Pairs Trading

Data Cognition Team[1]

Abstract

Pairs trading is a trading strategy that matches a long position with a short position. The strategy profits from the difference between the two instruments.[1]

Keywords

Machine Learning — Statistical Methods

[1] *BMO Capital Markets, Toronto, Canada*

Feel free to make assumptions, but make sure to explain the reasoning behind them.

Introduction

Following a machine learning approach: Design and evaluate a pairs trading strategy for securities **ABC** and **XYZ**.

The Data

In the data folder you will find *pairs.csv*. There are four columns:

- **DATE**: Starting from Jan 4, 2016 to May 5, 2018

- **ABC**: Price of *ABC*

- **XYZ**: Price of *XYZ*

- **RATIO**: XYZ/ABC

- **DIFF**: $XYZ - ABC$

Data Exploration

Perform any cleaning, exploratory analysis, and visualization using the data provided.

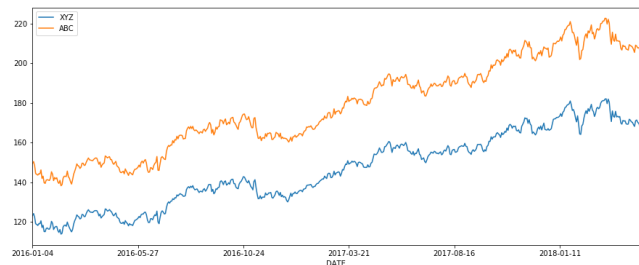Ensure to explain your approach at every step of the process.

Figure 1. Timeline.

Modeling

Build a model to execute the pairs strategy. Discuss why you chose your approach, what alternatives you considered, and any concerns you have about your model. How valid is your model? Include any indicators of model performance.

How would you quantify the risk associated with the trading? How would you determine when the enter or exit a position?

For your reference data captured in 15 minutes intervals is also provided in the data folder.

considerations

- Validate your model

- Pay special attention to the quality and organization of your code

- Feel free to use a notebook environment while doing the development, but ensure you orga-

nize your code as if it were going to production. At a minimum, explicitly state how you would organize the code.

Capital Allocation

Suppose you initial capital is $10 Million. Specify how you would utilize the capital.

Scaling

Explain how you would scale your model if instead of one observation per day, you have tens of thousands of observations per day. Also, consider the case where instead of having only two securities, you can choose from an universe of several thousand. How often would you train and replace the production model?

Market data can quickly add up. Storage considerations and data versioning are important. Assuming you have access to large cloud provider, explain what types of technologies you would use to store your data, model, and serve your model.

How do you launch, monitor, and maintain your model once it is deployed?

Deliverables

Code, plots, explanations.

Package all in one file (`.zip` or `.tar.gz`), place it in OneDrive link provided.

Questions can be send to armando.benitez@bmo.com

Thank you.

_____

_____

References

[1] Investopedia. Guide to pairs trading, 2018. [Online; accessed 04-May-2017].