MACHINE LEARNING ALGORITHMS IN PAIR TRADING

Sviridenko E.V., Filipchenkov V.D., Zuyonok R.V.

Belarusian State University of Informatics and Radioelectonics
Minsk, Rupublic of Belarus

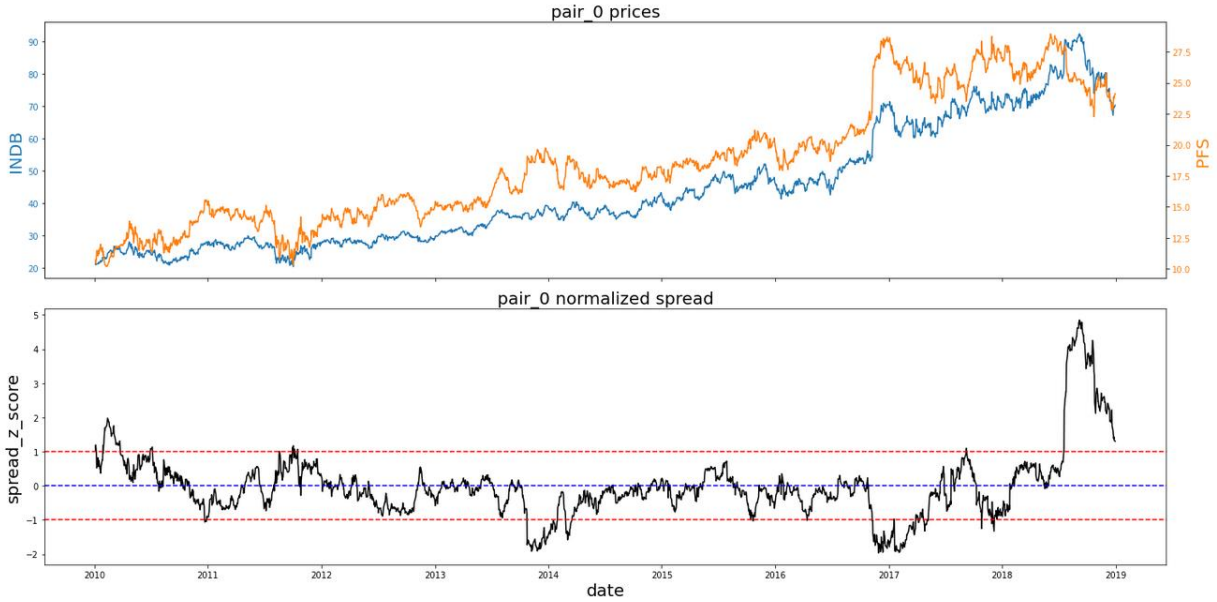
Alekhina. A.E – PhD (supervisor)

Annotation: Pairs trading is an approach of identification and construction of mean-reverting portfolio consisting of two or more assets. We have reviewed existing literature related to the usage of machine learning algorithms in statistical arbitrage trading strategy. Additionaly, we backtested trading-pairs, that were formed with clustering and dimension reduction algorithms, using 10 years (2012–2022) of S&P stock index time-series daily market data.

Keywords: Machine Learning, statistical arbitrage, OPTICS, PCA, pairs trading, clustering, dimension reduction, time-series, ADF, hiearchical agglomerative clustering, financial modeling.

Statistical arbitrage was created in Morgan Stanley in the early 1990s. The definition of statistical arbitrage brings together similar investment strategies that are based on finding mispricing between two or more assets and creating mean reverting portfolios. In the simplest form statistical arbitrage refers to trading only two assets but it can be extended to n-dimensional mean reverting portfolio. If two assets share the same characteristics and risk exposures, then we can assume that their behavior in the future would be similar as well [4]. In this case we don't have to estimate the intristic value of an asset but rather just if it is undervalued or overvalued relative to peer(s). If spread between securities diverges from its mean, we take advantage on mispricing and enter a short position on "winner" and long position on "loser".

The graph below shows the relationship between INDB and PFS. The spread bellow shows the mispricing between assets. The greater the price differens from 0 and ,hence, the spread, the greater the profit potential.



Picture 1 – Prices of INDB and PFS

Usually, statistical arbitrage answer the following questions: Firstly, given the asset universe, what are the long-short portfolios of similar assets?

Secondly, given those portfolios, what are the position entry points?

And finally, what weights should we assign for the trade? This paper focuses on answering the first question. Given the S&P asset universe e.g. 500 stocks, the simplest procedure [6] is applied to generate potential candidate pairs by considering the combination from every security to every other security in the dataset. It leads us to 124750 possible co-integration, meaning reverting tests, which can be time-consuming.

Proposed modified Sarment and Horta [2] Pipeline:

Step 1: Data collection

Daily closing prices were collected from 500 publicly traded U.S. securities and sampled for this paper. Data ranges from 2010-01-01 through 2022-01-27.

date	UVSP	CBSH	MMLP	LOGN	MTSL	CRDF	THYCY	CUB	CZNC	OPNT	...	TITN	CHMG	JLL	PARF	GPN	GURE	BCE	ROCK	NM	DVAX
2010-01-04	17.75	26.586	31.990	12.4	4.26	328.319	2.828	39.15	9.65	120.0	...	12.30	21.4	61.53	18.0	26.335	68.50	27.86	16.80	63.8	14.400
2010-01-05	16.86	26.539	32.370	12.4	4.74	323.999	2.828	41.58	9.55	120.0	...	12.72	21.4	63.04	18.0	26.280	68.25	27.65	17.76	67.0	15.600
2010-01-06	16.78	26.437	32.720	12.4	4.80	319.679	2.828	41.02	9.27	120.0	...	12.83	20.4	63.44	18.0	26.185	69.25	27.68	17.42	67.1	15.899
2010-01-07	17.12	27.060	32.950	12.4	4.80	323.999	2.828	40.26	9.03	120.0	...	12.88	20.4	63.67	18.0	26.020	67.75	26.89	17.42	67.3	15.400
2010-01-08	17.17	26.722	32.839	12.4	5.10	323.999	2.828	40.76	9.23	120.0	...	13.02	20.4	63.83	18.0	24.485	72.45	27.02	18.17	69.0	15.400

Picture 2 – Daily closing prices of 500 U.S. companies

Here is the formula for the daily return:

$$r_{t+1} = \frac{p_{t+1}}{p_t} - 1, \quad (1)$$

Where r_{i+1} is $(i+1)^{th}$daily return; and p_i is i^{th}close price of the stock.

date	UVSP	CBSH	MMLP	LOGN	MTSL	CRDF	THYCY	CUB	CZNC	OPNT	...	TITN	CHMG	JLL	PARF	GPN	GURE
2010-01-05	-0.050141	-0.001768	0.011879	0.0	0.112676	-0.013158	0.0	0.062069	-0.010363	0.0	...	0.034146	0.000000	0.024541	0.0	-0.002088	-0.003650
2010-01-06	-0.004745	-0.003843	0.010812	0.0	0.012658	-0.013333	0.0	-0.013468	-0.029319	0.0	...	0.008648	-0.046729	0.006345	0.0	-0.003615	0.014652
2010-01-07	0.020262	0.023565	0.007029	0.0	0.000000	0.013514	0.0	-0.018528	-0.025890	0.0	...	0.003897	0.000000	0.003625	0.0	-0.006301	-0.021661
2010-01-08	0.002921	-0.012491	-0.003369	0.0	0.062500	0.000000	0.0	0.012419	0.022148	0.0	...	0.010870	0.000000	0.002513	0.0	-0.058993	0.069373
2010-01-11	-0.019220	0.006324	0.009775	0.0	-0.011765	0.000000	0.0	0.030667	-0.028169	0.0	...	0.004608	0.000000	-0.015353	0.0	-0.023484	0.017253

Picture 3 – Daily returns of 500 securities

Step 2: Apply PCA

Principal component analysis is applied to the scaled return series. The below graphs plot the loadings on of each security on the first five principal components. It is unsurprising that nearly every security has a similar loading on the first principal component. This component is generally interpreted as the “market” component of financial instruments which explains much of the variation in price movements across securities. Because our sample data consist of publicly traded U.S. equity listings, we expect to see the presence of this first principal component.

First, normalize data

$$z = \frac{x-\mu}{\sigma}, \quad (2)$$

where μ=mean; σ-standard deviation.

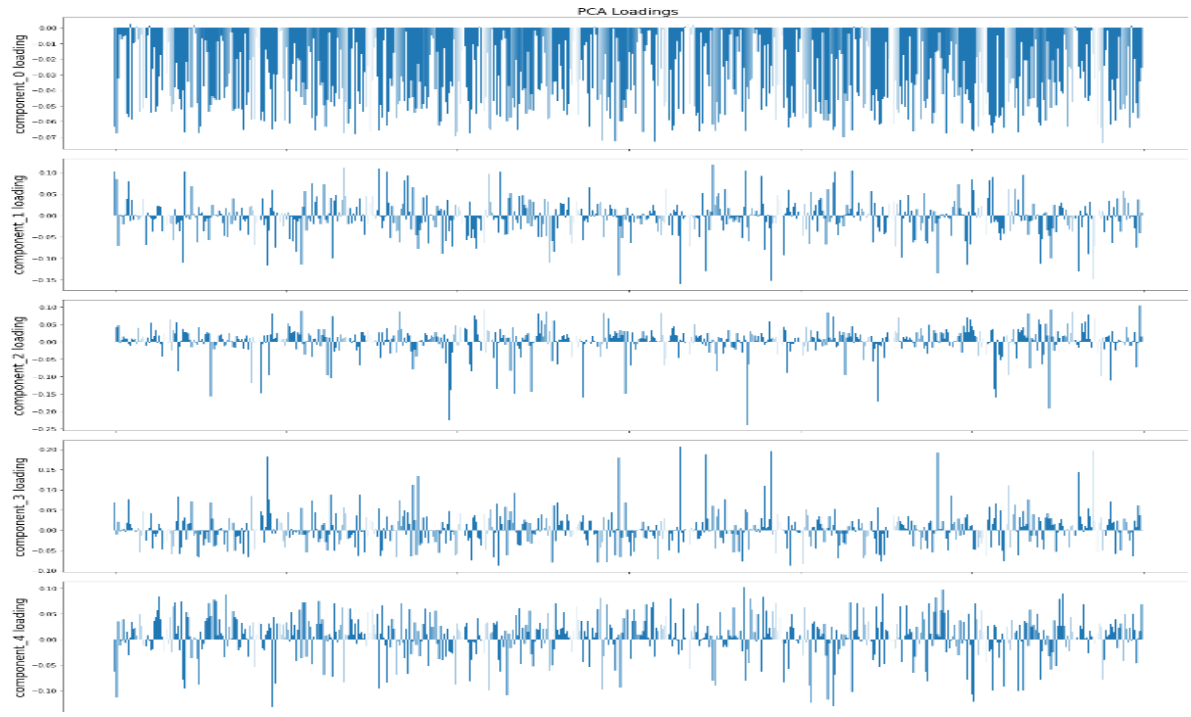
Then compute Covariance Matrix:

$$cov(X,Y) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{x})(Y_i - \bar{y}) \quad (3)$$

Compute Eigenvectors and Eigenvalues:

$$Av^{\rightarrow} = \lambda v^{\rightarrow} v^{\rightarrow}(A - \lambda) = 0, \quad (4)$$

where A is the covariance matrix; I is the identity matrix; $v^{\rightarrow}$ are eigenvectors; and λ are eigenvalues.



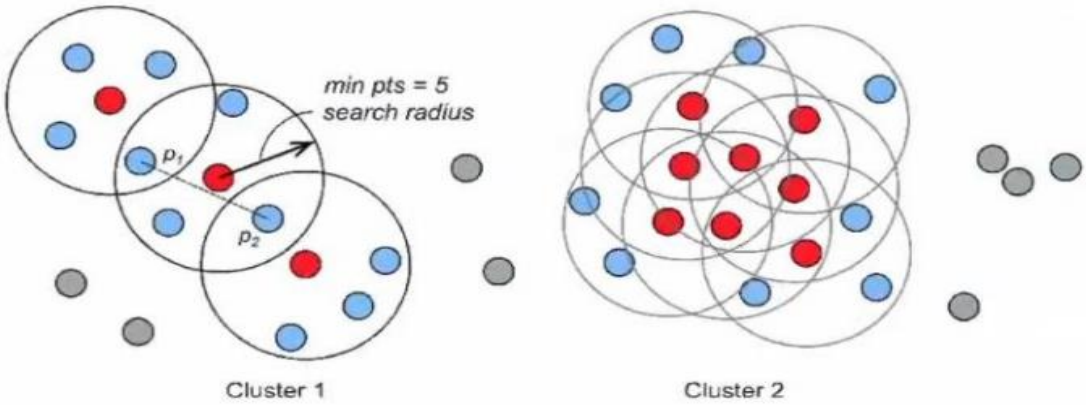
Picture 4 – Principal Component analysis Loadings

Implementing PCA reduced the dimension of the data from 3650x500 to 10x500

Step 3: Apply Clustering Algorithm (OPTICS)

Unlike other clustering algorithms, OPTICS computes an augmented cluster-ordering of the data rather than clustering the data explicitly. This frees the algorithm from relying on global parameters that might be heavily influenced by one cluster, but not accurately describe other clusters. Ankerst et. al state, “It is a versatile basis for both automatic and interactive cluster analysis” [1].

A point p is considered a core point if at least $MinPts$ are found with its $\epsilon - neighborhood$. Each point is given a core-dist which denotes the distance to the nearest $MinPts$ closest point.



Picture 5 Clustering Algorithm

Step 4: Select trading pairs

Sarment and Horta [6] suggest four criteria to filter the potential pair to increase the probability of selecting pairs of securities whose prices will continue to mean revert in the future.

Statistically significant t-stat for the Engle-Granger test p-value < 0.05 (5%).

Hurst exponent < 0.5.

Half-life between [1, 252].

Spread must cross the mean on average 12x time per year.

The Engle-Granger tests the pair for cointegration. A hurst exponent below 0.5 indicates that the pair of prices regress strongly to the mean. Pairs with extreme half-life values, below 1 or above 252, are excluded from selected pairs. Extreme half-life values indicate a price series that eitgher reverts too quickly of too slowly to be traded. Finally, the price series must cross the long-term spread mean on average 12 times a year. This enforces on average one trade per month.

Engle-Granger Test:

If $x_t \; and \; y_t$ are non-stationary and order of integration $d = 1$, then a linear combination of them must be stationary for some value of β and μ_t . [2]

$$x_t - \beta x_t = \mu_t , \quad (5)$$

where μ_t is stationary.

Cointegration:

Let $y_t = (y_{1t,} y_{2t,} y_{3t,\dots,} y_{kt})T$ is a set of time series, and each is $y_{it} \sim I(1)$. Those time-series called cointegrated if exists vector $a = (a_{1,} a_{2,\dots,} a_k)T$, such that $\varepsilon_t = a^T y_t = \sum_{(i=1)}^{k} a_i \; y_{it}$ is stationary process. a

is called cointegrated vector.

Hurst Exponent Calculation

The Hurst exponent, H, is used to measure the long-term memory of time-series. A value in the range of [0-0.5] indicates that a time series reverts strongly to the mean while a value of [0.5-1] indicates a time series with long-term positive autocorrelation and is likely to diverge. The Hurst exponent is calculates as:

$$\mathrm{E}\left[\frac{R(n)}{S(n)}\right] = Cn^H, \; n \to \infty \, R(n) \quad (6)$$

Where R(n) is the range of the first n cumulative deviations from the mean; $S(n)$ is the series of the first n standart deviations; E Is the expected value; n Is the number of observations in the time series; C Is a constant. [5]

Half-Life Calculation

The half-life of a series is the amount of time it takes for a series to return to a half of its initial value and is defined as:

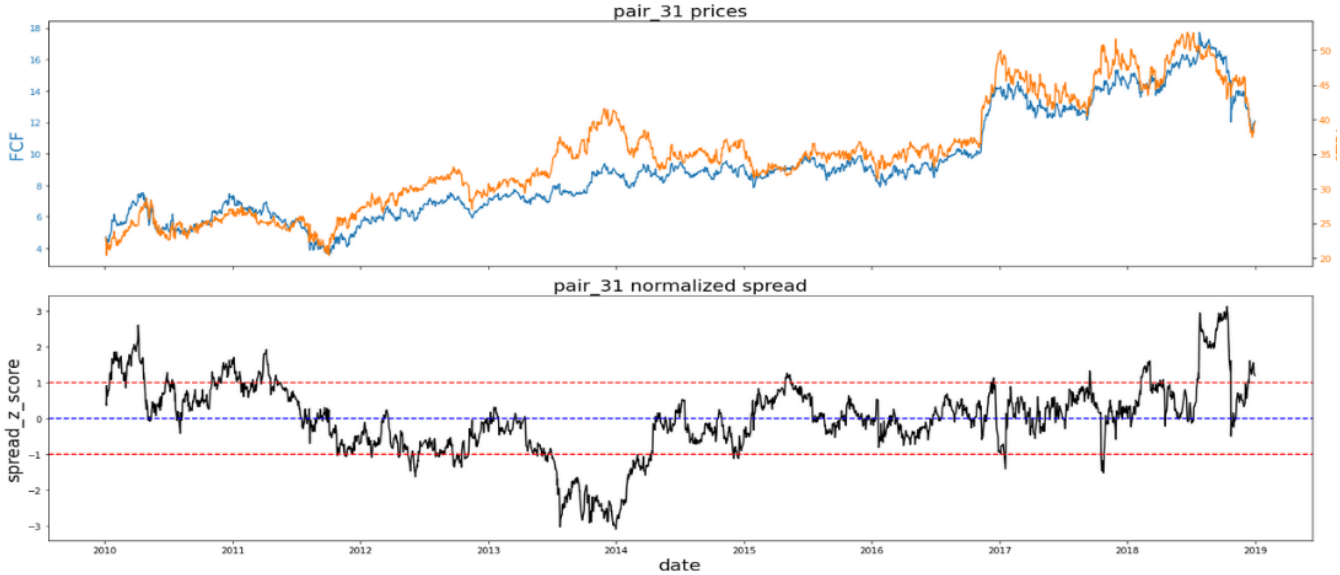
$$N(t) = N_0 e^{-\lambda t} \, N_0 \quad (7)$$

Where N0 is the initial quantity of the spread that will decay; $N(t)$ is the quantity that remains and has not yet decayed after a time t; λ is a positive number called the decay constant. [3] The half-life, $t_{\frac{1}{2}}$ is defined as: $t_{\frac{1}{2}} = \frac{\ln(2)}{\lambda}$.

Spreads with short half-lives indicate portfolios that revert to the mean and create many trading opportunities.

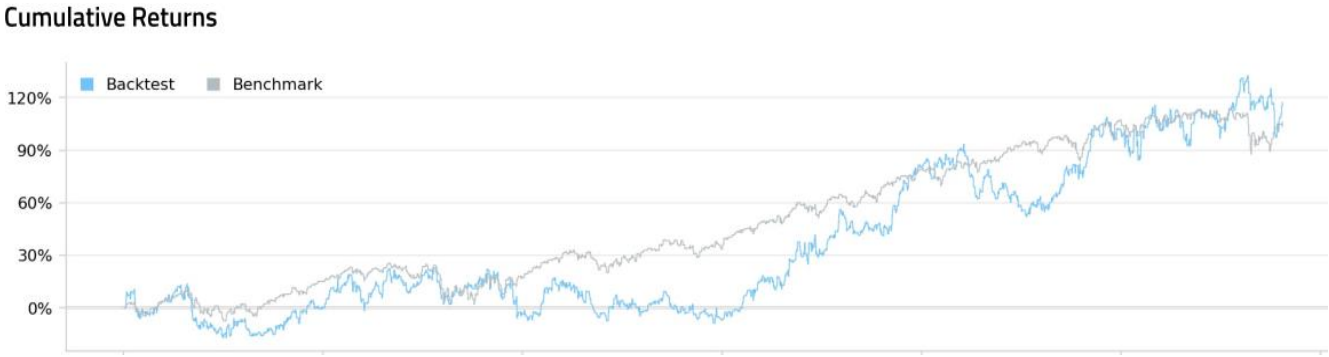
Results:

This paper demostrates an empirical example of the modified pairs selection process in [6] to efficiently reduce the search space and select quality trading pairs. Roughly 9 years of stock market price data for 500 securities were reduced to 10 dimensions through PCA algorithm. Next, over 300 potential trading pairs were identified using OPTICS clustering [1]. Seven pairs from the clusters met selection criteria.

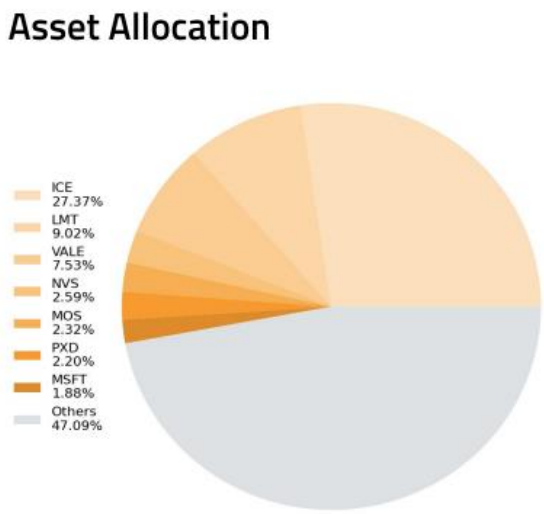


Picture 6 - Example of filtered pair

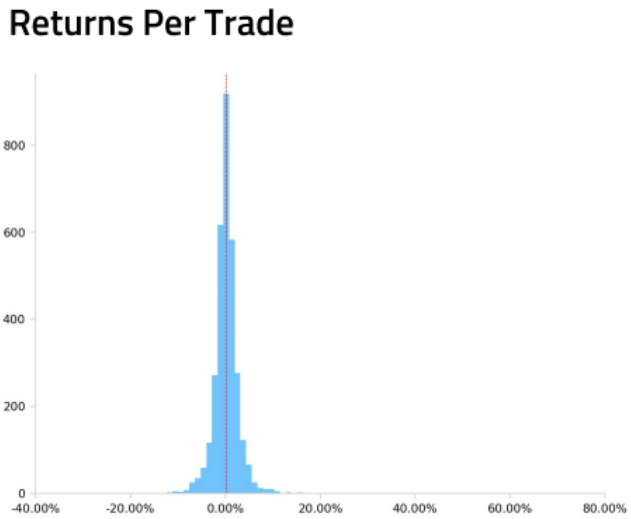
We constructed long-short portfolio for backtesting purposes. The graph below shows comparison of its returns with the Benchmark.



Picture 7 – Cumulative returns of constructed portfolio



Picture 8 – Asset allocation



Picture 9 – Returns per trade

Key Statistics

Days Live	-	Drawdown	27.0%
Turnover	49%	Probabilistic SR	8%
CAGR	14.3%	Sharpe Ratio	0.6
Markets	Equity	Information Ratio	0.1
Trades per Day	4.7	Strategy Capacity (USD)	4.3M

Picture 10 — Key statistics

References:

1. M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “Optics: ordering points to identify the clustering structure,” in ACM Sigmod record, vol. 28, no. 2. ACM, 1999, pp. 49–60.
2. Cointegration. (2004, September 17). Wikipedia. https://en.wikipedia.org/wiki/Cointegration
3. Exponential decay. (2003, September 28). Wikipedia. https://en.wikipedia.org/wiki/Exponential_decay
4. Hudson & Thames Pairs Trading Book
5. Hurst Exponent. (2006, November 9). Wikipedia. https://en.wikipedia.org/wiki/Hurst_exponent
6. Sarment and Horta 2020 A Machine Learing based Pair Trading Investment Strategy