Thibaud Bruyelle
Brian Lui
Tom Morvan
Julius Sterner

High Frequency Trading

Midterm Presentation

MS&E 448

Tuesday, 5th of May

Background

- High Frequency Trading
  - Large number of trade orders in a short time interval
  - Requires most granular type of data (order level)

- Constraints
  - Algos must be faster than the prediction time-scale
  - Slippage

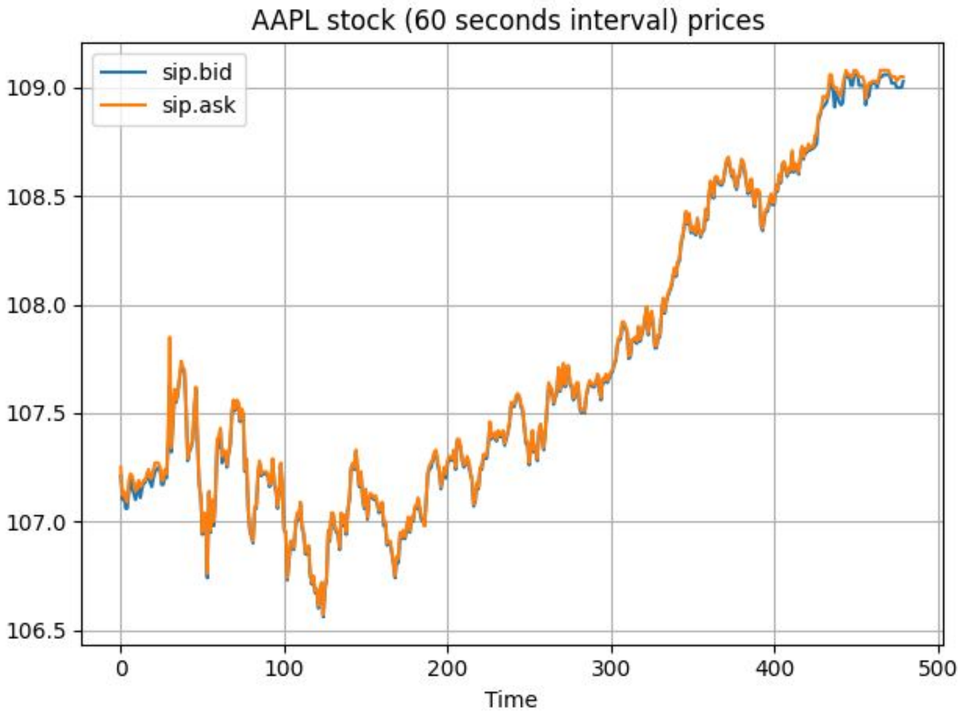
➡ MayStreet Simulator

→ Complete Market Replay (order by order)

→ Order execution incorporating slippage

Data Exploration & Statistical Models

- MayStreet provides us high-frequency data (order book events)
- Particularly, it gives access to bid/ask prices with very small time interval.



Data Exploration & Statistical Models

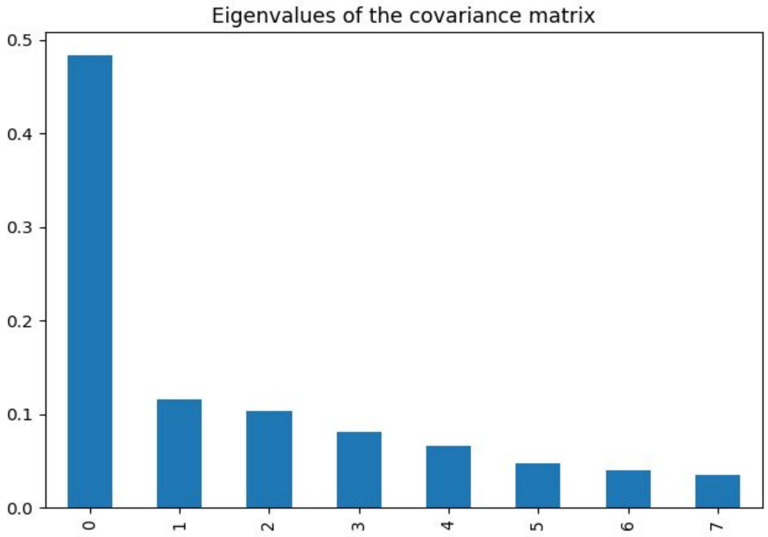
- ***How can we assess such high-frequency data ?***
- Simple/straightforward statistical tools in order to efficiently proceed with high-frequency data.
- Ex : estimators, linear regression, statistical tests (ADF test, Cointegration test), traditional clustering methods (k-means, NMF, affinity propagation)
- Illustration with PCA.
- PCA is visualization tool very useful to assess high-dimensional data. What we can do with PCA :
    - Time Series analysis
    - Build a portfolio
    - Make stocks associations (e.g. pairs selection in Pairs Trading Strategy)
    - Hedging strategies (e.g. using Principal Components as proxy)

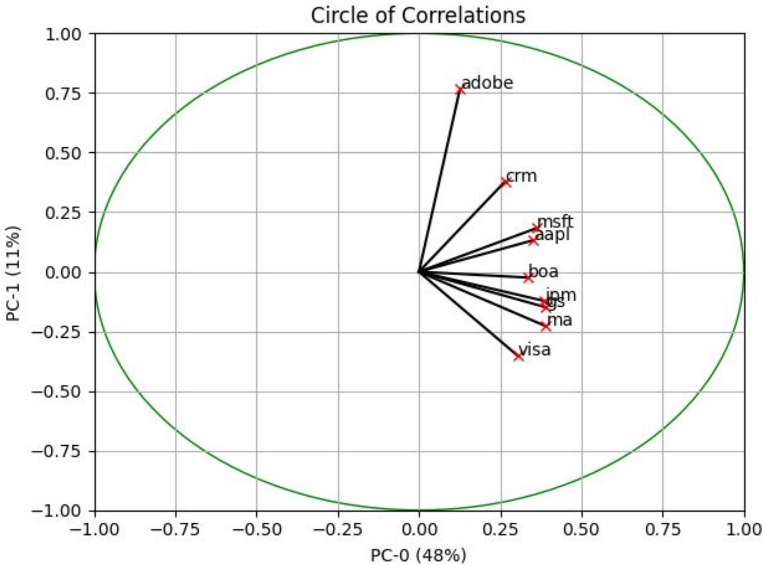
Data Exploration & Statistical Models

- Portfolio : AAPL, ADOBE, BOA, CRM, GS, JPM, MA, MSFT, VISA.
- Mid prices X_t over one day with 60s time interval.
- We show that : Returns R_t are stationary (ADF test)
- Augmented Dickey Fuller test :
  - Assumes that : X_t = rho * X_t-1 + u_t
  - Test for the presence of a unit root, *ie*, H_0 : rho =1
  - For the returns series, we can always reject H_0 at all confidence levels

Data Exploration & Statistical Models

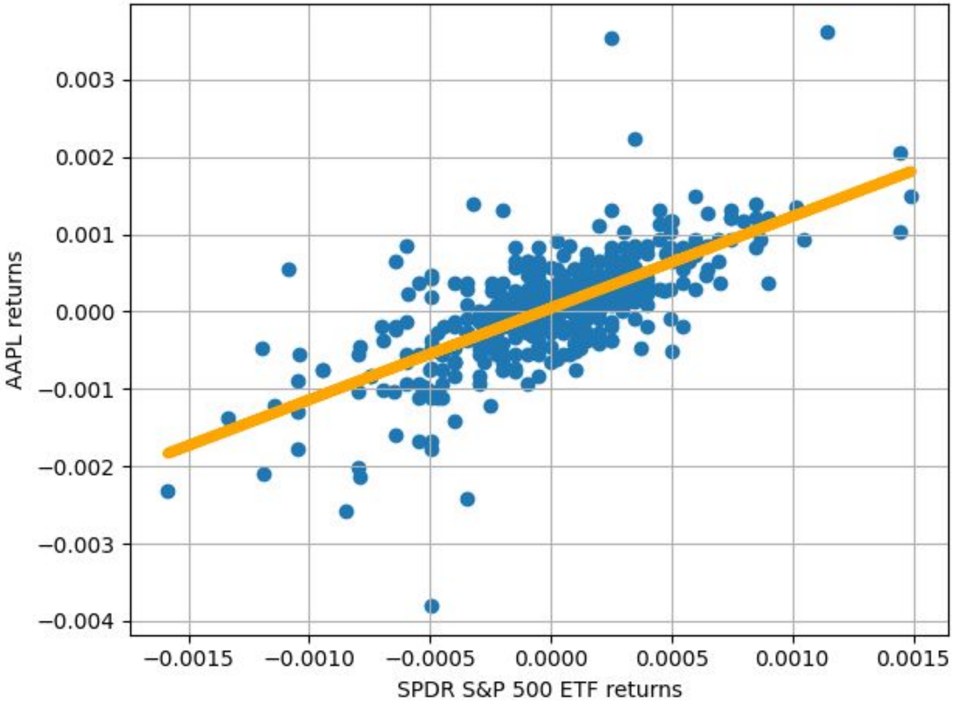
- PCA applied to the returns (R_t) of the portfolio.
- It does not make sense to apply PCA to the price series.
- Indeed, crucial assumption in PCA : observations must be iid





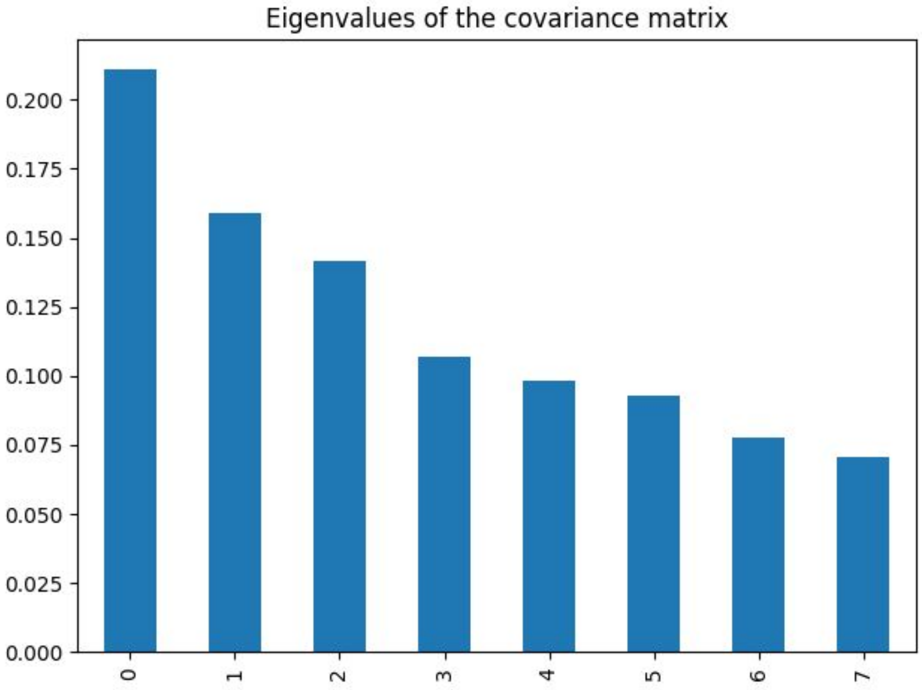
Data Exploration & Statistical Models

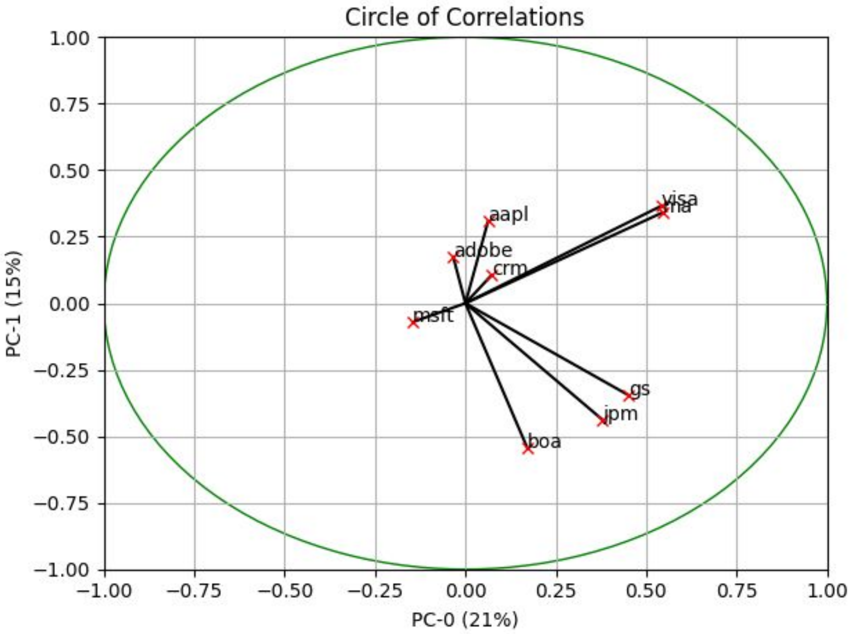
- All stocks strongly correlated to PC0 => market influence
- **To get rid of the market influence :** R_t = a * SPY_returns + r_t



Data Exploration & Statistical Models

- **PCA on residuals (r_t)** : results more useful and interesting to establish intrinsic relationships between stocks.





Data Exploration & Statistical Models

Conclusions :

- Traditional statistical tools are an efficient way to handle high-frequency data :
    - Preprocessing / Data exploration : it helps to understand patterns within the data.
    - Investing strategy using statistical models/metrics :
        - metrics can be computed quickly and used in the investment strategies (moving average, Z-score)
        - Simple models that can help the decisions (eg. hedging strategies with OLS estimate, Principal Components in PCA)

Next step :

Applying these methods to trading strategies in high-frequency data environment.

Mean Reversion - What is it?

- Mean Reversion algorithms trade on the belief that stocks, futures, indexes and other investment vehicles trade at a mean price set by the market over different periods of time
- The flow of the strategy is in essence:
  - If the current bid price is below the existing mean, buy the stock (i.e. buy low)
  - If the current ask price is above the existing mean, sell the stock (i.e. sell high)
- Mean Reversion, like Momentum Strategies, are worth considering as a first step in HFT because they have been time-tested as viable strategies **if they are implemented well

Mean Reversion - What's been attempted

- Tick-by-tick calculation of the mean over a variable number of ticks
- The trading was naive, meaning that it
  - Bought if the bid price was below the mean and sold if the ask price was above the mean
  - Traded a fixed quantity (100 shares), independent of the bid/ask orders
  - The algorithm did not care about the solvency of the fund itself
  - Calculated the mean of the mid price
- Traded exclusively on the S&P 500
- Results:
  - None significant on a tick-by-tick level
  - As discussed with Prof. Borland, tick-by-tick mean reversion (or any other strategy at that level of granularity) cannot have meanful calculations for sharpe ratio
  - Loss (to be plotted later) of the current algorithm over a one day trading period, trading a fixed 100 shares per trade, is -$329.24

Mean Reversion - What we’ve learned

- Time windows that are too granular (i.e. tick-by-tick) are subject to too many swings in the market for this strategy to produce effective results
- As discussed with Prof. Borland, the computational overhead of running a strategy at this time window exponentially increases the probability of:
  - Execution failure due to slippage
  - Prohibitively high transaction costs such that we are unable to achieve alpha
- Slippage does not need to be modeled within Thesys

Mean Reversion - What’s being done next

- Expanding time windows from tick level to a range between 100 ms. to 30 sec.
- Grid search to find the optimum mean window for each time window
- After a signal has been identified, incorporating the bid-ask spread as a viable measure of cost
- Optimize the generated trade to reflect the order size of the bid/ask orders
- Expanding the number of tickers experimented with beyond just the S&P 500
- Incorporate a solvency cutoff for the fund

Future - More Strategies to Consider

- In discussion with Prof. Borland, she has indicated that is it both acceptable and encouraged to explore both traditional statistical methods of alpha generation as well as try our methods of alpha generation (potentially ML based). Therefore, we will be devoting some time to fitting a model to the mid price
- Momentum strategy (the other major strategy in algo-trading)