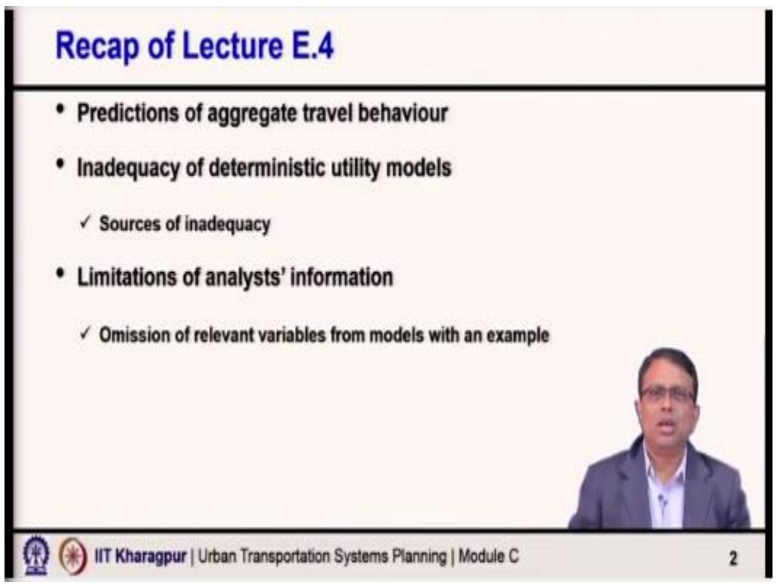
Urban Transportation Systems Planning
Prof. Bhargab Maitra
Department of Civil Engineering
Indian Institute of Technology-Kharagpur

Lecture-35
Disaggregate Mode Choice Models-IV

Welcome to module E, lecture 5. In this lecture, we shall continue our discussion about disaggregate mode choice models.

(Refer Slide Time: 00:25)



In lecture 4, we discussed about how we can do the prediction of aggregate travel behaviour using disaggregate models. We said that we can consider different segments based on the income or the characteristics what we are using. And then for different segments, what would be the choice that we can predict using the disaggregate mode choice model. And then if we know that in the overall population, what will be the shares for these different segments?

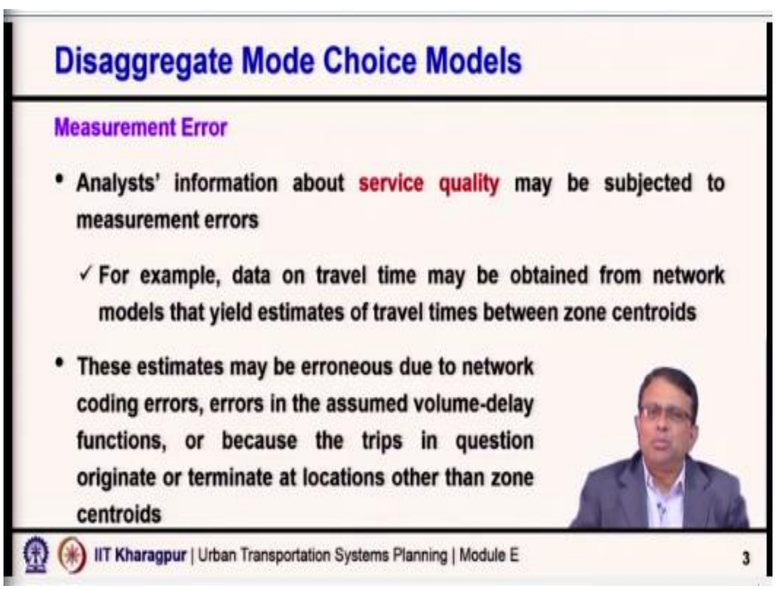
Say some income group 20%, some income group 10% and so on so forth. So, then we can predict how the aggregate mode choice will happen in the overall 100% population, maybe certain percentage will use one kind of mode, certain percentage will use another kind of option and so on. Then, we discussed about the inadequacy of deterministic utility model, what are the sources of inadequacy?

One aspect we highlighted that was the information available to all trip makers may not be same. The second we said there are certain limitations in terms of analyst's information. There are certain limitation, these limitations relates to the modeller who is doing the modeling part. Then we identified a few major areas to explain you further about the limitation of analyst's information.

And one such limitation namely omission of relevant variables from model that we discussed. And we also took an example to tell you that how because of omission of these relevant variables in a model; the choice may be different from the outcome, when we consider this irrelevant variable in the model. So, we have taken 2 models, one omitting these relevant variables and other including this relevant variable as well.

And then we said that, how just because of omission, the prediction may be very different or erroneous. So, we shall continue our discussion in this line regarding further limitations of analyst's information.

(Refer Slide Time: 03:12)



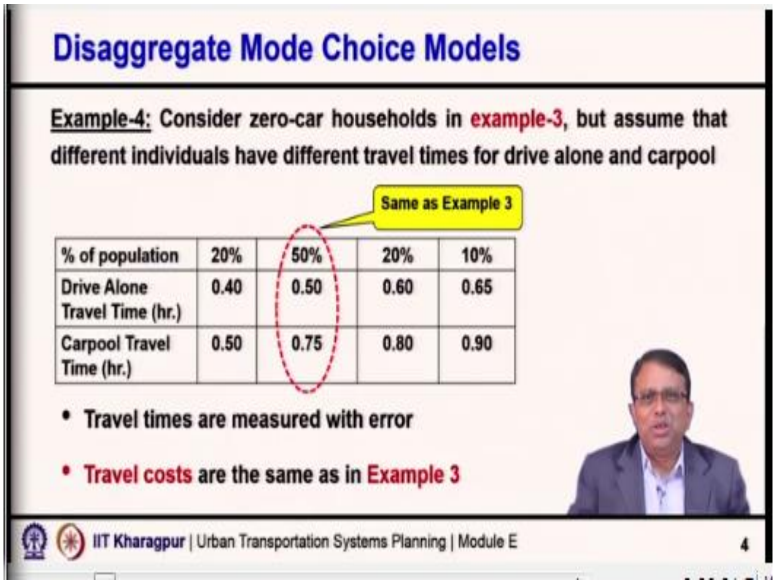
The second one relates to measurement error. Now, analyst's information about service quality may be subjected to measurement error. Every data we use, there could be some kind of error measurement error. For example, data on travel time may be obtained from network models that yield estimates of travel time between zone centroids. Whereas, all people living in a zone may not have same travel time.

Even if you consider your own locality, the distance to bus stop from every house is not same, they are different. So, when you do and get the data aggregate level data you use, you can (()) (04:02) what is the average distance to bus stop. And then that will take for everybody and try to predict it. But when you take the average distance and when you take maybe further small pockets and consider the different people living in different area.

And their actual travel time or time to bus stop or time to travel to travel to their respective destinations. The predictions or the outcome as per the utility maximization principle itself may be very different. So, these estimates may be erroneous due to network coding error, errors in the assumed volume delay function. These are all assumptions maybe because of the trips in question, originate or terminate at locations other than the zone centroid as I said.

So, there could be so many different reasons for this error. But the question is because of this measurement error the reality and whatever is given input to model may be different. So, the actual choice and the model choice also will be different. Again, this is not because utility maximization principle is something in question, but it is because of this measurement error. But this is kind of things which may come and probably is kind of unavoidable, you cannot say that in my model there is no measurement error, it could be, quite possible that there will be some measurement error.

(Refer Slide Time: 05:58)

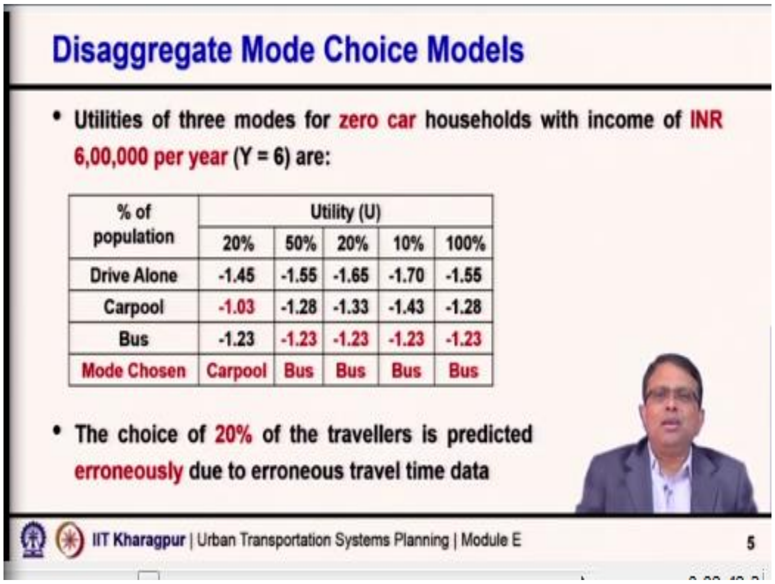


Let us take an example. Hopefully you remember the previous example what we took in the previous lecture. Now, we took 0 car, 1 car, 2 cars and let us take that you know zero car part no vehicle in the ownership, all values are 0, remaining everything is same. But now I am saying if you take that, take the drive alone and carpool travel times were giving us 0.5 and 0.75 hour, that was the time.

Now what we assumed there everybody will have this exact 0.5 and 0.75 travel time by drive alone and carpool that was the assumption. But now you consider that everybody may not staying in the same location, different bands are there. So, you can consider that maybe around the zone centroid, there are different rings or different zones around that. And depending on where you are actually there in that traffic analysis shown, accordingly your actual travel time will vary.

So, let us consider maybe for 20% people, the respective travel times are 0.4 and 0.5, for 50% it is yes, as per what you considered earlier say zone centroid to zone centroid 0.5 and 0.7 5 maybe for 20% people the travel times are 0.6 and 0.8 hour respectively. And for 10% it is 0.65 and 0.9 depending on within that same zone where they are located. So, travel times are method when we used say for all or 100% people 0.5 and 0.75 obviously my inputs are not correct for everybody we are giving erroneous input. And now let us see what is the consequence of that suppose if it is so as I said in this example, then what will be my predictions?

(Refer Slide Time: 08:44)

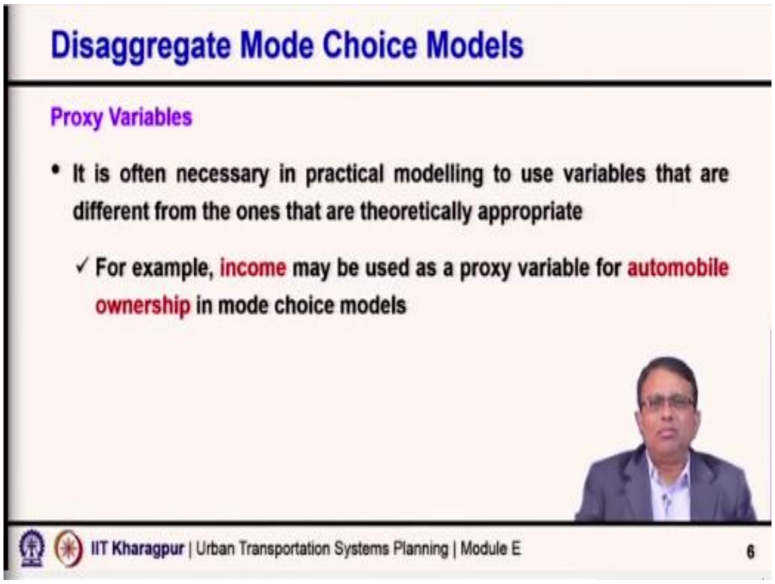


Now as per the actual travel time and then for each segment you calculate the utility of drive alone, carpool and bus and make the prediction as per the utility maximization principle. You will clearly say for the first 20% group, the choice is going to be carpool. Because of the actual travel time by drive alone, carpool and bus. For of all the 3 or the segments the choice is going to be bus. So in reality, if you considered first 4 columns 20%, 50%, 70%, + 20%, 90% + 10, 100% that is the distribution actually, actual travel time and travel cost.

So, as per actual travel time and travel cost, we can say 20% people will use carpool and the 80% will use bus. Now interestingly the fifth column the last one with 100%, what we are saying if we use that everybody would have this travel time from the zone centroid to zone centroid what was considered earlier in earlier example, then what you would predict? We would predict 100% people will use bus.

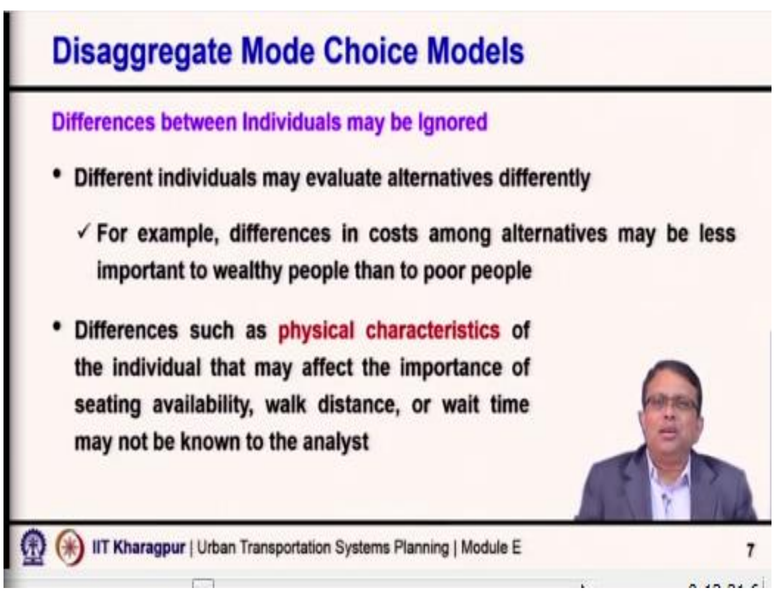
So, you can say or you can see it clearly that we shall make erroneous prediction for 20% of the traveler just because of the error on the measurement error in the travel time data. So, this tells us because of measurement error, our prediction from the model deterministic mode choice model disaggregate mode choice model. And the actual behaviour may be different, that is what is shown through this example. So, this is one more reason, why? The prediction from an deterministic mode choice model may not match with the reality or the real behaviour.

(Refer Slide Time: 11:35)



Third, use of proxy variables, sometimes we know that we should use this particular variable but simply maybe the data is not available. So, you use another data which we call it as proxy variable. So, for example, income may be used as a proxy variable for automobile ownership in mode choice model, you do not have the real automobile ownership data. In the absence of that, we use income as a proxy variable, higher the income, higher the vehicle ownership or automobile ownership. Now, what is the consequence of that?

(Refer Slide Time: 12:18)



Now because of that, again my actual prediction or actual behaviour and model prediction may be different. Next, difference between individuals are ignored, often ignored in the deterministic model. For example, difference in cost among alternatives may be less important to wealthy people than the poor people, not everybody will perceive the cost in the same manner. So, you are taken an example where you have not taken income maybe income is not there.

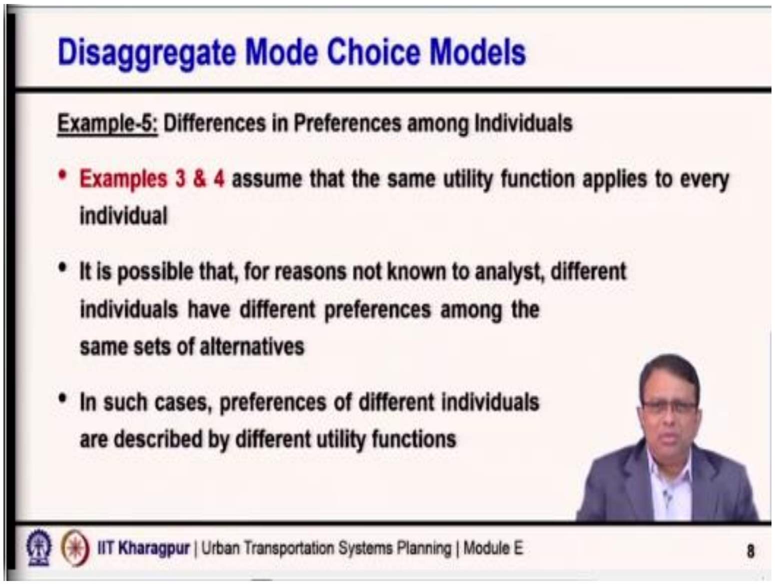
So, whatever you have taken that is not able to capture this person to person variation. There could be other reason, even the physical characteristics, the choice of mode, maybe the availability of seat in a public transport. A young person may not give it that much weightage, whereas for an elderly person that may be a major consideration. He may not or she may not use public transport simply the seat is not available and public transport is too crowded.

So, there are many such physical characteristics it may be age, it may be income, it may also be the gender, the security perception may vary. So, differences between individuals may be ignored. So, we are considering everybody same because 1 or 2 characteristics we have taken in the model, say income, or maybe age, or occupation something 1, 2, 3 variables you have taken maximum we can take.

And we are thinking that just by those 3 characteristics we are making everybody same, if 3 characteristics are matching, same then all individuals are going to be same it is in a way very strange, it cannot be so simple**.** So, because we are ignoring the differences between individual. So, we shall whatever characteristics we have considered only based on those characteristics, we shall make certain prediction that everybody will is going to make the same choice.

But in reality, because there are other attributes which we are ignoring, we are not considering the difference between individual. So, our model prediction and actual behaviour may be different. So, this is one more reason, why the disaggregate model may not be able to give absolutely correct prediction.

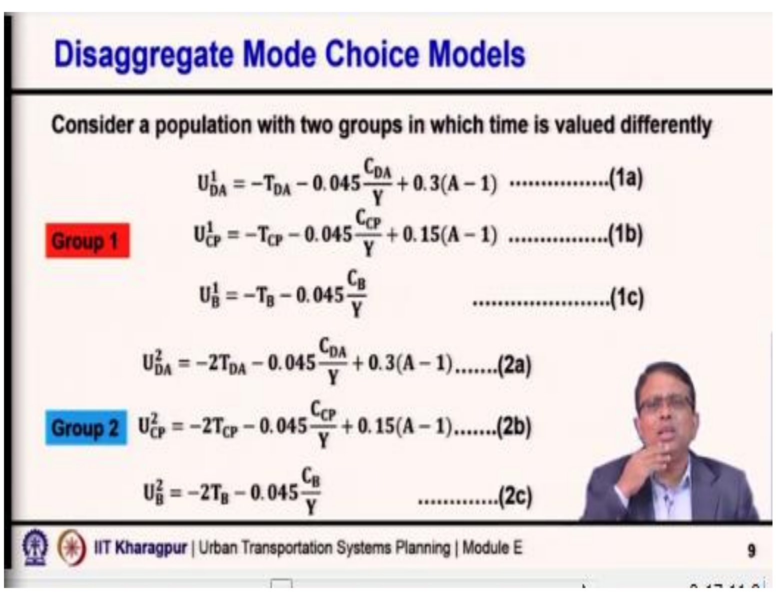
(Refer Slide Time: 15:51)



I have given a kind of example here. Difference in preference among individual let us take the example what we just discussed now, and also the example what we discussed in the previous lecture, where we assume that the same utility function applies to every individual. Now it is possible that for reasons not known to analyst, different individuals have different preferences among the set of alternatives because of some reason, some characteristics.

In such cases the preference of different individuals are described actually by different utility functions, but we are using the same utility function. So, naturally by model outcome and the real behaviour may not match always.

(Refer Slide Time: 16:57)

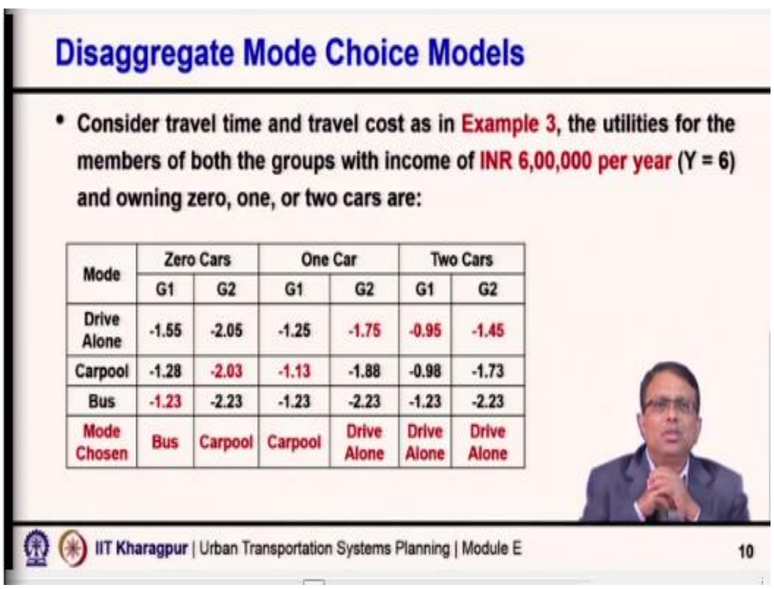


To take this example forward, now let us consider that we took the same utility equation weightages are all same, coefficients are all same. But now let us consider there are 2 groups of people, again simplicity. In reality it is not even 2, it may be even many groups, but just to convince you I am taking maybe 2 groups. These groups are different in terms of their perceived weightage on travel time.

You can see every other coefficient is same, only the travel time weightage on the travel time is not same in the 2 group. Group 2 give higher weightage to the travel time than group 1, it is possible for various reasons. I tell you maybe trip purpose wise the weightage to travel time could be different, people are going to office maybe different, gender wise male, female, the weightage could be different, income, of course here some variable is considered.

Otherwise also suppose, if you are not taking in (()) (18:26) because of income wise the high income and low income, the perceived value of travel time. And therefore at the backend the weightage on that travel time may not be really same may be different. So, we have actually 2 groups which are there in the population, not one group, and for one group the weightage on the travel time is higher than the other group.

(Refer Slide Time: 19:05)

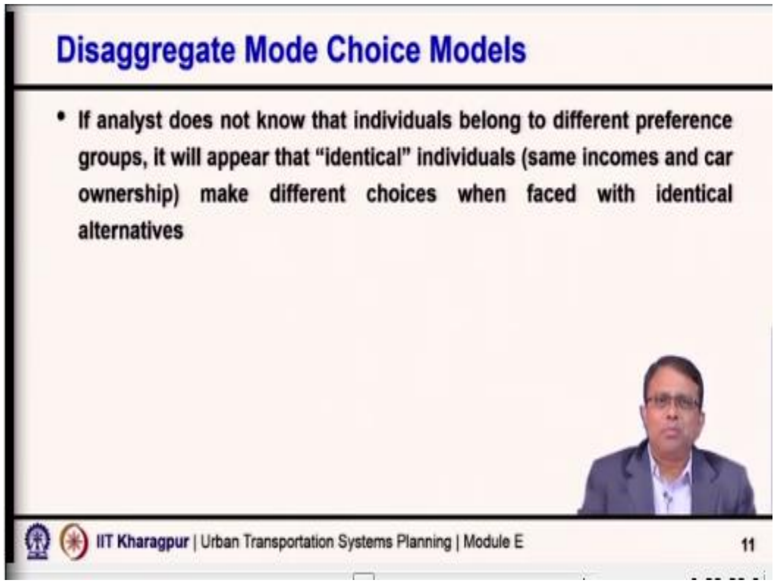


If this is true that we have 2 groups, then how our predictions will be, let us look at it. For people having no car that is 0 car, group 1 will actually use bus whereas group 2 will use carpool. Now group 2 is our inclusion here, we recognized here that actually not one group but 2 groups are there. So, we said that and we could get that from the model cell and from the utility maximization principle, that group 1 people they are going to use bus whereas group 2 people are going to use carpool.

But if I do not consider this in all the previous example, where we did not make this group 1 or group 2 consider them just as one group, there everybody was predicted to using bus. So, and you consider further grouping, choice is actually different. Similarly, for one car, group 1 would use carpool, group 2 will use drive alone, what we predicted was just one choice for everybody. Because we did not distinguish between group 1 and group 2, same thing you can see for group 3 also, we calculated the choices.

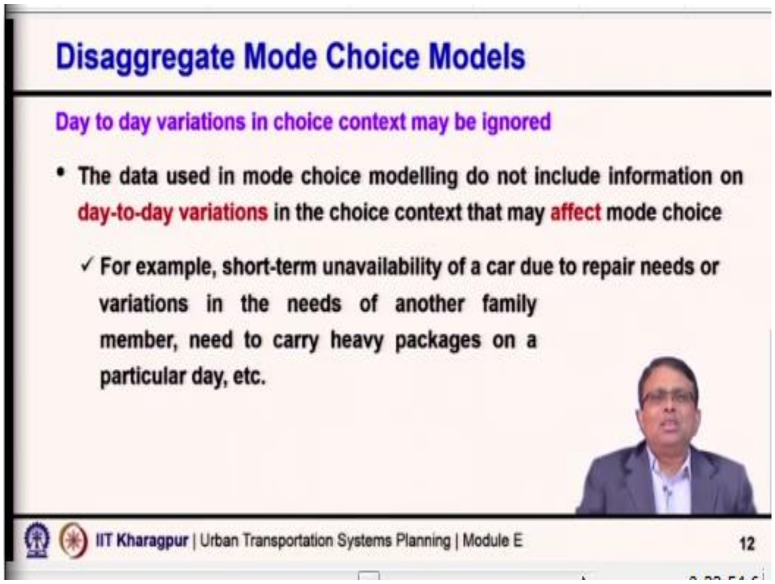
Here incidentally both group 1 and group 2 they are predicted to use drive alone. So, here you may not incur any error, but for 0 car and 1 car there would be an error. So, because this differences in preferences among individuals are not considered the prediction, in some cases would be erroneous, that is what is shown in this example.

(Refer Slide Time: 21:32)



If analyst does not know that individuals belong to different preference crew, it will appear that identical individual same income, same car ownership, but making different choices when faced with identical alternatives, it will be surprise, why? It will appear like same or identical individual, identical in terms of their income, in terms of their car ownership, facing same choices. But then some people are making this choice, some people will making another choice, it would be different, but it is because we did not consider that difference.

(Refer Slide Time: 22:30)



Last is day to day variation in the context may be ignored. Little bit give a discuss this in the previous lecture but this is again another interesting point. The data used in mode choice modeling do not include information on day to day variation in the choice context that also may affect the mode choice. For example, short term unavailability of a car due to repair needs or maybe variations in the needs of another family members.

Or need to carry heavy package on a particular day, all this may make some variation in the mode choice from one day to another day. So, maybe every day the person was using car, but for the moment the car is not available, because you have given the car for maintenance or repair. So, you are travelling by taxi, my model will not be able to predict that choice correctly, or everybody, everyday person probably travelled by car.

But that day, one particular day the family members probably they want to go somewhere and you leave the car and take a bus or take a taxi. So, my mode choice model will not be able to predict that correctly**.** Because I am not considering that variation or as I say somebody may have a health reason, simply not feeling well on a particular day, so using a different mode of transport for travel.

So, that kind of thing will not be captured in my model, I raise a question here, very important question and pertinent question, yes. As an analyst, we may not have all this information with us, it is impossible, can you just predict? Who will have his car in the garage or who will not feel well or fit on particular day or who has got family members have got an invitation somewhere on which date, it is impossible for us.

So, this kind of problems or this kind of limitations will remain, we cannot eliminate such kind of limitations completely. Now also if there is one more very relevant question here. Suppose, somebody tells you, ok, fine, I will give you all the data you require, anything whatever you missed. Due to all these reasons I discussed, somebody tells you that, I will give you all the data what you require, take it.

So, what will be your response? You will be very happy, that ok, now I have got all the data, so let me include everything in my model. Will that be our approach? Unfortunately, the answer is no, ok, if I can reduce my measurement error, I would definitely try to do that. If I can use the correct variable, I may be interested to do that rather than using a proxy variable, fine. But let us consider day to day variation, you tell me what I will do with that information even if I model it, what is the use of it?

I cannot, can I predict this? Suppose you get the present data and you make a model and you say, my model is giving very nice or very good prediction. But the question is model is to be applied in the future, can you predict in the future which day, which person will have some other reason not to use the mode or which day, somebody will not feel or how frequently people will become ill, can I predict that? I will not be able to predict it.

So, including anything in the model, in a way which I cannot control and which does not relate to my transport policy is a junk for me, I am not interested in those also, that is the another aspect. So, we have limitations that we may not get all that information, we may not get all the data, some error will be there, sometimes we may have to use proxy variable. Sometimes there will be some other issue, omission of variables all such limitations are there.

Some of these limitations you can overcome, all these limitations, you cannot overcome. The second is even if somebody tells you let us take it, granted I give you all the data you like. Then also as a modeller, I am not interested to include those things, you know ultimately why we make model? We make model because of the policy, I want to see how my mode choice is going to be affected, if I improve the public transport system in terms of my bus journey time.

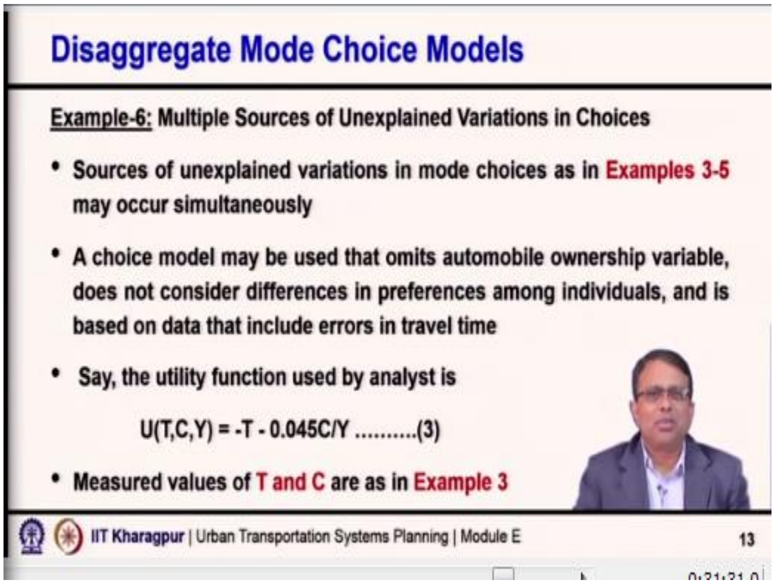
How then I can bring more people? How far can impact the choice of bus? How developing the bus network further? And thereby reducing the walking time can help me to get more people into the bus system. How by running more buses, more frequent services? So, the headway or the waiting time will get reduced, how that is going to help me to improve the bus system? That is where is our interest.

So, you want to include variables which are policy relevant, who will have a breakdown in the vehicle and who will have his car maintenance or who will be carrying a heavy luggage one day, what I will do in those? I cannot formulate a policy, and if I cannot use them for policy formulation, if I cannot use them for improving my bus system, then I am not interested to use those variables in my model.

So, you have limitations which will remain and all the limitations even if somebody gives you every data as a modeller also we are not interested. But at the same time we want to be accurate, more accurate in our predictions, so what is the solution? How we approach it? That brings us to the next step which is deviating from the deterministic model and then actually going to probabilistic mode choice model.

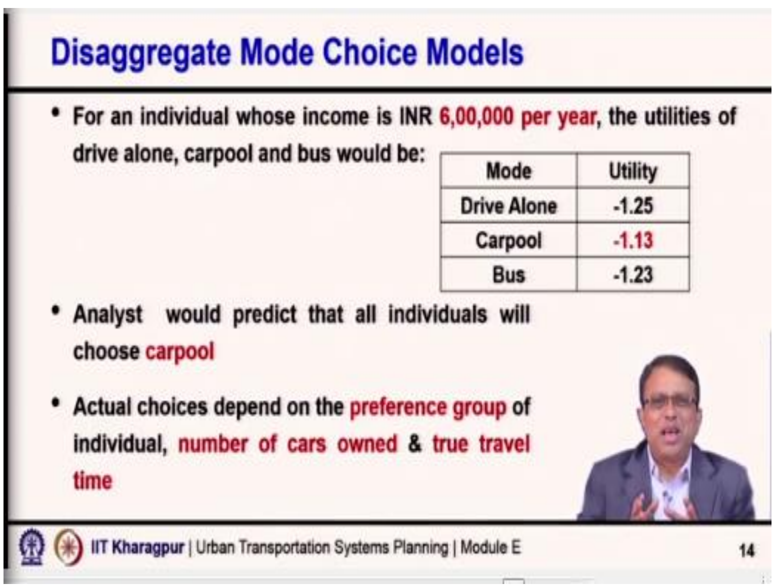
This is also disaggregate, that is also disaggregate, but what we discuss till now in the disaggregate mode choice model, we discussed about deterministic disaggregate mode choice model. So, in our next lecture, this will be our foundation, these are all the valid reason why we would deviate from deterministic disaggregate mode choice model and go to probabilistic disaggregate mode choice model. So, that will be our discussion in the next few lectures in the next week.

(Refer Slide Time: 31:19)



So, I have also taken an example here, where multiple sources of unexplained variations are considered. The sources of unexplained variation in mode choice as in example 3 to 5 may occur simultaneously. So, that I have say, all these I said 1, 1, 1, now I suppose all these may also occur combined not that always I will have only measurement error, always we will have proxy variable, it may omission of variable, all multiple things may be combined together.

(Refer Slide Time: 31:57)



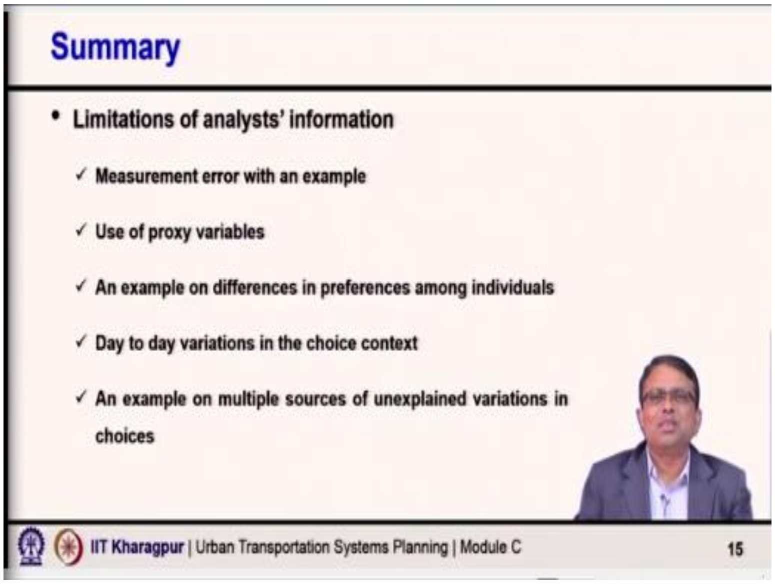
So, that is what we have shown that, whereas, things should be so different, but we will actually predict like everybody's choice will be same. Because everybody to us the way we have defined a group, the kind of measurements we have taken, the kinds of variables we are taken based on that we are defining a group. So, every individual in that group will appear to us identical, because we have no instrument, no other way to distinguish among those people.

So, the actual choice if you compare the original example with all the subsequent example, the actual choice is depends on preference group of individual, number of car owned and true travel time weightage to weightage on travel time. So, because we may miss all these our model prediction and real behaviour may be much different. And as I said, that is the reason these limitations are the reasons why we said all these because we wanted to communicate to you.

That these are the reasons why our model prediction and actual behaviour may be different. But then we say some of these errors with definitely try to reduce as far as possible, but every error we cannot reduce, we cannot eliminate every limitations we cannot eliminate. And even if suppose somebody comes and tell us, I give you all the data and everything correct, you consider everything, shall I start considering every variable and then build a model?

There also my answer was that no, I am only interested to enclose policy relevant variables in the model, not any junk variables. Anything other than policy which does not have any policy implication or operational implication, they are junk variables to me.

(Refer Slide Time: 34:14)



So, our next step would be with all this discussion is a ship to probabilistic disaggregate mode choice model. So, what we discussed in summary in this lecture are various limitations related to measurement error, proxy use a proxy variable, then we took some example also. Then day to day variation and that also we took an example, and then clearly told you why and how the deterministic mode choice models are deficient? And then, that gives us to now shift to the probabilistic disaggregate mode choice model. So, we shall meet you, I will meet you and discuss further in the next week, thank you so much.