

Step1: generate data set

Please generate a 1000x100 random data set with seed 2019. You can run the below script.

Step2: use data.csv (from step1) and calculate top 3 z-score pairs

Assume each column represents a unique stock's daily return series (not price). Each row is a daily return with 0 being

the earliest date and 999 being the latest.

We would like to find out the top 3 pairs trading strategy (long X, short Y) which has the highest Z-score of compounding return index. List the top 3 pairs.

Assume the positions are rebalanced in each period. i.e. for pair(X,Y) (long X, short Y), the pair return of any

period  $i$  is simply  $\text{return}(X,Y)_i = \text{return}(X)_i - \text{return}(Y)_i$

z-score is defined as  $(\text{latest value} - \text{mean}) / \text{std}$ .

Please note that the z-score of the **compounding return index**, not the daily return itself. E.g. if daily return is

0.01, 0.02, 0.03, ... then its compounding return index should be 1.01,  $1.01 \times 1.02$ ,  $1.01 \times 1.02 \times 1.03$ , ...