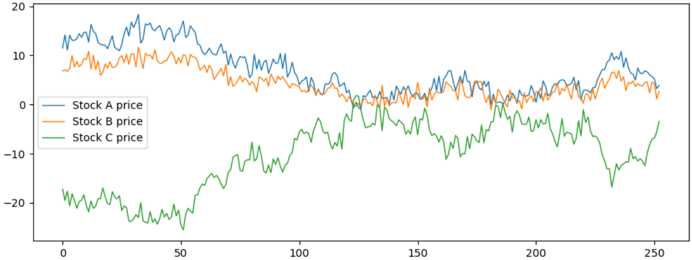
# Comparing Different Methods for Identifying Good Pairs Trade

Miao-Chin Yen

Department of Management Science and Engineering, Stanford University

# Pairs Trading

- Notion: identify a pair of stocks whose price performance behave similar by examining the historical data.
- Trading Strategy: "Buy Low and Sell High" strategy to make profit when the spread of the time series between the pair widens.
- Objective: identify a good approach to generate pairs



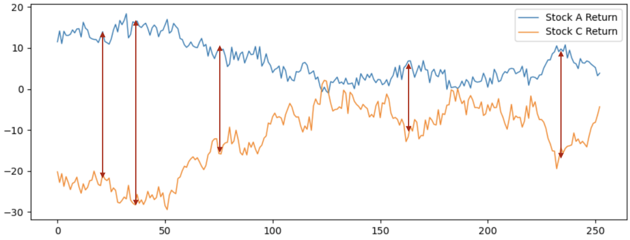
# Pairs Selection

1. Price time series $\{P_{i,t}\}_{t=1}^{T}$ for $i = 1, 2, ...., n$.
2. Accumulative return time series $\{R_{i,t} = \frac{P_{i,t} - P_{i,t-1}}{P_{i,t}}\}_{t=1}^{T}$ for $i = 1, 2, ...., n$.

- Selected by Distance (Distance Approach)
- Selected by Correlation (Correlation Approach)
- Selected by Cointegration (Cointegration Approach)
- Selected by Autoregressive Process (AR Approach)

# Distance Approach & Correlation Approach

- Distance approach selects the pair with the shortest euclidean distance between two time series.
- Correlation approach selects the pair with the highest linear correlation.




- Consider time series $\{R_{i*,t} - R_{j*,t}\}_{t=1}^{T}$ in our experiment.

# Cointegration Approach

- Cointegration enables us to describe the relationship between stock price time series instead of cumulative return time series.
- An $I(1)$ or integrated of 1 series is a random walk.
- An $I(0)$ or integrated of 0 series is a weakly stationary time series.

$$\exists \ \beta \text{ s.t. } \{z_t = P_{i,t} - \beta P_{j,t}\}_{t=0}^{T} \text{ is an } I(0) \text{ series}$$

- Consider time series $\{P_{i*,t} - \beta P_{j*,t}\}_{t=0}^{T}$ in our experiment.

# AR Approach

- Fit $\{X_t = R_{i,t} - R_{j,t}\}_{t=1}^{T}$ as AR(p) process and use mean-reverting(stationary) property to choose pairs.
- Half-Life: measure the mean-reverting speed

$$h = -\frac{\ln 2}{\ln \sum_{k=1}^{p} \phi_k}$$

- Shorter half-life means the process is expected to halve its distance to the stationary mean faster.

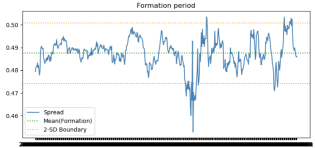
$$(i^*, j^*) = \underset{i,j,i\neq j}{\arg\min} -\frac{\ln 2}{\ln \sum_{k=1}^{p} \phi_k^{(i^*,j^*)}}$$

$$\text{s.t. } X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + W_t \quad \text{where } W_t \sim WN\,(0,1)$$

- Consider time series $\{R_{i*,t} - R_{j*,t}\}_{t=1}^{T}$ in our experiment.

# Data & Criteria

- Select 165 stocks from Vanguard Small Cap Value ETF (VBR), resulting in $\frac{165 \cdot 164}{2} = 13530$ different pairs
- Formation period: Jaunary 1, 2019 to December 31, 2020
- Trading period: January 1, 2021 to December 31, 2021
- Criteria
  1. Portfolio has high prob. within $[\hat{\mu}_{\mathsf{F}} - 2\hat{\sigma}_{\mathsf{F}}, \hat{\mu}_{\mathsf{F}} + 2\hat{\sigma}_{\mathsf{F}}]$
  2. $\text{MD} := \left|\frac{\hat{\mu}_{\mathbf{T}} - \hat{\mu}_{\mathbf{F}}}{\hat{\sigma}_{\mathbf{T}}}\right|$




(a) Formation


(b) Trading

# Experiment Results

| Distance | % Days within $[\hat{\mu}_{\mathrm{F}} - 2\hat{\sigma}_{\mathrm{F}}, \hat{\mu}_{\mathrm{F}} + 2\hat{\sigma}_{\mathrm{F}}]$ | MD | Correlation | % Days within $[\hat{\mu}_{\mathrm{F}} - 2\hat{\sigma}_{\mathrm{F}}, \hat{\mu}_{\mathrm{F}} + 2\hat{\sigma}_{\mathrm{F}}]$ | MD |
|---|---|---|---|---|---|
| EGP-FR | 87.2 | 0.58 | UA-UAA | 24.7 | 2.17 |
| NWE-BKH | 80.0 | 0.26 | PEB-PK | 100.0 | 0.67 |
| CENT-CENTA | 99.6 | 1.10 | CENT-CENTA | 99.6 | 1.10 |
| UA-UAA | 24.7 | 2.17 | GPMT-ARI | 100.0 | 0.94 |
| OGS-IDA | 9.9 | 1.91 | ROIC-SITC | 58.9 | 2.84 |

| Cointegration-Distance Sorting | % Days within $[\hat{\mu}_{\mathrm{F}} - 2\hat{\sigma}_{\mathrm{F}}, \hat{\mu}_{\mathrm{F}} + 2\hat{\sigma}_{\mathrm{F}}]$ | MD | Cointegration-Correlation Sorting | % Days within $[\hat{\mu}_{\mathrm{F}} - 2\hat{\sigma}_{\mathrm{F}}, \hat{\mu}_{\mathrm{F}} + 2\hat{\sigma}_{\mathrm{F}}]$ | MD |
|---|---|---|---|---|---|
| OGS-POR | 9.56 | 1.97 | PTEN-CLR | 53.3 | 0.66 |
| NWE-BKH | 78.0 | 0.25 | FULT-BRKL | 99.6 | 0.30 |
| IDA-WTM | 99.6 | 0.35 | SFNC-FCF | 7.2 | 3.58 |
| OGS-SR | 45.8 | 2.39 | PEB-ASB | 5.9 | 2.93 |
| FR-EGP | 99.6 | 0.63 | ASB-PK | 7.5 | 2.64 |

# Experiment Results

| AR(1) | % Days within $[\hat{\mu}_F - 2\hat{\sigma}_F, \hat{\mu}_F + 2\hat{\sigma}_F]$ | MD | AR(2) | % Days within $[\hat{\mu}_F - 2\hat{\sigma}_F, \hat{\mu}_F + 2\hat{\sigma}_F]$ | MD |
|---|---|---|---|---|---|
| GPMT-SABR | 36.25 | 1.35 | GPMT-SABR | 36.25 | 1.35 |
| IDA-WTM | 45.81 | 0.35 | OFC-MGEE | 99.60 | 0.69 |
| FOR-WH | 43.82 | 0.99 | H-BXMT | 90.83 | 0.35 |
| NWE-BKH | 78.08 | 0.25 | FOR-WH | 43.82 | 0.99 |
| H-UMPQ | 39.84 | 1.63 | H-UCBI | 59.76 | 2.78 |

| AR(3) | % Days within $[\hat{\mu}_F - 2\hat{\sigma}_F, \hat{\mu}_F + 2\hat{\sigma}_F]$ | MD | AR(4) | % Days within $[\hat{\mu}_F - 2\hat{\sigma}_F, \hat{\mu}_F + 2\hat{\sigma}_F]$ | MD |
|---|---|---|---|---|---|
| OFC-MGEE | 99.60 | 0.69 | OFC-MGEE | 99.60 | 0.69 |
| FOR-WH | 43.82 | 0.99 | FOR-WH | 43.82 | 0.99 |
| SBRA-FHI | 58.56 | 1.25 | IDA-WTM | 45.81 | 0.35 |
| GPMT-SABR | 36.25 | 1.35 | WH-HOMB | 75.29 | 0.36 |
| IDA-WTM | 45.81 | 0.35 | SBRA-FHI | 58.56 | 1.25 |

# Next Step & Summary

## Next Step

- Apply Granger Causality
- Analyze importance of different factors for selecting pairs

## Summary

By inspecting the experiments to date, it seems that mean-reverting property would be important for selecting pairs. However, other methods can sometimes provide strong and useful pairs. Therefore, it would be worth studying the importance of different factors.

# Supplementary

**Example. AR(1)**
Let $x_t = \phi_1 x_{t-1} + w_t$ where $|\phi_1| < 1$ denote a weak-stationary AR(1) process. We further let $E[x_t] = \mu$ for all $t$ because of stationarity. By computation, we derive

$$\mu = \frac{1}{1 - \phi_1}$$

We then plug $\mu$ into the AR(1) process and get

$$x_t - \mu = \phi_1 \left( X_{t-1} - \mu \right) + w_t$$

We define $y_t = x_t - \mathbb{E}[x_t]$ to be the the distance to the stationary mean. By definition of half-life, we want to find $h$ such that

$$\mathbb{E}_t \left[ y_{t+h} \right] = \frac{1}{2} y_t$$

Since $\mathbb{E}_t \left[ y_{t+h} \right] = \phi_1^h y_t = \frac{1}{2} y_t$,

$$h = -\frac{\ln 2}{\ln \phi_1}$$