Equity Portfolio Construction using Quarterly Institutional Holdings Data

Author: Oh Chunguan Darius
Project Number: H196350
Supervisor: Associate Professor Tan Wee Kek

Table of content

Research Motivation

SEC 13 Filings

- Large institutional funds (>US$100mil) are required to disclose their equity holdings each quarter
- Data is publicly available on the Securities and Exchange Commission (SEC) website
- Provides insights to what these top funds are investing in and how their investments change over time
- Level the playing field for small retail investors which do not have access to much resources and capital

Research Objectives

Research objectives

- Address a key limitation of 13F data where forms can be filed up to 45 days after the last day of the calendar quarter
  - ➢ Propose a set of rules to identify funds with low frequency trading strategies as their holdings should not change significantly during the short time lag

- Extract meaningful features in 13F Filings for stock-picking

- Build a Machine Learning model to predict stock returns from these features

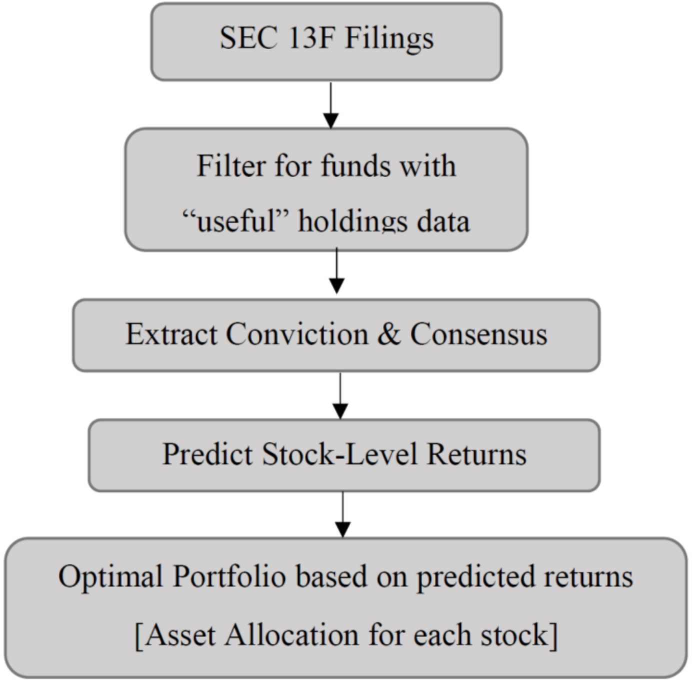
Research objectives

- Construct a portfolio of equity stocks which systematically rebalances itself every quarter and outperforms the benchmark S&P 500 Index

- Ensure that strategy is replicable by the average retail investor

- Address shortcomings in past literature and build on top of meaningful insights

Objective overview



Literature Review

INSIGHTS

1. Portfolio weights are traditionally correlated with features such as market-capitalization
2. Features combining conviction and consensus of funds led to the best performance
3. Proved relationship between fund characteristics and usefulness of 13F data
4. Introduced idea of splitting up funds into sub-funds based on the industry of the stock holdings

SHORTFALLS

1. Lookahead bias during the fund filter process

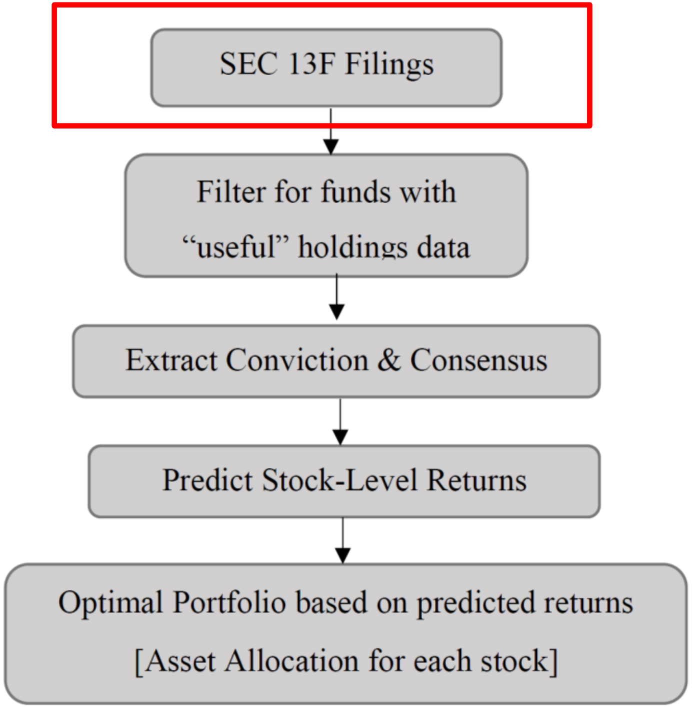
2. Unrealistic strategy backtest

3. Short time periods used for training and testing

4. Naïve approaches for portfolio construction such as equally weighing all stocks predicted to move in a positive direction

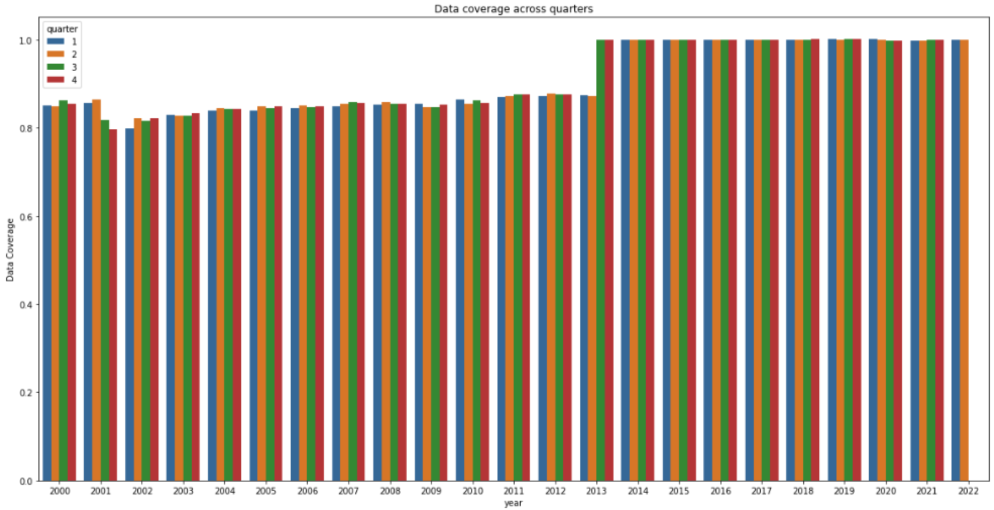
Data Methodology

OBJECTIVE OVERVIEW



Scraping of SEC 13F Filings

- External data sources require paid subscription and do not go far back in history
- Non-standardization of SEC 13F Filing format before mid-2013
- Dynamic web scraper allowed us to scrape ~80% of 13F Filings before 2013



CUSIP-Ticker information

- 13F filings identify financial instruments based on CUSIP number
- Missing other information required for the instrument
  - Listed Stock Exchange
  - Ticker symbol
  - Type of security (e.g. ETF, equity, option)
  - Industry sector of security
  - Delisted
- Obtained CUSIP-Ticker mapping through sec-api.io

Price information

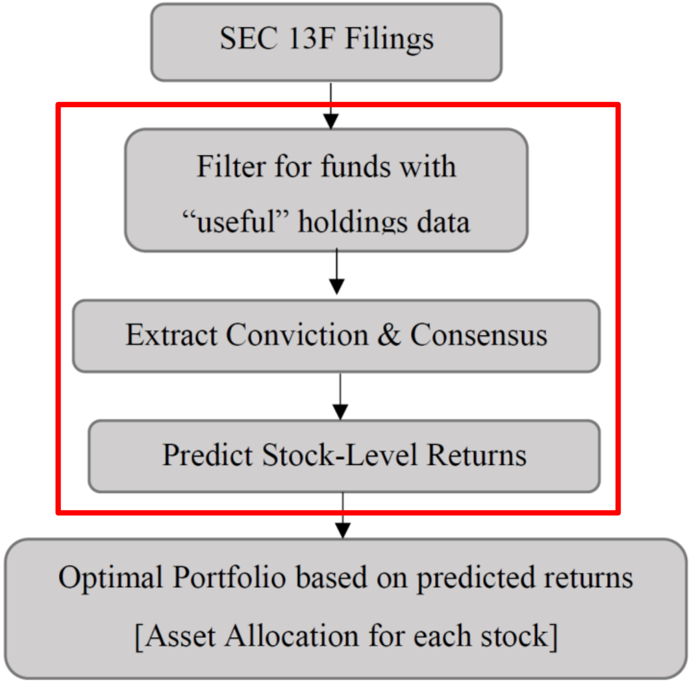
- Historical price time series of each stocks obtained via yfinance open-source python library
- Included daily, high, low, close and adjusted close price
- Does not provide prices of delisted tickers

Holdings Data format

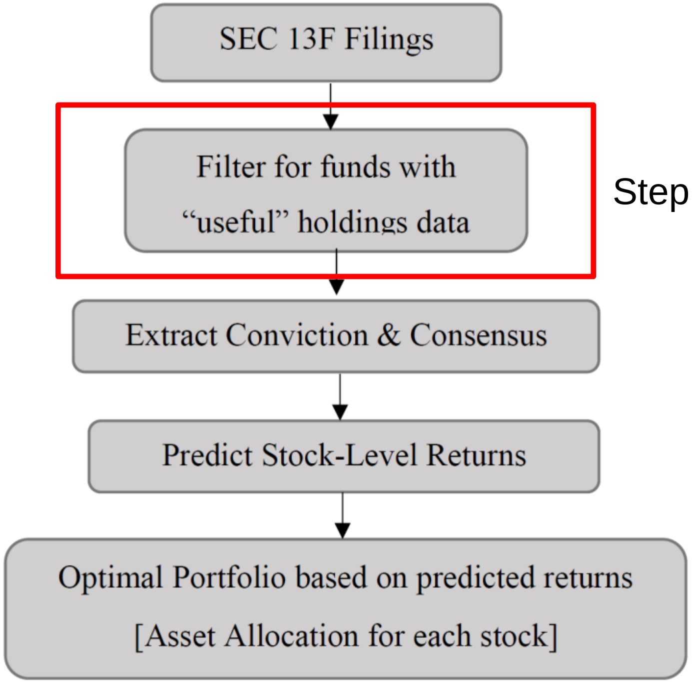
	cik	date_holding	date_filed	ticker	num_shares	buy_price	sell_price	value	quarter_return
0	886982	1999-12-31	2000-02-14	SRT	35000	28.74	58.32	1,005,835.44	1.03
1	854157	1999-12-31	2000-01-28	RCMT	246000	6.03	4.37	1,482,907.44	-0.28
2	854157	1999-12-31	2000-01-28	RF	20000	10.16	9.36	203,231.91	-0.08
3	854157	1999-12-31	2000-01-28	RHI	170000	10.38	17.24	1,765,052.67	0.66
4	854157	1999-12-31	2000-01-28	ROK	327400	11.02	9.68	3,608,587.29	-0.12

Model Methodology

Objective overview



Objective overview



Step 1: Fund filter

Fund Features

Key Assumption: Funds rebalance their positions at the end of each quarter and maintain these positions until the end of the next quarter

Fund Feature	Description
Age	Age of fund in quarters
Historical Performance	Annualized return of fund
Historical Volatility	Annualized volatility of fund
Size	Value of holdings reported in USD
Inflow	Inflow of new capital into fund
Turnover	Rate of which holdings are changed each quarter

Findings

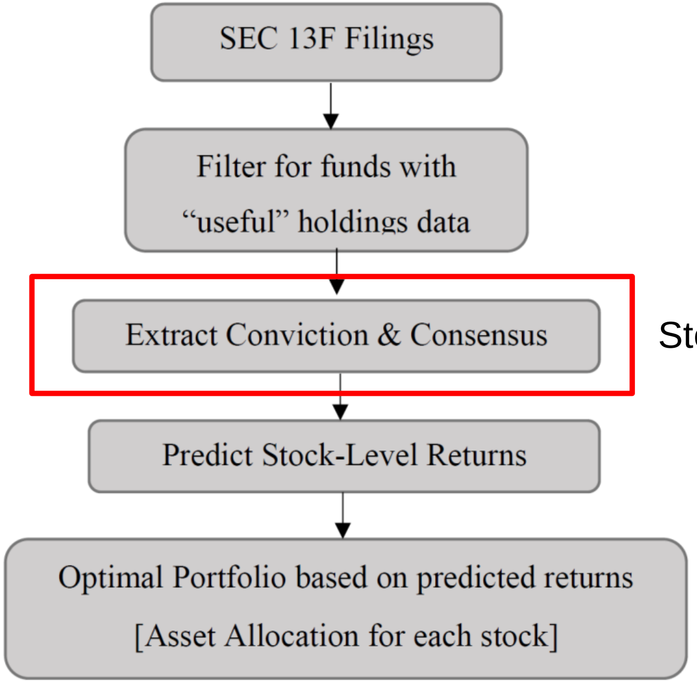
Fund Feature	Relationship with future performance	Proposed Filter
Age	No clear directional relationship	No filter
Historical Performance	Clear positive relationship	> 20th Percentile
Historical Volatility	Clear negative relationship	< 80th Percentile
Size	Clear negative relationship	< 80th Percentile
Inflow	Clear negative relationship	< 80th Percentile
Turnover	No clear directional relationship	No filter

Filter Review

- Approximately 50% of funds remained, opposed to study by Angelina et al. (2019) where funds with long-term view made up more than half of all funds

Fund	Fund Type	Included
Jane Street	Market Maker	No - Failed performance - Failed volatility - Failed size - Failed Inflow
D.E. Shaw	Quant Fund	No - Failed performance - Failed volatility - Failed size
Bridgewater	Asset Management	Yes

Objective overview



Step 2: Feature Engineering

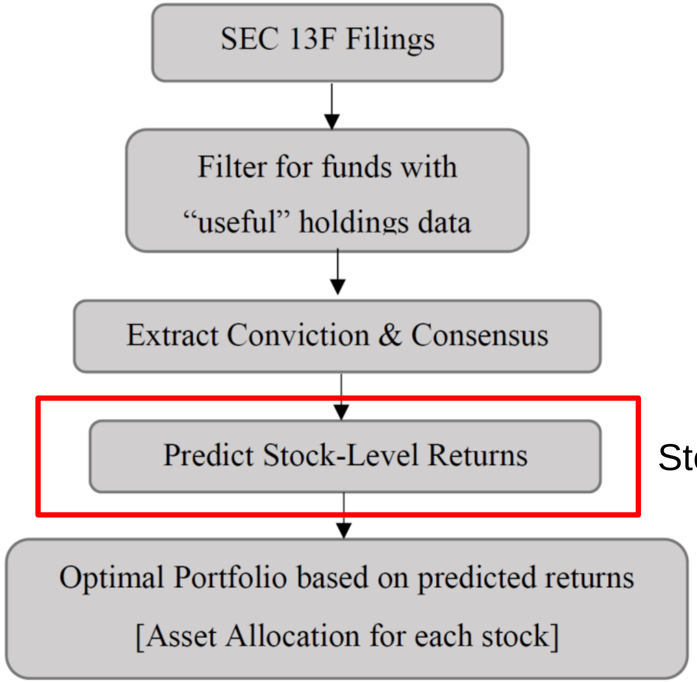
Feature engineering

Variable	Description
$x1$	Percentage of funds holding the stock
$x2$	Mean portfolio weights of funds holding the stock
$x3$	Median portfolio weights of funds holding the stock
$x4$	25th percentile portfolio weights of funds holding the stock
$x5$	75th percentile portfolio weights of funds holding the stock
$x6$	Last 21 trading day historical return of stock
$x7$	Last 42 trading day historical return of stock
$x8$	Last 63 trading day historical return of stock

Stock universe

- Unique constituents of S&P 500 Index from year 2000 up to year 2022
- Reduce survivorship bias

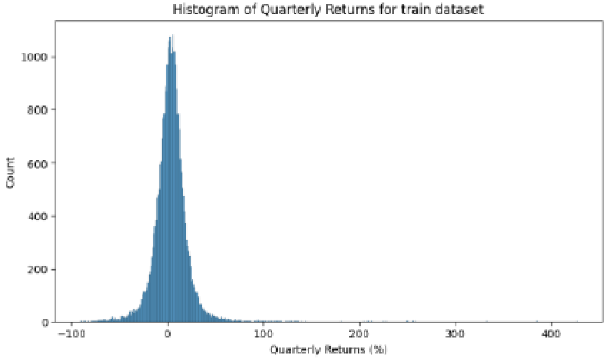
Objective overview

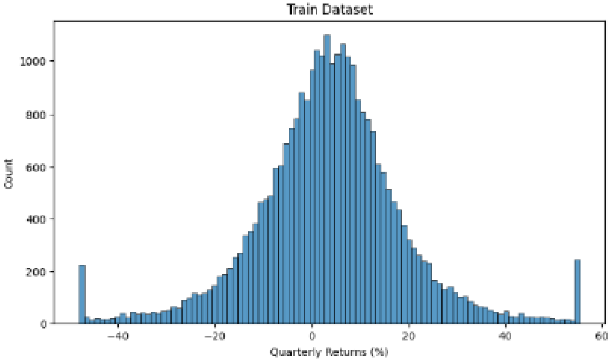


Step 3: Model Development

Outlier removal

- Winsorization by limiting stock returns of both train and test dataset to interquartile range of stock return distribution in the train dataset





Model Design

- Response variable is next quarter returns (Defined as returns from one filing date to the next)
- Train set from 2006Q1 to 2018Q4
- Test set from 2019Q1 to 2022Q2
- Evaluate results using RMSE, Spearman's Information Coefficient and Pearson's Information Coefficient
- Hyperparameter tuning via Bayesian Optimization
    - K-Fold Cross Validation during each step of tuning

MODEL RESULTS

- XGBoost is our best-performing model across all evaluation metrics

	Train Set			Test Set		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Baseline	0.2107	0.0632	0.0359	0.2659	-0.0180	-0.0118
Linear Regression	0.1530	0.1053	0.0830	0.1923	-0.1092	-0.0635
Tuned XGBoost	0.1497	0.2609	0.1734	0.1864	0.1048	0.0648
Tuned LightGBM	0.1460	0.3552	0.2659	0.1877	0.0647	0.0542

IC greater than 0.05 viewed as “good”
IC greater than 0.1 viewed as “very good”

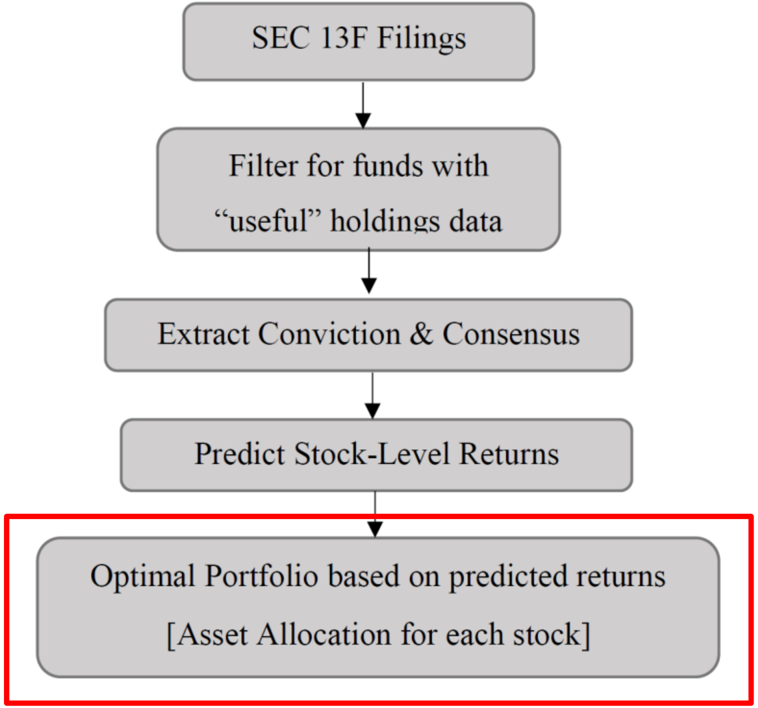
Testing Sub-fund hypothesis

- No performance improvement from dividing funds into sub-funds

	Train Set			Test Set		
	RMSE	IC (Pearson)	IC (Spearman)	RMSE	IC (Pearson)	IC (Spearman)
Tuned XGBoost (Fund)	0.1497	0.2609	0.1734	0.1864	0.1048	0.0648
Tuned XGBoost (Sub-fund)	0.1485	0.2904	0.1855	0.1873	0.0566	0.0457

Portfolio Construction

Objective overview



Benchmark portfolio

- SPY Portfolio: 100% allocation to buying and holding SPY shares which tracks S&P 500 Index
- Stock-Universe Portfolio: Equal weights to all stocks in our stock universe

Alpha portfolio

- Equally-Weighted Alpha Portfolio: Select top 200 stocks based on predicted returns and equally weigh them
- Alpha-Weighted Alpha Portfolio: Select top 200 stocks based on predicted returns and weigh them proportionately to the magnitude of their predicted returns

ALPHA VS BENCHMARK

- Alpha-Weighted Alpha Portfolio outperforms all other portfolios
  - ➢ Performance not due to stock universe definition
  - ➢ Magnitude of predicted returns value-adds to portfolio performance

	Test Period			
	Annualized Returns	Annualized Volatility	Return-Risk Ratio	Value-At-Risk 5%
SPY Portfolio	15.06%	22.50%	0.6693	-21.84%
Stock-Universe Portfolio	18.06%	26.21%	0.6890	-24.93%
Equally-Weighted Alpha Portfolio	24.63%	30.63%	0.8041	-25.60%
Alpha-Weighted Alpha Portfolio	27.54%	31.95%	0.8619	-24.86%

Alpha-Sigma Portfolio

- More complex method aims to maximize expected return for a given level of risk
- Considers volatility of constituent stocks as well as correlation between them

$$\max_w U = E(r_p) - \gamma \sigma_p^2$$

$$E(r_p) = \sum_{i=1}^{n} w_i E(r_i)$$

$$\sigma_p^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j Cov(r_i, r_j)$$

- Covariance matrix computed using rolling window of 1000 observations of realized returns

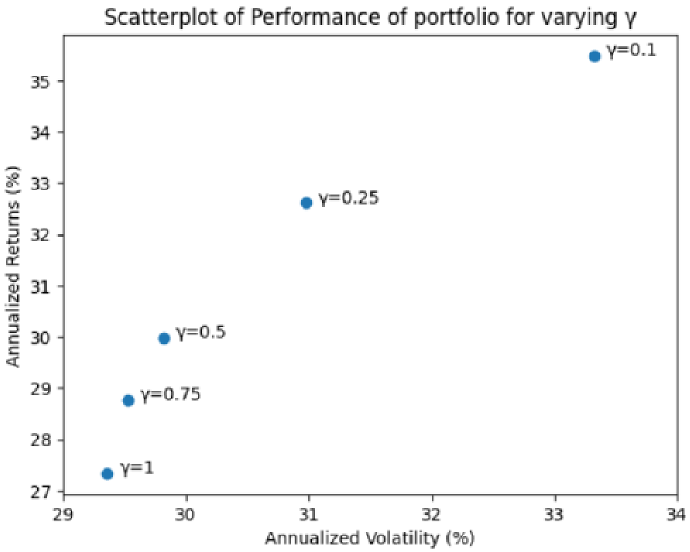
ALPHA-SIGMA VS ALPHA

- Alpha-Sigma portfolio outperformed Alpha portfolio
- Alpha-Sigma portfolio achieved twice the returns of SPY benchmark with a lower tail risk

	Test Period			
	Annualized Returns	Annualized Volatility	Return-Risk Ratio	Value-At-Risk 5%
SPY Portfolio	15.06%	22.50%	0.6693	-21.84%
Alpha-Weighted Alpha Portfolio	27.54%	31.95%	0.8619	-24.86%
Alpha-Sigma Portfolio	29.99%	29.82%	1.0058	-18.91%

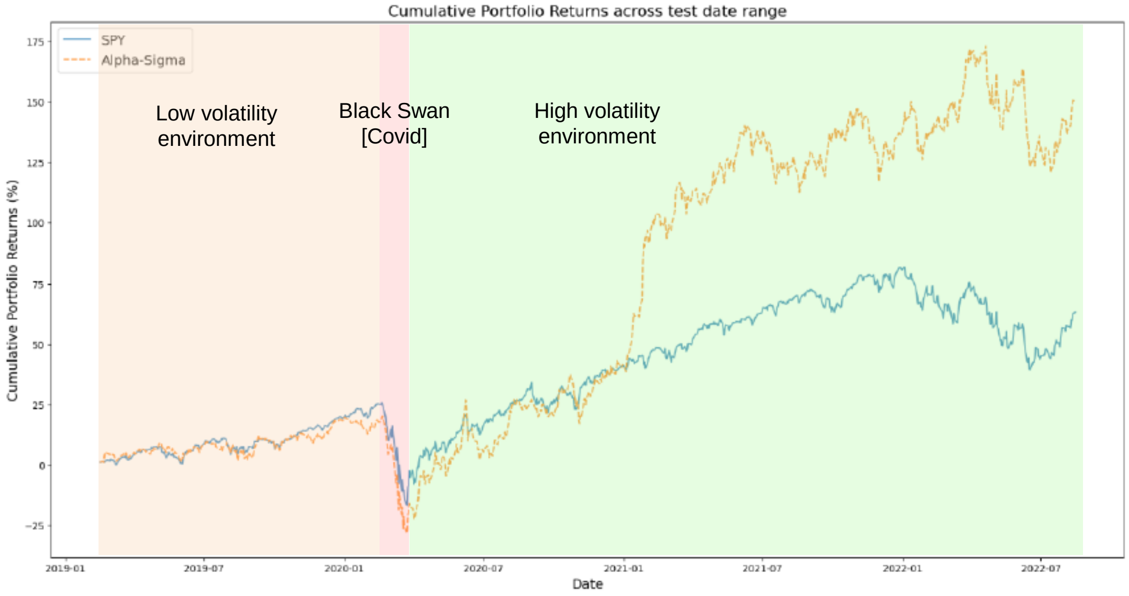
Risk appetite adjustment

- The risk aversion parameter of the Alpha-Sigma portfolio can be adjusted based on an investor's risk appetite



Time series evaluation

- We select the Alpha-Sigma portfolio with γ=0.5



Further Discussion

LIMITATIONS

- Poor quality of cost-free data sources
    - No access to historical price information of delisted tickers introduce survivorship bias
    - No access to historical market capitalization information prevented us from introducing features which control for a stock's market capitalization

- Assumed zero trading costs

- Strategy is unable to react swiftly to black swan events and underperforms in a low volatility environment

Recommendation for future works

- Perform cost-benefit analysis to explore inclusion of low-cost data sources
- Explore other objective functions for stock prediction model
- Explore a more complex covariance forecast model for portfolio optimization
- Include trading cost optimization in portfolio construction

Conclusion

CONCLUSION

- Proposed strategy significantly outperformed S&P 500 Index

- Implementation of strategy is simple and straightforward, it can be replicated by a retail investor
  - Quarterly updates for portfolio allocation
  - Cost-free data sources that are publicly accessible

- Addressed limitations of past literature by identifying and eliminating potential biases to the backtest results

- Recommend discretionary adjustments on top of our proposed systematic strategy

Thank you!