Machine learning for Pairs Trading in Utilities

Group 9: Chenyang Que, Haoxuan Su,
Hongyang Lu, Runyao Tang, Jiahao Zhang

TABLE OF CONTENTS

Motivation

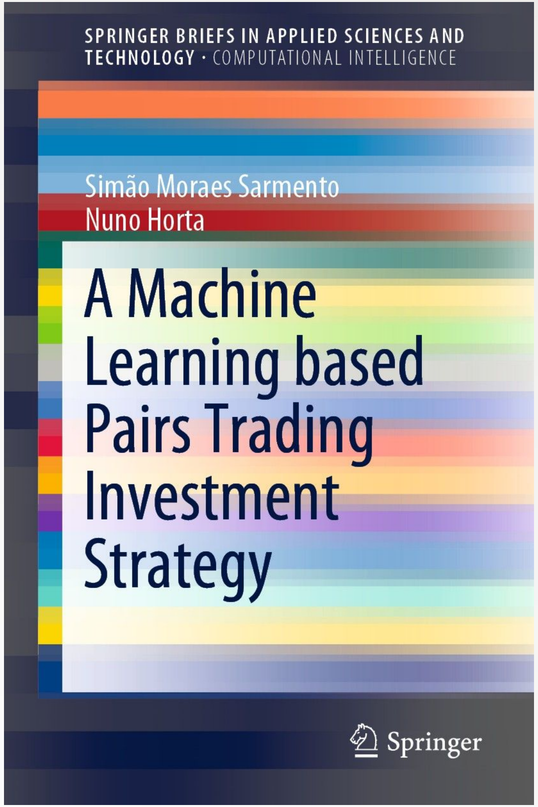
Machine learning for pairs trading

Assets Price & Return
- High dimensional
- Non-linear

Can Unsupervised Learning find more promising pairs?

Purpose:
- finding the appropriate candidate pairs
- selecting the most promising ones.

Compare with traditional method
- OPTICS
- Correlation



Sarmento, Simão Moraes, and Nuno Horta. *A machine learning based pairs trading investment strategy*. Springer, 2020.

Data

Data: Yahoo finance API (yahooquery)

- US market: 9502 asset
- Stock: 6010 (exclude ETF)
- Sector Utilities: 82 stocks
- Pairs: (82*81)/2 = 3321

Daily data

Train period 2023/03 - 2023/12
Test period 2023/01-2023/04

Ticker	Company	Sector	Industry	Country	Market Cap	P/E	Price	Change	Volume
ADN	Advent Technologies Holdings Inc	Utilities	Utilities - Renewable	USA	11.42M	0.00	0.15	-2.97%	848185.0
AEE	Ameren Corp.	Utilities	Utilities - Regulated Electric	USA	19.82B	17.00	74.36	0.96%	691501.0
AEP	American Electric Power Company Inc.	Utilities	Utilities - Regulated Electric	USA	45.60B	20.40	86.67	1.65%	1900397.0
AES	AES Corp.	Utilities	Utilities - Diversified	USA	12.66B	54.37	17.81	3.46%	5798232.0
AGR	Avangrid Inc	Utilities	Utilities - Regulated Electric	USA	14.15B	15.88	36.58	0.54%	334536.0
...	...	...	...	...	...	...	...	...	...
VST	Vistra Corp	Utilities	Utilities - Independent Power Producers	USA	27.04B	22.28	77.74	6.93%	3862044.0
WEC	WEC Energy Group Inc	Utilities	Utilities - Regulated Electric	USA	26.07B	19.60	82.63	1.40%	1160992.0
WTRG	Essential Utilities Inc	Utilities	Utilities - Regulated Water	USA	9.98B	19.58	36.47	1.05%	703016.0
XEL	Xcel Energy, Inc.	Utilities	Utilities - Regulated Electric	USA	30.27B	16.38	54.49	0.99%	2732624.0
YORW	York Water Co.	Utilities	Utilities - Regulated Water	USA	509.10M	21.39	35.53	0.44%	28265.0

Trading Execution Threshold-Based Trading Model

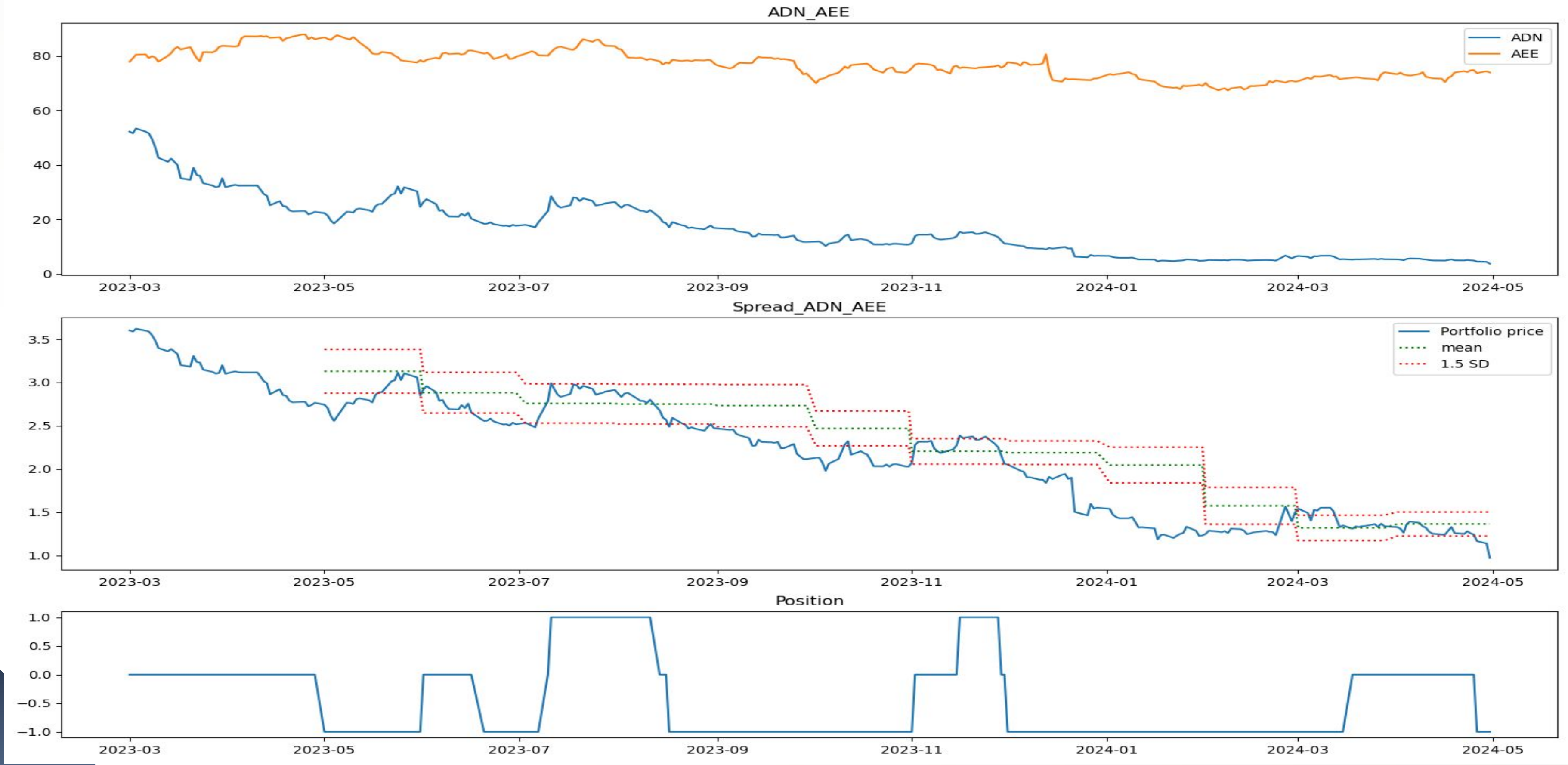
1. Calculate the spread's **(St = Yt - Xt)** mean **μ**, and standard deviation **σ** during the pair's formation period.

2. Define the model thresholds: the threshold that triggers a long position **L** ,the threshold that triggers a short position **S**, and the exit threshold exit that defines the level at which a position should be exited.

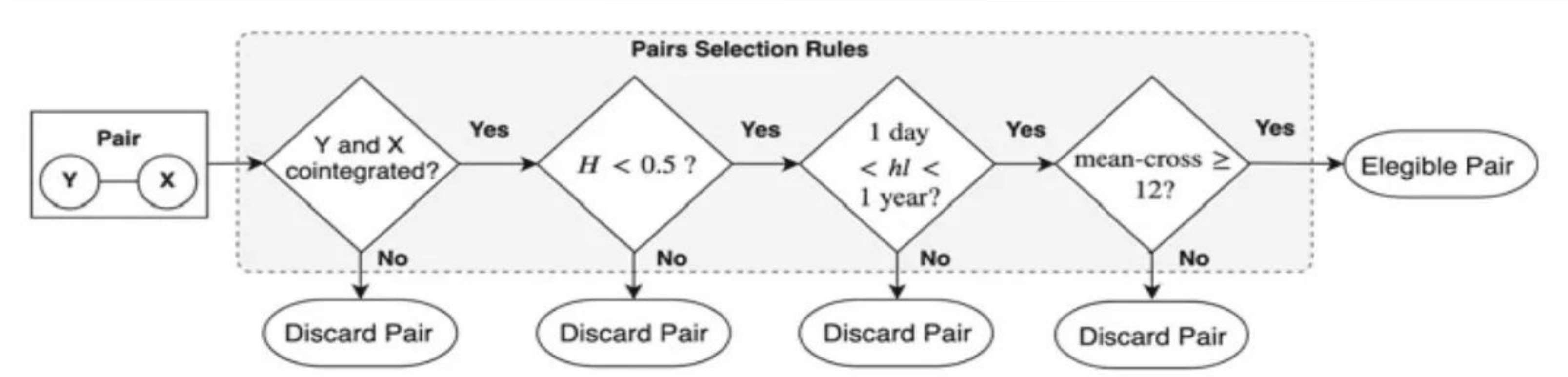
3. Monitor the evolution of the spread, St and control if any **threshold is crossed**.

4. In case L is crossed, go long the spread by **buying Y and selling X** .If S is triggered, short the spread by **selling Y and buying X** . Exit position when exit is triggered and a position was being held.

Pairs


ADN_AEE
ADN
AEE
Spread_ADN_AEE
Portfolio price
mean
1.5 SD
Position
2023-03
2023-05
2023-07
2023-09
2023-11
2024-01
2024-03
2024-05

Cointegration



Mean-Reversion and **Stationarity**

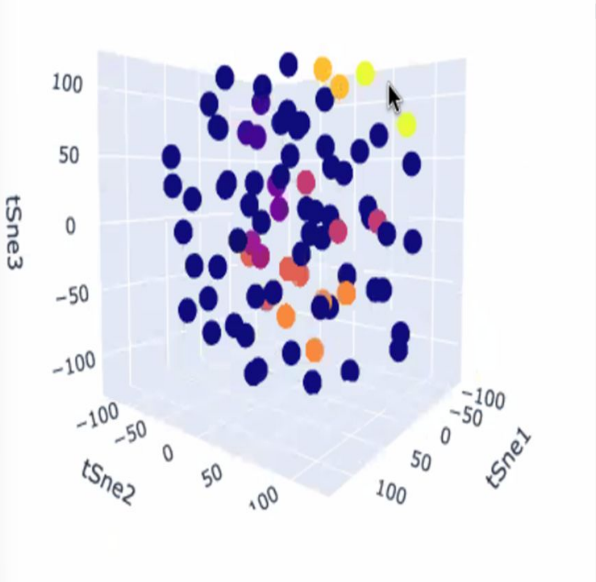
- Mean-reversion is the tendency of a time series to return to its historical average over time.
- Stationarity implies that the statistical properties of a time series, such as mean and variance, do not change over time.

  - **Cointegration Test p-value < 0.05**: Ensures a stable long-term relationship between the pair.
  - **Hurst Exponent < 0.5**: Confirms mean-reverting behavior.
  - **1 < Half-Life < 365**: Validates that the mean reversion occurs at a reasonable pace.
  - **Mean-Cross >=12**: Ensures sufficient trading opportunities through frequent mean reversion.

Sarmento, Simão Moraes, and Nuno Horta. *A machine learning based pairs trading investment strategy*. Springer, 2020.

Machine Learning

Identification of two securities, for example two stocks, for which the corresponding prices series display a similar behaviour, or simply seem to be **linked** to each other.



Normalization & Dimensionality reduction

- We imported,adjusted, standardized these data and find a compact representation for each security
- T-SNE:Non-linear method.

Unsupervised Learning—apply an appropriate clustering algorithm;

- OPTICS : A density-based algorith
- Handles Varying Densities
- Effectively with Noise
- Flexible Clustering

Select pairs—define a set of rules to select pairs for trading.

- Combined with Cointegration, we finally came up with 4 pairs that can meet all our requirements.

Result & Beaktest

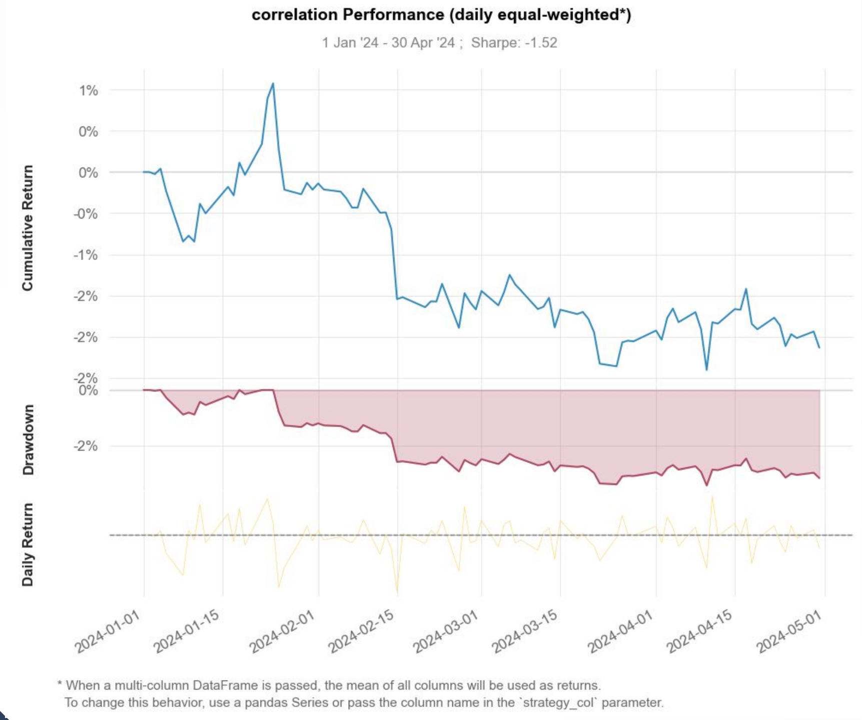
1. Correlation

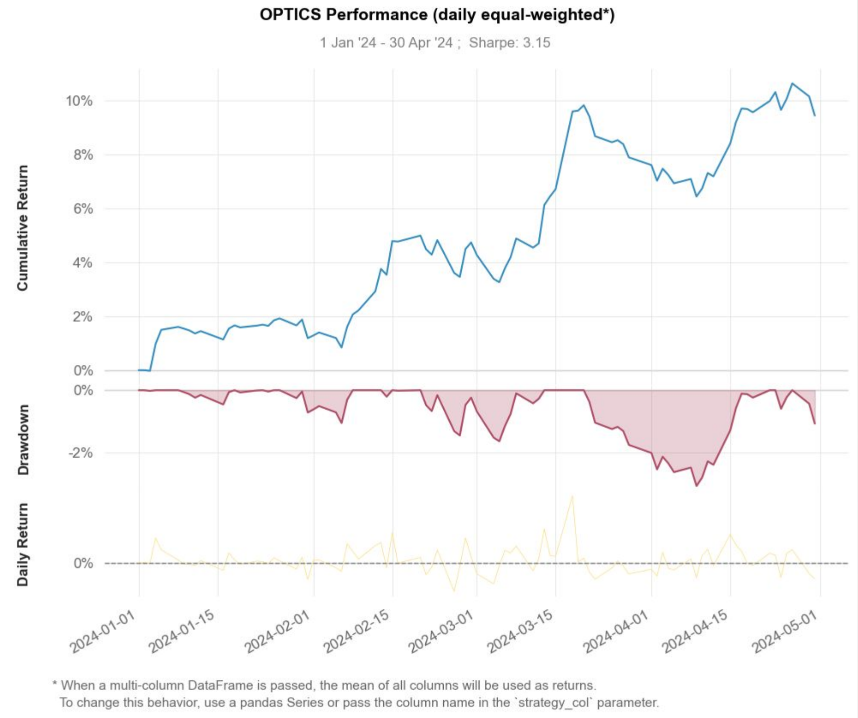
Pair	correlation	pvalue	hurst_exponent	half_life	crossovers	Return	Std Dev	Max Drawdown	Sharpe Ratio	Omega
AGR_ES	0.977222	0.014088	0.371316	6.513701	26	**-0.199022**	0.084834	-0.075814	**-2.346009**	**1.316523**
BKH_POR	0.975519	0.03509	0.230085	36.62348	33	**-0.047788**	0.049497	-0.023879	**-0.965465**	**1.186399**
ES_WTRG	0.974181	0.021631	0.412536	38.71063	26	**-0.136136**	0.087945	-0.090848	**-1.547979**	**1.34521**
CPK_ORA	0.973617	0.001601	0.14878	7.060058	46	**0.155662**	0.101842	-0.026966	**1.52847**	**1.443658**

2. OPTICS

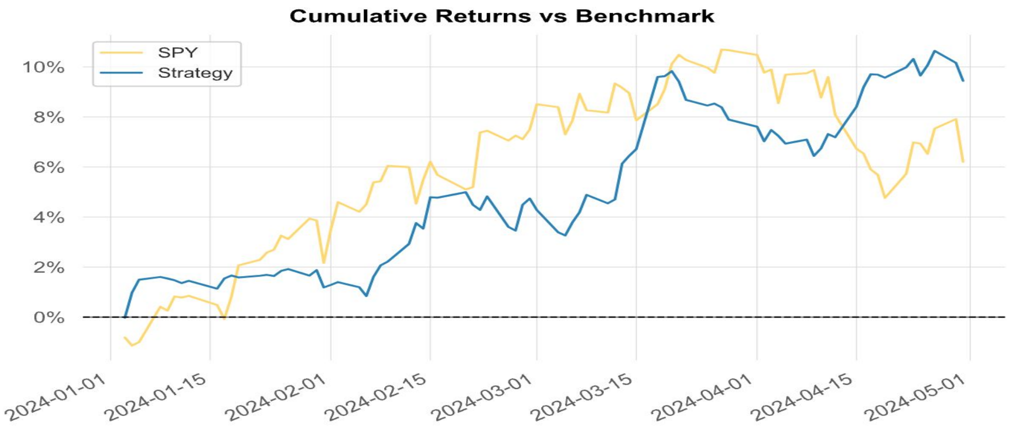
Pair	OPTICS	pvalue	hurst_exponent	half_life	crossovers	Return	Std Dev	Max Drawdown	Sharpe Ratio	Omega
AMPS_SWX	74.49906	0.001184	0.390763	319.1302	24	**2.384911**	0.286012	-0.049647	**8.338499**	**2.583018**
AES_D	44.41755	0.010932	0.361442	265.1005	26	**-0.049624**	0.154581	-0.105985	**-0.321022**	**1.627348**
ARTNA_NWN	49.96095	0.004631	0.349361	49.28517	34	**0.093354**	0.135373	-0.047971	**0.689605**	**1.455371**
NWN_UGI	58.28693	0.040827	0.326623	172.0943	18	**-0.192875**	0.103003	-0.068142	**-1.872517**	**1.339657**

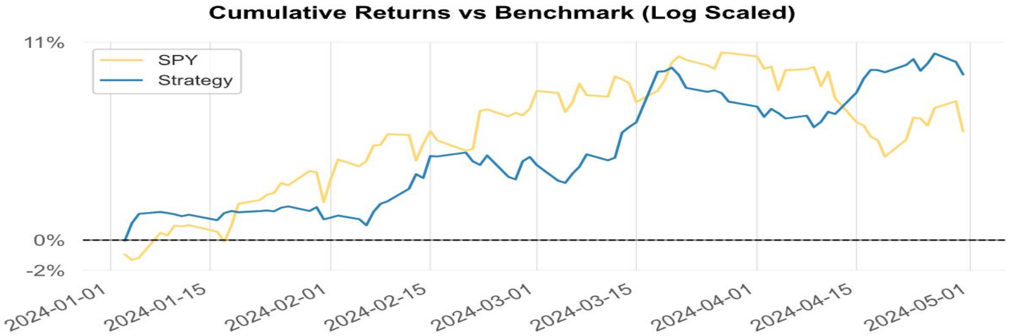
Result & Performance





Result & Performance

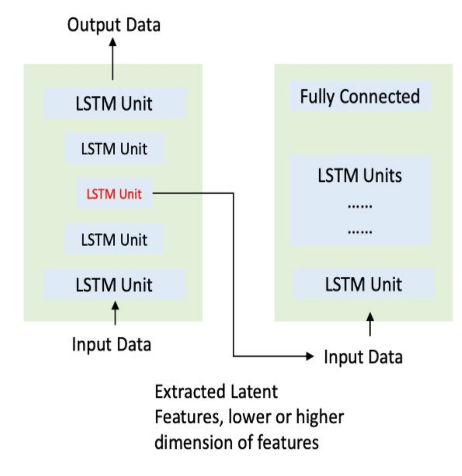




Key Performance Metrics

Metric	SPY	Strategy
Risk-Free Rate	0.0%	0.0%
Time in Market	100.0%	100.0%
Cumulative Return	6.21%	9.44%
CAGR﹪	13.72%	21.25%
Sharpe	1.62	3.19
Prob. Sharpe Ratio	82.03%	98.0%
Smart Sharpe	1.49	2.94
Sortino	2.48	6.42
Smart Sortino	2.29	5.91
Sortino/√2	1.75	4.54
Smart Sortino/√2	1.62	4.18
Omega	1.77	1.77
Max Drawdown	-5.35%	-3.09%
Longest DD Days	34	30
Volatility (ann.)	11.84%	8.83%
R^2	0.01	0.01
Information Ratio	0.04	0.04
Calmar	2.56	6.89
Skew	-0.09	1.32
Kurtosis	-0.07	4.97
Expected Daily	0.07%	0.11%
Expected Monthly	1.52%	2.28%
Expected Yearly	6.21%	9.44%
Kelly Criterion	9.15%	15.86%
Risk of Ruin	0.0%	0.0%

Future development



Refining Entry Points

- Threshold-Based Trading Model: One limitation is the lack of precision in defining entry points.
- Deep Learning Model: Implementing Autoencoder LSTM for Time Series Forecasting could achieve more robust performance.

Sentiment and Event Analysis

- News Impact: Analyzing how news events cause price jumps.
- Pair Selection: Selecting pairs based on mean-reversion speed and jump behavior to enhance performance.

Empirical Research

- Sector and Asset Variability: Conducting research across different sectors and asset classes.
- Frequency Analysis: Investigating from different frequencies, including microstructure market data

Sarmento, Simão Moraes, and Nuno Horta. *A machine learning based pairs trading investment strategy*. Springer, 2020.
Gu, Shihao, Bryan Kelly, and Dacheng Xiu. "Autoencoder asset pricing models." *Journal of Econometrics* 222, no. 1 (2021): 429-450.
Stübinger, Johannes, and Sylvia Endres. "Pairs trading with a mean-reverting jump–diffusion model on high-frequency data." *Quantitative Finance* 18, no. 10 (2018): 1735-1751.

Thanks!
Q&A