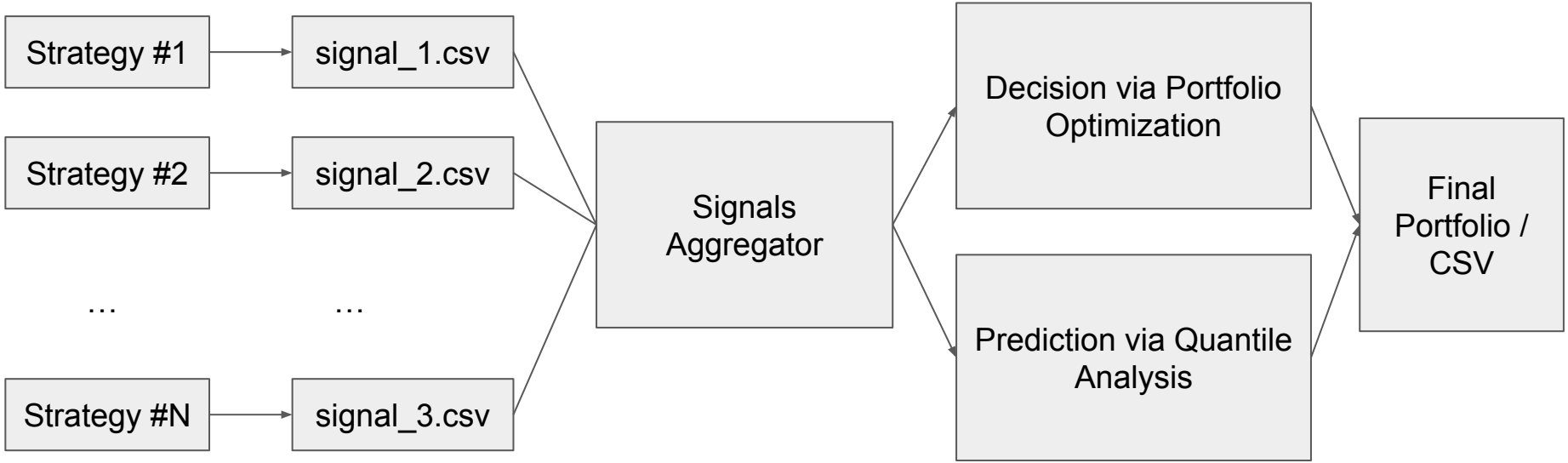
# Lions of Columbia

**Pipeline:**




Individual strategies by each members

Each .csv contains a DF with:

- ‘symbol’
- ‘pred_return’
- ‘weight’

Normalize predicted returns based on weights

# Signal Aggregation

Call_Put_Signals.csv

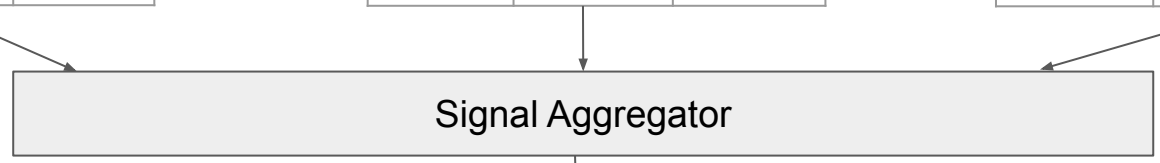
| | Prediction | Weight |
|---|---|---|
| **CARR** | 0.030638 | 0.112235 |
| **OGN** | -0.00796 | 0.065453 |
| **SHY** | 0.000305 | 0.042254 |
| **…** | … | … |

Features_Signal.csv

| | Prediction | Weight |
|---|---|---|
| **ABBV** | -0.0227 | 0.219722 |
| **ACN** | -0.06074 | 0.277259 |
| **AIZ** | 0.052159 | 0.321888 |
| **…** | … | … |

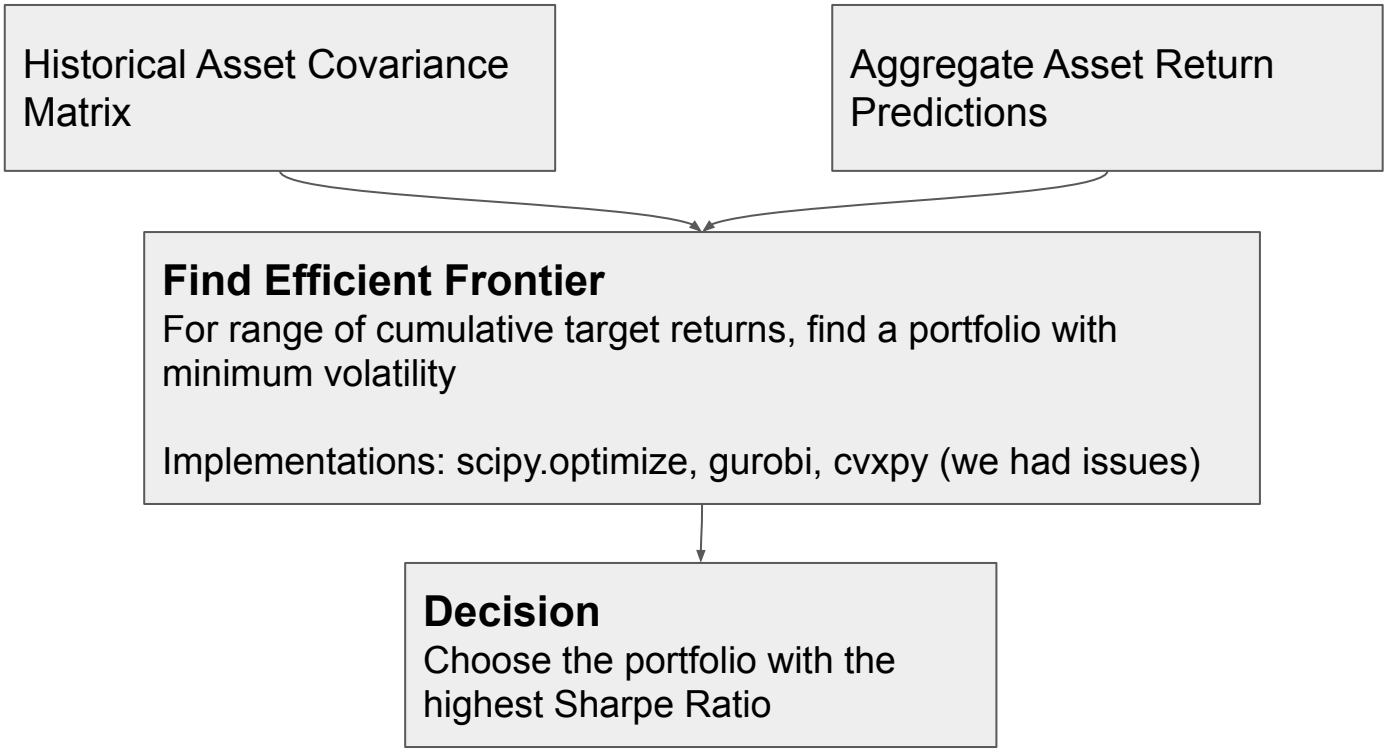
LSTM_Signals.csv

| | Prediction | Weight |
|---|---|---|
| **ABBV** | -0.04886 | 0.012812 |
| **ACN** | -0.01646 | 0.010753 |
| **AEP** | 0.007716 | 0.010066 |
| **…** | … | … |

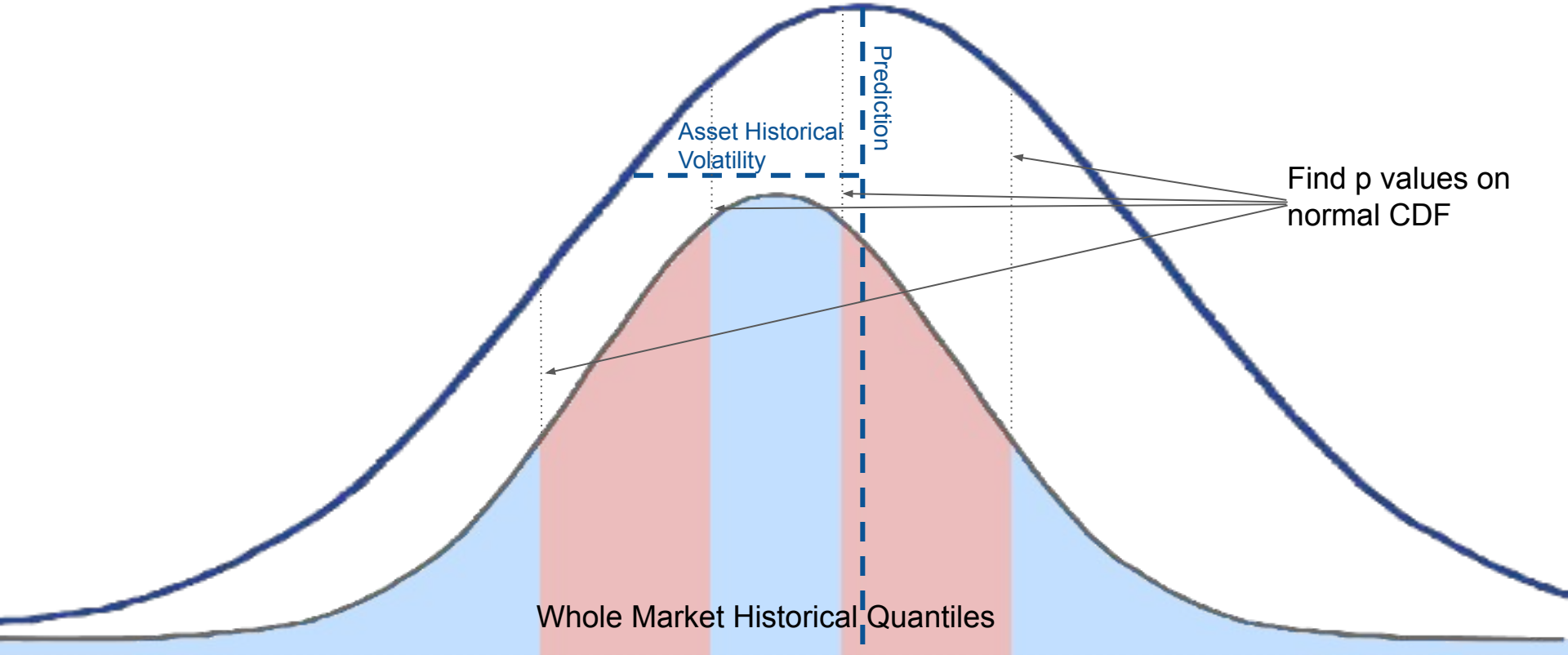



| | Prediction | Total Weight |
|---|---|---|
| **ABBV** | -0.013 | 0.344 |
| **ACN** | -0.0584 | 0.288 |
| **AIZ** | 0.052159 | 0.321888 |
| **…** | … | … |

# Decision via Portfolio Optimization



# Quantile Prediction



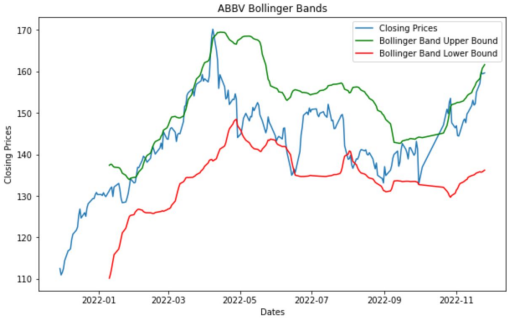
# Baseline Signal Generation: Momentum

Use historical average returns as baseline predictor of future returns

- 1 year average returns with weight 0.05
- 3 month average returns with weight 0.1

Other momentum factors (poor backtesting results - no longer considered)

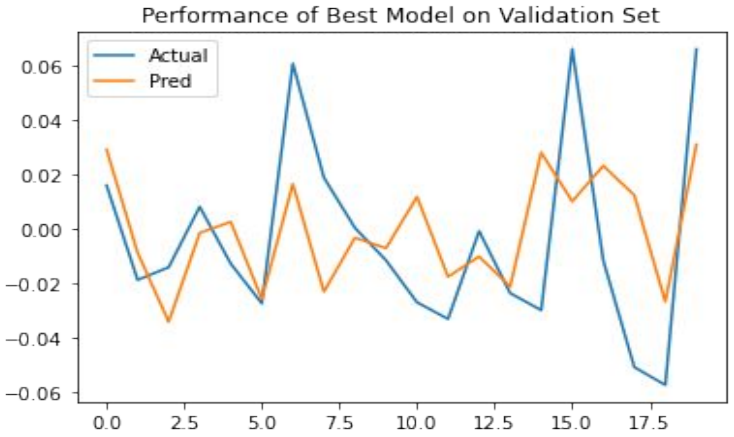
- Bollinger bands
- Golden/death cross
- Relative strength indicator




# LSTM Signals: Multivariate Time Series Forecasts

- Deploy all assets in the STOCK and ETF categories as a factor except for ones subjected to data shortage
- Predict the next week's cumulative return of an asset using the past 20 daily returns of the all factors
- **Models in Considerations:**
  LSTM, Stacked LSTM, Bi-directional LSTM, CNN-LSTM
- Find the best model using S&P500 ETF and conduct hyperparameter tuning for the selected model
- Make the prediction for every asset and assign different weight based on their MAE values for the aggregation

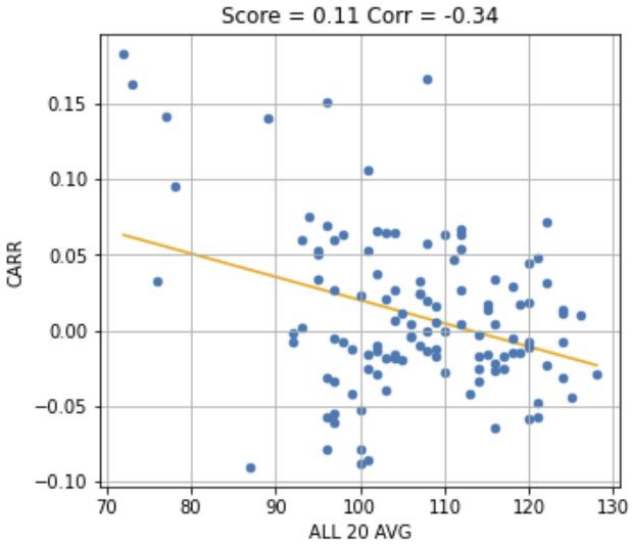
| Model | Mean Absolute Error (Validation Set) |
|---|---|
| **Stacked CNN-LSTM** | 0.031 |
| **Bi-directional LSTM** | 0.033 |
| **Vanila_LSTM** | 0.034 |
| **Stacked LSTM** | 0.035 |



# Sentiment Signals: Historical ISEE Index

- Chart historical put-call ratios (Securities, 10-day average, 20-day average, …) against next week's asset returns
- Do a linear regression on each asset with each possible series
- Choose the regression with the highest score (p-hacking)
- Apply regression coefficient to latest ISEE value to get asset prediction, using score (r-value/correlation coefficient) as weight

| DATE | ALL SECURITIES | ALL EQUITIES | EQ 10 AVG | EQ 20 AVG | ALL ETFS | ETF 10 AVG | ETF 20 AVG |
|---|---|---|---|---|---|---|---|
| **11/1/2006** | 121 | 156 | 175 | 174 | 49 | 55 | 52 |
| **11/2/2006** | 132 | 193 | 177 | 176 | 37 | 49 | 52 |
| **11/3/2006** | 118 | 145 | 173 | 176 | 51 | 49 | 53 |
| **11/6/2006** | 162 | 235 | 180 | 177 | 34 | 48 | 54 |
| **11/7/2006** | 165 | 238 | 187 | 179 | 40 | 47 | 54 |



# Surprise Signals: Historical Surprise Returns

| Date | Symbol | EPS Estimate | Reported EPS | Surprise(%) | Friday + 1w | Friday + 2w | Friday + 3w | Friday + 4w |
|---|---|---|---|---|---|---|---|---|
| **10/19/2022** | PG | 1.54 | 1.57 | 0.0182 | 10/28/2022 | 11/4/2022 | 11/11/2022 | 11/18/2022 |
| **10/25/2022** | GOOG | 1.25 | 1.06 | -0.1506 | 11/4/2022 | 11/11/2022 | 11/18/2022 | 11/25/2022 |
| **10/26/2022** | META | 1.89 | 1.64 | -0.1309 | 11/4/2022 | 11/11/2022 | 11/18/2022 | 11/25/2022 |
| **10/27/2022** | AMZN | 0.22 | 0.28 | 0.3023 | 11/4/2022 | 11/11/2022 | 11/18/2022 | 11/25/2022 |
| **11/2/2022** | VRSK | 1.47 | 1.46 | -0.0068 | 11/11/2022 | 11/18/2022 | 11/25/2022 | 12/2/2022 |
| **…** | … | … | … | … | … | … | … | … |

- Chart historical surprise returns against the +1w, +2w, +3w, +4w returns of an asset
- Do a linear regression on each asset
- If an asset's latest earnings report is within the last 4 weeks, apply the corresponding regression coefficient to surprise value to get asset prediction, using score (r-value/correlation coefficient) as weight
- Result - r-values were generally pretty low here, thus low effect on prediction

# Implemented Too Late - Features Signals

| Asset | Price | # Shares | Market Cap | Earnings Per Share | Book Per Share | P/B | P/E | # Analysts |
|---|---|---|---|---|---|---|---|---|
| **ABBV** | 151.87 | 1.77E+09 | 2.69E+11 | 6.526509 | 9.043925 | 16.79249 | 23.26972 | 12 |
| **ACN** | 291.51 | 6.30E+08 | 1.84E+11 | 10.914755 | 35.084588 | 8.308777 | 26.70788 | 18 |
| **AEP** | 89.74 | 5.14E+08 | 4.61E+10 | 4.841943 | 47.246353 | 1.899406 | 18.53388 | 14 |
| **AIZ** | 125.76 | 5.28E+07 | 6.64E+09 | 25.976976 | 77.584543 | 1.620941 | 4.84121 | 6 |
| **ALLE** | 111.05 | 8.78E+07 | 9.76E+09 | 5.498333 | 9.007932 | 12.32802 | 20.19703 | 10 |

- Find P/E, P/B, # Tracking Analysts, Market Cap for each stock
- Create buy signals for 1st quintiles, sell signals for 5th quintiles, amplified if an asset is in an edge quintile for multiple features
- Use average recent positive returns for buy signals, average recent negative returns for sell signals

# Conclusion: Things We Learned

- Quality assurance is important
- Finding good predictive data is hard
- Managing risk can be just as important as maximizing returns
- Good data gathering/processing practices