HUDSON AND THAMES

# Simulating Cointegrated Pairs and Minimum Profit Optimization

Yefeng Wang

# About Me

- PhD candidate in Health Informatics at University of Minnesota
- Research focus: Natural language processing (NLP) on Twitter data
- Joined Hudson and Thames Apprenticeship Program to pursue a career path of quant
- Twitter: @stochastic_adv
- E-mail: stoch_adv@protonmail.com

HUDSON
AND THAMES

# Key Takeaways

- Understand the definition and the properties of cointegration
- Simulate cointegrated time series with ArbitrageLab
- Optimize the minimum total profit of a mean-reversion trading strategy on cointegrated asset pairs with ArbitrageLab

HUDSON AND THAMES

# What is cointegration?

And why is it important for pairs trading?

HUDSON AND THAMES

# Pairs Trading: High Correlation?

## Pairs Trade

By JAMES CHEN | Updated Sep 6, 2020

### What Is a Pairs Trade?

A pairs trade is a trading strategy that involves matching a long position with a short position in two stocks with a high correlation.

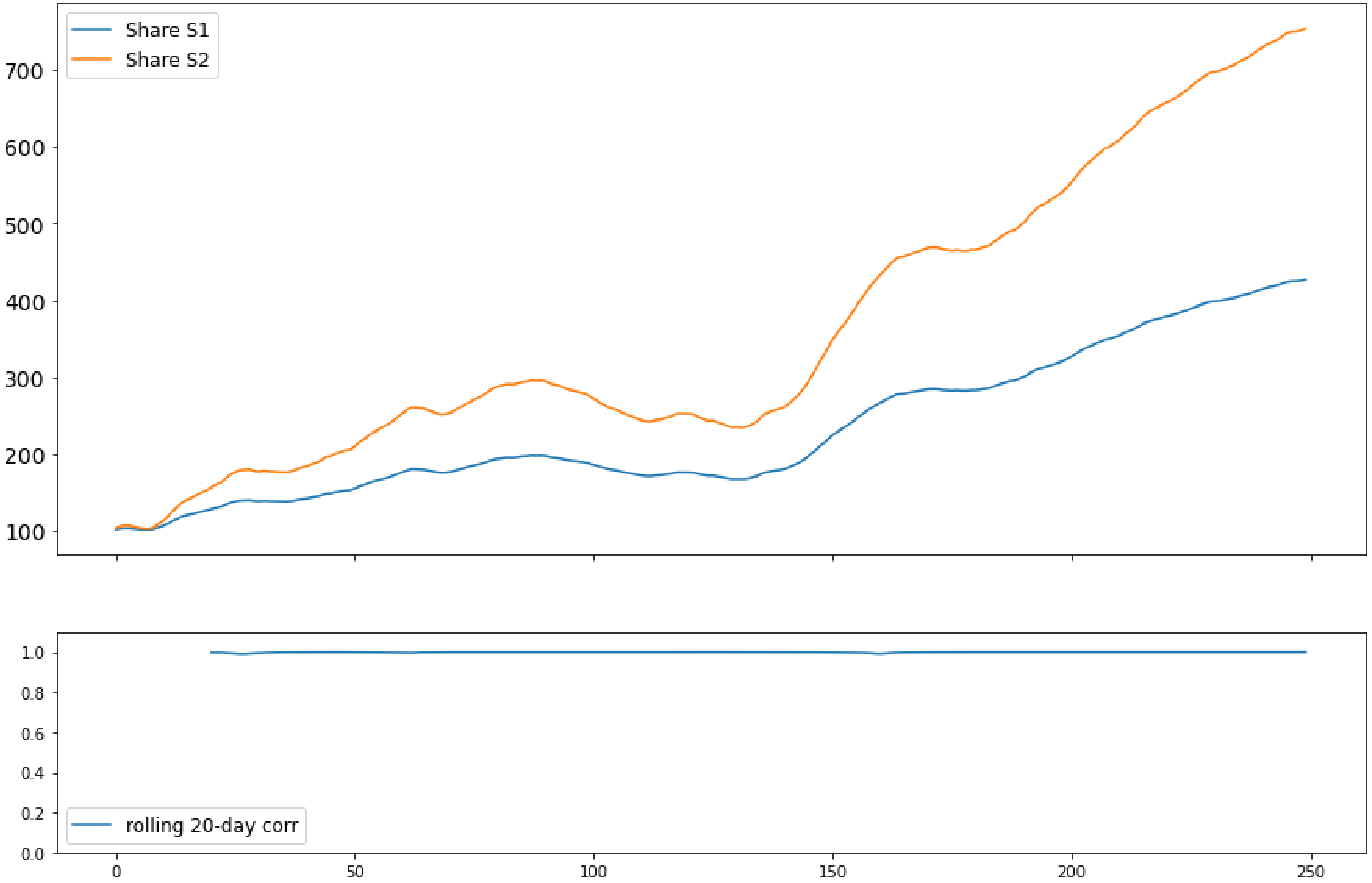
**KEY TAKEAWAYS**

- A pairs trade is a trading strategy that involves matching a long position with a short position in two stocks with a high correlation.
- Pairs trading was first introduced in the mid-1980s by a group of technical analyst researchers.
- A pairs trade strategy is based on the historical correlation of two securities; the securities in a pairs trade must have a high positive correlation, which is the primary driver behind the strategy's profits.

https://www.investopedia.com/terms/p/pairstrade.asp

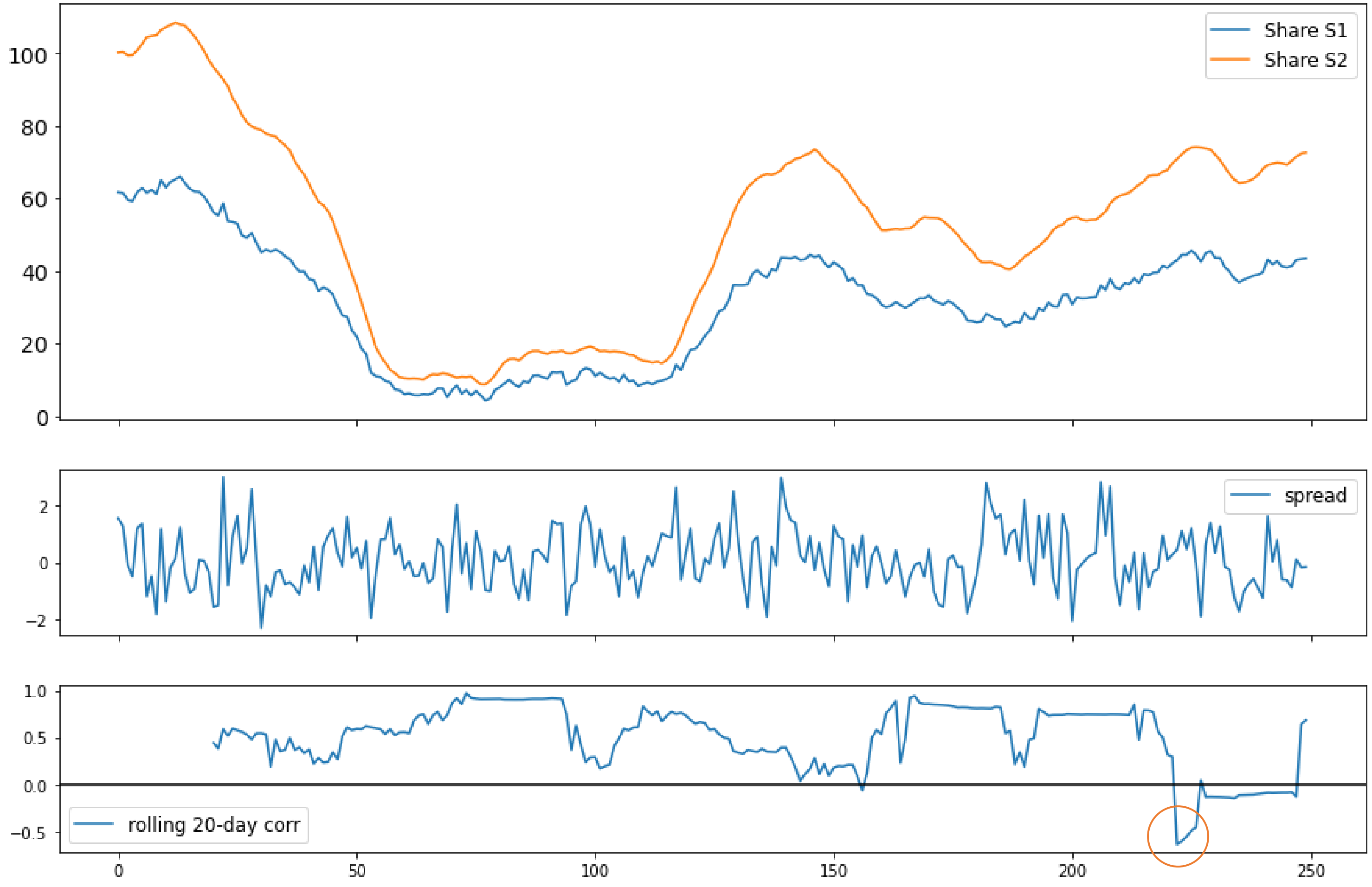
HUDSON AND THAMES

# Pairs Trading: High Correlation?




Asset pair that has perfectly correlated returns but diverging spread

# Pairs Trading: High Correlation?




Asset pair that has negative correlation but stationary spread

HUDSON AND THAMES

# Pairs Trading: Introducing Cointegration

- *Q: Correlation between what? Prices or returns?*
- A: Correlation between **returns.** Asset prices are often non-stationary and their correlation will lead to spurious conclusions.

- *Q: Is there a way to describe the relationship between asset prices?*
- A: Yes. Cointegration.

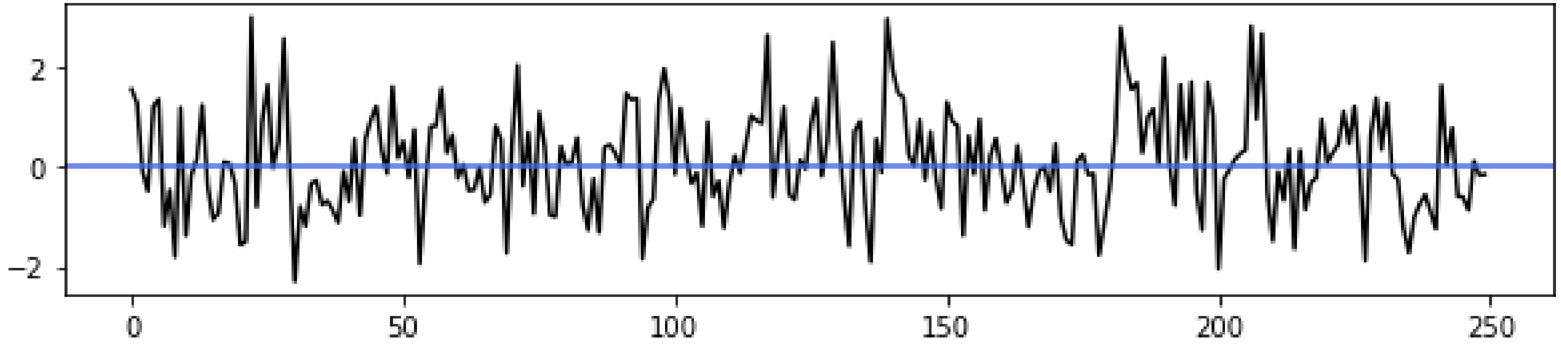
HUDSON AND THAMES

# Cointegration: $I(d)$ series

- $I(d)$ series means "**integrated series of order *d***"
- $I(1)$ series:
  - Price
  - Yield
  - Exchange rate
- $I(0)$ series:
  - Returns
- $I(0)$ series can be obtained by differencing an $I(1)$ series

# Cointegration: $I(0)$ series

- $I(0)$ series is weak-sense stationary
  - Mean function is finite and time-invariant
  - Variance function is finite and time-invariant
  - Covariance function depends only on time lag
- The time-invariant mean of $I(0)$ series is the basis of a **mean-reversion** trading strategy



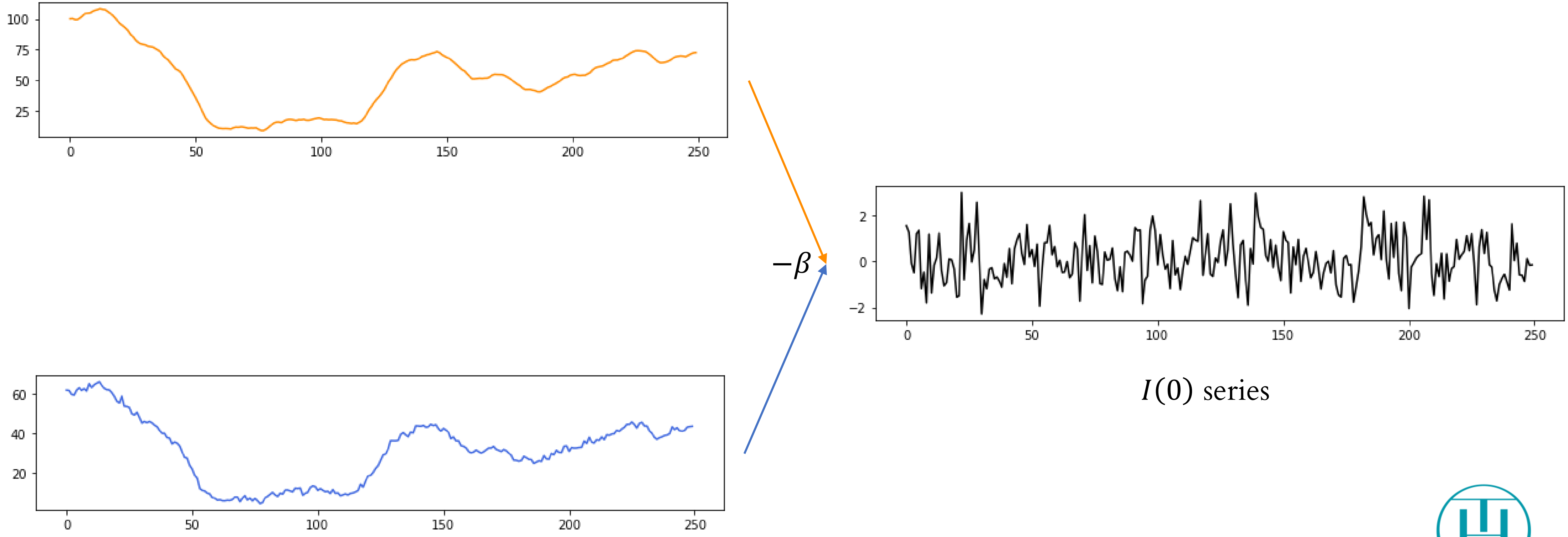

HUDSON AND THAMES

# Cointegration: Definition

**$x_t$ and $y_t$ are cointegrated, if $x_t$ and $y_t$ are $I(1)$ series and $\exists\, \beta$ such that $z_t = x_t - \beta y_t$ is an $I(0)$ series.**

# Cointegration: Definition

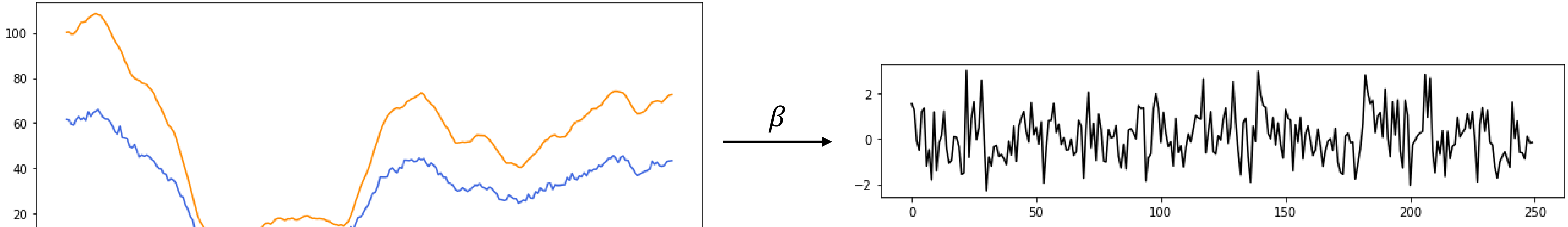
$I(1)$ series




$I(1)$ series

HUDSON
AND THAMES

# Cointegration: Definition




$I(0)$ series

# Cointegration: Properties

- The prices of cointegrated assets are tethered due to the stationarity of their spread
- Cointegration is a measure of similarity of assets in terms of risk exposure profiles
- Cointegration describes a long-term relationship between asset **prices**; correlation is a short-term relationship between **returns**
- Cointegration specifies the ratio of the long leg to the short leg, $\beta$

# Cointegration: Properties

- Cointegration specifies the ratio of the long leg to the short leg, $\beta$
- How to find $\beta$?
  - Engle-Granger test (suitable for a pair of assets)
    - Linear Regression on asset prices and test if the residual is stationary
    - Drawback: Which asset to choose as the dependent variable?
  - **Johansen test (suitable for two or more assets)**
    - Vector Error Correction Model (VECM)
    - Preferred method to test for cointegration and derive $\beta$.
  - ArbitrageLab has both test implemented

HUDSON AND THAMES

# Simulating Cointegrated Pairs

When you do not have data but still want to test the properties of cointegration

# Why simulation?

- You might not know where to find a cointegrated asset pair
  - S&P 500 has 505 constituents
  - 127,260 pairs to consider → Not feasible!
- Simulation is pervasive in trading strategy development
  - Options pricing
  - Monte Carlo backtesting
  - Cointegration is no exception

HUDSON AND THAMES

# Simulation: Stationary AR(1) process

- If a time series is stationary, then it must be an $I(0)$ series
- Simulate $I(0)$ series with stationary AR(1) process

$$a_t = \varphi a_{t-1} + \varepsilon_t, \qquad |\varphi| < 1$$

Where $\varepsilon_t \sim N(0, \sigma^2)$

- Two components are $I(0)$ series
  - Returns of one asset
  - Spread of the cointegrated pairs

HUDSON AND THAMES

# Simulation: Algorithm in ArbitrageLab

**Algorithm: Cointegrated Spread Simulation**

1. Simulate the log-price of asset Y, $y_t$, such that $y_t - y_{t-1} = a_t + c_1$, where $a_t$ is stationary AR(1).
2. Sum up $a_t$ and obtain the log-price of asset Y.
3. Simulate the cointegrated spread between asset X and asset Y, such that $x_t + \beta y_t = b_t + c_2$, where $b_t$ is stationary AR(1).
4. Retrieve the log-price of asset X, $x_t$, by calculating $x_t = b_t - \beta y_t$.

HUDSON AND THAMES

# Simulation: Algorithm in ArbitrageLab

**Algorithm: Cointegrated Spread Simulation**

1. Simulate the log-price of asset Y, $y_t$, such that $y_t - y_{t-1} = a_t + c_1$, where $a_t$ is stationary AR(1).
2. Sum up $a_t$ and obtain the log-price of asset Y.
3. Simulate the cointegrated spread between asset X and asset Y, such that $x_t + \beta y_t = b_t + c_2$, where $b_t$ is stationary AR(1).
4. Retrieve the log-price of asset X, $x_t$, by calculating $x_t = b_t - \beta y_t$.

The choice of using log-price is preferable, for its difference is log-return. However, according to Alexander *et al*., it is acceptable to use raw price in this simulation.

# Simulation: Results

```
coint_simulator = CointegrationSimulation(20, 250)
```

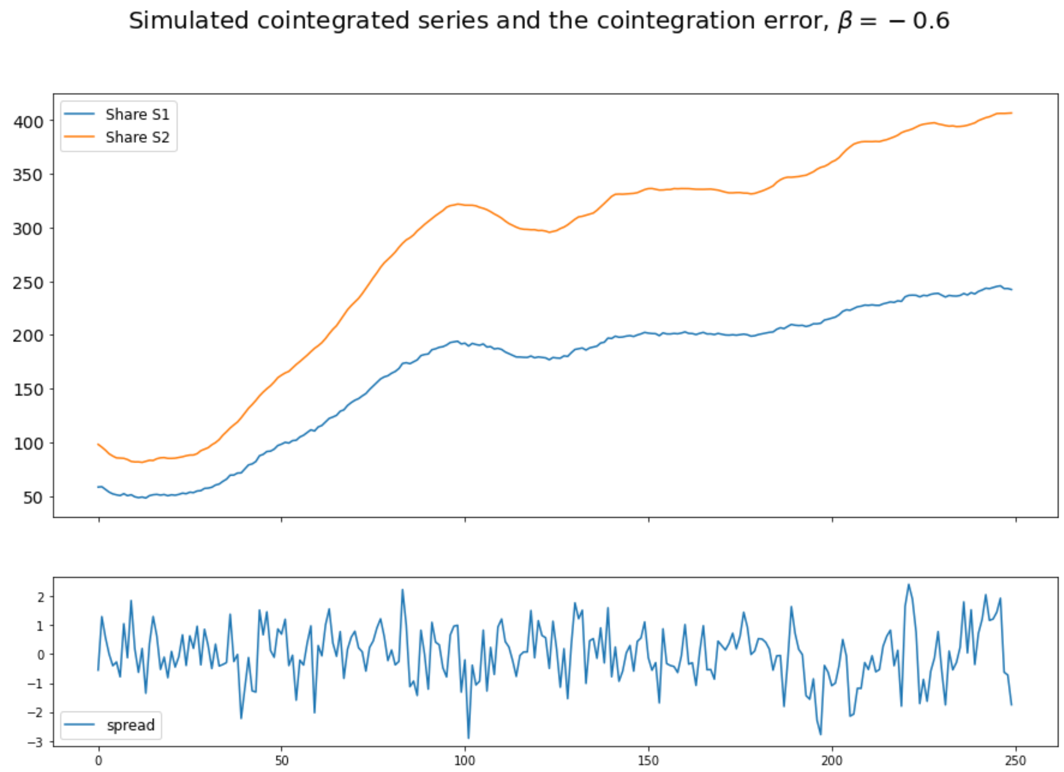
```
# Set the parameters for a_t
price_params = {
    "ar_coeff": 0.95,
    "white_noise_var": 0.5,
    "constant_trend": 0.5
}

# Set the parameters for b_t
coint_params = {
    "ar_coeff": 0.2,
    "white_noise_var": 1.,
    "constant_trend": 0.,
    "beta": -0.6
}

# Load the parameters
coint_simulator.load_params(price_params, target='price')
coint_simulator.load_params(coint_params, target='coint')
```

```
s1, s2, coint_errors = coint_simulator.simulate_coint(initial_price=100.,
                                                       use_statsmodels=True)
```

```
plot = coint_simulator.plot_coint_series(s1[:,5], s2[:,5], coint_errors[:,5])
```




Simulating 20 cointegrated pairs in one batch. Each cointegrated series have 250 time points.

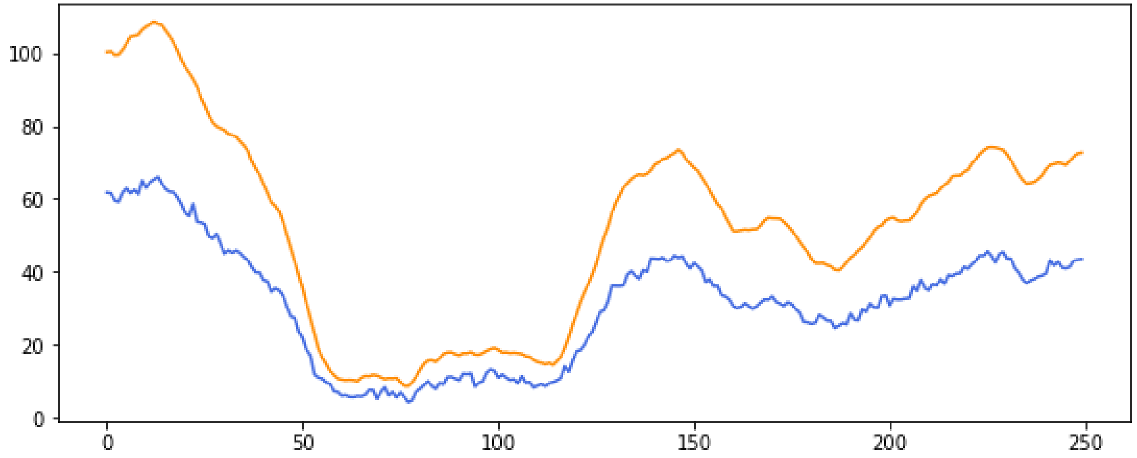
HUDSON AND THAMES

# Minimum Profit Optimization

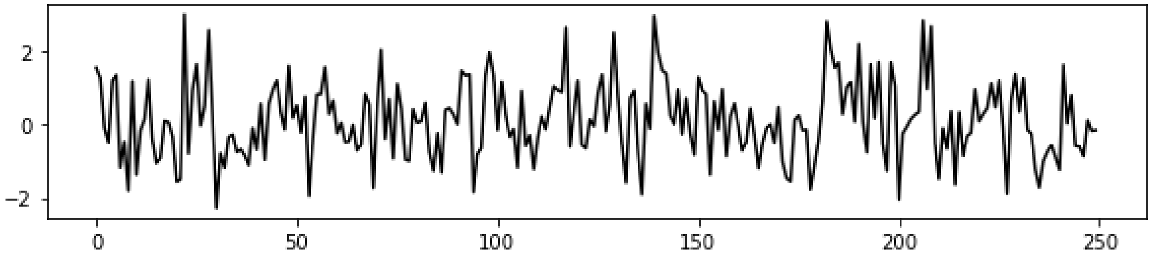
Time to develop the mean-reversion trading strategy on cointegrated assets

HUDSON AND THAMES

# Brief Recall: What are we trading?




$\beta$




$I(0)$ series

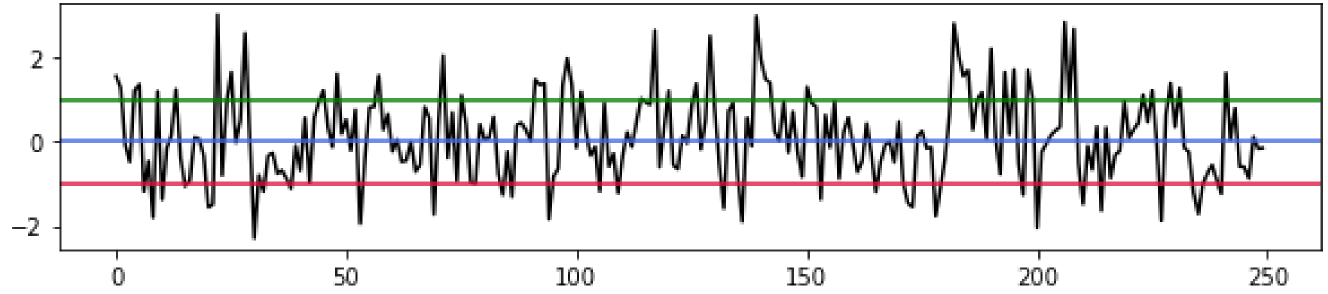
# What are we optimizing?

- The spread between the cointegrated assets is mean-reverting
- The trading strategy is intuitive: fade the extremes, i.e., “buy low, sell high”
- **Factor 1: Where to initiate the trades? (Trade Location)**
  - Determines the minimum P&L per trade
- **Factor 2: How frequent are the trades? (Trade Frequency)**
  - Determines the number of trades over a certain time period
- **Optimization Target:**

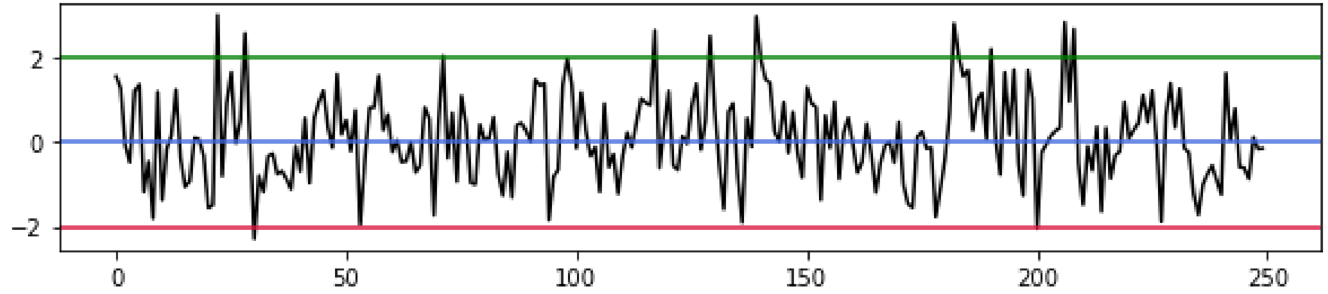
$$\text{Trade Location} \times \text{Trade Frequency}$$

HUDSON AND THAMES

# What are we optimizing?




Tighter boundaries, smaller minimum profit per trade, higher trade frequency




Looser boundaries, larger minimum profit per trade, lower trade frequency

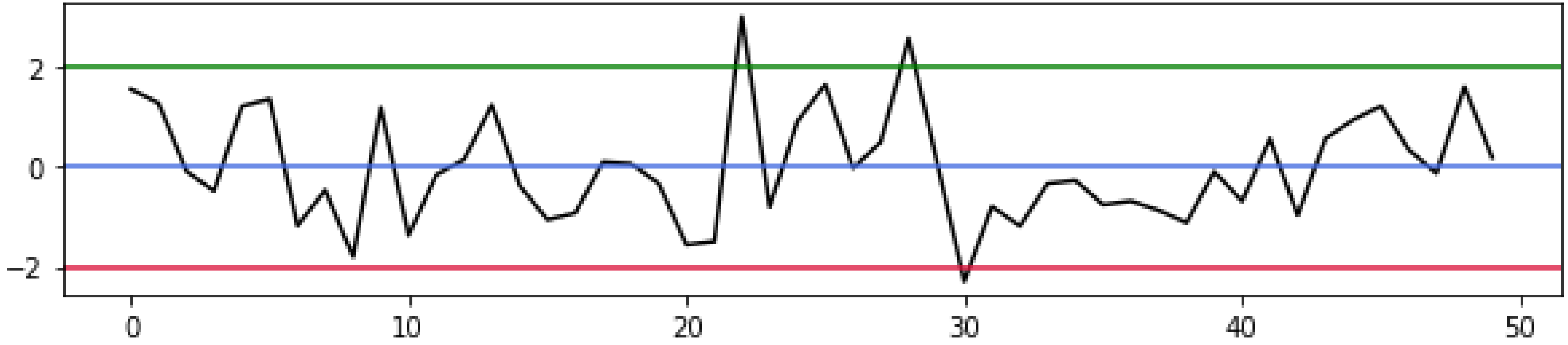
HUDSON
AND THAMES

# Trade Location: Minimum Profit Per Trade

- Minimum profit per trade
  - Assume the upper boundary is $U > 0$, and the cointegrated coefficient of asset $X$ and $Y$ is $\beta$, i.e., $\varepsilon_t = X - \beta Y$ is an $I(0)$ series.
  - When $\varepsilon_t \geq U$, open a trade by selling $N$ of asset $X$ and buying $\beta N$ of asset $Y$
  - When $\varepsilon_t \leq E[\varepsilon_t]$, close the trade
  - Set a symmetric lower boundary for more trades

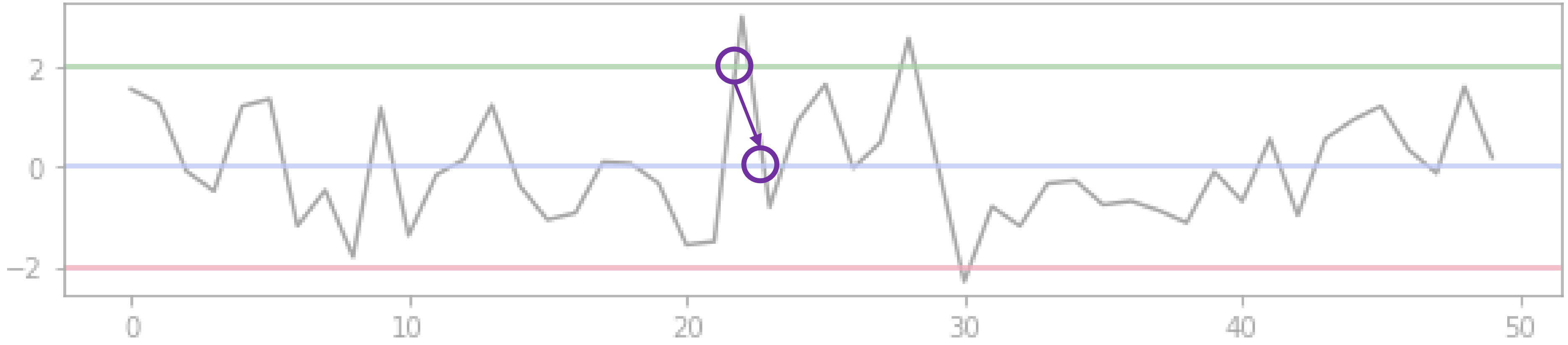
HUDSON AND THAMES

# Trade Location: Minimum Profit Per Trade



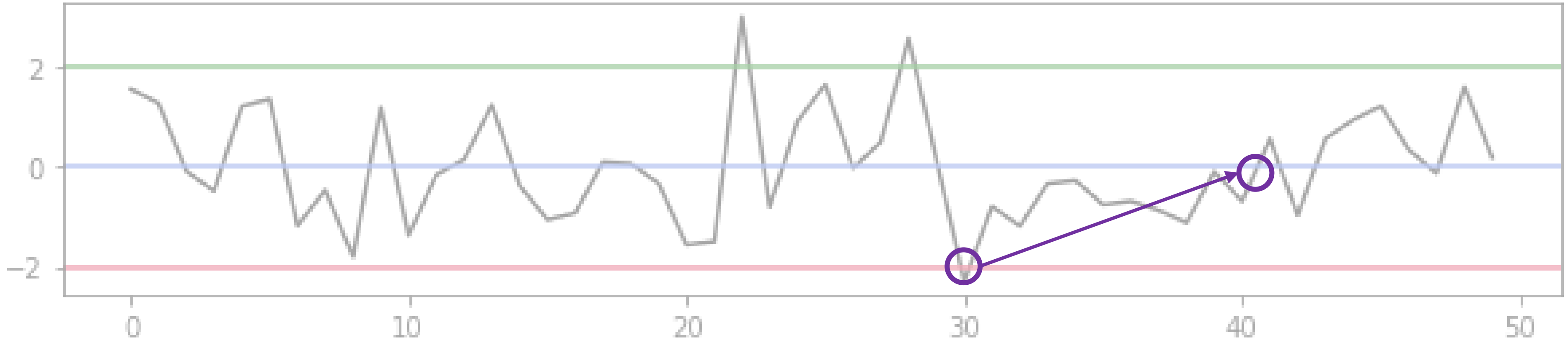

HUDSON AND THAMES

# Trade Location: Minimum Profit Per Trade




HUDSON
AND THAMES

# Trade Location: Minimum Profit Per Trade



# Trade Location: Minimum Profit Per Trade

| Trading Strategy | Minimum Profit per Trade |
|---|---|
| • If $\varepsilon_t \geq U$, sell $N$ of asset $X$, buy $\beta N$ of asset $Y$<br>• If $\varepsilon_t \leq E[\varepsilon_t]$, close the trade | |
| • If $\varepsilon_t \leq U$, buy $N$ of asset $X$, sell $\beta N$ of asset $Y$<br>• If $\varepsilon_t \geq E[\varepsilon_t]$, close the trade | |

HUDSON AND THAMES

# Trade Location: Minimum Profit Per Trade

| Trading Strategy | Minimum Profit per Trade |
|---|---|
| • If $\varepsilon_t \geq U$, sell $N$ of asset $X$, buy $\beta N$ of asset $Y$<br>• If $\varepsilon_t \leq E[\varepsilon_t]$, close the trade | |
| • If $\varepsilon_t \leq U$, buy $N$ of asset $X$, sell $\beta N$ of asset $Y$<br>• If $\varepsilon_t \geq E[\varepsilon_t]$, close the trade | |

Profit per trade

$$
\begin{aligned}
P &= N(X_o - X_c) + \beta N(Y_c - Y_o) \\
&= N(X_o - \beta Y_o) - N(X_c - \beta Y_c) \\
&= N\varepsilon_{t_o} - N\varepsilon_{t_c} \geq NU
\end{aligned}
$$

HUDSON AND THAMES

# Trade Location: Minimum Profit Per Trade

| Trading Strategy | Minimum Profit per Trade |
|---|---|
| • If $\varepsilon_t \geq U$, sell $N$ of asset $X$, buy $\beta N$ of asset $Y$<br>• If $\varepsilon_t \leq E[\varepsilon_t]$, close the trade | $NU$ |
| • If $\varepsilon_t \leq U$, buy $N$ of asset $X$, sell $\beta N$ of asset $Y$<br>• If $\varepsilon_t \geq E[\varepsilon_t]$, close the trade | $NU$ |

**The minimum profit per trade, when trading 1 unit of the spread, is exactly the boundary value $U$!**

Caveat: The minimum profit is terminal.

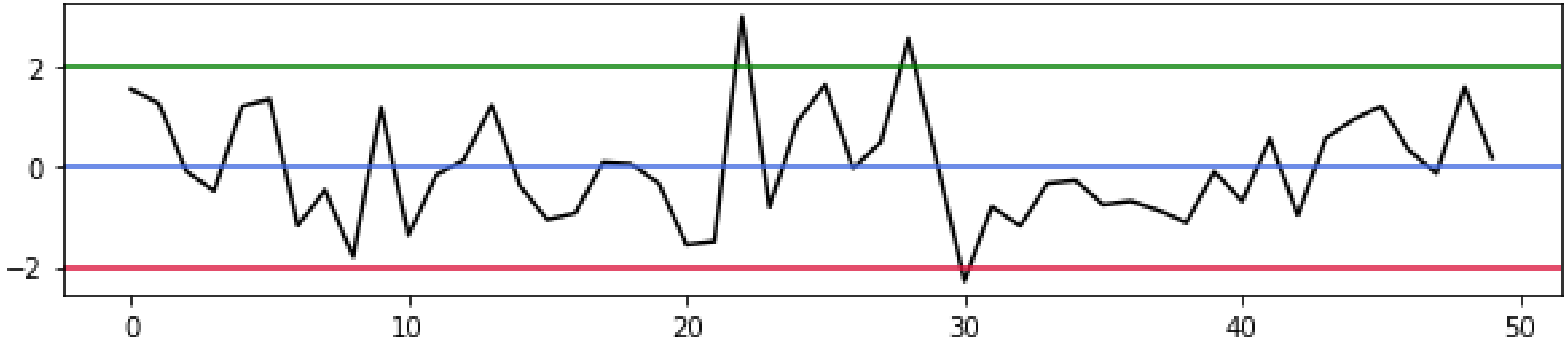
HUDSON AND THAMES

# Trade Frequency: Mean First-passage Time

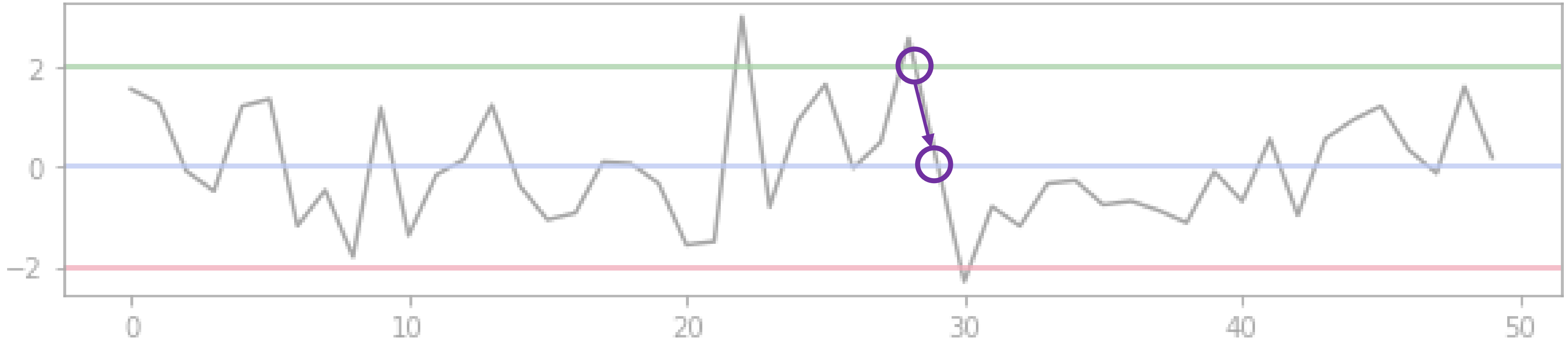
- Assume the cointegrated spread, $\varepsilon_t$, between asset $X$ and asset $Y$ is a stationary AR(1) process
- Use **mean first-passage time** of the AR(1) process to optimize the trade duration and inter-trade interval
  - Trade duration: the mean first-passage time of $\varepsilon_t$ to pass the mean $E[\varepsilon_t]$, if the starting point is at $U$
  - Inter-trade interval: the mean first-passage time of $\varepsilon_t$ to pass $U$, if the starting point is at the mean $E[\varepsilon_t]$
  - Zero the mean for easier calculation

HUDSON AND THAMES

# Trade Frequency: Mean First-passage Time



# Trade Frequency: Mean First-passage Time

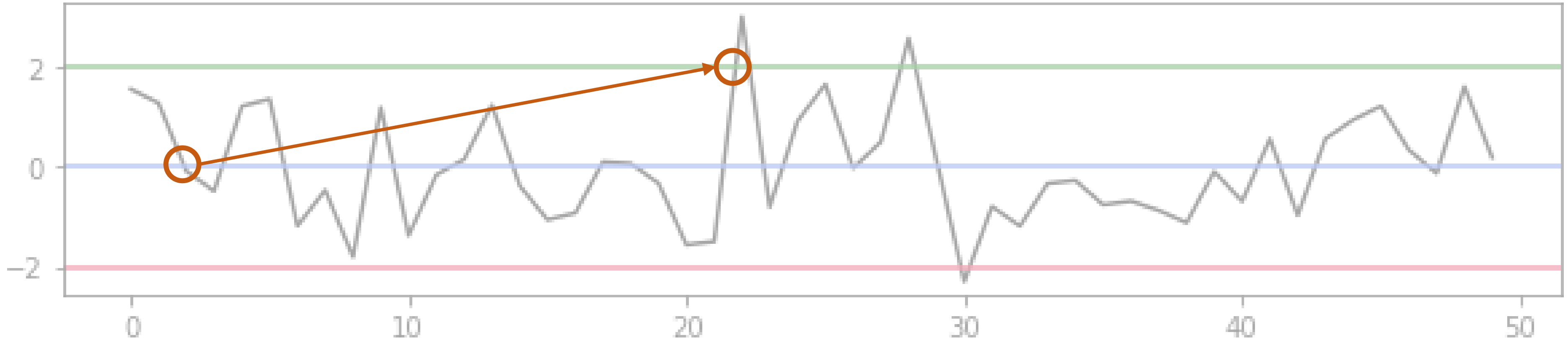



Trade Duration

$$E\left(\mathrm{T}_{0,\infty}(U)\right) = \lim_{b\to\infty}\int_0^b E\left(\mathrm{T}_{0,b}(s)\right)\exp\left(-\frac{(s-\phi U)^2}{2\sigma_a^2}\right)ds + 1$$

HUDSON AND THAMES

# Trade Frequency: Mean First-passage Time

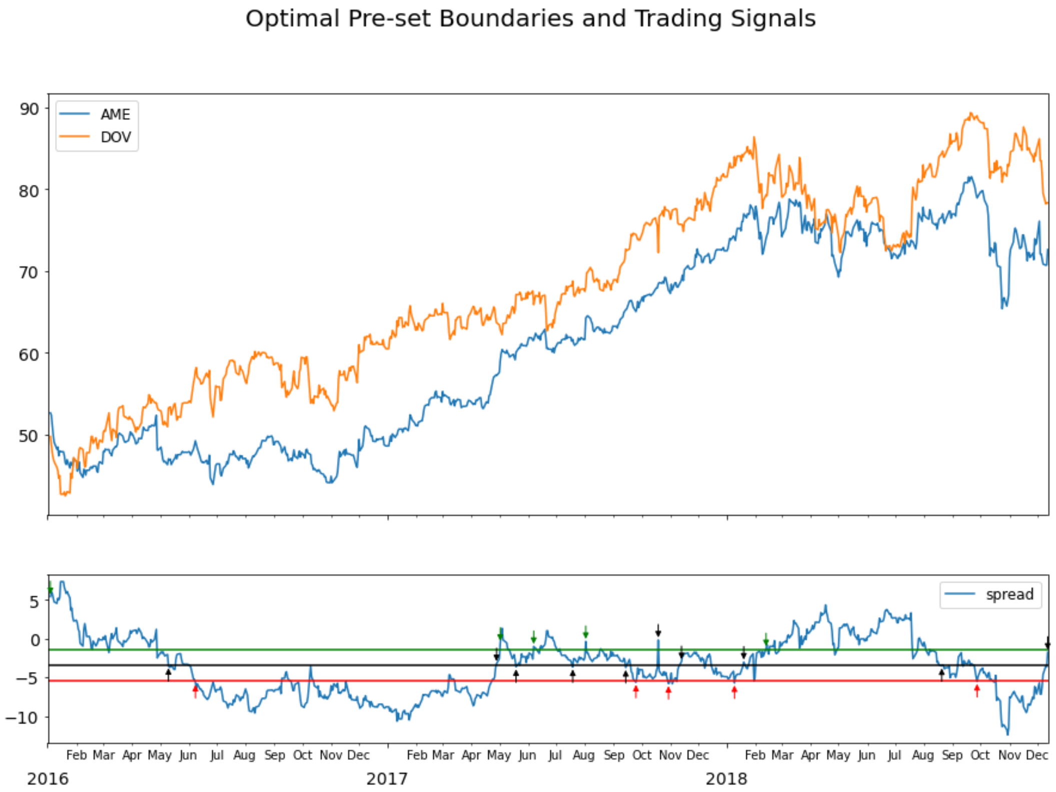



Inter-trade interval

$$E\left(\mathrm{T}_{-\infty,U}(0)\right) = \lim_{-b\to-\infty} \int_{-b}^{U} E\left(\mathrm{T}_{-b,U}(s)\right) \exp\left(-\frac{s^2}{2\sigma_a^2}\right) ds + 1$$

HUDSON AND THAMES

# Minimum Profit Optimization: Results

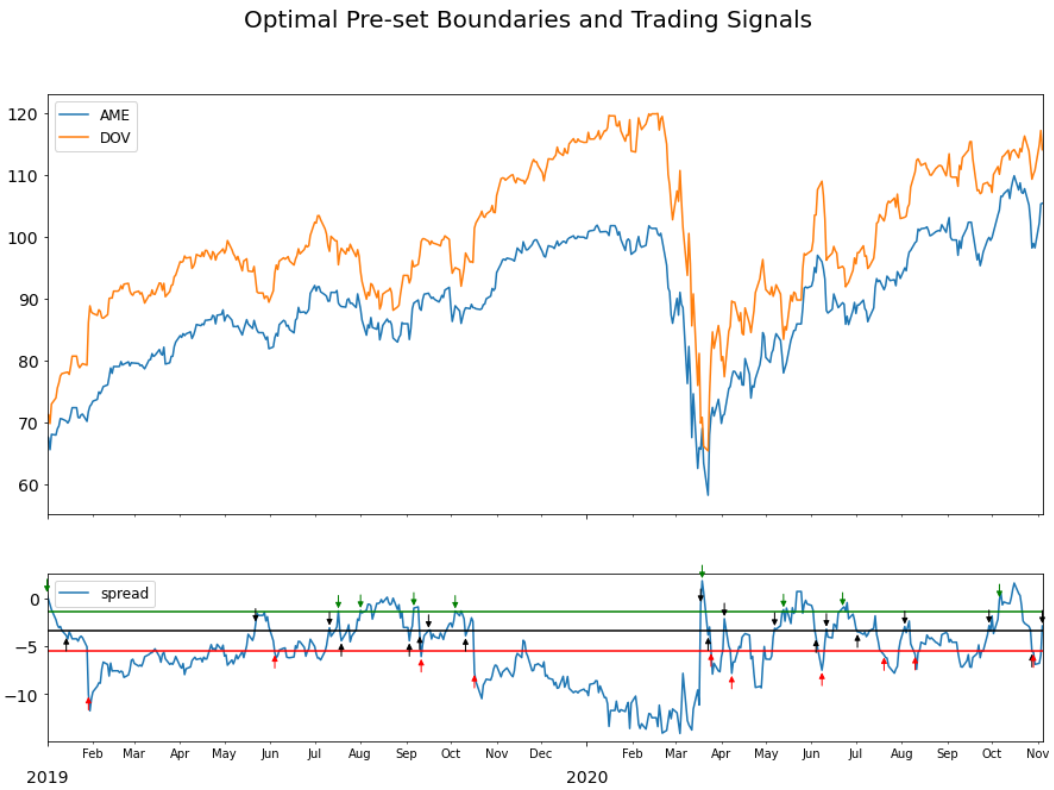


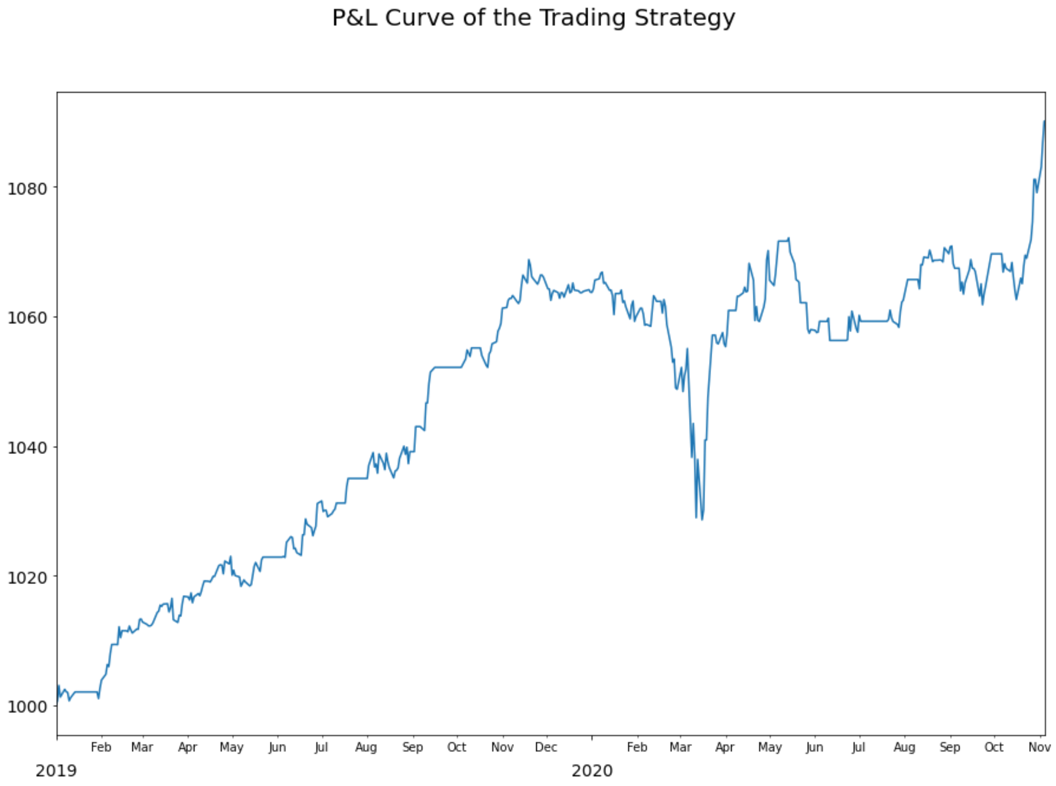


In-sample test of the trading strategy on a cointegrated S&P 500 stock pair, AME and DOV.
Date range: Jan 1st, 2016 to Dec 31st, 2018

HUDSON AND THAMES

# Minimum Profit Optimization: Results




Out-of-sample test of the trading strategy on a cointegrated S&P 500 stock pair, AME and DOV.
Date range: Jan 1st, 2019 to Nov 4th, 2020

HUDSON AND THAMES

# Summary

- Cointegration defines a long-term equilibrium relationship between a pair of assets
- The cointegrated spread between the asset pair is stationary, providing the basis of a mean-reversion trading strategy
- ArbitrageLab can simulate cointegrated asset pairs
- ArbitrageLab can optimize the minimum profit of the mean-reversion trading strategy

HUDSON AND THAMES

# References


- Alexander, C., Giblin, I. and Weddington, W., 2002. Cointegration and asset allocation: A new active hedge fund strategy. *Research in International Business and Finance*, 16(5), pp.65-90.
- Galenko, A., Popova, E. and Popova, I., 2012. Trading in the presence of cointegration. *The Journal of Alternative Investments*, 15(1), pp.85-97.
- Lin, Y.X., McCrae, M. and Gulati, C., 2006. Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Advances in Decision Sciences*, 2006.
- 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), pp.391-419.
- Stock, J.H. and Watson, M.W., 1988. Testing for common trends. *Journal of the American statistical Association*, 83(404), pp.1097-1107.
- Vidyamurthy, G., 2004. Pairs Trading: quantitative methods and analysis (Vol. 217). John Wiley & Sons.


HUDSON AND THAMES

# Questions?

HUDSON
AND THAMES