Pair Trading incorporating Lead-Lag relationship

Documented by Jackson SHIH

May 14, 2023

1 Introduction

This repository contains the code for a quantitative trading project developed by the author and his collaborators. The strategy is based on a paper by Gupta, K., and Chatterjee [1] and the code is designed to be run on QuantConnect for backtesting, as it relies on packages specific to that platform. This document provides a brief synopsis of the strategy's workings.

Color codes:

- Parameters that can be tuned for optimizing/modifying the strategy are highlighted with blue
- Terminologies for which definitions are explained later in Appendix are highlighted in red

2 Strategy Overview

The aim of the section is to give a quick synopsis of the strategy. While making clear the overall logic of the strategy, we leave some of the details, such as the explanation of concept an alignment path, portfolio weighting etc. to later sections to maintain the overall clarity.

The strategy is composed of 2 parts: 1. Pair selection and 2. Making trading decisions. The major innovation of the referenced papers was the use the Dynamic-Cross-Correlation-Type (DCCT) measure in the pair selection process (Part 1), which is also the major highlight of this strategy. The trading decisions after pair selections (Part 2) are made using a pretty standard approach in pair trading: fitting regression, computing spread of stock pairs and also its z-score. Nevertheless, we would include the illustration of both parts for the sake of completeness.

2.1 Pair selection

2.1.1 Normalisation

Given a stock pair consisting of stock X and stock Y:

The stock prices time series are normalised by dividing their initial price:

$$\begin{aligned} &\{X_0, X_1, ..., X_{N-1}\}, \{Y_0, Y_1, ..., Y_{N-1}\} \\ \mapsto^{normalise} &\{x_0, x_1, ..., x_{N-1}\} = \{1, \frac{X_1}{X_0}, ..., \frac{X_{N-1}}{X_0}\}, \\ &\{y_0, y_1, ..., y_{N-1}\} = \{1, \frac{Y_1}{Y_0}, ..., \frac{Y_{N-1}}{Y_0}\} \end{aligned}$$

2.1.2 Computation of DCCT and SSD

The normalised data $\{x_t\}_{t=0}^{N-1}$, $\{y_t\}_{t=0}^{N-1}$ are used to calculate the DCCT measure:

$$DCCT_{X,Y} = \min_{\text{all alignment paths } \{(p_i,q_i)\}_{i=1}^{L}} \sum_{i=1}^{L} CR(p_i, q_i, p)$$

where p is a predetermined window size (we take $p = 25, 51, 101$ as proposed in the paper) , and (formula given in [2] written by the same authors, there were some mistakes on indexing in the original paper)

$$CR(p_i, q_i, p) := 2 \cdot (1 - \frac{\sum_{j=-p}^{p} (x_{p_i+j} - x_{p_i+j-1})(y_{p_i+j} - y_{p_i+j-1})}{\sqrt{\sum_{j=-p}^{p} (x_{p_i+j} - x_{p_i+j-1})}\sqrt{\sum_{j=-p}^{p} (y_{p_i+j} - y_{p_i+j-1})}})$$

note that p zeros are appended at both ends of x_i's and y_i to ensure that terms such as x_{-1}, x_{N+1} are well-defined and $CR(p_i, q_i, p)$ can be computed for all $i = 1, 2, ..., L$. Here L simply denotes the length of alignment paths which is not fixed but subject to the length of two time series we consider.

Remark 2.1. *In practice, it is not computationally feasible to find the above quantity by brute force (i.e. trying all alignment path and search for the maximum) since the number of alignment path between two time series grows in an exponential manner with respect to the legnth of the time series. Thus, we need to apply the Dynamic Time Warping(DTW) algorithm, and the optimal alignment path can be found when we replace the usual Eucildean metric by $CR(p_i, q_i, p)$ in the alogorithm.*

Apart from the DCCT, we shall also consider the SSD (sum of squared deviation) as proposed by the paper, the SSD measure is also computed using normalised stock prices, given by:

$$SSD_{X,Y} = \frac{1}{n}\sum_{t=1}^{n}(x_t - y_t)^2$$

2.1.3 Finding optimal pairs

The above process is repeated over all possible pairs in the stock universe of our choice. For example, if we choose the universe to be 500 stocks in S&P500, then there will be $\binom{500}{2} = 124750$ possible pairs.

Having computed DCCT and SSD values for all these pairs, the pairs with the lowest DCCT and lowest SSD will be picked.

2.2 Making trading decisions

2.2.1 Fitting moving regression

Given the current time t, where we have the historical data of (un-normalised) price data $\{X_{t-1}, X_{t-2}, ...\}$, $\{Y_{t-1}, Y_{t-2}, ...\}$ of stocks X, Y. We would do the following:
Choosing a certain lookback period Q (=10 in our implementation), we fit a regression using data from time $= t - Q$ to time $= t - 1$.

2.2.2 Normalisation:

The stock price data are normalised by dividing their price data from their respective **previous time frame**, more precisely:

$$\{X_{t-Q}, X_{t-Q+1}, ..., X_{t-1}\} \mapsto^{\text{normalise}} \{x_{t-Q}, x_{t-Q+1}, ..., x_{t-1}\} = \{1, \frac{X_{t-Q+1}}{X_{t-Q}}, ..., \frac{X_{t-1}}{X_{t-2}}\}$$

$$\{Y_{t-Q}, Y_{t-Q+1}, ..., Y_{t-1}\} \mapsto^{\text{normalise}} \{y_{t-Q}, y_{t-Q+1}, ..., y_{t-1}\} = \{1, \frac{Y_{t-Q+1}}{Y_{t-Q}}, ..., \frac{Y_{t-1}}{Y_{t-2}}\}$$

2.2.3 Regression fitting and calculation of spread

The data $\{x_i\}_{i=t-Q}^{t-1}, \{y_i\}_{i=t-Q}^{t-1}$ are used to fit a regression, giving:

$$\hat{y} = s_t x + c_t$$

note that we have put a subscript t for the slope s and intercept c to emphasize that they depend of the current time t.

The **spread** at time t, ϵ_t, is defined as the difference between actual value of y at time t and the predicted value of y at time t, given by:

$$\epsilon_t = y_t - (\underbrace{s_t x_t + c_t}_{\hat{y}_t})$$

2.2.4 Computation of z-score

Repeating the above computation for P times, we get the spreads $\{\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-P}\}$ for time $= t-1, ..., t-P$, where P is called the moving window size which is fixed (taken to be 500 in the paper). The z-score of ϵ_t is then computed as:

$$z_t = \frac{\epsilon_t - \mu_{t,P}}{\sigma_{t,P}^2}$$

$$\text{where } \mu_{t,P} := \frac{1}{P}\sum_{k=1}^{P} \epsilon_{t-k},$$

$$\sigma_{t,P} := \sqrt{\frac{1}{P}\sum_{k=1}^{P}(\epsilon_{t-k} - \mu_{t,P})^2}$$

Trading decision is then made based on z-scores, for example, the paper suggested a z-score cut-off, denoted by z_{cutoff} of 2 be used. In other words, if $z_t > 2$, then (recall that since $\epsilon_t = y_t - \hat{y}_t$ so a positive z-score means that the portfolio with 1 share of Y and s_t shares of X is overvalued) we would short the portfolio. i.e. setting the weight of stock X and stock Y to be

$$w_X = \frac{1}{\#\text{pairs}} \cdot \frac{-s_t}{1+|s_t|} \quad \text{and} \quad w_y = \frac{1}{\#\text{pairs}} \cdot \frac{1}{1+|s_t|}$$

respectively. The position is closed once we first observe $z_t \leq 0$. The opposite position is taken when we have $z_t < -2$. On the other hand, it is also possible to try with other cut-off for z-score or design some function for weight based on z-score, since the absolute value of z-score in some sense reflects the strength of the signal for divergence. One example that we have tried was to take $z_{cutoff} := 3$ and, after computing w_X and w_Y as above, we multiply them by an extra signal-strength-multiplier function, say $\frac{2(|z_t|-2)}{|z_t|}$.

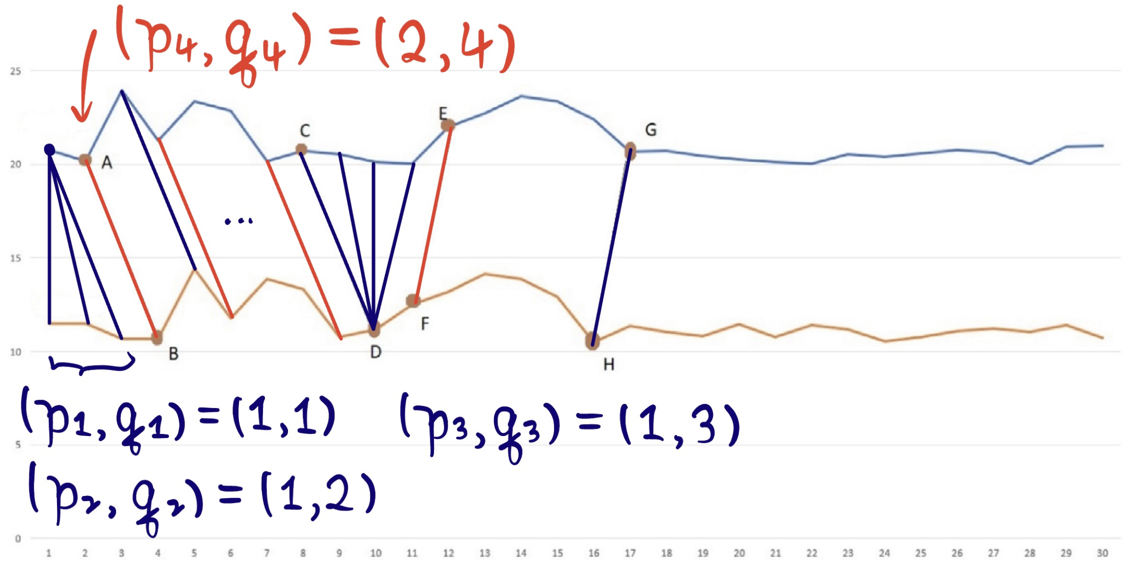
2.3 Implementation

A rolling window approach is conducted in the process of experiment. We would first determine a length for training period (say 1 month) and testing period (say half a month). Under the rolling window approach, the 1-month period would be used for computing DCCT and SSD and use that for pair selection, and the following half a month would be used to trade using the pair we have selected. The whole process is rolled over and repeated as we completed the half month trading.

For example, we may use the data in Jan1-Jan31 for pair selection, and Feb1-Feb15 for trading. After that, we re-select the pairs using the data from Jan16-Feb15 and use them for trading in Feb16-Feb28, and so on.

3 Appendix: Alignment Paths

Alignment path is a concept in time series analysis which is often used to identify the lead-lag relationship between two series. Intuitively, one might be able to sense that the following two time series ”looks alike”, in a sense that we might associate the points on the two time series by drawing ”lines” to connect them, as seen below. (the figure comes from [1])



Indeed, these lines that we draw are exactly the components an alignment path. In this case, denoting the blue time series as $\{x_1, x_2, ..., x_n\}$ and brown time series as $\{y_1, y_2, ..., y_n\}$, we see that the first ”line” we draw associates x_1 with y_1, so the first member of the alignment path is $(p_1, q_1) = (1, 1)$; the second line we draw associates x_1 with y_2, so the second member of the alignment path is $(p_2, q_2) = (1, 2)$. In general, the i-th member of the alignment path, denoted by (p_i, q_i), is equal to (j, k) if it associates the x_j with y_k.

Additionally, the paper[1] also requires some extra boundary conditions for alignment paths:

Definition 3.1. *An alignment path (of length L) is a sequence $P = \{(p_i, q_i)\}_{i=1}^{L}$ with $(p_i, q_i) \in \{1, 2, ..., n\}^2$ satisfying the following conditions:*

1. Boundary Condition: Given a parameter psi (which stands for ”Post Suffix Invariant”), (say 100), we require $p_1, q_1 \leq psi + 1$, $p_L, q_L \geq n - psi$

2. Monotonicity Condition:$p_1 \leq p_2 \leq ... \leq p_L$ and $q_1 \leq q_2 \leq ... \leq q_L$

3. Step size Condition: $(p_{i+1}, q_{i+1}) - (p_i, q_i) \in \{(1, 0), (0, 1), (1, 1)\}$ for all $i \in \{1, 2, ..., n - 1\}$

In this strategy, we need to compute the DCCT measure which is a generalised correlation-like quantity that incorporates alignment paths in indexing the terms. The computation of DCCT requires optimizing a certain quantity among all possible alignment path, implemented using the dynamic time warping algorithm, examples of illustration could be found in, for example, [3] and [4].

References

[1] Gupta, Kartikay, and Niladri Chatterjee. 2020a. "Selecting Stock Pairs for Pairs Trading While Incorporating Lead–Lag Relationship." *Physica A: Statistical Mechanics and Its Applications*, January, 124103. https://doi.org/10.1016/j.physa.2019.124103.

[2] Gupta, Kartikay, and Niladri Chatterjee. 2020b. "Examining Lead-Lag Relationships In-Depth, With Focus On FX Market As Covid-19 Crises Unfolds." April. arXiv:2004.10560.

[3] Wikipedia contributors. (2021, May 3). Dynamic time warping. In *Wikipedia, The Free Encyclopedia*. Retrieved May 13, 2023, from https://en.wikipedia.org/wiki/Dynamic_time_warping

[4] ForecastEgy. (n.d.). How to Measure Time Series Similarity in Python. Retrieved May 13, 2023, from https://forecastegy.com/posts/how-to-measure-time-series-similarity-in-python/