High Frequency Statistical Arbitrage Model

Pair and cluster trading using price movement per second in correlated companies

Dottie, Luisa, Cedrick, Vidushi, Tyler

Background

Statistical arbitrage:

- Pairs and cluster trading: trade based on the linear combination of assets
- Rooted in *mean-reversion principles*

High frequency trading:

- Trade orders down to a fraction of a second

Our model:

- Combine HFT and statistical arbitrage strategies based on an optimal band strategy
- Universe: NASDAQ 100 companies
- Timescale: seconds
- Data: Thesys

Outline

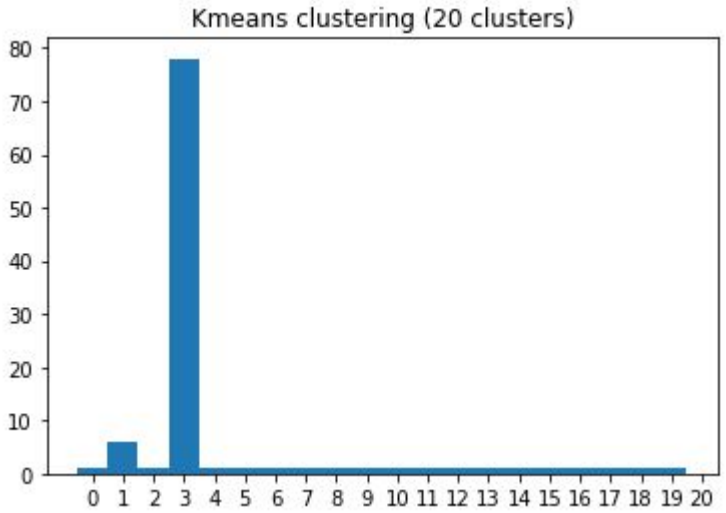
1. Company selection
2. Our approach
3. Future steps

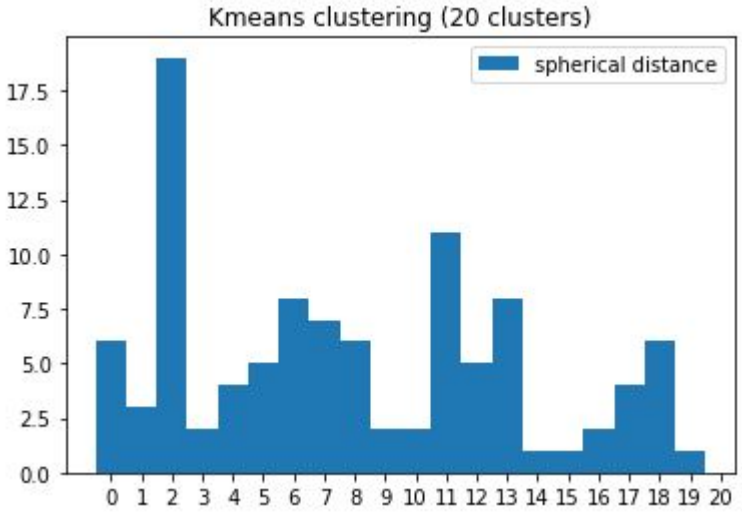
Company Selection: Methodology

- Naive method: select pairs according to our intuition
- Automated selection: clustering.
  - On which data ? All residual history or residuals at particular time stamps?
- Data preprocessing:
  - Remove market impact by subtracting beta coefficient from the returns

Company Selection: Results

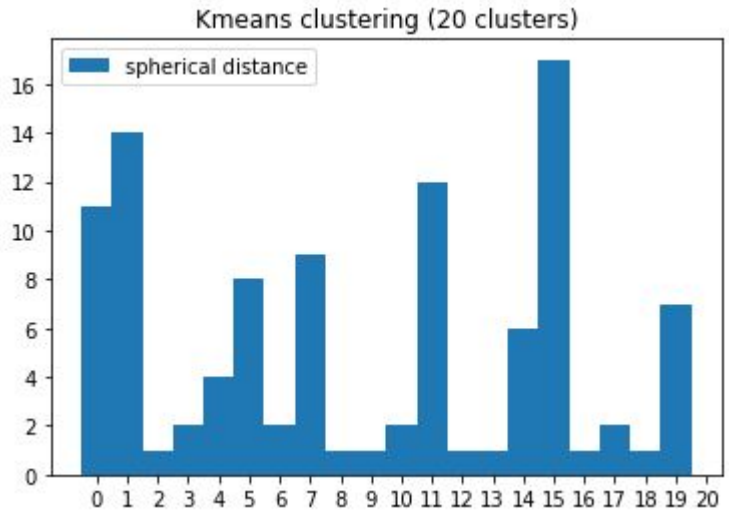
- **Method 1:** K-means on the history of residuals (d=1260)

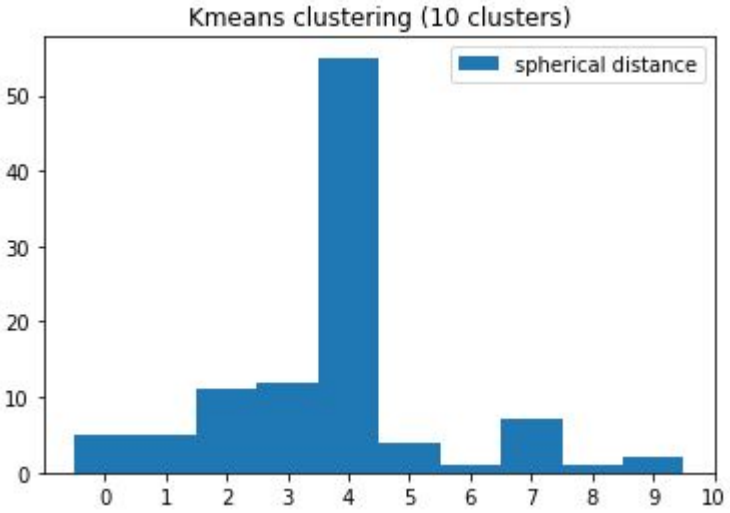




Company Selection: Results

Importance of removing market effect

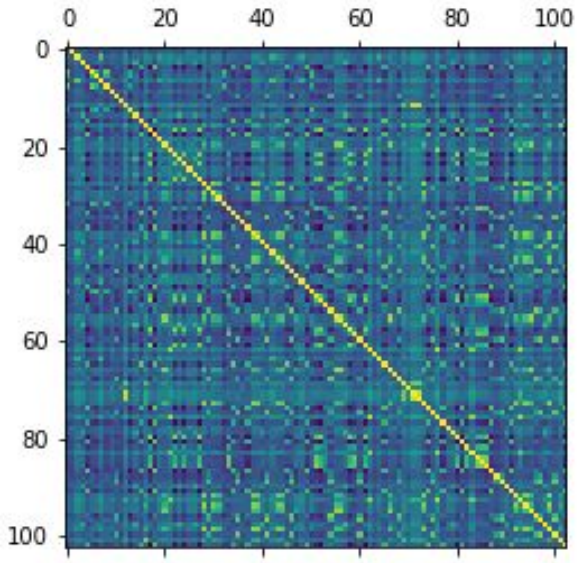




Company Selection: Results

- **Method 2**: Track evolution of clusters at each time stamp (d=1)
  - Select the pairs with the highest correlation

- Next steps:
  - Check the hypothesis
  - Compare the methods

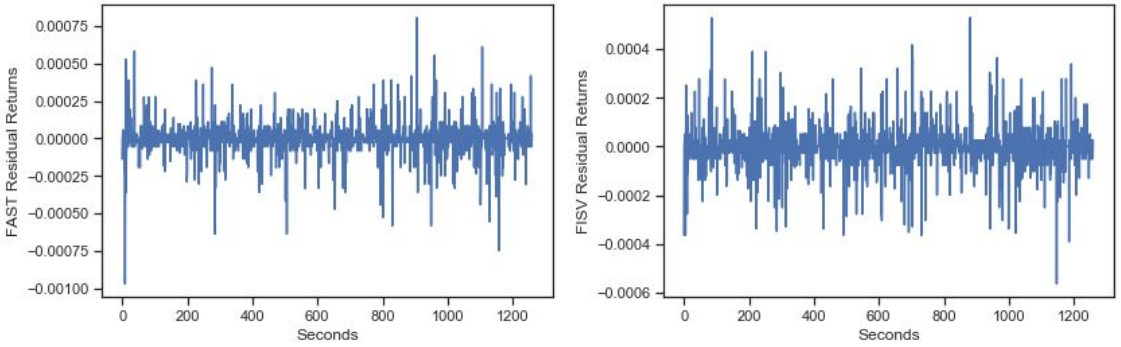


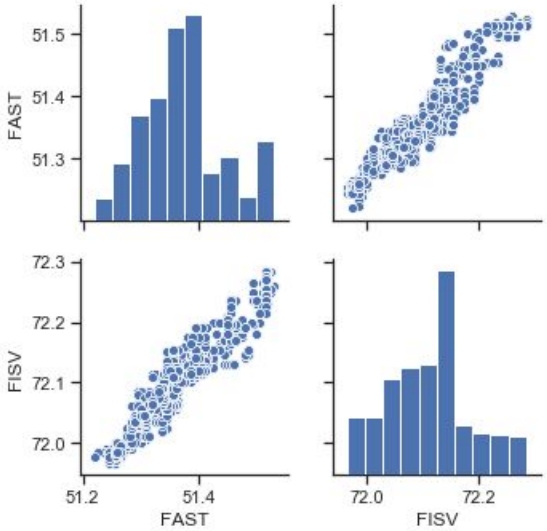
Cointegration of Pairs: Methodology

- Determines relationship between non-stationary time series variables
- Engle-Granger Method

$$x_t \text{ and } y_t \text{ are non-stationary time series variables}$$
$$\text{if } y_t - \beta x_t = u_t \text{ where } u_t \text{ is stationary, then cointegration}$$

- Cointegration test run on residual returns





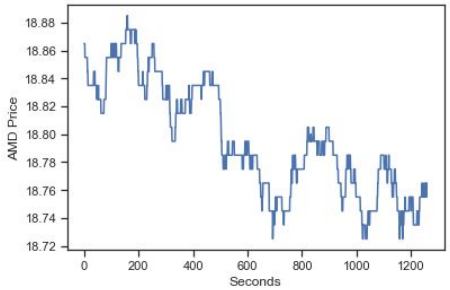
Cointegration of Clusters: Methodology

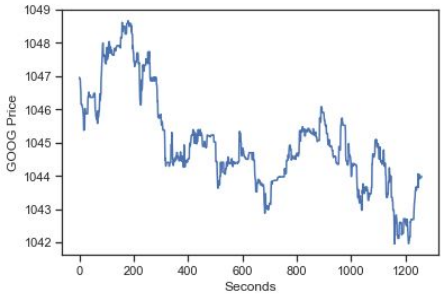
- Johansen Test for more than 2 time series
  - Verifies relationship between multiple stocks returned by k-means clustering

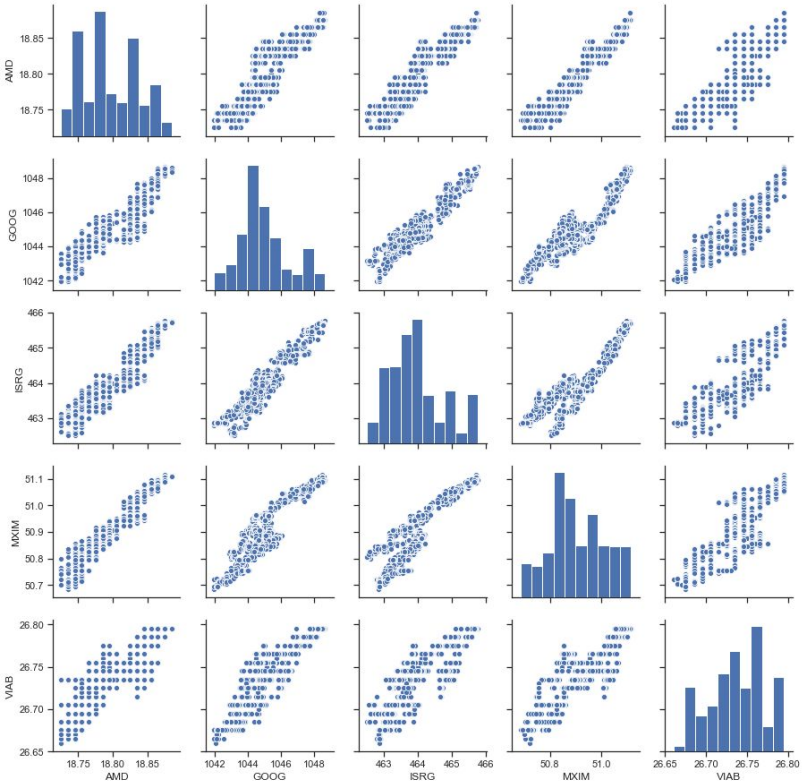
$x_{1_t}, \ldots, x_{k_t}$ are non-stationary time series variables

if $x_{j_t} - \sum_{i=1, i \neq j}^{k} \beta_i x_{i_t} = u_t$ where u_t is stationary, then cointegration

- Extension of pair trading to clusters of stocks?

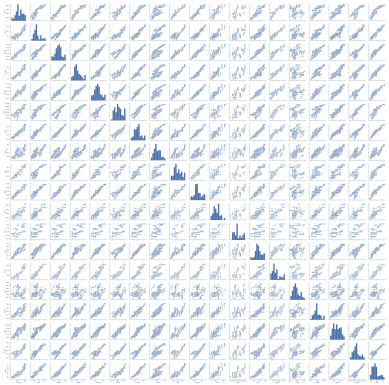






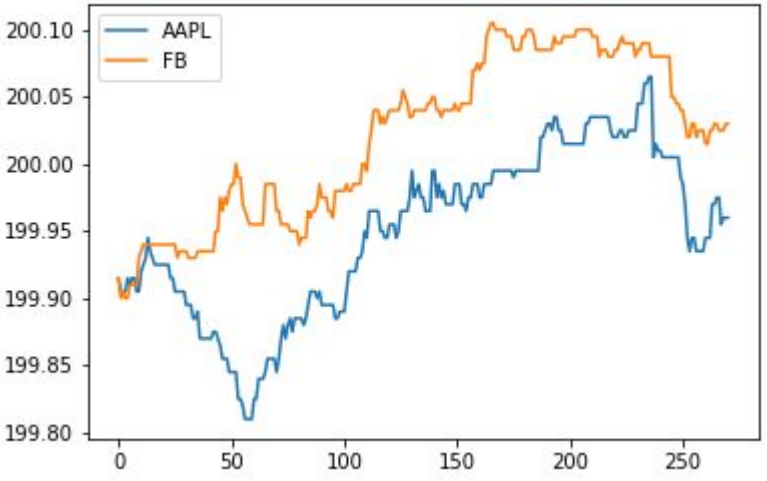
Cointegration of Pairs and Clusters: Discussion

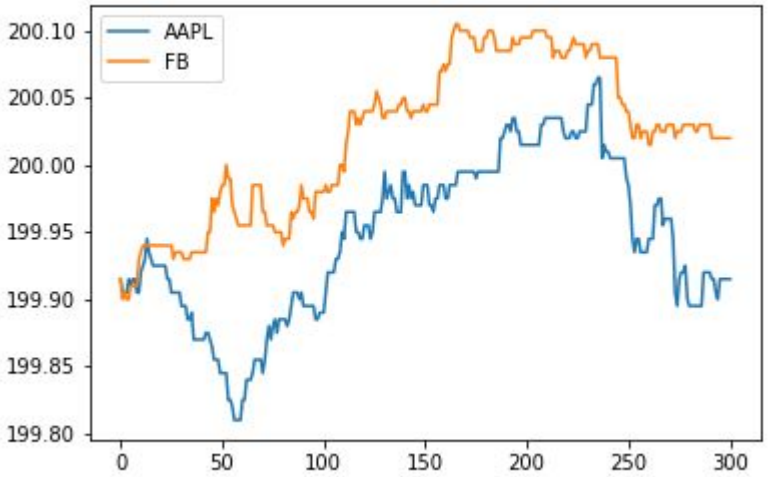
- Highly dependent on k-means clustering to produce good results
  - All clusters returned by k-means are highly correlated
- Increasingly difficult to determine cointegration with larger clusters
  - More computationally expensive (matrix inverse)
  - Lower accuracy due to more inaccurate critical value approximation (Mackinnon et al. 1999, Onatski et al. 2018)
- Future steps: develop a trading strategy using clusters rather than pairs



Running Simulations on Cointegrated Clusters

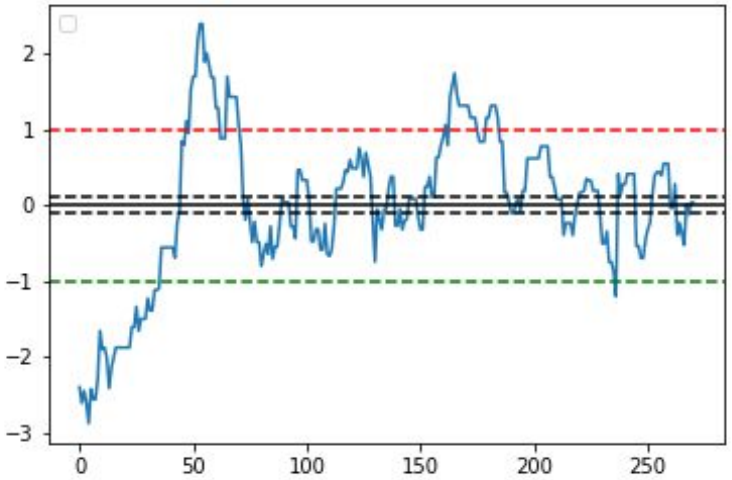
- Used Thesys for Simulations
- Used data from 04/12/2019 from 12:00-12:05 pm and 1s intervals

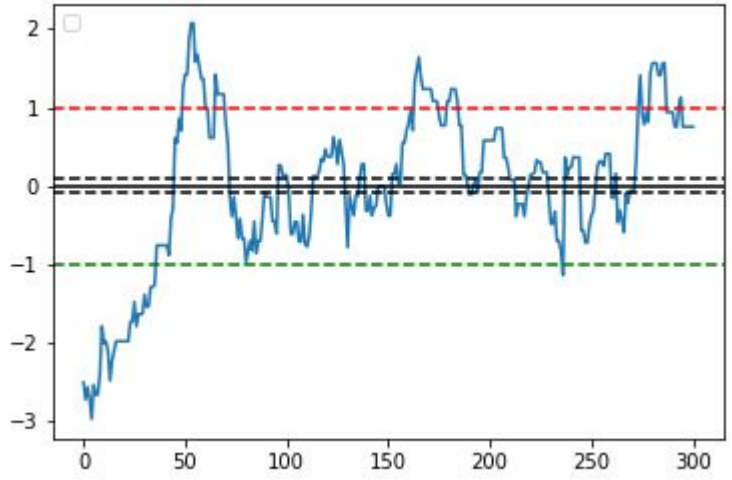




Running Simulations on Cointegrated Clusters

- Linear Regression on the mid prices of the stocks
- Calculated the running average and running standard deviation

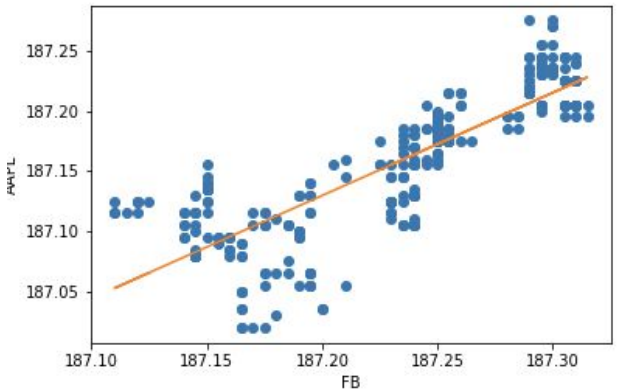




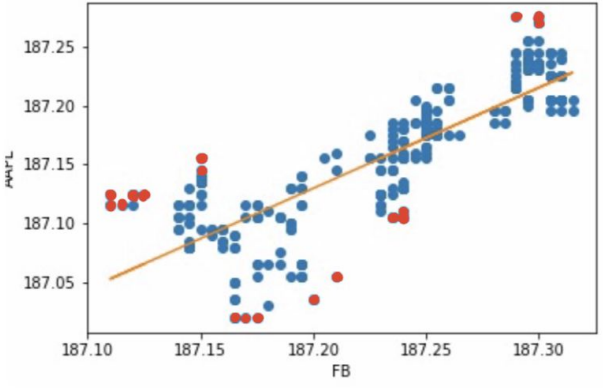
Future Steps: Modeling Residuals

- Modeling residuals beyond linear regression using midprices
  - Adding variables to regression model (e.g. bid, ask, volume, lags of midprices)
    - Autocorrelation and Partial Autocorrelation Functions
  - Classification Methods

Linear Regression

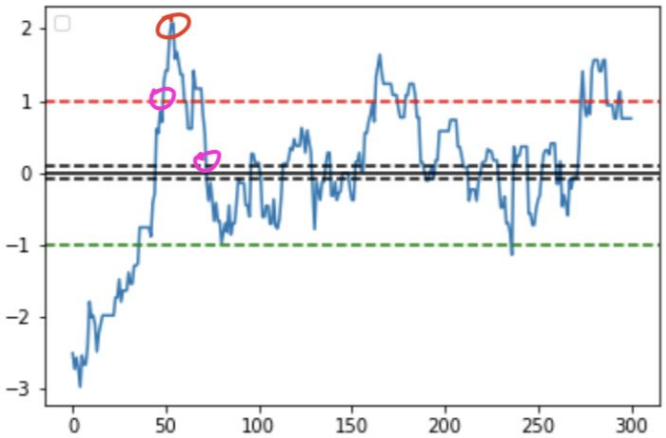


Classification Method Idea



Future Steps: Optimal Band Selection

- Stochastic Differential Equations in order to optimize: [1]
  - Optimal Band Selection
  - Optimal Entry and Exit Strategy



Can be thought as Maximizing a value/utility Function

Maximization for exiting a long position:

$$H_{+}^{(\tau)}(t,\epsilon) = \mathbb{E}_{t,\epsilon}[e^{-\rho(\tau-t)}(\epsilon_\tau - c)]$$

Maximization for entering a long position

$$G_{+}^{(\tau)}(t,\epsilon) = \mathbb{E}_{t,\epsilon}[e^{-\rho(\tau-t)}(H_{+}(t,\epsilon_\tau) - \epsilon_\tau - c)]$$

Other Steps and Summary

Our steps:

1. Optimization of company selection
2. Cointegration of pairs & clusters
3. Modeling residuals
4. Optimal band selection
5. Backtesting and executing trades

Questions?

References

[1] Cartea Alvaro, Jaimungal Sebastian, Penalva José(2015). Algorithmic And High-Frequency Trading.

[2] Almgren Robert, Chriss Neil(1999). Optimal Execution of Portfolio Transactions.

[3] Elliott, Robert & van der Hoek, John & P. Malcolm, William. (2005). Pairs Trading. Quantitative Finance.