# Statistical Arbitrage Pairs Trading

A presentation series by Hudson & Thames

HUDSON
AND THAMES

WE BELIEVE THAT THE SCIENTIFIC METHOD

IS THE BEST WAY TO APPROACH

INVESTMENT MANAGEMENT

https://hudsonthames.org

HUDSON
AND THAMES

# OUR MISSION

Our mission is to promote the scientific method within investment management by codifying frameworks, algorithms, and best practices to build the world's first central repository of ready to use implementations and intellectual property.

HUDSON
AND THAMES

# OUR PRODUCTS

## MlFinLab

Power of machine learning with interpretable and easy to use tools.

## PortfolioLab

Landmark implementations regarding portfolio optimization.

## ArbitrageLab

Key pairs trading statistical arbitrage algorithms.

Get a 2-month trial with your Business School email address.
Including documentation and research notebooks.

https://forms.gle/jy7mGjVksqgqT1rU6

HUDSON AND THAMES

# APPRENTICESHIP

## APPLICATIONS REOPEN IN FEBRUARY 2021

£600
Cost per month

3 Months
8 hours per day

3 Phases
Expert Knowledge, Tangible Results, Authority Positioning

100%
Online

https://hudsonthames.org/apprenticeship-program/

HUDSON AND THAMES

# PRESENTATION SERIES PLAN

- Distance Approach In Pairs Trading
- Simulating Cointegrated Pairs and Minimum Profit Optimization
- Introduction to Copula-Based Pairs Trading Strategy
- Applying Machine Learning to Statistical Arbitrage

25 Minutes per Presentation

5 Minutes Q&A after each Presentation

HUDSON
AND THAMES

# ABOUT ME

- Quantitative Research Team Lead at Hudson & Thames
- M.Sc. in Computer Science and Econometrics at the University of Warsaw

Twitter: @IllyaBarziy
GitHub: @PanPip
LinkedIn: www.linkedin.com/in/illyabarziy/

HUDSON
AND THAMES

# DISTANCE APPROACH IN PAIRS TRADING

HUDSON
AND THAMES

# PLAN OF THE PRESENTATION

01 **PAIRS TRADING FIELD**
Classification of approaches and history

02 **DISTANCE APPROACH**
Where the pairs trading started

03 **IMPLEMENTATION AND RESULTS**
How to use the tools and what results can be obtained

04 **UPSIDES AND DOWNSIDES**
What can be done to improve the results of the original method

HUDSON
AND THAMES

# PAIRS TRADING CONCEPT

## FINDING A PAIR

Two securities whose prices have moved together historically

## MONITOR THE SPREAD

Monitor the spread between prices in a subsequent period

## TRADE

If the prices diverge and the spread widens, short the winner and buy the loser

## VARIATIONS

Multiple ways to use the concept. Possible extensions to multivariate frameworks

HUDSON AND THAMES

# PAIRS TRADING APPROACHES (according to C. Krauss)

- DISTANCE
- COINTEGRATION
- TIME SERIES
- STOCHASTIC CONTROL
- OTHER
  - MACHINE LEARNING
  - COPULA
  - PRINCIPAL COMPONENT ANALYSIS

Completion Status in ArbitrageLab

(as of January 2021)

HUDSON AND THAMES

# BASELINE DISTANCE APPROACH (by Gatev et al.)

- Picking securities
- Price normalization
- Euclidean distance calculation
- Pair selection
- Position entry/exit logic
- Examples
- Upsides of the strategy
- Downsides of the strategy

HUDSON
AND THAMES

## Preparation

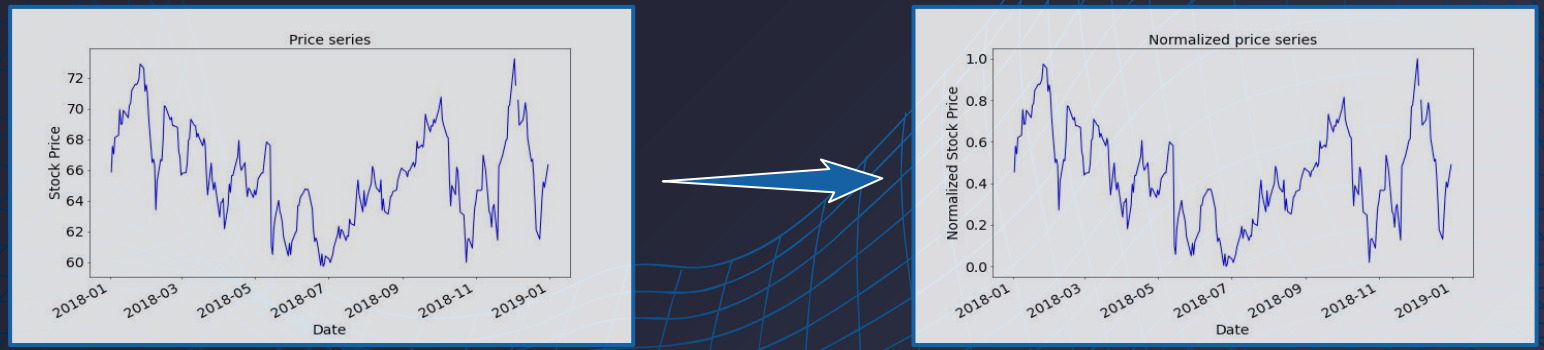
- Picking liquid U.S. stocks (period 1962-2002)

## Pairs Formation Period I

- Normalizing all stocks prices
  (or cumulative total return index)

$$P_{normalized} = \frac{P - min(P)}{max(P) - min(P)}$$

- And saving the normalization values $min(P)$ and $max(P)$




HUDSON
AND THAMES

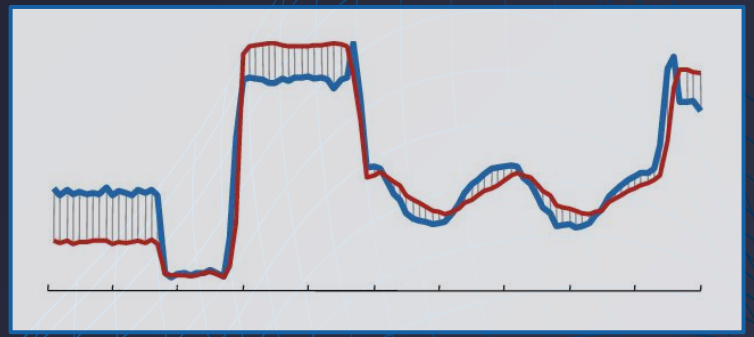
# Pairs Formation Period II

- Calculating distances between each pair of normalized prices

$$SSD = \sum_{t=1}^{N} (P_t^1 - P_t^2)^2$$

(Euclidean square distance - SSD)

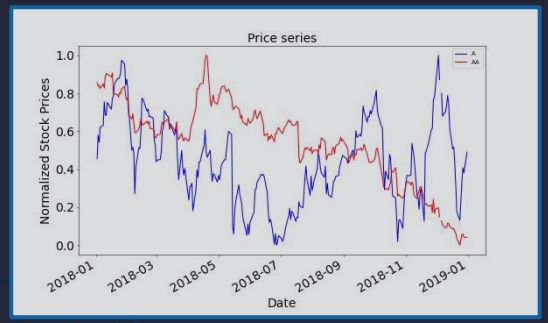
- Picking top n pairs with smallest distances



HUDSON
AND THAMES

## Pairs Formation Period III

- Calculating historical spread volatility for picked pairs





$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N - 1}}$$

## Trading Period I

- Normalizing stocks prices in the trading period using saved values

$min(P)$ and $max(P)$

- Calculating pair spreads in the trading period

HUDSON AND THAMES

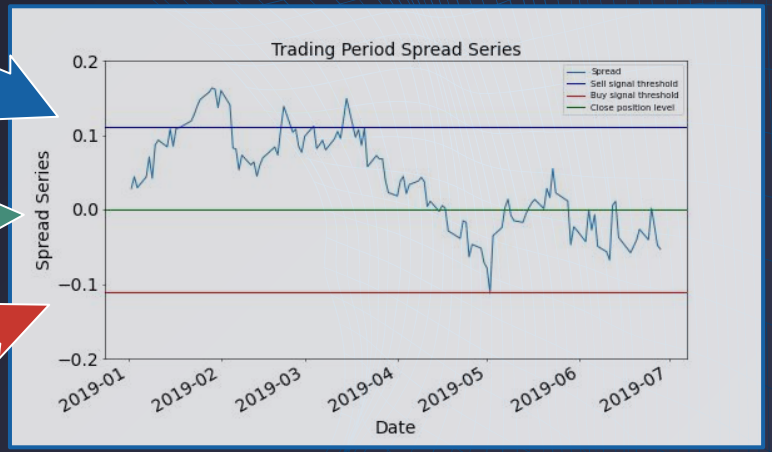
# Trading Period II

- Generating trading signals

If the spread value exceeds two historical deviations, generate a **sell** signal

**Close** open position when spread crosses the zero mark

If the spread value is below minus two historical deviations, generate a **buy** signal




HUDSON AND THAMES

## Strategy Usage

| 12 months pairs formation period | 6 months trading period |
| --- | --- |

Picking top 20 pairs of stocks

Using 2 historical deviations threshold to enter a trade

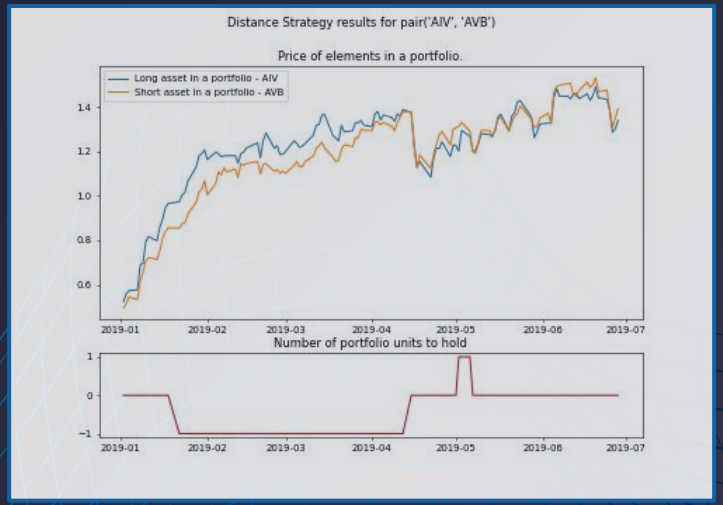
## Strategy Variations

- Pairs formation / trading period duration
- Number of chosen pairs
- Sensitivity of the signal generation
- Weights in the portfolio of pairs
- Weights of assets in each pair

HUDSON
AND THAMES

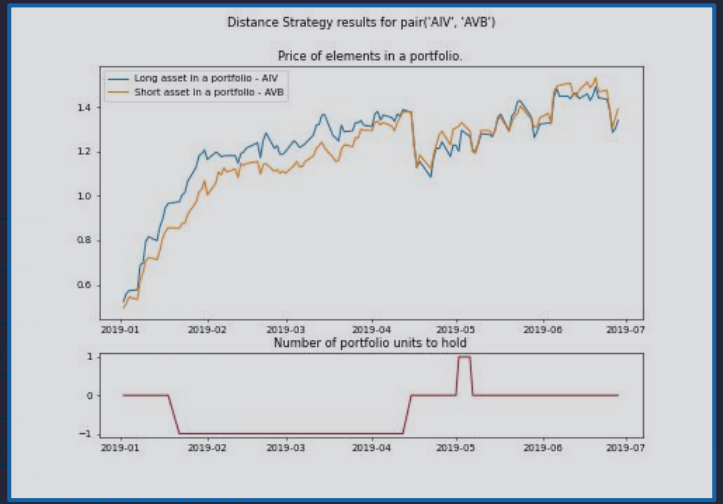
# SHOW ME THE CODE

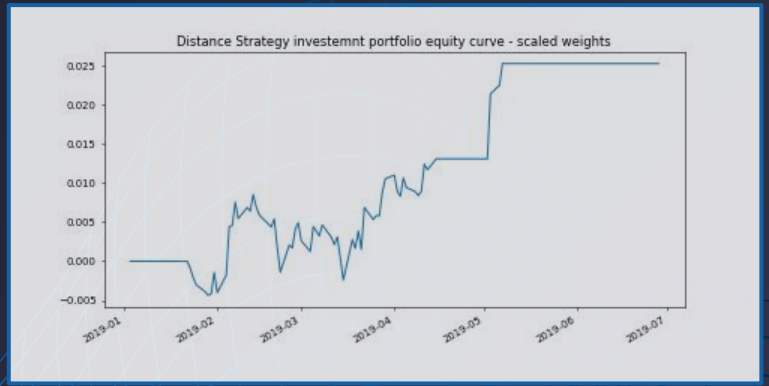
```python
# Importing packages
import pandas as pd
from arbitragelab.distance_approach.basic_distance_approach import DistanceStrategy

# Getting the dataframe with price time series for a set of assets
data = pd.read_csv('prices.csv', index_col=0, parse_dates = [0])

# Dividing the dataset into two parts - the first one for pairs formation
data_pairs_formation = data.loc[:'2019-01-01']

# And the second one for signals generation
data_signals_generation = data.loc['2019-01-01':]

# Performing the pairs formation stage of the DistanceStrategy
# Choosing pairs 5-25 from top pairs to construct portfolios
strategy = DistanceStrategy()
strategy.form_pairs(data_pairs_formation, num_top=20, skip_top=5)

# Now generating signals for formed pairs, using (2 * st. variation) as a threshold
# to enter a position
strategy.trade_pairs(data_signals_generation, divergence=2)

# Checking portfolio values for pairs and generated trading signals
portfolios = strategy.get_portfolios()
signals = strategy.get_signals()

# Plotting price series for elements in the second pair (counting from zero)
# and corresponding trading signals for the pair portfolio
figure = strategy.plot_pair(1)
```




HUDSON AND THAMES

# RESULTS




HUDSON
AND THAMES

# Upsides

- Established pairs trading concept in the literature
- Baseline strategy generated excess returns over a long period
- Method robust to data snooping
- Easy to interpret
- Nonparametric
- Economic model-free

# Downsides

- Using Euclidean squared distance for picking pairs is not optimal (see Krauss et al. 2015)
- No cointegration tests are performed, so observed correlations may be spurious
- 32 percent of all identified pairs based on the distance metric do not converge (see Do and Faff 2010)
- Worse mean reverting behavior compared to cointegration relationships (see Huck 2015)
- Worse volatility properties compared to cointegration relationships (sdd Huck 2015)
- Profitability of the method is declining on tests after 2009 (see Do and Faff 2012)

# Improvements

- Choosing different pair distance measure
- Add high historical spread variance criterion to pairs selection
- Add high historical spread mean reversion criterion to pairs selection

- Limit matching securities within industries (see Do and Faff 2012)
- Favor pairs with high number of zero-crossing in the formation period (see Do and Faff 2012)
- Adding a familywise error rate (see Cummins and Bucca 2012)

# Other Strategies

- Using Pearson correlation (see Chen et al. 2012)
- Moving to quasi-multivariate pairs trading (see Chen et al. 2012)
- Expanding the strategy to high frequency applications (see Nath 2003)

# REFERENCES


- **Krauss, C., 2017.** Statistical arbitrage pairs trading strategies: Review and outlook. Journal of Economic Surveys, 31(2), pp.513-545. [Available here]

- **Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G., 2006.** Pairs trading: Performance of a relative-value arbitrage rule. The Review of Financial Studies, 19(3), pp.797-827. [Available here]

- **Do, B. and Faff, R., 2010.** Does simple pairs trading still work?. Financial Analysts Journal, 66(4), pp.83-95.

- **Do, B. and Faff, R., 2012.** Are pairs trading profits robust to trading costs?. Journal of Financial Research, 35(2), pp.261-287. [Available here]

- **Huck, N., 2015.** Pairs trading: does volatility timing matter?. Applied economics, 47(57), pp.6239-6256.

- **Cummins, M. and Bucca, A., 2012.** Quantitative spread trading on crude oil and refined products markets. Quantitative Finance, 12(12), pp.1857-1875.

- **Chen, H., Chen, S., Chen, Z. and Li, F., 2019.** Empirical investigation of an equity pairs trading strategy. Management Science, 65(1), pp.370-389. [Available here]

- **Nath, P., 2003.** High frequency pairs trading with US treasury securities: Risks and rewards for hedge funds. Available at SSRN 565441. [Available here]


HUDSON
AND THAMES

# THANK YOU!

Does anyone have any questions?

Twitter: @IllyaBarziy
GitHub: @PanPip
LinkedIn: www.linkedin.com/in/illyabarziy/

HUDSON
AND THAMES

# CREDITS

- Presentation by Illya Barziy

- Presentation template by Slidesgo
- Icons by Flaticon
- Infographics by Freepik
- Imagen by natanaelginting - Freepik
- Author introduction slide photo created by Freepik
- Text & Image slide photo created by Freepik.com
- Big image slide photo created by Freepik.com

HUDSON
AND THAMES