Team Questions

November 15 – Answers Due November 22

Question 1:

Firstly, what is the physical meaning behind the alpha score? The largest alpha score in the time series data set is 1.92e-5, which corresponds to Hasbro Inc in January 2007. Does this mean that the stock is expected to rise only 0.00192% in 12 months?

Answer:

The Alpha score is a z-scored rank that has been scaled into the same units as the risk model, to treat the scores as % one has to multiply them by 10,000. We rescaled it in the data to not introduce another parameter in the objective function. This can be used without any further additions as part of the objective function.

Question 2 :

Secondly, how is a stock pair's covariance calculated? Specifically, how many past values are used in the calculation? Are these daily, weekly, monthly etc... ?

Answer:

It is a long term risk model, 12 M outlook, calculated over a 6.15 year data history per stock, we build weekly risk models and average over 4 week rolling window. Which leads to 80 4-weekly returns being used to calculate the covariance matrix.

Question 3:

Is there any more specific information you can give regarding what is a standard 8-core machine?

Answer:

A standard multi-core desktop workstation, not a cluster.

Question 4:

We recently have troubles on analyzing formula (2) that provided in the problem statement. Could you please provide us a detailed explanation of that or any reference paper that can give us a better understanding?

Answer:

Your objective function minimizes the trade-off between risk and expected return, as a risk measure tracking error has been chosen, the expected returns are given by alpha scores that have been z-scored and scaled back into risk model units. A good resource to read up would be "Robust portfolio optimization and management" by Fabozzi, Kolm, Pachamanova and Focardi

Question 5:

What does data '$$CASH0' stand for in the SEDOL index? We cannot see it in the Timeseries data. How can we deal with it?

Answer:

The Cash entry in the risk model has been removed in the most recent version, you probably haven't downloaded the latest version. As general guidance, in the unlikely case that there are

data problems in the risk model, you can fill in 0.5 for the diagonal entries and 0 for the off-diagonal values. You have to have the same dimensions in data set and covariance matrix.

Question 6:

We are considering a deep learning-based approach. Should the training time be included in the total computational time? More specifically, can we use a pre-trained neural network that was trained using the data BEFORE the target period? (We would like treat the pre-trained neural network as some kind of solving framework that can be built a priori

Answer:

If you retrain your network every time you rebalance your portfolio the time has to be part of the total time spent per iteration. You are essentially simulating consecutive rebalances, therefore you only have data available up to your present point in time. If you only train it once exclude it but please state both times we are interested to see those numbers.