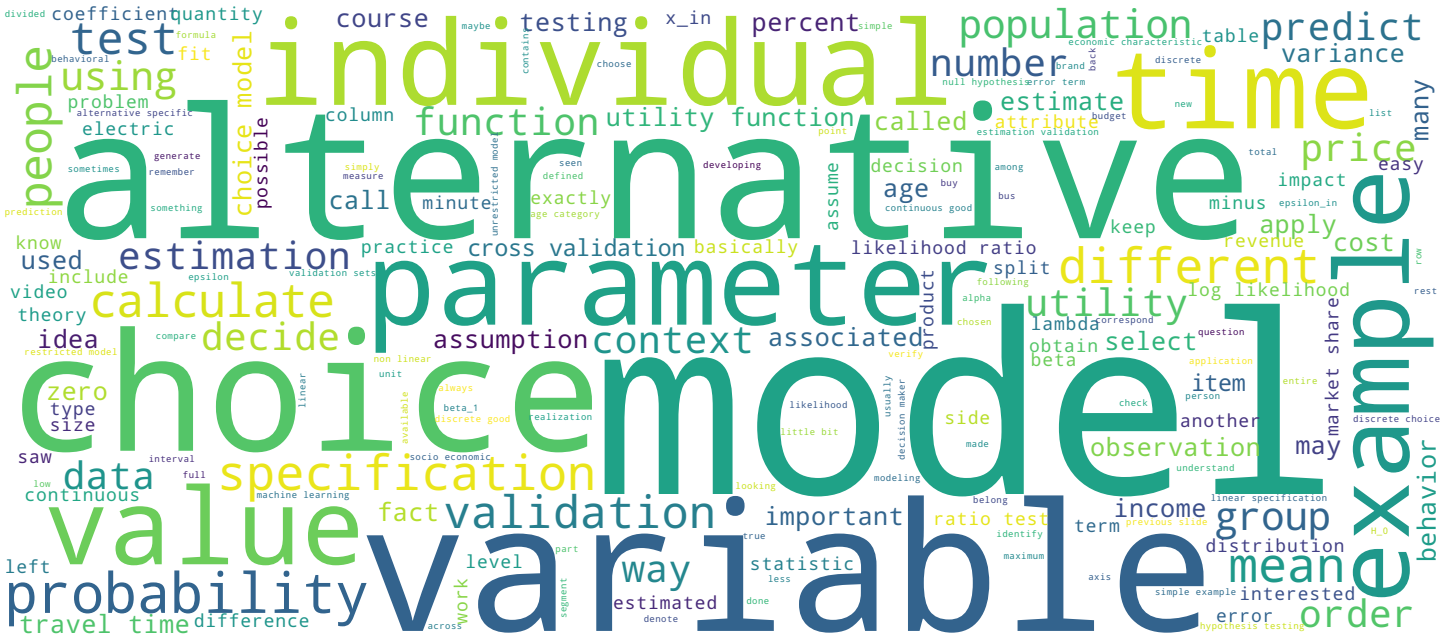
# Testing

## Prediction tests

Michel Bierlaire

Introduction to choice models

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE




Search MOOC

Video

EPFL

# Out of sample validation

In the previous tests that we have analyzed, we are using the dataset both for estimation and for testing. Now, what we would like to do is to say: let's challenge the model a little bit more, and let's hide some observations from it, and see if the model is able to predict these observations that we are hiding from it at the estimation stage. This is called "out of sample validation". And let's see how it works now.

Notes

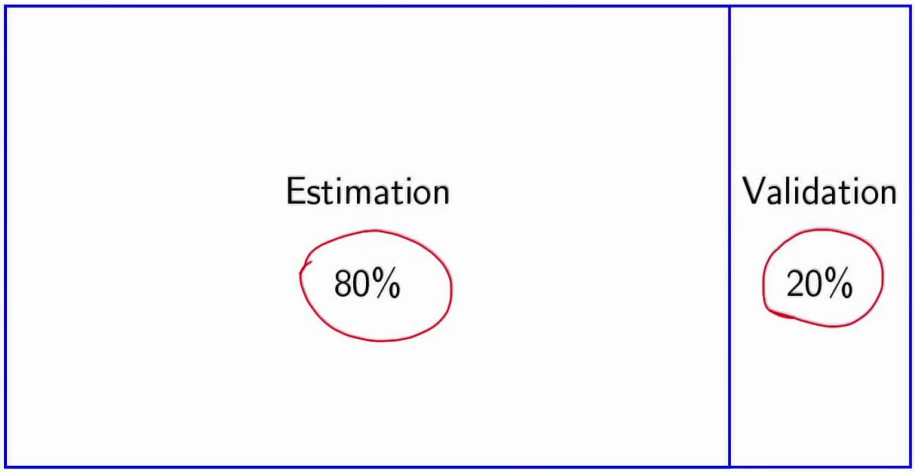
Summary

0m 04s

# Cross-validation

## Motivation

- Purpose of the model: prediction.
- Is the model able to predict?




The motivation of doing out of sample validation, which is sometimes called cross-validation as well is that, because we are developing these models for prediction, we would like to check if the model is able to predict, actually. So let's look at this rectangle as the entire sample. The way we do cross-validation is that we split this sample into two parts, let's say, for example, one part containing 80% of the data, that will be used for the estimation of the parameters of the model, and the rest, the remaining 20%, we keep them on the side, and they will be used for validation. I mean, 80%/ 20% is purely arbitrary. You decide if it is 50/50, 80/20 or 70/30. It is not really important.

Notes

Summary

0m 34s

# Methodology

## Split the sample

- Decide the size of the validation set (e.g. 20%)
- Draw randomly an estimation set and a validation set.
- Repeat $R$ times.

## Evaluate

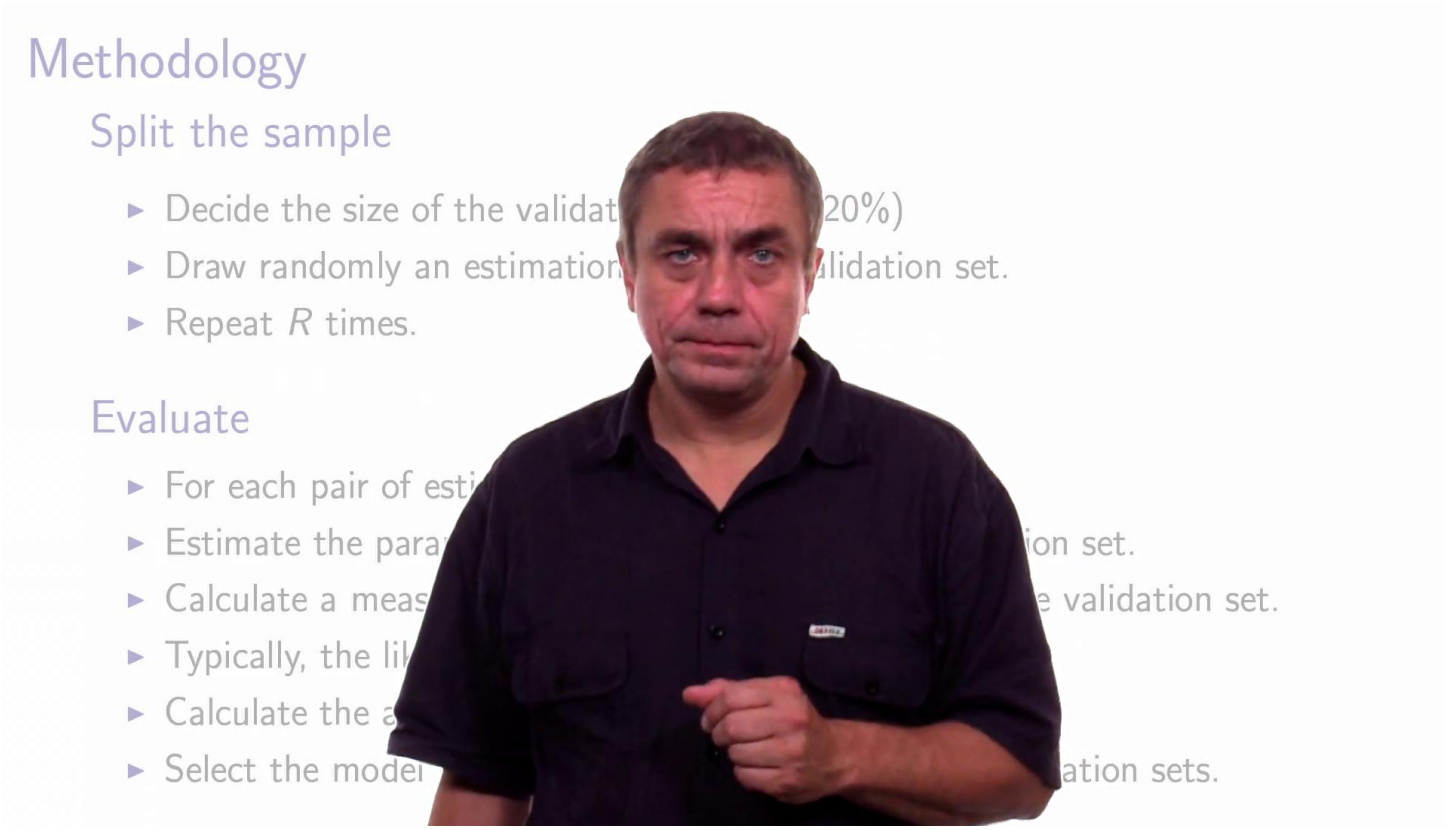
- For each pair of estimation/validation set...
- Estimate the parameters of the model with the estimation set.
- Calculate a measure of fit of the estimated model on the validation set.
- Typically, the likelihood.
- Calculate the average measure of fit.
- Select the model with the highest average fit on the validation sets.

So the methodology is the following. You split the sample, as we saw in the previous slide, we decide about the size of the validation set, and you draw randomly an estimation set and a validation set. OK, so, this splitting is done several times. Let's say R times. And then, for each realization, for each pair of estimation/validation sets, you estimate the parameters of the model with the estimation set, and then you calculate a measure of fit of the estimated model on the validation set. Well, the typical measure of fit that you can use is the likelihood, but you can calculate anything. Then, because you have done it R times, you have repeated this R times. with estimation/validation sets that have been randomly selected from the full sample, and then you calculate the average measure of fit across these R times. And then, among the models that you are testing, you select the one with the highest average fit on the validation set.

Notes

Summary

1m 20s




This method is actually inspired by the machine learning community. This is what they do a lot in machine learning. They verify that the model is actually able to predict a sample which has been left on the side, and we'll select the model which is doing the best performance onto this validation set. So it's very easy to do, and it is extremely important in order to test a model. So I really encourage you to test your model using cross-validation or out of sample validation.

Notes

Summary

2m 21s