Pairs Trading with Machine Learning - Part I

Vince Siow, Elroy Haw, Andreas, Ryan Phua, Rishi, Shane Suxin

***Abstract*—In this paper, we would like to present a way to profit from a universe of more than 1800 stocks using a well known statistical arbitrage technique - pairs trading. This paper focuses on pairs selection and evaluation of the selected pairs performance in a portfolio or pairs. Our initial result shows above average Sharpe ratio with acceptable range of drawdown. Also, the work on this paper is a precursor to our future work which will incorporate different machine learning (ML) techniques to this pairs trading strategy.**

I. Introduction

Pairs trading is a market neutral trading strategy which involves trading assets which usually highly correlated. These correlations can stem from the assets being in the same industry, sector, market beta or even P/E ratios. Execution of this strategy involves a long position in one asset and a short position in the other, with the emphasis that both long and short positions have the same market value. The idea of pairs trading was first introduced in the mid-1980s by Morgan Stanley's quant division and was used to great success.

This paper is divided into several parts. Section II talks about the literature review and some initial work. In section III, we introduce formally the idea of pairs trading which cover pairs selection and trade execution. This is followed by our final methodology in section IV and evaluation in Section V. Section VI wraps up the paper with some discussion on the next phase of our project.

II. Literature Review & Initial Work

It is worth noting that they have been many extensions to the original pairs trading which incorporates ideas from different fields. In our literature review, we focus on the ML extensions of pairs trading. Our initial idea for ML on pairs trading root from a Stanford CS229 project paper which incorporates Kalman filter techiniques to estimate the spread - trading signal [1]. Despite the use of novel ML techniques, we are not entirely confident in the following two aspects: (1) the use of Kalman filter on a erratically-behaved time series and (2) the use of previous period spread as a feature to predict the next period's spread as a trading signal. We follow up with a later CS229 project which includes other features for ML such as technical indicators [2]. We attempted to reimplenet what [2] did but faced a problem of overfitting in our model which resulted in below par out-of-sample performance. The disappointing intial results are presented in Section 3 of this paper.

Upon getting bad initial results, we did some rethinking and understanding of the problem. We realized that most of the literature that we had seen thus far did not take into account the slight possibility of forward bias. One notable ignorance of forward bias is the estimation of the hedge ratio β. Most literature computed the value of β using data from the training period. This would introduce a more accurate than possible estimation of β which is used in the later ML model training. To mitigate this, we used a sliding window approach to avoid any chance of forward bias.

We also observed that ML cannot be blindly utilized in this case due to the high occurrence of noise in the price movement of assets in the pairs portfolio which can easily cause the problem of overfitting. Our initial model included technical features such as exponential moving average, relative strength index but the result showed huge disparity between the in-sample and out-of-sample performance. Also, the use of ML to predict trade directions might be contradictory to the theoretical basis of pairs trading. If we try to learn all the subtleties in market movements using ML, we are deviating from our initial hypothesis of price reversion between the pairs. These two reasons made us reconsider the role of ML in our pairs trading strategy. Instead of using ML for trading signal prediction, we propose using ML as a portfolio construction tool. The main idea on this will be presented in Part II of this paper.

III. Pairs Trading

A. Cointegration

We consider a spread model as follows. Let X_t and Y_t be two time series which represent the stock prices in our paper. To serve the purpose of pairs trading, X_t and Y_t must be uni-root non-stationary. Suppose they share some common source of non-stationarity.

$$X_t = a_t + \beta_X W_t$$
$$Y_t = b_t + \beta_Y W_t$$

If we manage to solve for some linear combination of X_t and Y_t, then we can get a trend-stationary pairs portfolio by

$$\beta_Y X_t - \beta_X Y_t.$$

This gives rise to the cointegration between X_t and Y_t. To get a pairs portfolio, we simply need to find some linear combination β_X and β_Y as weights for the two assets.

It is important to note that there is a difference between correlation and cointegration. Two time series can be correlated but the spread between them can still be diverging. if the expected value of the spread changes our time, then the two time series are not cointegrated and using this pair may cause us to lose a lot of money. This can be illustrated in the diagram below.

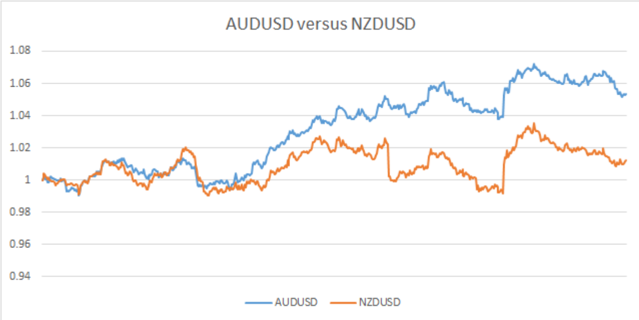

Figure 1. *Despite being correlated, currency pairs AUD/USD and NZD/USD can be seen to be cointegrated initially but diverge away from one another at the latter stages. If we take the spread as constant and execute the pairs trading on short AUD/USD and long NZD/USD, we will lose a lot of money.*

B. Engle and Granger Procedure

A prominently used technique to test for cointegration between two time series is the Engle and Grange method [3]. The algorithm can be summed up in the following steps using the aforementioned time series X_t and Y_t.

1) Test X_t and Y_t for unit-root nonstationarity $I(1)$ using the Augmented Dickey-Fuller test. If they are, proceed to the next step.
2) Run a regression of Y_t on X_t.
3) Test the residuals of the regression for unit root stationarity $I(0)$ using the Augmented Dickey-Fuller test. If they are, then X_t and Y_t are cointegrated.

The above procedure can be simply done using the built-in coint method of the statstool package in Python.

C. Trade Execution

Once we obtain a pair of cointegrated stocks X_t and Y_t, we first run an OLS regression to get the model

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

where ϵ_t is regression error which can be seen as the spread between X_t and Y_t. The next step is to normalize ϵ_t into ϵ_t. Our trading strategy is as follows: (1) long \$1 worth of X_t and short \$1 worth of Y_t when $\epsilon_t > 2$ and hold this position until $\epsilon_t < 0.5$, (2) short \$1 worth of X_t and long \$1 worth of Y_t when $\epsilon_t < -2$ and hold this position until $\epsilon_t > -0.5$. Note that we hold no position when the spread is close to zero [4].

D. Single Pair Example

In this subsection, we perform the above procedures and choose one pair which meets the cointegration test and evaluate its pairs trading performance. Firstly, to avoid any forward bias, we must split our data into a train (2012-2014) period and a test (2015-2017) period. Note that train data must always come before test data. Then we perform the cointegration test using Engle and Granger on the train data only. The selection threshold is a p-value of smaller than 5%. From this test, we chose the pair J.P. Morgan (JPM) and Morgan Stanley (MS). Intuitively, these two companies are fundamentally relatable and hence, the cointegration between their stock prices. Subsequently, we perform a regression on the train data to get the spread from the residuals which is used as a trading signal.

The train (in-sample) and test (out-of-sample) performances can be seen in the following diagrams.

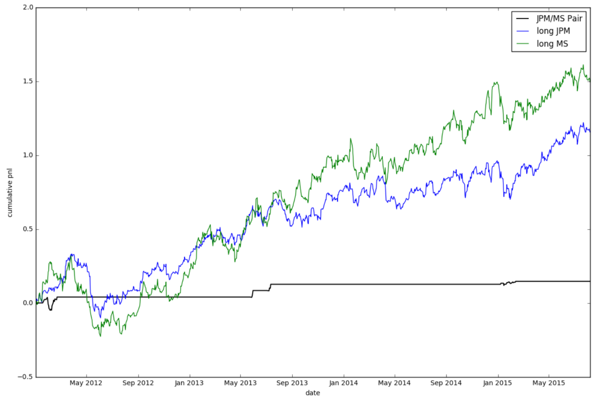

Figure 2. *Cumulative PnL during the train period of JPM/MS pair with respect to long 1 unit of JPM or MS.*

Figure 3. *Cumulative PnL during the test period of JPM/MS pair with respect to long 1 unit of JPM or MS.*

Figures 2 and 3 do not show superior performance from our pairs trading strategy in terms of absolute returns. However, if we look beyond absolute returns and consider the shape of the PnL graph, we can see that we manage to avoid getting into negative profits. A constant long position in these two stocks should be considered as "fortunate" cases because equities in general experienced a long-term bullish run during this period. A look into Figure 3 shows that our strategy manage to avoid the huge drawdown in Dec 2015.

One drawback of pairs trading is the high probability of having periods of inactivity in trade for the benefit of being less risky. To post more consistent returns from more trade, we can consider a portfolio of pairs where different pairs will have different periods of trading, allowing us to profit throughout. The final methodology will be further elaborated in the next section.

IV. Final Methodology

A. Data Preprocessing

The selection of our trading universe is paramount to the strategy. In our case, we focus on the US equity market. Our dataset contains 1824 stock tickers and the coverage starts from mid 1990s to 2018. Due to the long time frame in the dataset, there is a high chance of survivor-bias. Hence, we filter stocks which are fully traded between 2012 to 2018. Let us call this sample S_1.

It would be computationally expensive to do a pairwise cointegration test among all the stocks in S_1. Hence, we focus on stocks which have higher liquidity by filtering the universe with average daily volume greater than 1,000,000 and call this sample S_2. In S_2 we performed the Engle and Granger method outlined in Section 3 in a pairwise manner to find out potentially pairs for the trading strategy which we note as S_3.

B. Randomized Portfolio

It is easy and trivial to just include all pairs generated from the preprocessing step. However, doing this we include the possibility of non-independent pairs in the portfolio. The consequence of this can be detrimental as there is a chance that we hold long and shot position on certain assets at the same time. Also, we might have the chance of having too much position in a particular asset which would increase the risk of our portfolio.

To mitigate the above problem, we use a randomized portfolio. The construction of this portfolio is outlined using the following steps.

1) From S_3, pick one pair randomly and add into the portfolio.
2) Remove any pairs that contain any of the two stocks in the randomly chosen pair from S_3.
3) Repeat Steps 1 and 2 until S_3 is empty.

For backtesting purposes, we fix the seed of our random choice picker. The following pairs are generated from seed 7.

FOXA/JCP	BSX/HBAN	CMCSA/MRK
NVDA/WFC	AIG/GILD	AMAT/HPQ
KO/VZ	FTR/ORCL	GE/RF
C/SIRI	QCOM/XOM	JPM/MS
KEY/RAD		

To keep things simple, we assume that we whole equal weights on each pair in our portfolio. A more optimized portfolio on better weightings for each pair is possible which we will reserve for further work in Part II.

V. Results

A. Evaluation Metrics

Before we can proceed with evaluation results, we must consider the different metrics at play. In our work, we adopt WorldQuant's propriety evaluation metrics using Sharpe ratio, average drawdown, average turnover and absolute returns.

1) Sharpe ratio: We define Sharpe ratio as the information ratio scaled by the number of trading days in a year.

$$SR = \frac{mean(\text{daily returns})}{std(\text{daily returns})}\sqrt{252}$$

We take a Sharpe of above 1.5 to be good.

2) Average turnover: Daily turnover is the dollar amount of position changed or traded with respect to the total dollar value of position.

$$DT = mean\left(\frac{\text{\$ position traded}}{\text{total \$ position}}\right)$$

We take average turnover below 0.1 to be good.

3) Fitness score: A combination of the above scores

$$Fitness = SR\sqrt{\frac{|\text{absolute returns}|}{DT}}$$

We take a fitness score of above 1.5 to be good.

B. Evaluation

Using the portfolio generated by random seed 7, evaluation results are illustrated in the following diagrams.

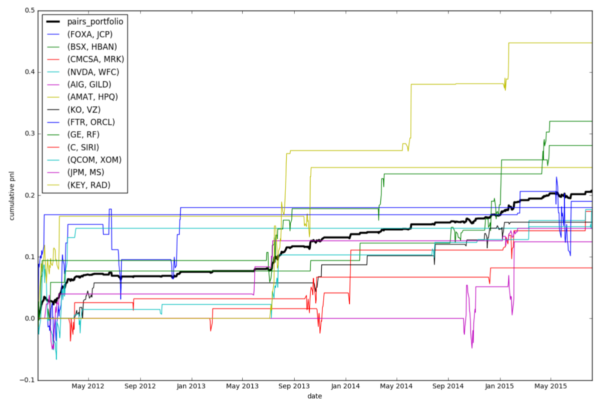

Figure 4. *Cumulative PnL during the train period of pairs with respect to individual pairs.*

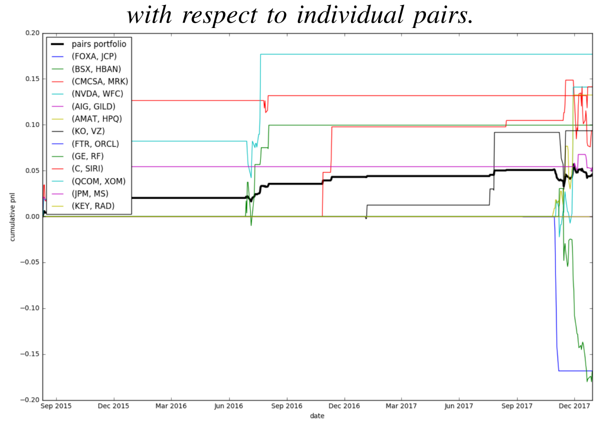

Figure 5. *Cumulative PnL during the test period of pairs with respect to individual pairs.*

The evaluation statistics are tabulated below.

	Sharpe	Turnover	Returns	Fitness
Train	4.35	0.0180	0.0539	7.53
Test	1.46	0.0121	0.0188	1.82