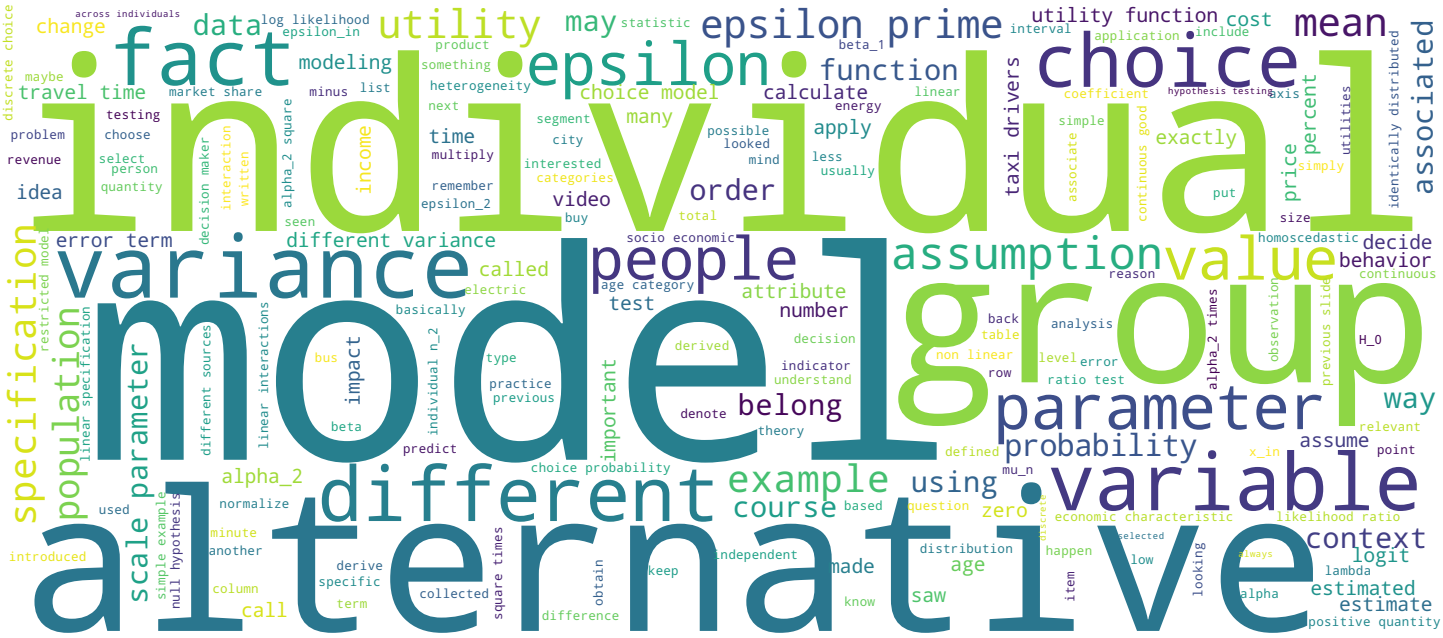
Choice with multiple alternatives

Specification of the deterministic part

Michel Bierlaire

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE



EPFL

Search MOOC

Video

Nonlinear specifications: heteroscedasticity

As you could see, we are spending a lot of energy in modeling the fact that people are different: the heterogeneity of tastes in the population. We looked at it using these interactions between the attributes of the alternatives and the socio-economic characteristics of the individuals. We saw that we could do linear interactions, and non linear interactions. Here, we will look at another dimension of the heterogeneity. The fact that the error term, which is basically containing everything that we were not able to put in the model, may be different from one individual to the next. The fact that these error terms may be different is actually contradictory to the assumption we made for the derivation of logit. So let's see how we can actually relax this assumption in a way that is compatible with the mathematical properties of the model.

Notes

Summary

0m 04s

Heteroscedasticity

Logit is homoscedastic

- ε_{in} i.i.d. across both i and n.
- In particular, they all have the same variance.

Motivation

- People may have different level of knowledge (e.g. taxi drivers)
- Different sources of data

So logit, has we derived it, is based on the assumption that the epsilon_in are i.i.d., so independent and identically distributed, across both the alternatives i and the individuals n. In particular, they all have the same variance. In technical terms, the fact that we have the same variance, is referred to as "homoscedastic". Now, we would like to model the fact that people actually in reality, may have different variance. Why? Because they may have a different expertise, a different level of knowledge for the choice context. If you consider the choice of an itinerary in a city, taxi drivers are more skilled, they have more experience. So the epsilon that they have will be different from a newcomer in the city who has no clue about the city, and is looking for his way on the fly. So that's what we would like to do. To model the fact that these two categories of people have different variances. But this is also relevant in the context where we combine different sources of data. Suppose that you have collected a survey last year, and now you collect new data this year, you would like to combine them.

Notes

Summary

0m 55s

Heteroscedasticity

Data

- G groups in the population.
- Each individual n belongs to exactly one group g.
- Characterized by indicators:

$$\delta_{ng} = \begin{cases} 1 & \text{if } n \text{ belongs to g,} \\ 0 & \text{otherwise} \end{cases}$$

and $\sum_g \delta_{ng} = 1$, for all n.

Of course, they have been collected in different contexts, and you cannot assume that the epsilon that will be associated with each of these piece of data is exactly the same. So we would like to be able to model the fact that the epsilons have different variances across individuals. This is called "heteroscedasticity", so different variances. So let's set up the context. We have G groups in the population. This could be taxi drivers and non taxi drivers. Or this could be G different sources of data. And the idea is that these groups define a partition, meaning that each individual n belongs to exactly one group g. And this, belonging to groups, can be characterized using an indicator delta_ng which is equal to one if individual n belongs to group g, and zero otherwise. So the fact that each individual n belongs to exactly one group is characterized by the fact that these delta's, these indicators, when you sum over all the groups, they are equal to one for each individual n.

Notes

Summary

2m 11s

Heteroscedasticity

Assumption: variance of error terms is different across groups

Consider individual n_1 belonging to group 1, and individual n_2 belonging to group 2.

$$\begin{aligned} U_{in_1} &= V_{in_1} + \varepsilon_{in_1} \\ U_{in_2} &= V_{in_2} + \varepsilon_{in_2} \end{aligned}$$

and $\text{Var}(\varepsilon_{in_1}) \neq \text{Var}(\varepsilon_{in_2})$

Modeling

Without loss of generality:

$$\text{Var}(\varepsilon_{in_1}) = \alpha_2^2 \, \text{Var}(\varepsilon_{in_2})$$

So the assumption that we would like to model is that the variance of the error terms is different across groups. So let's take two individuals. Let's take individual called n_1 who belongs to group 1, and individual n_2 who belongs to group 2. And here I have written the utility of each of these two individuals using the different indices. And here, we have the two epsilons associated with each of these individuals. And we would like to model the fact that the variances of these epsilons are different. So this is what we call "heteroscedasticity". One way to do it, in order to derive the model, is to say: let's assume that the variance of the first epsilon is equal to the variance of the second multiplied by a positive quantity. And we call this alpha_2 square, this positive quantity. If alpha is different from one, then the two variances will be different.

Notes

Summary

3m 20s

Heteroscedasticity

Modeling: scale parameters

$$\begin{aligned} U_{in_1} &= V_{in_1} + \varepsilon_{in_1} = V_{in_1} + \varepsilon'_{in_1} \\ \alpha_2 U_{in_2} &= \alpha_2 V_{in_2} + \alpha_2 \varepsilon_{in_2} = \alpha_2 V_{in_2} + \varepsilon'_{in_2} \end{aligned}$$

Variance

$$\begin{aligned} \mathrm{Var}(\varepsilon'_{in_2}) &= \mathrm{Var}(\alpha_2 \varepsilon_{in_2}) \\ &= \alpha_2^2 \, \mathrm{Var}(\alpha_2 \varepsilon_{in_2}) \end{aligned}$$

So this alpha_2 parameter that we have introduced can be seen as the scale of the utility. And, if you remember from what we saw in the beginning of this course, when you multiply all utilities for a given individual by a positive quantity, it does not change the choice probability, because what matters is the order of these utilities. The choice probability is the probability associated with the highest utility. So here, we can do that. For individual n_2, we can multiply the utility of each alternative by alpha_2. It will not change the model. What happens if we do that? We calculate alpha_2 times the utility, which is equal to alpha_2 V plus alpha_2 epsilon. But now we can redefine a new error term that we call epsilon prime, for individual one, epsilon prime will be the same as epsilon. We don't change it. But for individual two, epsilon prime is equal to alpha times epsilon. So it means that the variance of epsilon prime for individual 2 is actually equal to the variance of alpha_2 times epsilon_2. By the property of the variance, this is equal to alpha_2 square times the variance of epsilon_2. There is no alpha_2 here.

Notes

Summary

4m 20s

Heteroscedasticity

Modeling: utility function

$$\mu_n V_{in} + \varepsilon_{in}$$

where

$$\mu_n = \sum_{g=1}^{G} \delta_{ng} \alpha_g$$

and α_g, $g = 1, \ldots, G$ are unknown parameters to be estimated from data.

Remarks

- Even if $V_{in} = \sum_j \beta_j x_{jin}$ is linear-in-parameters, $\mu_n V_{in} = \sum_j \mu_n \beta_j x_{jin}$ is not.
- Normalization: one α_g must be normalized.

And by definition, by the assumption we made in the previous slide alpha square times the variance of epsilon_2 is actually the variance of epsilon one. And the variance of epsilon one is the variance of epsilon prime one. So now, the new model, with the new epsilons here, is a model where the variance of epsilon prime one is the same as the variance of epsilon prime two. It means that it is "homoscedastic". And this is consistent now with the assumption we made for logit. We can apply logit for the new formulation. We can assume that epsilon prime one and epsilon prime two can be assumed identically distributed. And again, independent. So the impact of this analysis on the modeling is that, if you want to model heteroscedasticity in the utility functions, you need to associate a scale parameter to groups of individuals. And the scale parameter is defined as a function of the group membership. We define mu_n, the scale parameter of individual n, as equal to the sum over all the groups, of the delta_ng, which is the indicator that we defined, times alpha_g. So alpha_g will be a scale parameter associated with group g. And the alpha parameters are unknown parameters to be estimated from data.

Notes

Summary

5m 40s

Heteroscedasticity

Modeling: scale parameters

$$\begin{aligned} U_{in_1} &= V_{in_1} + \varepsilon_{in_1} = V_{in_1} + \varepsilon'_{in_1} \\ \alpha_2 U_{in_2} &= \alpha_2 V_{in_2} + \alpha_2 \varepsilon_{in_2} = \alpha_2 V_{in_2} + \varepsilon'_{in_2} \end{aligned}$$

Variance

$$\begin{aligned} \mathrm{Var}(\varepsilon'_{in_2}) &= \mathrm{Var}(\alpha_2 \varepsilon_{in_2}) \\ &= \alpha_2^2 \mathrm{Var}(\varepsilon_{in_2}) \\ &= \mathrm{Var}(\varepsilon_{in_1}) \\ &= \mathrm{Var}(\varepsilon'_{in_1}) \end{aligned}$$

ε'_{in_1} and ε'_{in_2} can be assumed i.i.d.

Of course, even if the utility is linear in the parameters, the fact that we have introduced this mu_n now, the scale parameter, well we don't have a linear in parameters specification anymore. And we have to keep in mind as well that all alpha's cannot be estimated. We have to normalize one of the alpha's to one, for the same reason that we had to normalize the scale parameter of the logit to one. In the previous case, we had only one group. And this group had to be normalized.

Notes

Summary

7m 04s