Improving Pairs Trading Strategy with Machine Learning

Feb 2021

Presentation Outline

- Introduction
- Research Question
- Data Collection and Preprocessing
- Methodology
- Results
- Conclusions
- References

Narrative of Pairs Trading

- It follows a simple 2-step process:
  - Find two stocks whose prices have moved historically together in a formation (training) period, and
  - monitor the spread between them in a subsequent trading (test) period.

- Common approaches finding promising pairs of stocks are based on:
  - Distance metrics [1, 2, 3], and
  - Cointegration metrics [4, 5, 6, 7].

- The study focuses on the Cointegration approach to identify pairs because it is a parametric way possessing forecasting ability in terms of convergence [8].

Cointegration Approach

- It uses the 2-step Cointegration test developed by Engle and Granger (1987) [9].

- Engle-Granger test's steps:

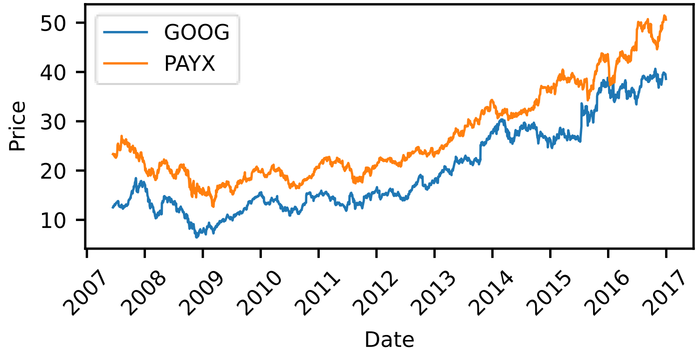
[Step 1]: Linearly combine two Non-stationary Price series (P_i, P_j) of stocks (i, j):

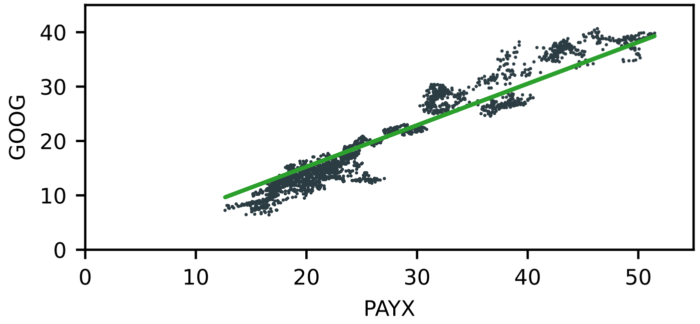
$$spread_{ij,t} = P_{i,t} - \gamma \cdot P_{j,t}$$

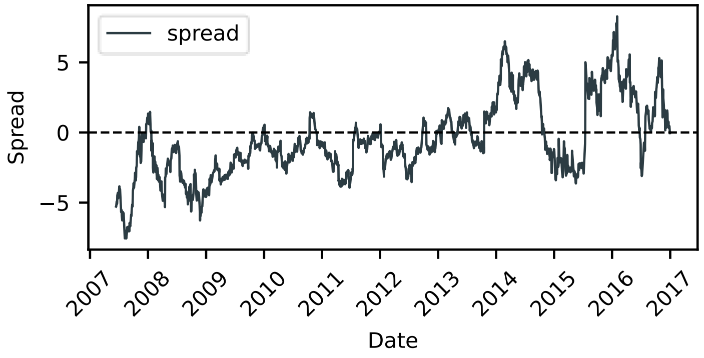
where γ should be a positive cointegration coefficient [7]. An intercept is neglected [3].

[Step 2]: If their combined time series ($spread_{ij,t}$) is Stationary according to the augmented Dickey-Fuller test [10], then the two stocks are cointegrated.

- Essentially, spread implies residuals of linear regression when fitting a line.







Machine Learning Applications in Pairs Trading

- A few studies have deployed Machine Learning techniques to Pairs Trading [19].
- Supervised learning:
  - Neural Networks / Deep Learning [11, 12, 13, 14, 15, 16, 17, 18].
- Unsupervised learning:
  - Principal Component Analysis (PCA) and Clustering [19].

Research Question

- In 1993, Fama and French published their seminal work on Asset Pricing [20]. The model suggested:

$$r_{A,t} - r_{f,t} = a + b_1 \cdot \left(r_{m,t} - r_{f,t}\right) + b_2 \cdot SMB_t + b_3 \cdot HML_t$$

where,

- $r_{f,t}$ risk-free rate,
- $r_{A,t} - r_{f,t}$ excess return of Asset *(time series)*,
- $r_{m,t} - r_{f,t}$ excess return of market *(time series)*,
- b_1 measures the level of exposure an asset has to market risk,
- SMB_t Small minus Big factor constructed to measure the Size premium *(time series)*,
- b_2 measures the level of exposure an asset has to size risk,
- HML_t High Minus Low factor constructed to measure the Value premium *(time series)*,
- b_3 measures the level of exposure an asset has to value risk, and
- α is an intercept (should be statistically insignificant).

Research Question

- Since its publication, the Fama/French 3-factor model (3FF) has been repeatedly employed by Academics and Professional Investors to explain stock returns.

- Therefore, the following **question** may arise:

  Can beta coefficients of 3FF model improve pair identifications and in turn portfolio profitability compared to Cointegration approach (baseline)?

- Following [19] study, Machine Learning techniques are going to be applied in order to answer the above question. However, instead of PCA data of Prices [19], this study makes use of beta coefficients to form clusters.

- Beta coefficients are produced by Multiple Linear Regression models [20], and

- The DBSCAN* is chosen as the clustering algorithm because:
  - it detects core samples in regions of high density [19],
  - it discovers clusters of arbitrary shapes [19, 23], and
  - domain knowledge about the number of clusters is not required [19, 23].

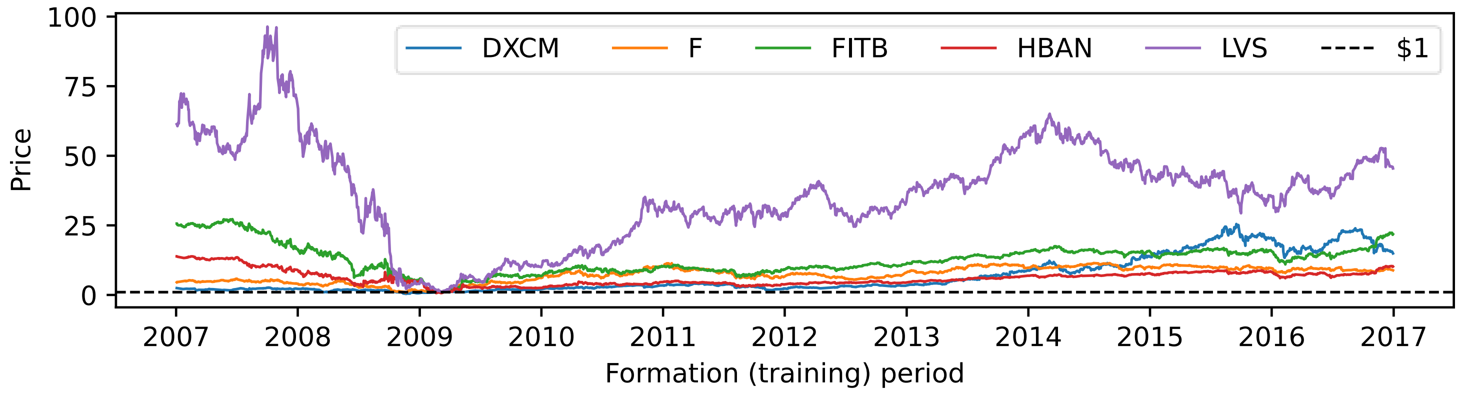
**Density-Based Spatial Clustering of Applications with Noise*

Data Collection

- Daily adjusted closing prices of stocks listed on Standard & Poor's 500 index (S&P500) were used and downloaded via Yahoo Finance open API.
- Kenneth French's website was used for downloading Fama/French 3 factors: r_m-r_f, SMB, and HML (time series).
- The daily data was from 3rd of January 2007 to 30th of December 2020.
- The dataset sample period was 3,524 days (14 years) and included 503 stocks.
- Besides the entire sample, two sub-periods were investigated:
  - The formation (training) period was between 3rd of January 2007 – 30th of December 2016 and consisted of **2,516** days (10 years).
  - The trading (test) period was between 3rd of January 2017 – 30th of December 2020 and consisted of **1,006** days (4 years).
- Both formation and trading period had been chosen arbitrarily and remained the same since the beginning of the study [1].

Data Collection

- S&P500 was chosen because it fulfilled the below criteria [5]:
  - It contained liquid stocks.
  - There were only 5 (five) stocks with prices less than $1 (one dollar) in the formation period. Though, this happened mainly at the end of bear market that United States faced with between 2007 – 2009 due to financial crisis [21]. Also, splits might cause prices to be lower than $1 after adjustments.



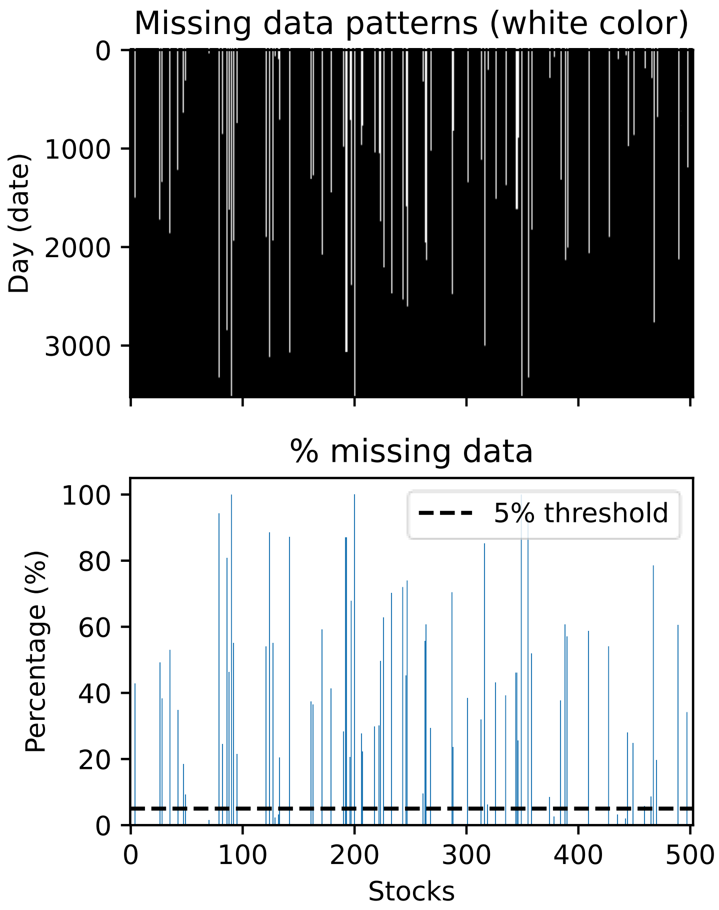
  - Stocks were traded daily with adequate trading volumes.

Data Preprocessing | Handling Missing data

- Stocks with higher than 5% [22] missing data were not considered.

- Missing data between 0% - 5%:
  - If it was located at the beginning of formation period, the corresponding dates of missing data were excluded.
  - If it was located elsewhere, last valid observations were propagated forward to next valid ones (imputation). Though, there were very few cases.

- After handling missing data, the sample period was 3,412 days and included **431** stocks. Regarding subperiods,
  - Formation (training) period, 14 Jun 2007 – 30 Dec 2016, **2,406** days.
  - Trading (test) period, 03 Jan 2017 – 30 Dec 2020, **1,006** days.



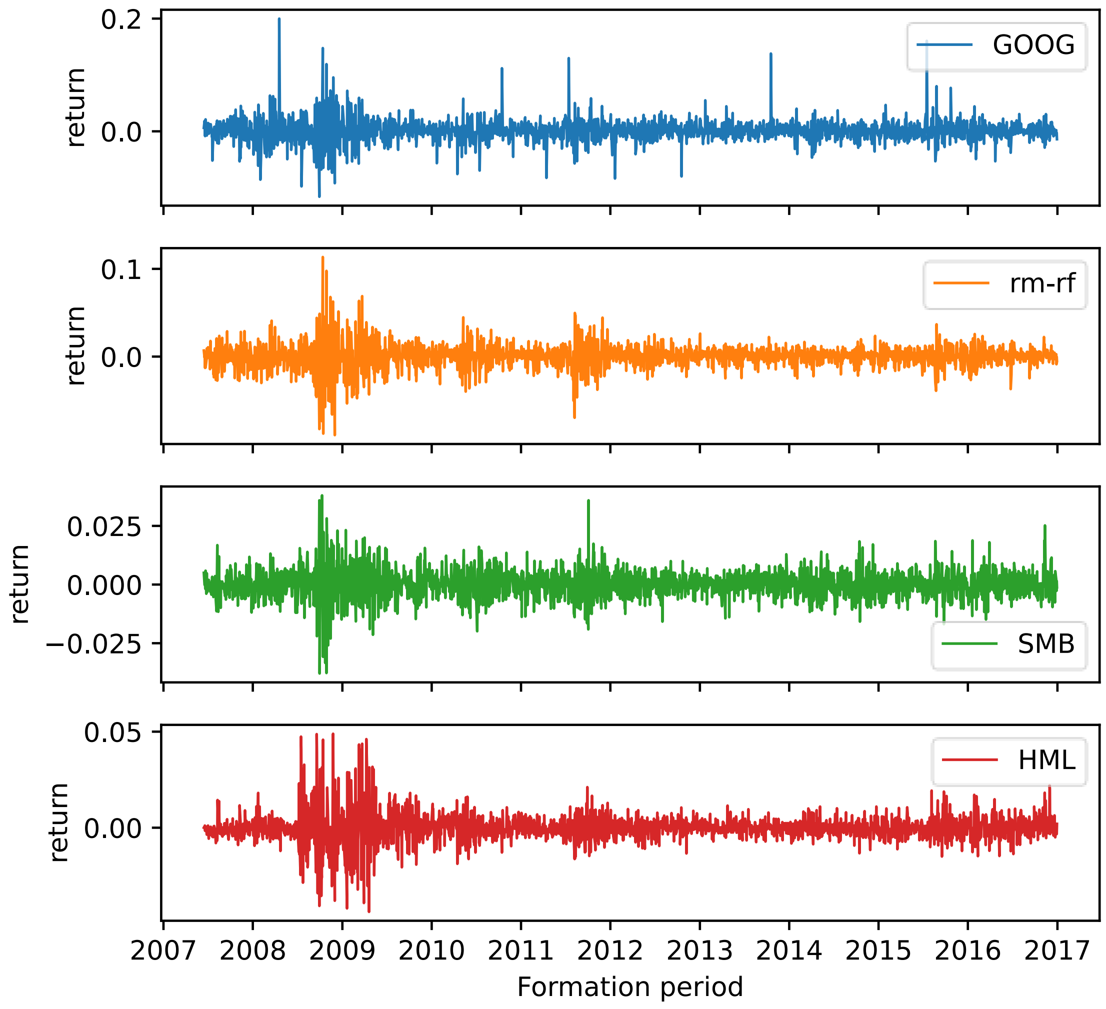
Data Preprocessing | Calculating returns

- Daily returns were calculated per asset (universe) [13, 19] in the formation period,

$$r_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t-1}}$$

where P_i was the price series of asset *i* at time *t*.

- Dates (indices) of 3FF returns were matched with Stock returns' dates.



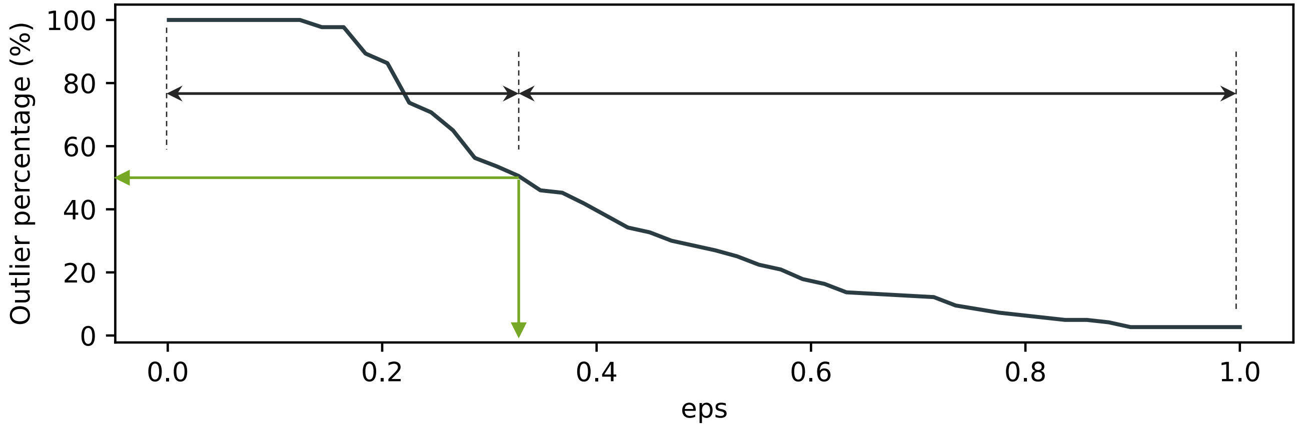
Methodology | Clustering

- Multiple Linear Regression models were applied between Stock returns and 3FF returns and statistically significant **beta** coefficients of stocks were kept. **263** stocks were found among 431.

- Beta coefficients were a 3D dataset (dimensions = 3) and used in DBSCAN clustering algorithm.

- Beta coefficients were standardised by removing the mean (m = 0) and scaling to unit variance (std = 1).

- eps = **0.33** parameter of DBSCAN was selected by elbow method. Default was 0.5.



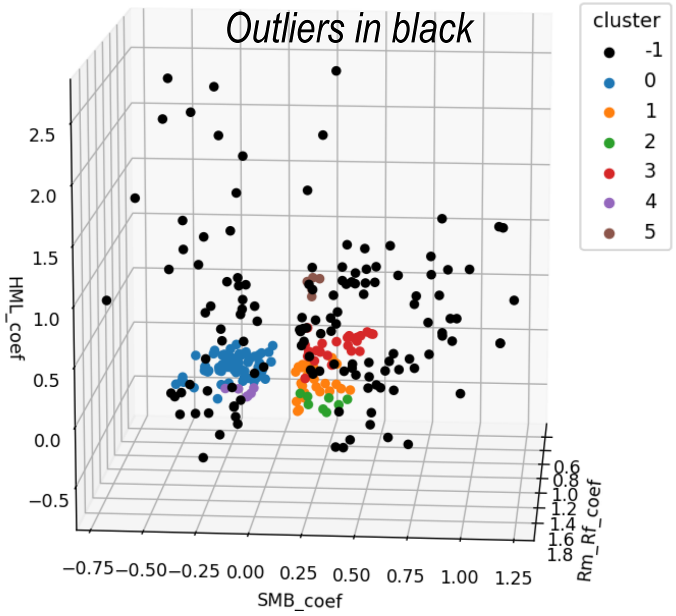
- min_samples = $2 * dimensions - 1 = 2 * 3 - 1 =$ **5** parameter of DBSCAN was selected by [24] study. Default was 5 as well.

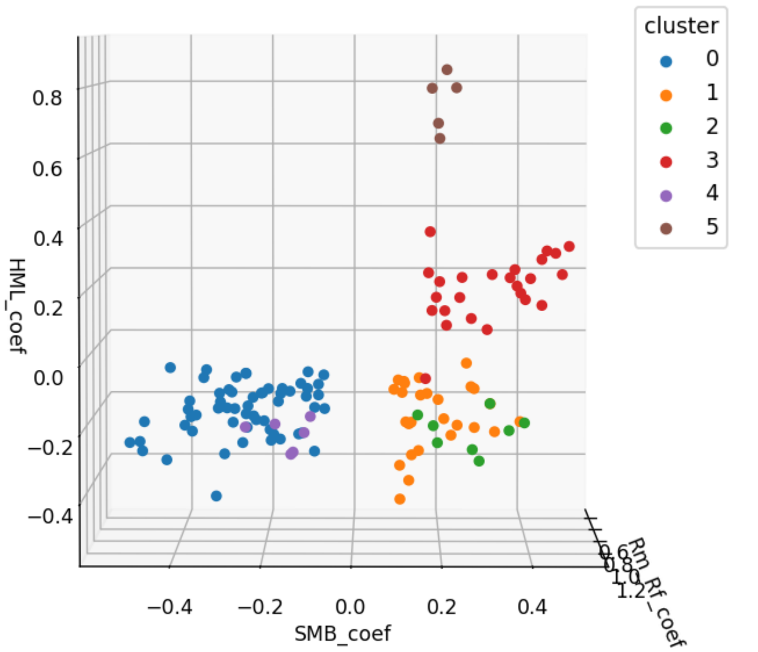
Methodology | Clustering

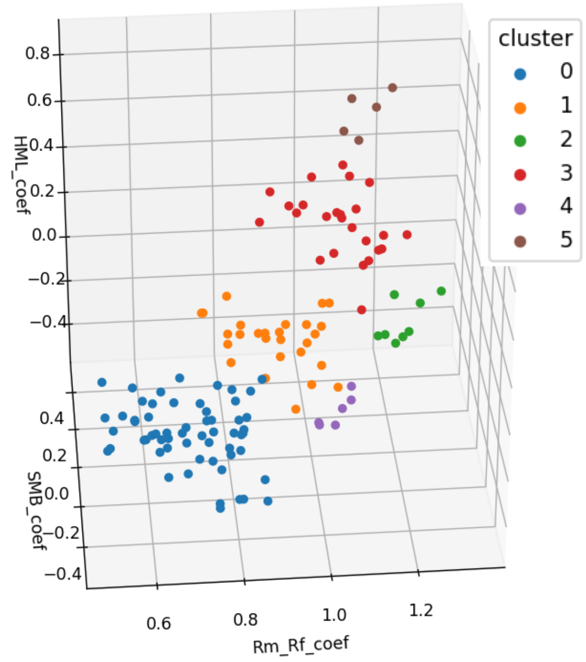
- DBSCAN was applied to scaling beta coefficients to find clusters. Each set of betas was associated with an asset.

cluster	-1 (outliers)	0	1	2	3	4	5
# stocks	133	59	27	25	8	6	5

= 263







Methodology | Cointegrated Pairs

- The Cointegration approach was considered for picking out all possible pairs from stock universe (exhaustive matching) [3, 4, 12] in the formation (training) period,

$$all\ possible\ pairs\ (combinations) = \frac{n \cdot (n-1)}{2} = \frac{431 \cdot (431-1)}{2} = 92{,}665$$

where *n* was the number of stocks (universe).

- From all possible pairs, the cointegrated ones (statistically significant) were kept and sorted by p-values (≤ 0.05). **11,414** in total (universe).

- From cointegrated pairs above, the pairs of stocks included in clusters were found:

cluster	0	1	2	3	4	5
# pairs	1515	675	231	640	152	57

Methodology | Trading Strategy

- Portfolios were formed with top *N* pairs of universe and clusters identified in the formation period [1, 3, 5].
- Various top N pairs were investigated [1, 2], ***N*** = [5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120].
- Results of portfolios of universe were considered as Baseline.
- Rolling cointegration coefficients ($\gamma_{ij,t}$) were calculated and in turn rolling spread for each pair (*i*, *j*) of baseline and clusters in the **trading (test) period**:

$$\gamma_{ij,t} = \frac{\sum_{\tau=t-W}^{t} P_{i,\tau} \cdot P_{j,\tau}}{\sum_{\tau=t-W}^{t} P_{j,\tau}^{2}} \quad without\ an\ intercept\ [25], \qquad spread_{ij,t} = P_{i,t} - \gamma_{ij,t} P_{j,t}$$

where $\gamma_{ij,t}$ was the rolling cointegration coefficient of stocks *i* and *j* at time *t*, *W* was the window size, and *P* was the price series of stocks *i* and *j*.

- Various Window sizes were investigated [26, 27], ***W*** = [50, 150, 200].

Methodology | Trading Strategy

- Rolling z-scores of spread were calculated for each pair (*i*, *j*) [3]:

$$zscore_{ij,t} = \frac{spread_{ij,t} - m_{ij,t}}{s_{ij,t}}$$

where $m_{ij,t}$ was the rolling average spread of pair (*i*, *j*) at time *t* and $s_{ij,t}$ was the rolling standard deviation.

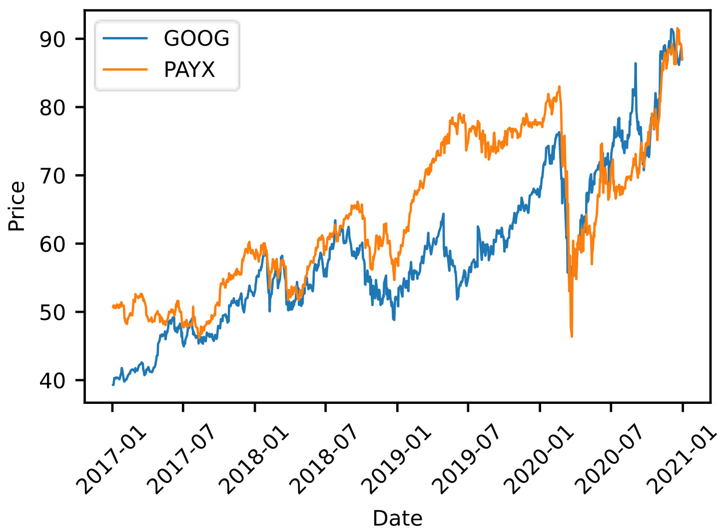
$$m_{ij,t} = \frac{1}{W} \sum_{\tau=t-W}^{t} spread_{ij,\tau}$$

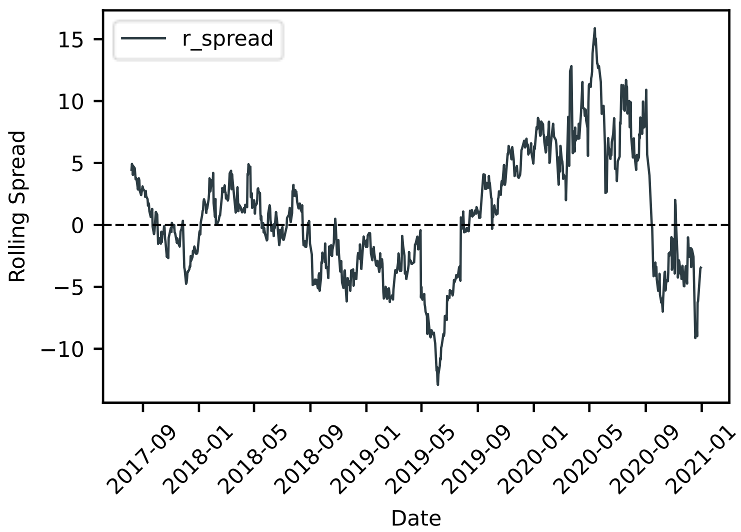
$$s_{ij,t} = \left[\frac{1}{W-1} \sum_{\tau=t-W}^{t} \left(spread_{ij,\tau} - m_{ij,t} \right)^2 \right]^{1/2}$$

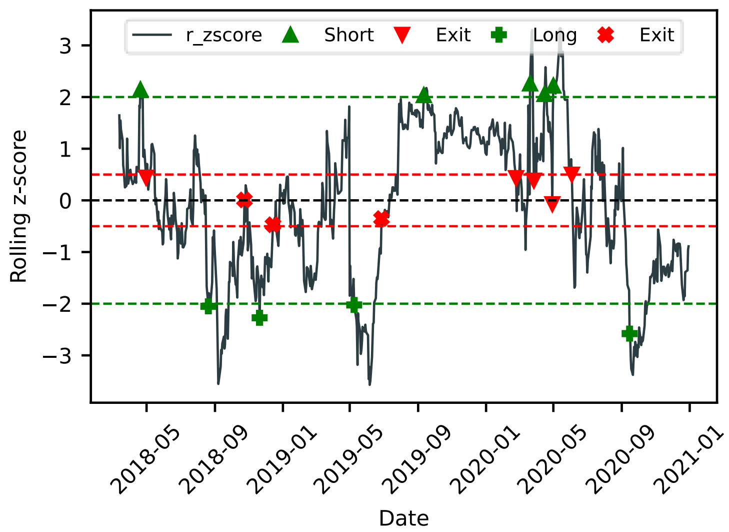
where *W* was the window size.

Methodology | Trading Strategy | Signals

- [Entry to market] Open long (buy) and short (sell) positions when z-score diverged beyond 2 [1, 3] which was the entry **threshold**. Specifically,
    - [Long (buy) z-score] If pair = (i, j) and z-score dropped below -2, then long i and short j, and
    - [Short (sell) z-score] if pair = (i, j) and z-score moved above 2, then short i and long j.
- [Exit market] Close both positions once z-score returned to a certain exit threshold. Various exit **thresholds** were examined: [0, 0.1, 0.2, 0.3, 0.4, 0.5] and [0, -0.1, -0.2, -0.3, -0.4, -0.5].







Methodology | Trading Strategy

- Rules (Hyperparameters) considered: ***N***, ***W***, and ***Thresholds***. In total, **234** combinations of Rules were examined. For instance, the rule of (20, 50, 2, 0.5, -2, -0.5) indicated 20 top pairs, 50 observations used for window, and the entry/exit thresholds accordingly.

- All trades had a one-day delay [1, 2]. It meant that when z-score generated a signal, trading was triggered the next day. In this way, returns that might be biased upwards due to bid-ask bounce were not considered [1, 28, 29, 30].

- Pair (*i*, *j*) daily returns were calculated based on [13, 14, 15]:

$$[\,Long\ (buy)\ zscore\,]\quad r_{long,ij,t} = \left(\frac{P_{i,t}}{P_{i,t-1}} - 1\right) + \left(\frac{P_{j,t-1}}{P_{j,t}} - 1\right)$$

$$[\,Short\ (sell)\ zscore\,]\quad r_{short,ij,t} = \left(\frac{P_{i,t-1}}{P_{i,t}} - 1\right) + \left(\frac{P_{j,t}}{P_{j,t-1}} - 1\right)$$

where *P* was the prices of stocks *i* and *j* at time *t* and *t-1*.

Methodology | Trading Strategy

- The performance of signals from z-scores along with their daily returns were investigated in accordance with [26] methodology.

- In-market daily returns of top pairs were kept separately and calculated their averages:

$$\bar{r}_{ij} = \frac{1}{n_m} \sum_{k}^{n_m} r_{ij,k}$$

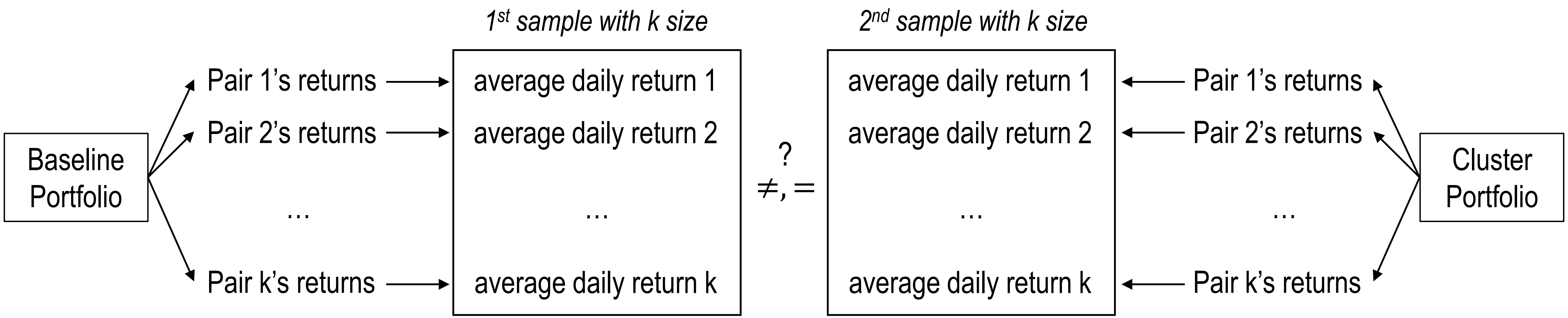
  where $\bar{r}_{ij}$ average daily return of pair of stocks *i* and *j*, and n_m number of in-market (trading) days.

- A Portfolio consisted of top N pairs:
  - A Baseline portfolio contained pairs from universe of stocks.
  - A Cluster portfolio contained pairs from a cluster defined by DBSCAN and beta coefficients.

- For instance, if N = 20, the Baseline portfolio should include 20 average daily returns ($\bar{r}$) of top 20 pairs.

Methodology | Trading Strategy

- Standardised test statistics (t-statistics) for each Rule (hyperparameter set) were calculated.
- The unpaired two-sample Welch's t-tests with unequal variances were used for examining the following hypotheses:
  - Null Hypothesis: the Baseline portfolio and the Cluster one had equal average daily returns.
  - Alternative Hypothesis: the above two portfolios had **unequal** average daily returns ($pvalue \leq 0.05$).



Results

	Count stat significant portfolios	# trading days	% trading days	# positive returns	% positive returns	mean	std	min	median	max	skew	kurtosis
Baseline	47	385	0.4308	199	0.5190	0.0011	0.0204	-0.0934	0.0007	0.1224	0.6209	8.539
Cluster 0	13	341	0.4423	175	0.5135	0.0011	0.0229	-0.1129	0.0005	0.128	0.3992	9.0085
Cluster 1	136	316	0.4802	163	0.5175	0.0014	0.0226	-0.0934	0.0007	0.1267	0.6739	6.8878
Cluster 2	81	310	0.4645	160	0.5177	0.0012	0.0225	-0.0973	0.0007	0.1185	0.4259	6.5374
Cluster 3	2	414	0.456	207	0.5015	0.0012	0.0237	-0.1260	0.0002	0.1548	0.5705	9.9915
Cluster 4	**141**	305	0.4557	159	0.5227	**0.0016**	0.0240	-0.1007	0.0009	0.1289	0.5255	6.5724
Cluster 5	15	345	0.4279	180	0.5218	0.0013	0.0216	-0.1013	0.0008	0.1376	0.8819	11.8974

- There are three Cluster portfolios that outperform the Baseline ones generating statistically higher average daily returns which in turn means higher profitability. Specifically,
  - Cluster 4 portfolios are statistically significant in 141 Rules (hyperparameter sets) out of 234 with 0.16% average daily return.
  - Cluster 1 portfolios are statistically significant in 136 Rules (hyperparameter sets) out of 234 with 0.14% average daily return.
  - Cluster 2 portfolios are statistically significant in 81 Rules (hyperparameter sets) out of 234 with 0.12% average daily return.

Conclusions

- This study adopted a **new approach** to uncover pairs based on the application of DBSCAN coupled with beta coefficients of Fama/French 3-factor model (3FF).

- The results of the empirical analysis suggested that pairs displaying similar characteristics associated with 3FF factors **outperformed** the baseline.

- Therefore, it seems that the performance of Pairs Trading can be **enhanced** with the integration of Machine Learning techniques.

- Also, the analysis showed the **profitability** of Pairs Trading Strategy seems to depend on the 3FF factors.

- The proposed methodology was capable of increasing the average daily return up to **45%**.

- As many Rules (hyperparameter sets) as possible were examined in an attempt to eliminate **data-snooping**. In addition to the Rules, the examined period remained the same since the beginning of the study.

- In the trading period, when the proposed methodology was tested, rolling measurements were calculated to avoid **look-ahead biases**. In other words, all the data used was available at the required time.

Next Steps

- Backtest various periods.
- Examine other universes of stocks.

References

1. Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. The Review of Financial Studies, 19(3), 797–827.
2. Do, B. and Faff, R. (2010) Does simple pairs trading still work? Financial Analysts Journal 66(4): 83–95.
3. Krauss, C. (2016). Statistical arbitrage pairs trading strategies: Review and outlook. Journal of Economic Surveys, 31 (2), 513–545.
4. Vidyamurthy, G. (2004) Pairs Trading: Quantitative Methods and Analysis. Hoboken, NJ: JohnWiley & Sons.
5. Rad, H., Low, R.K.Y. and Faff, R.W. (2016) The profitability of pairs trading strategies: distance, cointegration, and copula methods. Quantitative Finance 16 (10): 1541-1558.
6. Lin, Y.-X., McCrae, M. and Gulati, C. (2006) Loss protection in pairs trading through minimum profit bounds: a cointegration approach. Journal of Applied Mathematics and Decision Sciences 2006: 1–14.
7. Puspaningrum, H., Lin, Y.-X. and Gulati, C.M. (2010) Finding the optimal pre-set boundaries for pairs trading strategy based on cointegration technique. Journal of Statistical Theory and Practice 4(3): 391–419.
8. Do, B., Faff, R. and Hamza, K. (2006) A new approach to modeling and estimation for pairs trading. In Proceedings of 2006 Financial Management Association European Conference.
9. Engle, R., and C. Granger (1987). Co-integration and Error Correction: Representation, Estimation and Testing. Econometrica, 55 (2), 251–276.
10. Dickey D, . A., and Fuller W, .A . (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. Journal of the American Statistical Association, 74 (366), 427-431.
11. Huck, N. (2009) Pairs selection and outranking: an application to the S&P 100 index. European Journal of Operational Research 196(2): 819–825.
12. Huck, N. (2010) Pairs trading and outranking: the multi-step-ahead forecasting case. European Journal of Operational Research 207(3): 1702–1716.
13. Dunis, C.L., Laws, J. and Evans, B. (2006) Modelling and trading the gasoline crack spread: a non-linear story. Derivatives Use, Trading & Regulation 12(1/2): 126–145.
14. Dunis, C.L., Laws, J. and Evans, B. (2008) Trading futures spread portfolios: applications of higher order and recurrent networks. The European Journal of Finance 14(6): 503–521.
15. Dunis, C.L., Laws, J.,Middleton, P.W. and Karathanasopoulos, A. (2015) Trading and hedging the corn/ethanol crush spread using time-varying leverage and nonlinear models. The European Journal of Finance 21(4): 352–375.
16. Krauss, C. , Do, X. A. , & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259 (2), 689–702 .

References

17. Fischer T. & Krauss, C. (2017). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270 (2), 654-669.
18. Kim, T. & Kim, H.Y. (2019). Optimizing the pairs-trading strategy using deep reinforcement learning with trading and stop-loss boundaries. Complexity.
19. Sarmento M. S., & Horta N. (2020). Enhancing a Pairs Trading strategy with the application of Machine Learning. Expert Systems with Applications, 158, Article 113490.
20. Fama F. E. & French R. K. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33 (1), 3-56.
21. Beber A., & Pagano M. (2013). Short-Selling Bans Around the World: Evidence from the 2007–09 Crisis. The Journal of Finance 68 (1), 343-381.
22. Lin W. C., & Tsai C. F. (2019). Missing value imputation: a review and analysis of the literature (2006–2017). Artificial Intelligence Review, 53, 1487–1509.
23. Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining.
24. Sander J., Ester M., Kriegel, H. P., & Xu X., (1998). Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. Data Mining and Knowledge Discovery, 2, 169–194.
25. James G., Witten D., Hastie T., & Tibshirani R. (2013). An introduction to statistical learning. Springer, vol. 112.
26. Brock W., Lakonishok J. & LeBaron B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. The Journal of Finance, XLVII (5), 1731-1764.
27. Wang, J., Rostoker C., & Wagner A. (2009). A high performance pair trading application. Parallel & Distributed Processing, IEEE.
28. Jegadeesh, N. (1990). Evidence of Predictable Behavior of Security Returns. Journal of Finance, 45, 881–898.
29. Jegadeesh, N., & S. Titman (1995). Overreaction, Delayed Reaction, and Contrarian Profits. Review of Financial Studies, 8, 973–993.
30. Conrad, J., & G. Kaul (1989). Mean Reversion in Short-horizon Expected Returns. Review of Financial Studies, 2, 225–240.