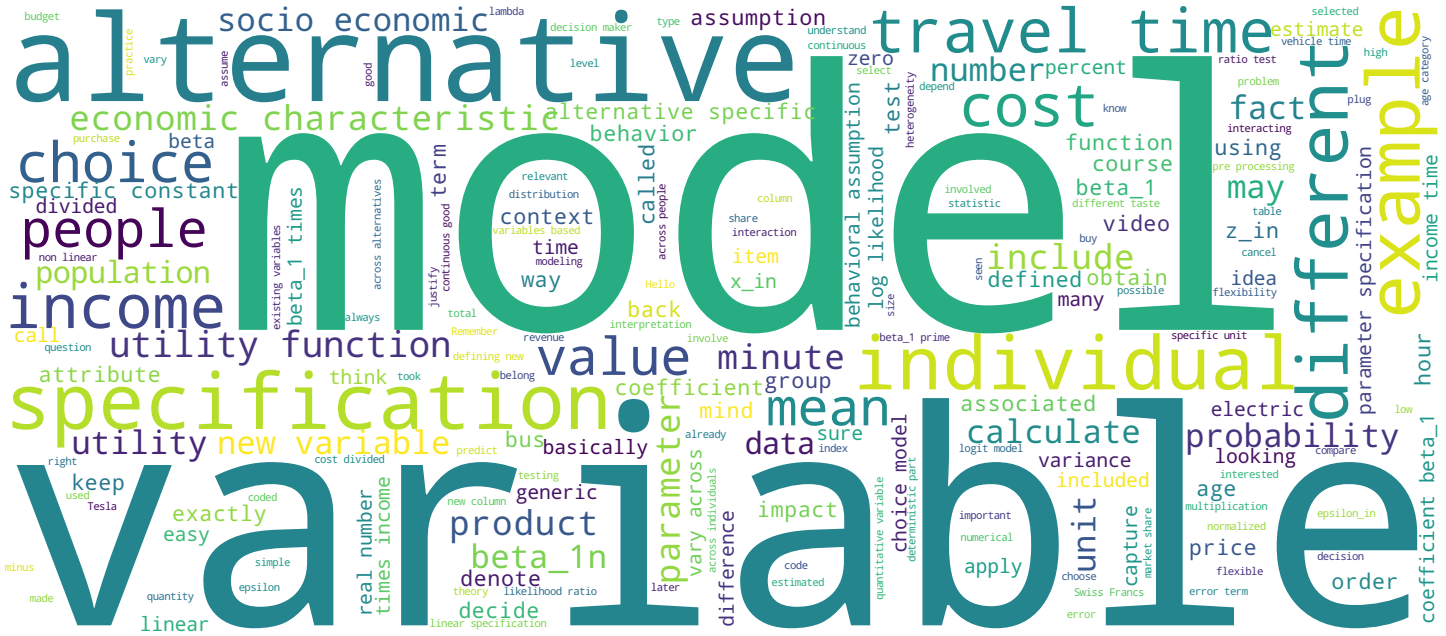
Choice with multiple alternatives

Specification of the deterministic part

Michel Bierlaire

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE



Search MOOC

Video

EPFL

Quantitative explanatory variables

Hello. In this video, we will now look at how to specify the deterministic part of the utility function, in order to model the fact that people behave differently. Everybody is different. We have a heterogeneous population and we would like to make sure that we're able to capture that in the model. That's the topic of this video.

Notes

Summary

0m 04s

Quantitative attributes

Numerical and continuous

- $(z_{in})_k \in \mathbb{R}, \forall i, n, k$
- Associated with a specific unit
- Vary across both i and n.

Examples

- Auto in-vehicle time (in min.)
- Transit in-vehicle time (in min.)
- Auto out-of-pocket cost (in cents)
- Transit fare (in cents)
- Walking time to the bus stop (in min.)

Straightforward modeling

So let's consider the variable number k for individual n and alternative i. It's represented by a real number that we denote by (z_in)_k. And this real number is associated with a specific unit in order to have an interpretation. As an attribute, it varies both across alternatives i and individuals n. So there are plenty of examples that we can cite. Again, if we look at the example of the transportation mode choice, you can think about "Auto in vehicle time" which is a quantity expressed in minutes. Or the "transit in vehicle time" which is also expressed in minutes. You can have the cost represented in cents, or dollars, or Swiss Francs, or what have you. Or the transit fare in the same units. You can consider the walking time to the bus stop, in minutes. So everything is very straightforward. OK. So if you want to involve a variable, an attribute, which is quantitative and numerical, the only thing you need to do is to think about the units, which is a natural way to think about it already at the data collection stage, and include this as such in the model.

Notes

Summary

0m 26s

Quantitative attributes

- V_{in} is unitless
- Therefore, β depends on the unit of the associated attribute
- Example: consider two specifications

$$\begin{aligned} V_{in} &= \beta_1 \mathrm{TT}_{in} + \cdots \\ V_{in} &= \beta'_1 \mathrm{TT}'_{in} + \cdots \end{aligned}$$

- If TT_{in} is a number of minutes, the unit of β_1 is $1/\mathrm{min}$
- If TT'_{in} is a number of hours, the unit of β'_1 is $1/\mathrm{hour}$
- Both models are equivalent, but the estimated value of the coefficient will be different

$$\beta_1 \mathrm{TT}_{in} = \beta'_1 \mathrm{TT}'_{in} \Longrightarrow \frac{\mathrm{TT}_{in}}{\mathrm{TT}'_{in}} = \frac{\beta'_1}{\beta_1} = 60$$

About the units, what you need to keep in mind is that the utility itself has no unit. Therefore, the coefficient beta of the variable will actually cancel out the units. So the value of beta will depend on the unit that has been selected for the attribute. So let's take an example. Suppose that you want to include travel time in your model and you have two possibilities: you can code the travel time in terms of number of minutes, but you can also do it in terms of number of hours. In the first example, where travel time is coded in minutes, the beta_1 has also a unit. Remember it cancels the unit of minutes. So beta_1 will be 1 divided by minute. In the second specification where travel time prime is actually the same travel time but coded in hours, well beta_1 prime now is 1 divided by hours. Both models are completely equivalent. They are exactly the same. It does not matter if you code it in minutes, hours or what have you. But, the estimated value of the coefficient will be different because the units are different. So if you realize that the contribution of travel time to the utility function is the same in the two models, because the ratio of the two travel times is 60 (between hours and minutes), well the ratio between the two coefficients will be also 60. OK, so beta_1 prime will be 60 times bigger than beta_1.

Notes

Summary

1m 38s

Quantitative attributes

Generic vs alternative specific

$$\begin{aligned} V_{in} &= \beta_1 TT_{in} + \cdots \\ V_{jn} &= \beta_1 TT_{jn} + \cdots \end{aligned}$$

or

$$\begin{aligned} V_{in} &= \beta_1 TT_{in} + \cdots \\ V_{jn} &= \beta_2 TT_{jn} + \cdots \end{aligned}$$

Modeling assumption: a minute has/has not the same marginal utility whether it is incurred on the auto or bus mode

Another important feature that you need to keep in mind when you include a quantitative variable in a specification, is to decide if this variable will be generic or alternative specific. What does it mean? Let's compare these two specifications. The top specification, the travel time is involved in two alternatives. For example, alternative i can be the car, driving by car, and alternative j can be taking the bus. So TT_in is the travel time by car, TT_jn is the travel time by bus. And if you look correctly at the specification, the coefficient beta_1 is the same for car and for bus. So it's the same for i and j. In the second specification, we have made a different assumption. We have decided that the impact of travel time on the utility will be different, for alternative i and for alternative j. To do so, we have defined two different coefficients. We have the coefficient beta_1 which calculates the impact of travel time on alternative i. And the coefficient beta_2 which captures the impact of the travel time for alternative j. And actually, as we will see later on, you can then test these two specifications against the data and then decide which one is the most appropriate for your context.

Notes

Summary

3m 16s

Quantitative socio-eco. characteristics

Numerical and continuous

- $(S_n)_k \in \mathbb{R}, \forall n, k$
- Associated with a specific unit
- Vary only across n, not across i.

Examples

- Annual income (in KCHF)
- Age (in years)

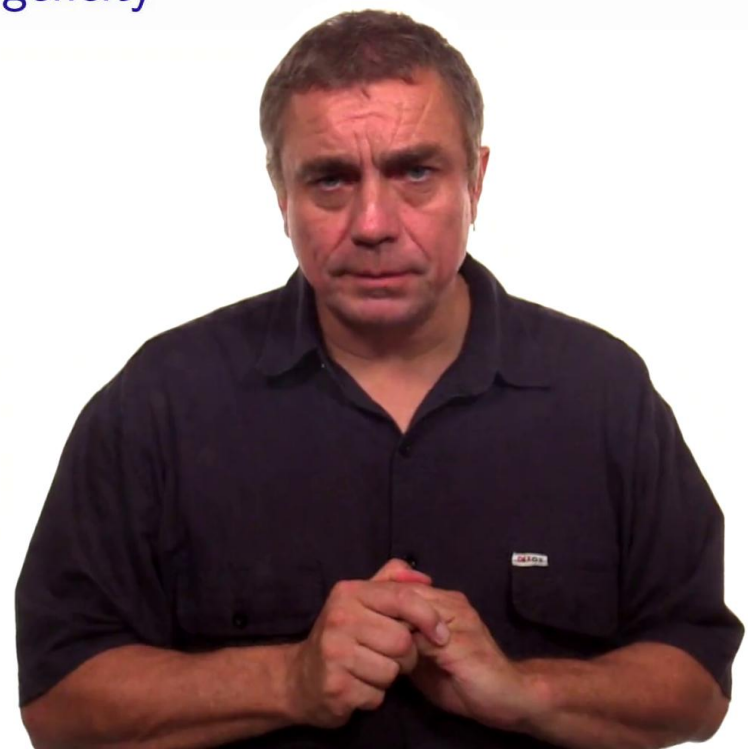
Now, let's move from attributes to socio-economic characteristics. Well, as you will see, it will be basically the same. So, because we are looking at quantitative variables, they are again numerical and continuous. They are represented by a real number that we denote here by (S_n)_k. Remember, a characteristic does not vary across i, across alternatives. I mean, if you take the bus or the car, your age will always be the same, right? So the index is n, the individual. It's again associated with a specific unit, like Swiss francs for annual income, or number of years for age. These are the two classical example of continuous social economic characteristics.

Notes

Summary

4m 42s

Modeling heterogeneity



OK so far so good. It was easy. We take these continuous variables and we include them in the model. We may want to model the fact that they are generic or specific. But let's see if we can go further in the modeling of behavior. An important part of what we want to do is to be able to model the fact that people are different. People have different tastes. People do not react the same way in the same situation. So this is what we call "heterogeneity". The population is a heterogeneous. People are different and we would like to be able to model this mathematically in our framework. This is what we will be looking at now.

Notes

Summary

5m 26s

Modeling heterogeneity

Interaction

Typical definition of β_{1n}:

$$\beta_{1n} = \beta_1 \text{income}_n$$

$$V_{in} = \beta_{1n} z_{in} + \cdots = \beta_1 \text{income}_n z_{in} + \cdots = \beta_1 x_{in} + \cdots$$

where

$$x_{in} = \text{income}_n z_{in}$$

So I write my utility function V_in and the difference between what I did before and this specification is that I have included an index n to the coefficient. So by doing this, I say this beta_1n, which captures the impact of a variable Z_in on the utility, this beta_1n will vary across people. Now the question is how does it vary. Well, for example, it can vary as a function of income. A typical example is that, if Z_in is the cost variable, the way the cost of a product, or an option, influences the utility, will vary with income. Maybe rich people will be less sensitive to cost than poor people. That's the behavioral Assumption. So beta_1n now becomes a function of a socio-economic characteristic like income. This is called interaction. We are interacting an attribute, which is a variable that describes the alternative (like cost), with a socio- economic characteristic (like income). So, for example, one possibility to relate the beta_1n and income is to say: beta_1n is equal to the product of a coefficient beta_1, which is now generic for all individuals, times income. So now we plug this specification in the utility function to have this. We have beta_1 times income times z, which is the variable.

Notes

Summary

6m 06s

Modeling heterogeneity

Interaction

Typical definition of β_{1n}:

$$\beta_{1n} = \beta_1 \text{income}_n$$

$$V_{in} = \beta_{1n} z_{in} + \cdots = \beta_1 \text{income}_n z_{in} + \cdots = \beta_1 x_{in} + \cdots$$

where

$$x_{in} = \text{income}_n z_{in}$$

And actually, there is a very interesting way to see this specification. Let's denote income time z, let's denote it by x, x_in. So x_in is defined as the product of income and z_in. If we do that, then we are back to the original specification of a linear parameter formula. It's beta_1 times the variable. So these interactions can be seen as defining new variables. In this case, we have defined a new variable which is a product of income and cost, for example. This is easy to do. You can do it using your Excel or whatever spreadsheet you are using. You include a new column and you calculate these variables based on the existing variables. So you go to this column, that you call x_in here, and you calculate the product between income and cost and you have a new variable. And this is this new variable, that is in your data file, that you include in the model. So this is very flexible. It gives you a way to include complex behavioral reality. Capturing the heterogeneity of people using a simple linear specification by pre-processing the data, by defining new variables based on existing variables.

Notes

Summary

7m 39s

Modeling heterogeneity

Behavioral assumption

- Individuals have different taste parameters.
- The difference is explained by several socio-economic characteristics.

$$V_{in} = \beta_{1n} z_{in} + \cdots$$

where

$$\beta_{1n} = \beta_{1n}(\text{income}_n, \text{age}_n).$$

So this behavioral assumption that people have different taste parameters can be generalized. So the difference across people can be characterized by one or several socioeconomic characteristics. So the example that I took where beta_1 was a function of income, you could do exactly the same by deciding that beta_1 varies across course both income and age. Everything generalizes very nicely.

Notes

Summary

8m 55s

Modeling heterogeneity

Interaction

Typical definition of β_{1n}:

$$\beta_{1n} = \beta_1 \text{income}_n \text{age}_n$$

$$V_{in} = \beta_{1n} z_{in} + \cdots = \beta_1 \text{income}_n \text{age}_n z_{in} + \cdots = \beta_1 x_{in} + \cdots$$

where

$$x_{in} = \text{income}_n \text{age}_n z_{in}$$

So this is the example. So I define beta_1n as the product of beta_1 times income, times age. I plug it again in the definition of the utility function and I obtain this term. And now in order to be back to the interpretation of the linear-in-parameter specification, I define this variable x_in as the product of income, age, and z_in. And now my specification is a linear-in-parameter specification. So again, this more advanced modeling of heterogeneity involving two socio-economic characteristics amounts to create a new variable, so a new column in your Excel file, that can be included in the model.

Notes

Summary

9m 24s

Modeling heterogeneity

So this is very flexible. I mean, you can combine all these variables in various ways, by pre-processing the data. You can decide to interact several variables together, attributes, socio-economic characteristics, to define new variables and include them in the linear-in-parameter specification. OK. So the danger with this flexibility is that you can do anything. Well, you may want to make sure that whatever you do is relevant, so that it has a behavioral explanation. OK? And in this case, when I took the income and the cost, well this made sense, right? The cost parameter would be different for people with different income. It's something that I can justify from a behavioral point of view. So that's what you need to keep in mind when you do that. You have a lot of flexibility, but you need to make sure that you can justify your specification based on some behavioral assumptions.

Notes

Summary

10m 12s

Modeling heterogeneity

Creativity and relevance

- Several functional forms can be investigated.
- For instance, if z_{in} is the cost variables, we write

$$\beta_{cn} = \beta_c / \text{income}_n$$

- Indeed, in this case, the new variable can be interpreted as the share of the income dedicated to this purchase:

$$x_{in} = z_{in} / \text{income}_n$$

Actually, if we go back to this cost variable, it's very common to replace the multiplication of the cost by income, by calculating the ratio: cost divided by income. The reason why you want to use such a specification instead of the multiplication is that, the new variable (cost divided by income) represents something meaningful: it's the share of the cost of the purchase of this alternative in your budget, in your household budget. If you go back to the example that we looked at in the beginning of the course, the purchase of an electric car, well if you decide to buy a Tesla, which is quite an expensive car, well it may make sense to look at the share of the cost of this Tesla in your income. if you have a low income, this variable will be very high. And if you have a large income, this variable will be low. So this is a common way to interact cost and income.

Notes

Summary

11m 06s

Modeling heterogeneity

Creativity and relevance

- Several functional forms can be
- For instance, if z_{in} is the cost va
- Indeed, in this case, t as the share of the income dedicated to

Interactions can be done with any variable. Okay you can interact socio-economic characteristics with any variable. But you can also interact them with the alternative specific constants. So if you remember, the alternative specific constants capture the mean of the error term and by assumption in the logit model, they should be the same across n, because of the i.i.d. assumption. But still, we would like to model the fact that they may be different across individuals. And we will use the same trick as for the interaction with normal variables.

Notes

Summary

12m 08s

Modeling heterogeneity: alternative specific constants

Heterogeneous specification

$$\begin{aligned} V_{1n} &= \beta_x x_{1n1} + \beta_{1n} + \cdots \\ V_{2n} &= \beta_x x_{2n1} + \beta_{2n} + \cdots \\ V_{3n} &= \beta_x x_{3n1} + \cdots \end{aligned}$$

where

$$\beta_{in} = \beta_i \text{income}_n$$

$$\begin{aligned} V_{1n} &= \beta_x x_{1n1} + \beta_1 \text{income}_n + \cdots \\ V_{2n} &= \beta_x x_{2n1} + \beta_2 \text{income}_n + \cdots \\ V_{3n} &= \beta_x x_{3n1} + \cdots \end{aligned}$$

If you look at the first specification on this slide, that I called the base model, it's the classical specification where I have included the alternative specific constants to each of the three alternatives except one that I have normalized to 0. Now, what I would like to do, like we did before, is to say: this beta_1 and beta_2, the alternative specific constants, vary across individuals. So I denote them by beta_1n and beta_2n. And like we did before, I can be more specific about the definition of beta_1n. I can say it's the product between a coefficient beta_i and the income of the individual n. In this case, we obtain this specification. We obtain the fact that beta_1 income and beta_2 income appear in the utility function.

Notes

Summary

12m 39s

So what you have to keep in mind here is that, when you include a socio-economic characteristic in the model like this, because it's interacting with the constant, you cannot include them in all alternatives. One of them must be normalized to 0 like we do for the alternative specific constants. Now, I suggest that you go to the practice quiz which is provided online so that you get more familiar with these specifications. And then in the next video, we will look into the qualitative variables that will be modeled using discrete variables and we will see how to do that.

Notes

Summary

13m 29s