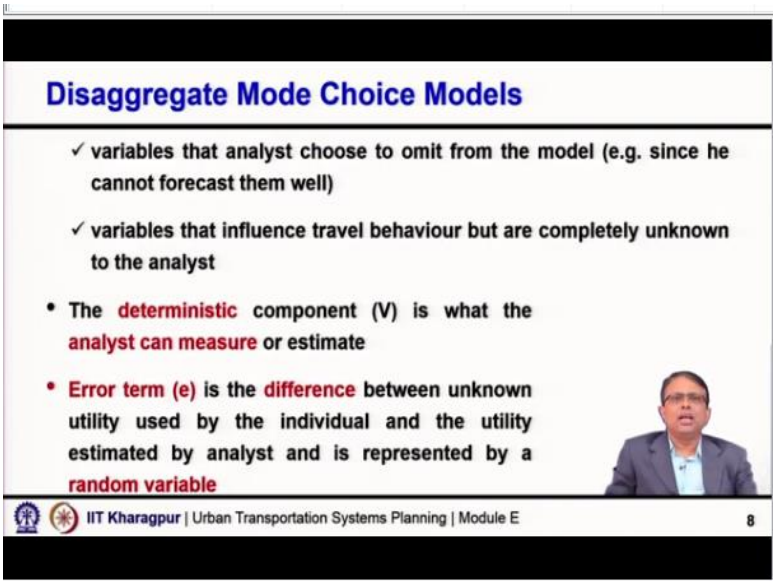
Then there was no need to use such kind of equation, that the true utility is the specified utility plus an error term, but in reality the analyst never knows the true utility function, maybe in these 3 examples we showed you so now you know what is the actual error? Such kind of errors could be there in probably every utility function we specify in a real world scenario or in a practical scenario.

So, there will always be an error term which will be present. And whatever we are considering or whatever we are specifying as the utility the true utility is not going to be exactly the same for all the alternatives what we are considering. So, the error term accounts for what 3 things it actually accounts for. Number 1 variables known to analyst that influenced travel behaviour but are not included in that data set.

That means as an analyst maybe I know that this variable influenced the choice decision and I should have actually considered it but I did not consider it for some reason, maybe one possible reason that data is not available and you know so I could not consider it. Like for example the proxy variable or omission of relevant variable car ownership I know should be considered but I did not consider.

That kind of problems maybe there because it is not only the model develop and you want to calibrate the model. So, you want to know basically that relevant data should be available. So, one part is variables known to analysts that influence the travel behavior very well known but are not included in that data set and in the model that kind of problem whatever error will be there. So, this error term accounts for such thing.

(Refer Slide Time: 17:44)



Second variables that analysts choose to omit from the model. That means intentionally we choose the model since we cannot forecast them. For example let me tell you that day to day variation I gave you one day I am not feeling well, so maybe instead of travelling by bus I travelled by car or by taxi if I do not have car or car is not available. So, that means I take it different mode or one day my family wants the car so instead of travelling in car I am travelling by bus.

Now such kind of variations even if somebody gives you the data can we forecast this that what will happen after 2 years or after 1 year or after 5 years, can we predict this? We cannot predict it. And if we cannot predict them correctly, cannot forecast them correctly we do not want to include them. Because development of the model is not for the only current scenario, we develop the model for the current scenario, so that the variables and the outcomes all are known and we can calibrate the model.

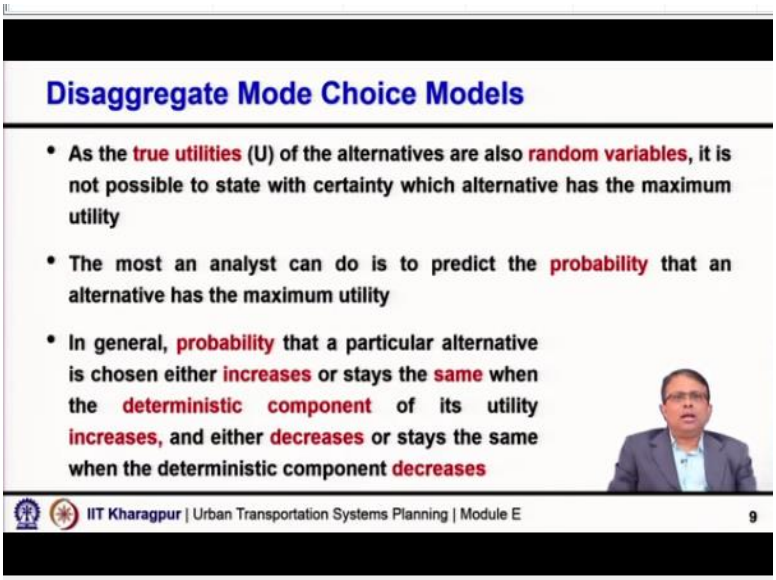
But the model is developed for the future. So, we have to make sure that we can forecast those variables which are included in the model. So, we may choose to omit because simply

we cannot forecast. The third variables that influence travel behaviour but are completely unknown to the analyst. Maybe there are something which some individual or for some alternatives or impacting the choice.

But as an analyst I may not know that because that behaviour or that thing is not known. So, I may not even know that that is influencing and I may not consider because what we consider in the model? Looking at the literature, looking at the discussion with the experts and the traveller we try to select some variables. So, there is no guarantee that every variable which are actually impacting the choice are to be included.

So, the third component as I said that it is actually the deterministic component of the utility which we take and which may not include all such components. So, the error term is the difference between the unknown utility used by individual and the utility estimated by analysts and is represented by a random variable that is what we are saying that is an error term which is a random variable.

(Refer Slide Time: 20:47)



Now as the true utility of the alternatives are also random variable it is not possible to state with certainty which alternative has the maximum utility? When I am considering only the deterministic component of the utility I can calculate the value that for alternative 1 it is –10 alternative 2 –15 alternative 3 maybe –20. So, I know clearly that which one is the maximum value of utility.

And then I can say confidently that this is going to be the choice because this is the alternative where for which the utility is maximum. So, according to the utility maximization principle we can tell certainly that this is the mode that is to be chosen because the utility is maximum here. In this case we only know that deterministic component of the utility, the other true utility includes an error term and the error term is a random variable.

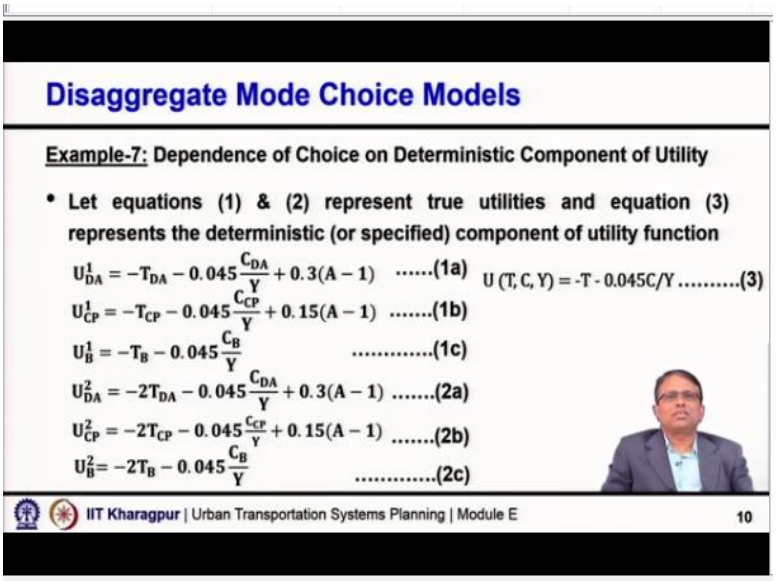
So, once the error term is random variable and the whole utilities random, based on only deterministic component we cannot say that which one is going to be the maximum, it is not possible because there is an error term which can take any value following the certain distribution. So, in general what we can say at best the probability that an alternative has the maximum utility.

So, I know the deterministic component of the utility and based on that I can say what is the probability that alternative 1; whatever utility we got it is going to be based on that knowledge that it is going to be the maximum utility, that alternative. So, in general what is found that the; probability that a particular alternative is chosen either increases or stays same when the deterministic component of its utility increases.

That means if the deterministic component of the utility increases the probability of choosing that thing will either increase or maybe remain same or either decrease it may decrease or stay same when the deterministic component decreases. So, if the deterministic component increase then the utility will increase or the probability will increase of choosing that mode will increase.

And if the utility decreases then the deterministic the probability will decrease because the deterministic component of the utility decreases. So, now the remaining time in this lecture we shall try to explain you this statement with some example, that whatever statement we said whether that is true.

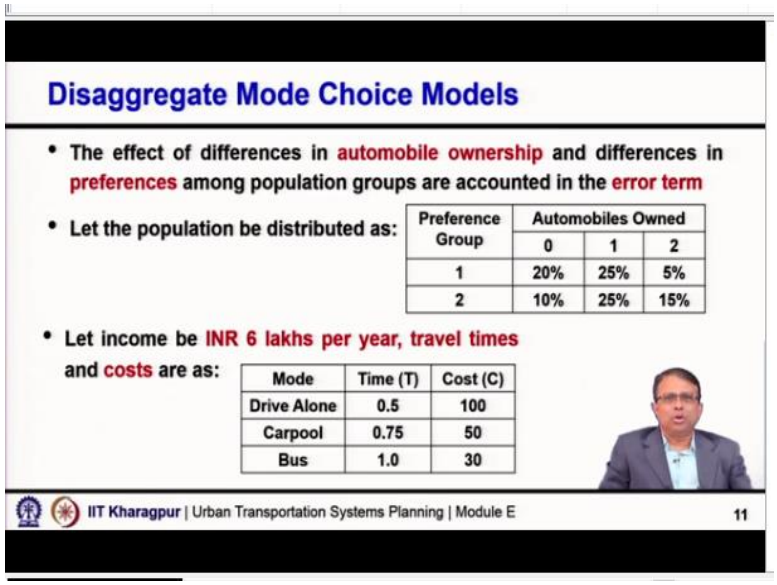
(Refer Slide Time: 24:10)



So, let us take an example to tell you the dependence of choice on deterministic component of utility. We have shown here 2 sets of equation, equation 1 of course 1 is further into 1a, 1b, 1c for 3 modes. Equation 2 is again for 3 modes or 3 alternatives 2a, 2b, 2c and 3 is basically what the analyst consider. That is equation 3. So, let this equation 1 and 2 represent the true utility and equation 3 represent the deterministic component of the utility what we know and what we considered. And 1a, 1b, 1c, 2a, 2b, 2c are the true utility. Now why we considered them?

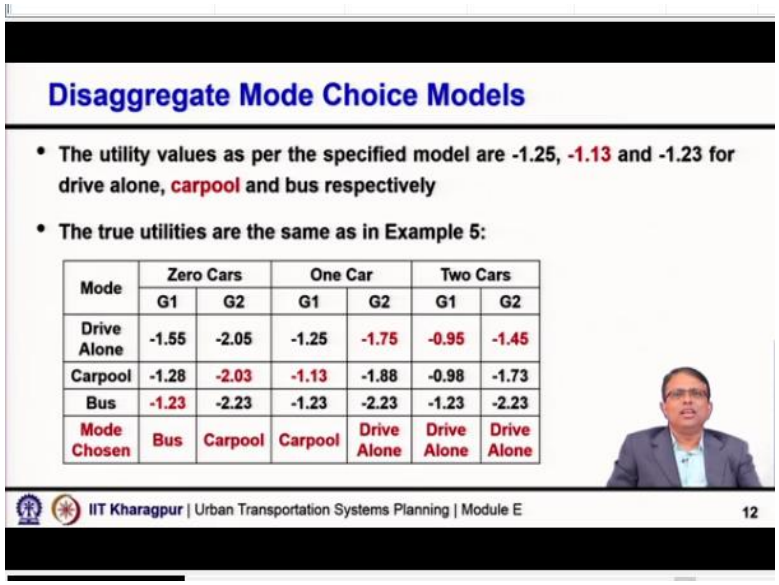
Because there; is an additional term that was not considered in the utility equation specified by the modeler in the deterministic component. So, you can very well connect these to the previous example. You can say 1a, 1b, 1c, they are different in terms of considering the car ownership which equation 3 does not include and equation 1 and 2 is different from 3 also in a term that the coefficient associated with travel time in equation 1 and equation 2 are different. Equation 2 case it is basically double.

(Refer Slide Time: 25:51)



So, the effect of difference in automobile ownership and the difference in the preference among population groups are accounted in this error term. Let the population be distributed as you know automobile ownership considerations maybe 0 car ownership, 1 car ownership, 2 car ownership. And let us consider that the population is heterogeneous and there are 2 groups in terms of their weightage to the travel time that is group 1 and group 2 what is the distribution? And if we considered that income is 6 lakhs Indian rupees per year and the travel time and travel costs are like this.

(Refer Slide Time: 26:37)



Then you can easily actually calculate that as per the deterministic utility equation what the modeler considered the utility values are -1.25, -1.13 and -1.23 for drive alone and carpool and bus respectively. So, as a modeller I said what I said that the choice will be carpool

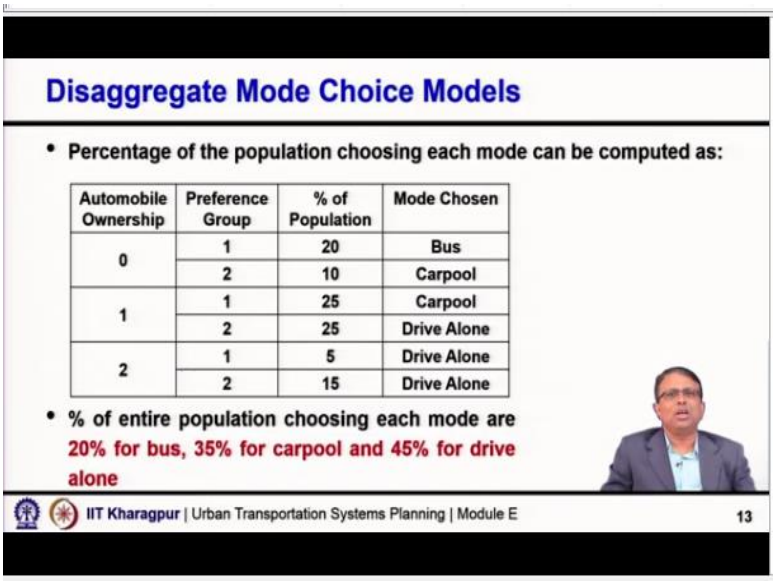
because carpool is maximizing the utility and in this case this negative value, so the minimum one is taken.

But the true utility using equation set 1 and 2 you can calculate for each user group, group 1 and group 2 G1 G2 indicate group 1 and group 2 for 0 car, for 1 car and 2 car, each case we have calculated the utility using this equation 1a, 1b, 1c for user group 1 and 2a, 2b, 3c for user group 2, using this travel time and travel cost values for 3 different alternatives.

And we can see clearly the chosen modes are going to be for 0 car user group 1 bus, group 2 carpool for 1 car ownership group 1 carpool group 2 drive alone and for 2 car ownership group 1 drive alone and group 2 also drive alone. So, what we will find we know how much percentage of population belong to 0 car 1 car 2 car and also based on group 1 and group 2. So, that distribution is known from this example.

You know that say group 1 with 0 car 20%, group 1 with 1 car 25% of the overall population. So, we know it and we know how much percentage of population are there in under each group 1, group 2 and considering the 0 car, 1 car and 2 car ownership. So, and we know for each group or each segment what is going to be the choice?

(Refer Slide Time: 28:54)



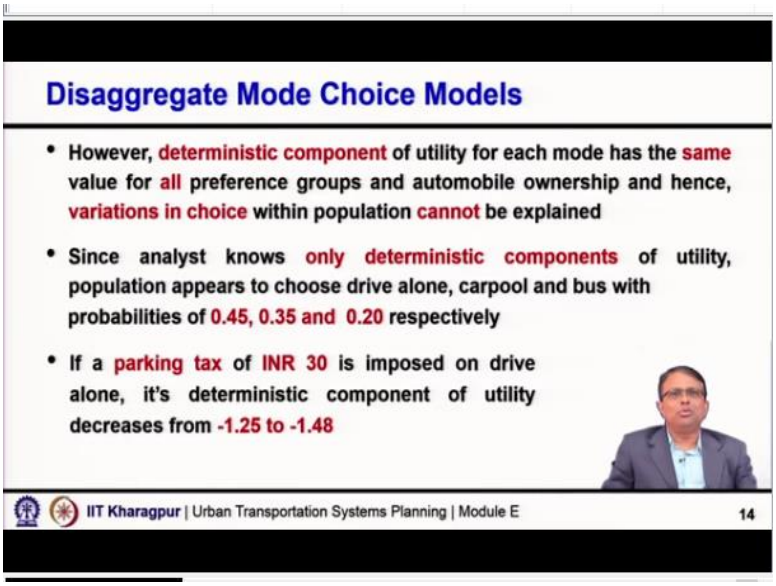
So, accordingly we know that what will be the actual real behaviour? The real behaviour what we shall then expect to see 20% people will use bus, 30% will use carpool and 45% will drive alone. Remember that this will be the actual distribution because actual choice depends

on car ownership and actual choice is different for group 1 and group 2 people in terms of their perceived weightage on the travel time.

So, that is why we are getting this distribution in reality, but the modeller will predict what based on this utility value, modeller will predict that everybody is going to use carpool. So, this difference is due to the 2 reasons, one is that we have not considered in our model the car ownership that is one part. The second is we did not consider heterogeneity in the group in terms of their perceived weightage on the travel time.

So, he did not consider group on group 2 separately. So, while the modeler will predict that everybody use will use carpool, actually absurd thing would be that 20% will use bus, 35% uses carpool and 45% will use drive alone. So, that is what it is.

(Refer Slide Time: 30:27)

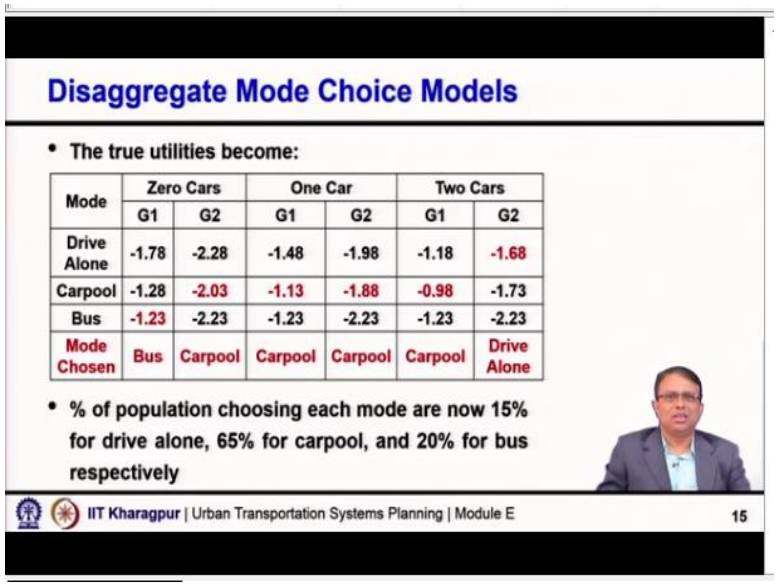


Now, let us consider one more little bit addition. Addition is that let us consider we are now imposing a parking tax of 30 rupees for the drive alone. If deterministic component of the utility what will happen then decrease. Decrease from -1.25 to -1.48 why, because that posture will change in the equation, look at this equation 3, 3, 1a, 1b in all 1 and 2 also for the drive alone, this utilities are going to change.

And for my equation 3 also as a modeller, what do I have specified in that equation also the utilities are going to change. So, what I will find? I will find now, that my as a modeller, what is my specified model, there also the utility of drive alone has decreased from -1.25 which

was earlier value and now because we have considered 30 rupees parking for drive alone. So, my disutility becomes 1.48 or utility becomes -1.48.

(Refer Slide Time: 31:55)

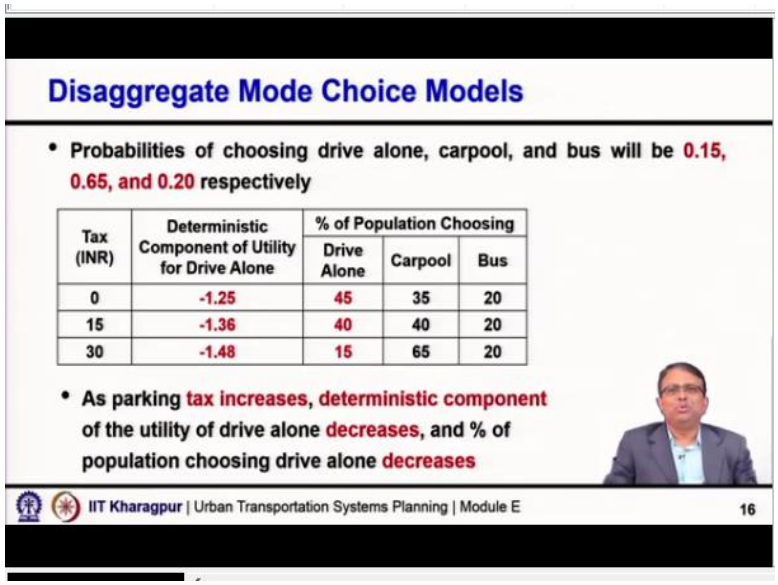


Similarly, this will also have impact on the utility of drive alone for 0 car ownership, 1 car ownership, 2 car ownership people and also for income group 1 and income group 2. So, those utilities also will change somewhere and because the utilities will change the utility maximization principle wise choice also in some cases will change. So, that; change choices what actually would happen that is shown here in this table.

Again, I know the distribution of population across car ownership and across income group 1 and group 2 people. So, knowing that distribution and knowing the choice for each group, I can now find out that 15% people will use drive alone, 65% will use carpool and 20% will use bus. What was the result? True result, result was earlier 20% bus, 35% carpool, 45% drive alone.

So, 20, 35, 45 now what are the values? Now 15, 65, 20. So, what has happened? The drive alone portion, percentage of people who are using drive alone that has dropped, that has dropped from earlier it was 25, 20, now it has become 15, why this has happened? Because I have increased the parking charges, I have imposed the parking charges of 30 rupees.

(Refer Slide Time: 33:41)



Now, like that I am showing you here one table where 0 parking charge the first case, second case 15 rupees, third case 30 rupees. Now, in each case, what is the deterministic utility of drive alone as considered by the modeller, I do not know all these errors, I do not know that I should have considered the car ownership, I could have considered 2 groups of user, I do not know all these.

So, in my consideration also the utility will change first case it was -1.25, second case -1.36, 3 case -1.48. Similarly, the actual choice considering setup equation 1 and 2 and considering 2 income group user group, group 1, group 2 and 3 car ownership level, percentage, their individual choice and overall distribution, the drive alone percentage will be how much? First is 45% in reality actual choice.

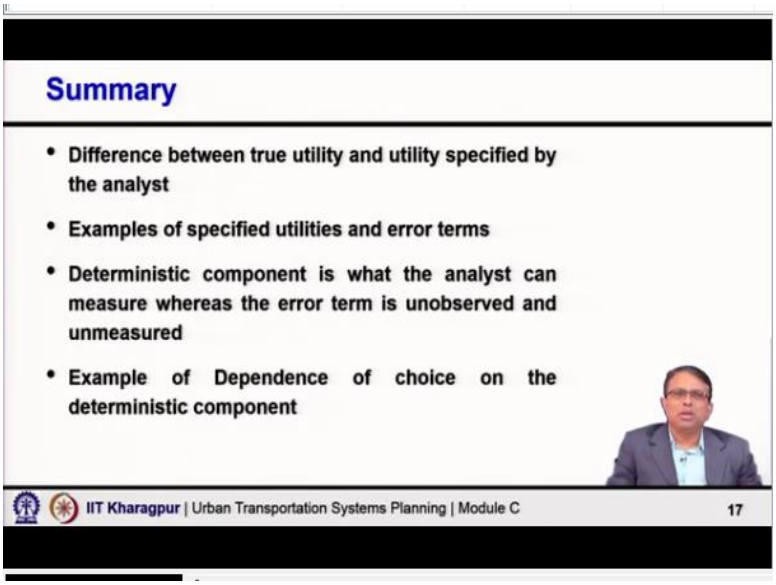
Second case only 40%, third case only 15%. So, what you are seeing here as the disutility is decreasing for a mode in this case, it is driving alone, as the utility is decreasing the percentage of people using that alternative drive alone is also decreasing, that is what the statement we meant, that if the disutility of a mode increases, it decreases then the probability of using that mode will also decrease, that is shown here.

So, that all what we did? We tried to explain you this particular statement what we made earlier, we tried to explain you with some example. So, altogether what is happening as you said that the probabilities is decreasing of choosing the driver alone because you have as you are increasing the parking charge or as you are increasing that utility of that particular alternative. So, utility decreasing so probability is also decreasing. So, as parking charge

increases deterministic component of the utility decreases and percentage of population choosing drive alone also decreases.

That justify this previous statement what we said about deterministic component and its change and the corresponding likely change in the probability of choosing the mode.

(Refer Slide Time: 36:31)



So, what we discussed here in summary, we discussed the difference between the true utility and the utility specified by the analyst strata modular we gave you an example of specified utility and error term, several examples we discussed. We explained you deterministic component is what the analyst can measures, whereas the error term is unobserved and unmeasured, we cannot observe, we cannot measure also in reality. And we gave you an example of how the utility is dependence of choice on the deterministic component with this I close this lecture, thank you so much.