**Strat – Pair Trading - Cointegration**

Cointegration

Considering two series

Two series are called cointegrated if there exist s.t. is Stationary

Given , consider the following spread

where are the return of asset j

If the spread is stationary, we expect the long-term stability and mean-reversion after the divergence  
We would use this property for the pair trading

(Cointegration Related Model)

1.Vector Error Correction Model

This set of equation is equivalent to cointegration in some cases

Therefore, if , is stationary and two series are cointegrated

Vector Form (In our case : )

In general, follow

2.Common Trend Model

If two series are governed by the same trend (), two series are cointegrated

Consider where

Therefore, is stationary as are white noise

Trading Strategy Procedure

In the Formation Period:

(0)Pair Selection

Method (1)  
Distance Measure on APT  
Step 1 : PCA for One-Factor Model  
Step 2 : Distance Measure on the asset factor return

Method (2)  
Step 1 : TLS Regression to find the and   
Step 2 : Regress the difference of noise at time t over the noise at t-1 and the difference of noise at time t-1 to t-p -> finding the optimal time lag p  
Step 3 : ADF Test on the noise of the regression based on the model defined on Step3

Selecting the one with the most negative one

(1)Finding

Method (1) : The One factor Model

Model 1 :

By considering the linkage between common trend model and APT

Model 2 :

By considering the linkage between common trend model and APT

Note that would be the risk exposure return of asset i

Then we would select the model with larger coefficient (due to precision purpose)

Then we would determine the equilibrium level () by taking the period average of spread

After this step, we would obtain the equation : or

Method (2) : TLS Method

(2)Our Error is not stable in variance

Standard Regression : Uniform Weighted Average Error vs Our Case : Varying Weighted Average Error

In standard regression, the error terms are constant variance while the error in our case is varying variance. When the error has a greater variance, they are in high chance to be a relatively large value which means that it is less informative on indicating the general structure. Therefore, we would like to compute the objective function (Error) with weight inverse to variance. The problem of varying error variance may be solved by using the VWAP as Price at higher liquidity is expected to be more stable

Therefore,  
In our regression, we like to find the parameter by minimizing the following function:

Picking the referencing point to scale log-price and keeping the non-stationarity  
and we selected log price at time 0

In this case, let and

Similar to Approach one : we would like to make the regression based on two spread model and pick the one with larger coefficient

(2)Testing Stationary on the Spread (ADF Test)

The key of this strategy would be stationary spread -> we have to test the stability of -> ADF Test

In general, For the following model

ADF would test whether . However, testing would be difficult in some sense.  
We can modify the model as follow:

Then ADF would test

In general , follow AR(p)

Test Statistic =   
More Negative -> Higher Stability (At least <-2)

(\*\*\* To determine the lag order p -> we would fit regression and select the model by AIC)  
(\*\*\* For TLS, )

(3)Estimating volatility () of the spread

In the Trading Period :

(1)Normalized Spread

(2)Trading Signal

-Long A Short B if   
-Long B Short A if

(3)Exit Signal

(3.1)The normalized spread back to 0

We would conduct a research prior to the backtrader backtesting system

Research Procedure

Goal :   
(1)Determining how is the trading pair fitted the assumption of cointegration  
(2)Determining the trading threshold  
(3)Determining the average zero-crossing time spread as the trading window  
(4)Determining whether the trading signal is profitable

For Goal (1) : Determining how is the trading pair fitted the assumption of cointegration

-Measuring the Signal-to-Noise Ratio

-Target Measure

Relation between Signal to Noise Ratio and Trading Signal Profitability

For Goal (2) : Determining the trading threshold

Non-parametric Approach  
Given the realized spread sample, we may measure the empirical threshold hit rate and average trade profit. However, we are under a dilemma of the spread sample size. Short spread sample size is not enough for us to study the whole spread dynamics while the long spread sample would include the sample that may not relevant to current situation as the fundamental or the risk exposure would change over time. To adapt this difficulties, we would deploy regularization on the empirical threshold hit rate curve over threshold to adjust the curve

Given the empirical threshold hit rate curve where x is the threshold and y is the corresponding empirical hit count. Under regularization, we have the cost function:

is the cost of fitting error  
 is the cost of property error (smoothness)  
 is trade-off rate between two cost  
The cost at is obtained by minimizing the function with the input ()  
Selecting at the point of balancing two cost  
Then , we select the threshold with highest trading porfit

For Goal (3) : Determining the average zero-crossing time spread as the trading window

Step 1 : Bootstrapping on the observed spread  
Step 2 : Measuring the time spread of zero-crossing

Shorter the average time spread -> More Stability  
Taking the average time spread as the trading window

For Goal (4) : Determining whether the trading signal is profitable

-Labelling scheme (Three-Barrier Labeling)

2 horizontal barriers as TP and SL (based on the current volatility of the spread)  
1 vertical barrier (based on the Result of Goal (2))

-Target Measures

(1)TP and SL hit rate  
(2)The performance it Exit Signal if not hitting TP and SL