Final Presentation ZibratQuant

Théo Bazille, Ziad El Assal, Ruben Illouz

Data-driven Methods for Finance
April 24th, 2023

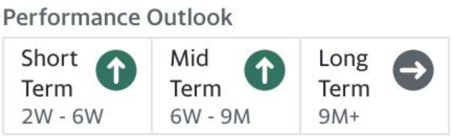
Discretionary decisions *(first weeks submissions):*

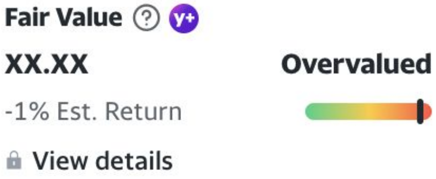
Used equal weights for 90% of the stocks.

Decisions Using : Intuition, News, Yahoo Finance

For the rest : +2%/-2% if Long/Short Decision

Fail : Lack of finance background





id	rank1	rank2	rank3	rank4	rank5	decision
ABBV	0.2	0.2	0.2	0.2	0.2	0.008380952381
ACN	0.2	0.2	0.2	0.2	0.2	0.008380952381
AEP	0.2	0.2	0.2	0.2	0.2	0.008380952381
AIZ	0.2	0.2	0.2	0.2	0.2	0.008380952381
ALLE	0.2	0.2	0.2	0.2	0.2	0.008380952381
AMAT	0.2	0.2	0.2	0.2	0.2	0.008380952381
AMP	0.2	0.2	0.2	0.2	0.2	0.008380952381
AMZN	0.1	0.15	0.2	0.25	0.3	0.02

Predictions using alternative indexes

Prediction using past performance of indexes as external features.

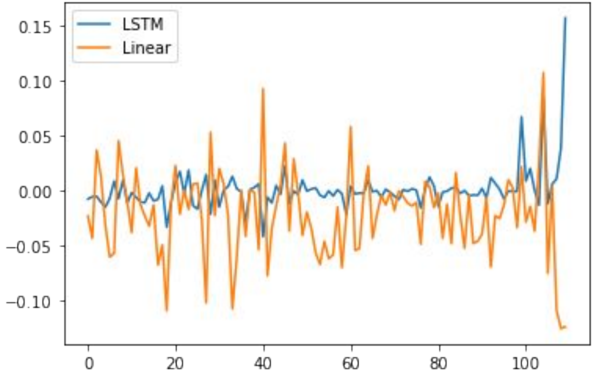
__Alternative indexes__: S&P GS, US Dollar Index, Nasdaq Composite, S&P 500, Treasury Bill Index, Dow Jones Index

For each stock : $r_{t+1} = f(r_t, \text{indexes}_t)$

Time series prediction on the last 10 months, comparing:

- Linear regression , Lasso, Ridge
- Random Forest
- LSTM : $r_{t+1} = f(r_s, \text{indexes}_s, \text{for } s>t\text{-}7)$

Lowest MSE using **Random Forest**



__Example of the comparison of one week predictions__

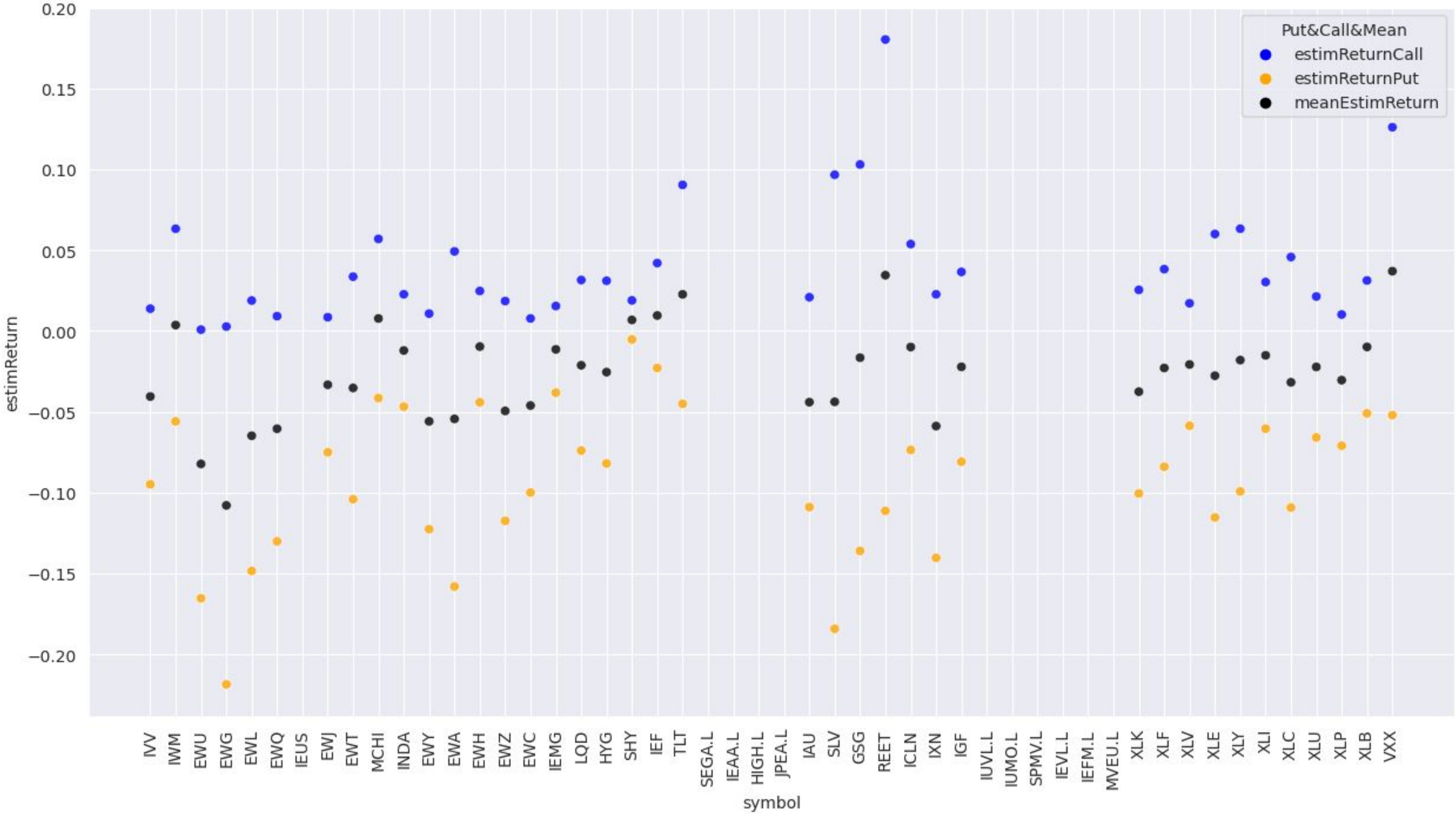
Predictions using options

Idea: if an investor is ready to pay c_i to have the right to buy a security at strike price s_i at time $t+1$, it means that they expect the price to be $s_i + c_i$.

Predicted price given by $$\frac{\sum_i (s_i + c_i) w_i}{\sum_i w_i}$$ for call options

and $$\frac{\sum_i (s_i - p_i) w_i}{\sum_i w_i}$$ for put options.

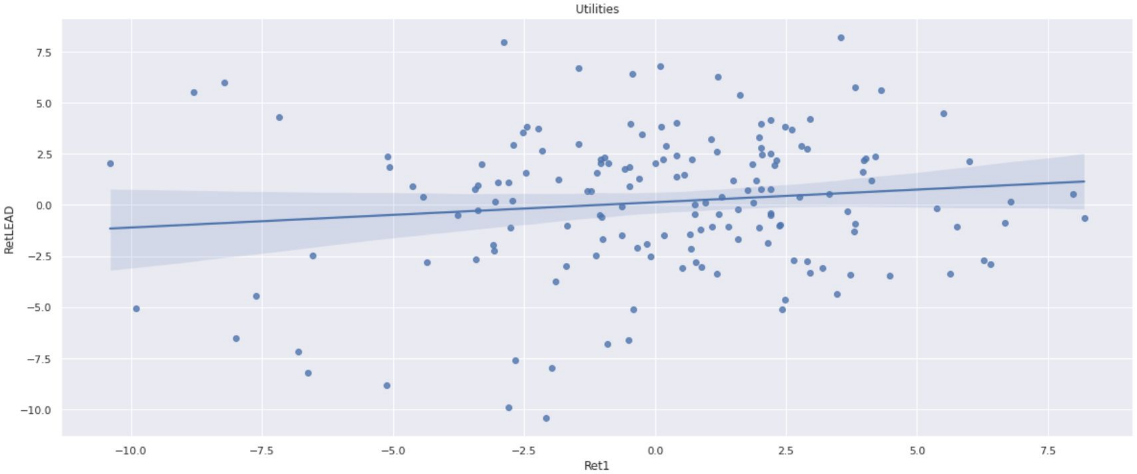
- s_i: strike price,
- p_i: put option price,
- c_i: call option price
- w_i : open interest, i.e. total number of options of type i held by investors.


Put&Call&Mean
estimReturnCall
estimReturnPut
meanEstimReturn
estimReturn
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
symbol
IVV IWM EWU EWG EWL EWQ IEUS EWJ EWT MCHI INDA EWY EWA EWH EWZ EWC IEMG LQD HYG SHY IEF TLT SEGA.L IEAA.L HIGH.L JPEA.L IAU SLV GSG REET ICLN IXN IGF IUVL.L IUMO.L SPMV.L IEVL.L IEFM.L MVEU.L XLK XLF XLV XLE XLY XLI XLC XLU XLP XLB VXX

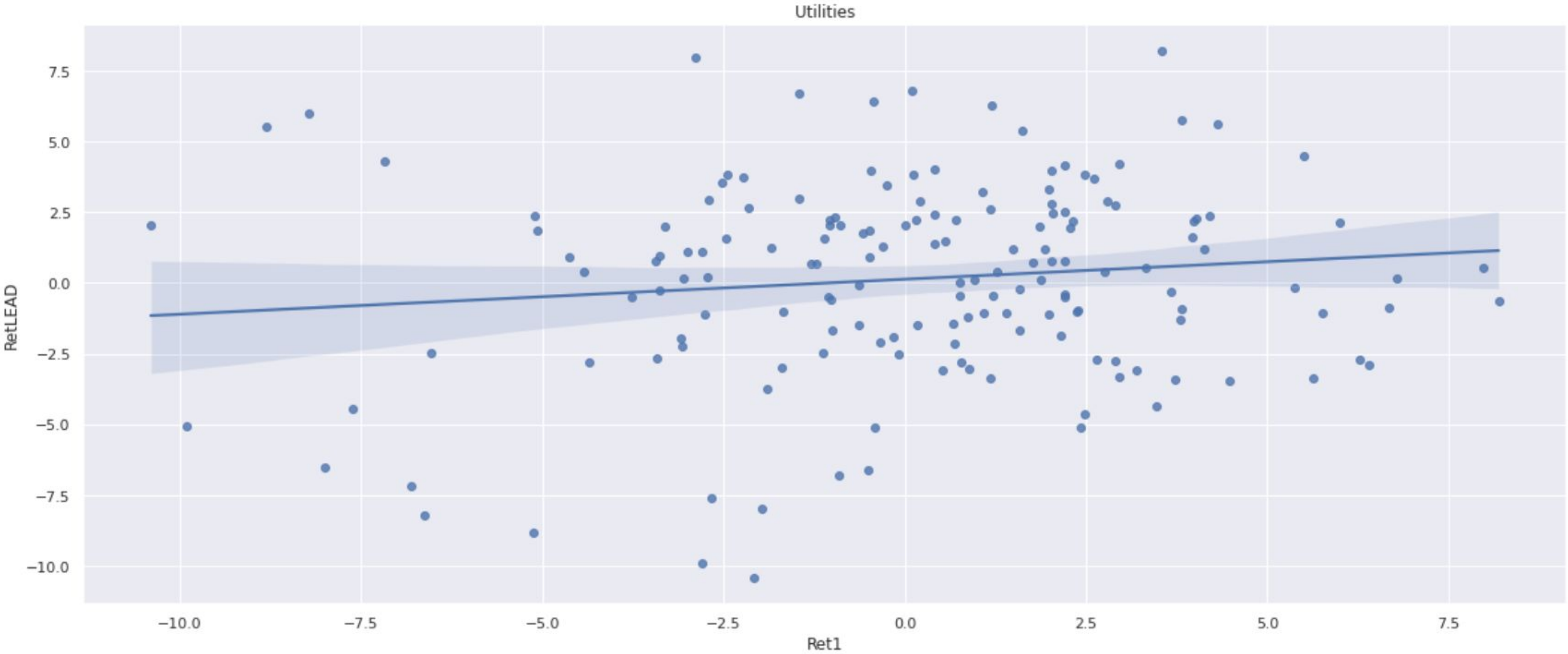
Autoregression on returns, by sector

Autoregression on returns instead of prices.

$$r_t = \alpha_1 r_{t-1} + \dots + \alpha_p r_{t-p} + \varepsilon_t$$



Some sectors have a very strong (positive or negative) autocorrelation, with a 1, 2 or 3-period lookback.



Utilities
RetLEAD
Ret1
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5

Autoregression on returns, by sector

- Sectors like Utilities, Real Estate and Energy have a strong autocorrelation.
- Mainly 1-period lookback
- Positive coefficients
  → momentum pattern
- Negative coefficients
  → mean reversion patterns

period	sector	beta	p-value	rsquared
1	Utilities	-0.227285	1.011799e-08	0.052171
1	Real Estate	-0.163057	4.712270e-05	0.026675
1	Crypto	0.127575	5.707084e-07	0.016399
1	Energy	0.122992	1.270535e-02	0.015123
1	Communication Services	-0.104760	1.059565e-02	0.010608
2	Volatility	0.102662	1.544653e-02	0.028686
1	Volatility	0.093633	1.803775e-01	0.008823
2	Energy	0.086448	1.008528e-02	0.016192
4	Utilities	-0.082969	6.729461e-04	0.018977

More systematic: ARIMA(p, 0, q)

$$r_t = \alpha_1 r_{t-1} + \dots + \alpha_p r_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Momentum

Idea: assets that have performed well in the past tend to continue to perform well in the future (Jegadeesh and Titman 1993).

Why it works:

- Behavioral biases (overconfidence, herding)
- Delayed reactions
- Positive feedback loops

Limitations:

- Momentum strategies may not work well in volatile markets
- May be affected by sudden changes in market conditions or news events

Mean reversion

Idea: prices tend to revert to their mean or average levels over time (Fama and French, 1988 over long term horizons).

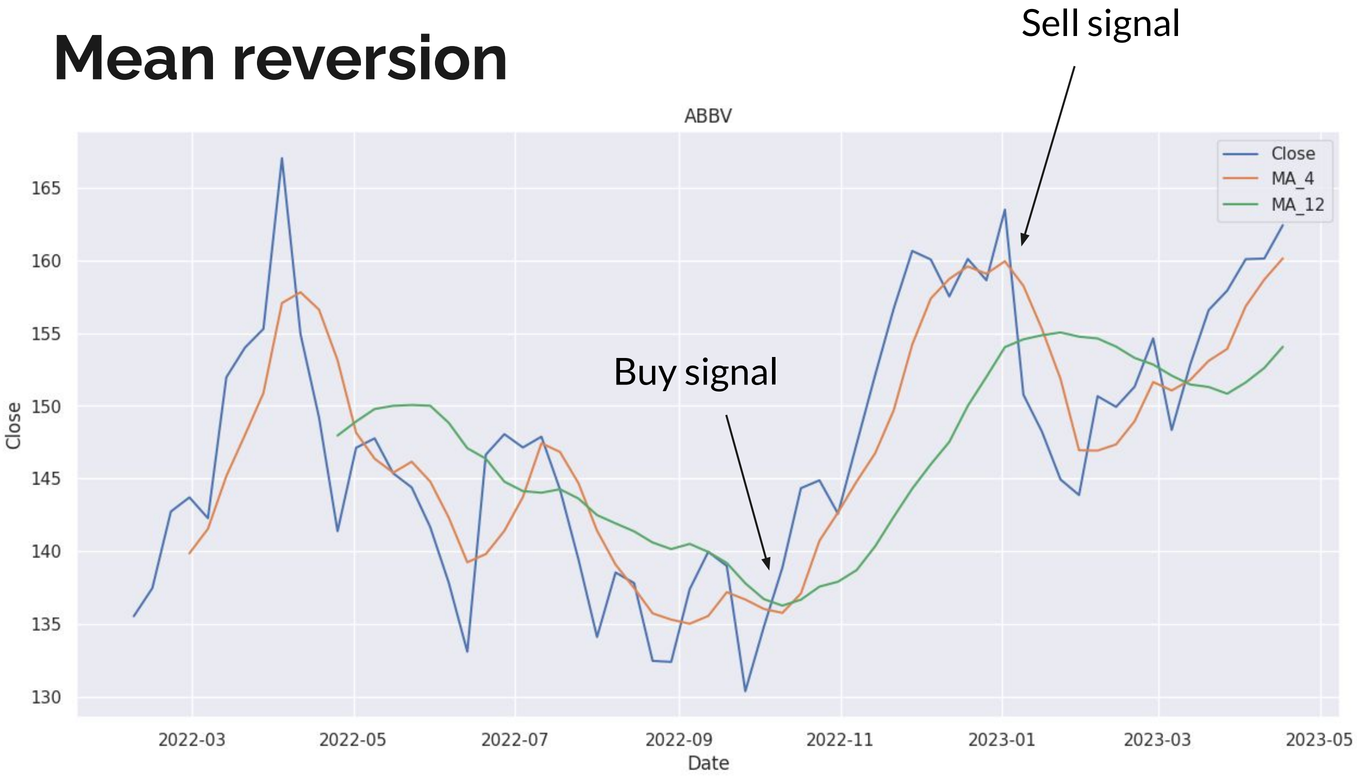
Why it works:

- Market overreaction, delayed reaction
- Value investing (buying undervalued assets increases their price)
- Risk management (avoid assets which prices are high compared to historical level)

Limitations:

- Timing (hard to know exactly when the mean will revert)

Mean reversion



Incorporated in the discretionary decision process.

But could be added as a feature in the technical analysis:

$$d_t = p_t - MA_{t,k}$$

$MA_{t,k}$ = moving average at time t with a k-week window

Regression coefficient expected to be negative

Analyst Ratings Updates

Scraped from MarketWatch.

Updates are very rare. Insightful updates (upgrades or downgrades) even more.

→ useful for discretionary investing, but hard to embed in a model.

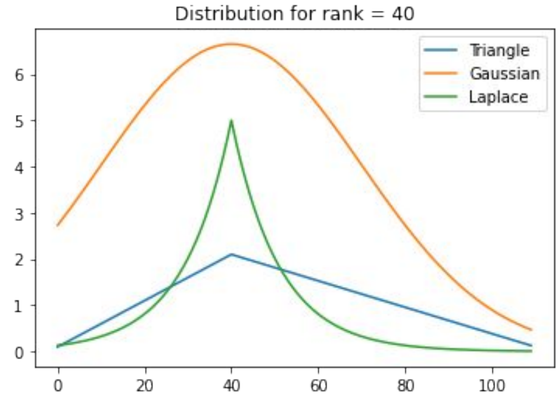
Interesting behaviour: often, the price follows the opposite direction of the update. Among the potential reasons:

- Buy the rumor, sell the news
- Higher expectations
- Profit-taking

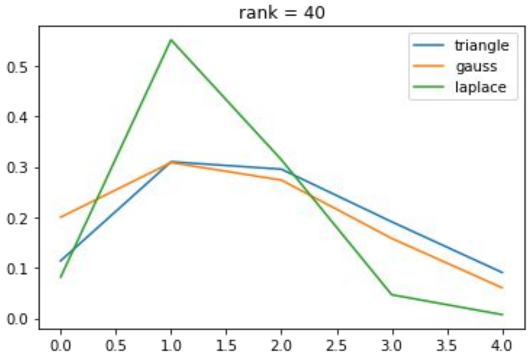
	Stock	Date	Category	Firm
0	ABBV	2023-04-05	Downgrades	Argus Research
5	ACN	2023-04-05	Upgrades	Edward Jones
25	AMAT	2023-04-06	Downgrades	Exane BNP Paribas
40	AVB	2023-04-10	Upgrades	Evercore ISI Group
85	CHTR	2023-04-04	Downgrades	Keybanc
100	COP	2023-04-05	Upgrades	Societe Generale
190	PRU	2023-04-04	Upgrades	JP Morgan
215	UNH	2023-04-05	Upgrades	Raymond James

Forecast Method *(first weeks)*

1. Method based on the **estimated performance rank** on the 110 returns: gives a distribution on the 110 ranks. **Renormalisation on the 5 quintiles.**
2. Test of three distribution families: triangle, Gaussian and Laplace.
3. Select the distribution that minimizes the RPS for **simulated estimated ranks** (following a gaussian error with variance = 70).
4. We choose gaussian method with variance = 30.
5. Didn't work well : 30 is a too low variance: needed a flatter distribution



Three distributions for estimate rank = 40



Renormalisation of these distributions

Forecast Method *(following weeks)*

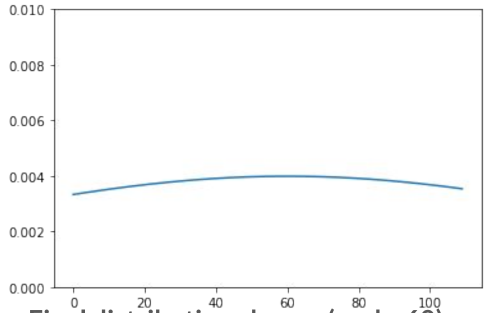
1. We noticed: forecast performance of other teams most of the time in [0.158, 0.17]

2. Our reaction: select a distribution that doesn't exceed 0.18 (no need to go under 0.15)

3. We choose a gaussian distribution with a high variance (variance = 100), closer to equal weights.

	group_name	forecast_performance
0	ew	0.16
1	quantstratsllc	0.164
2	sudeep	0.17603
3	random	0.176709
4	Lihui	0.16074
5	gambling	0.168336
6	S	0.171403
7	index	0.173255
8	QUANTify	0.204983
9	opt	0.345455
10	ZibratQuant	0.257841

Forecast performance for all of the teams week 3



Final distribution chosen (rank=60)

Optimal Portfolio :

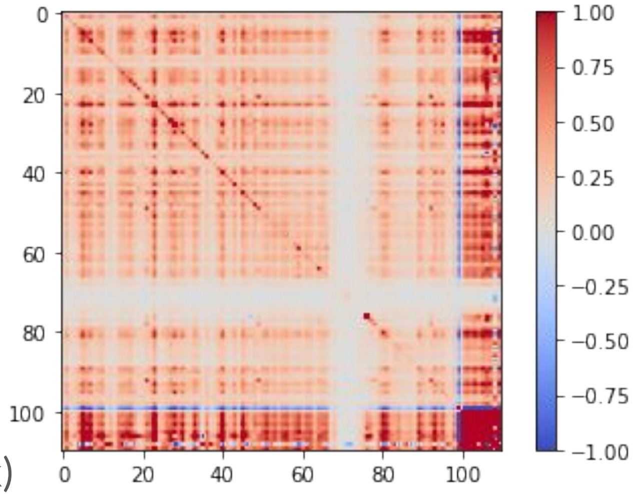
Used PyPortofolioOpt (Python library) to get an **optimal portfolio**

Needed: covariance matrix and forecasted returns

Issue: for the options based method, no data for crypto

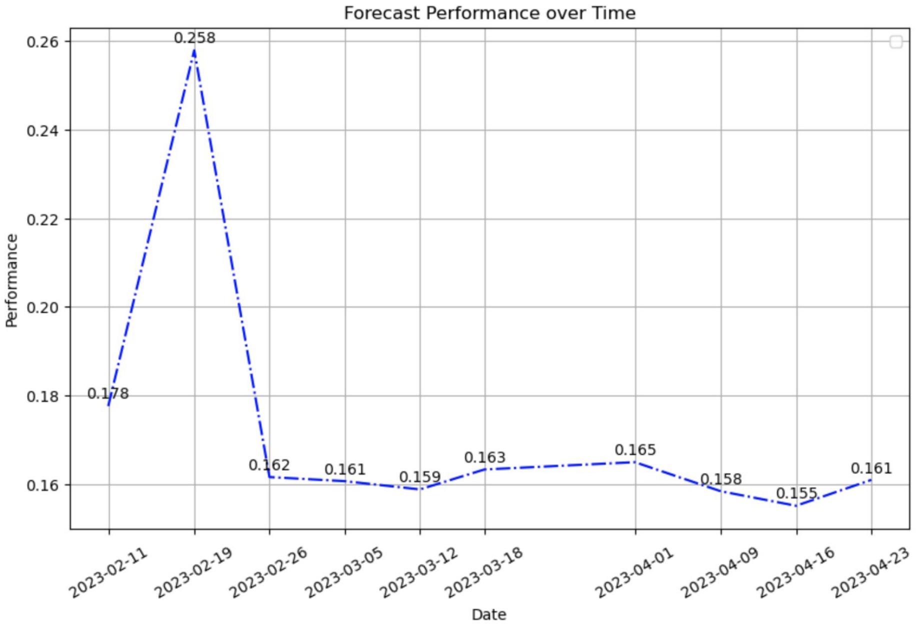
Solution : In crypto, forecast = median of the forecasted returns

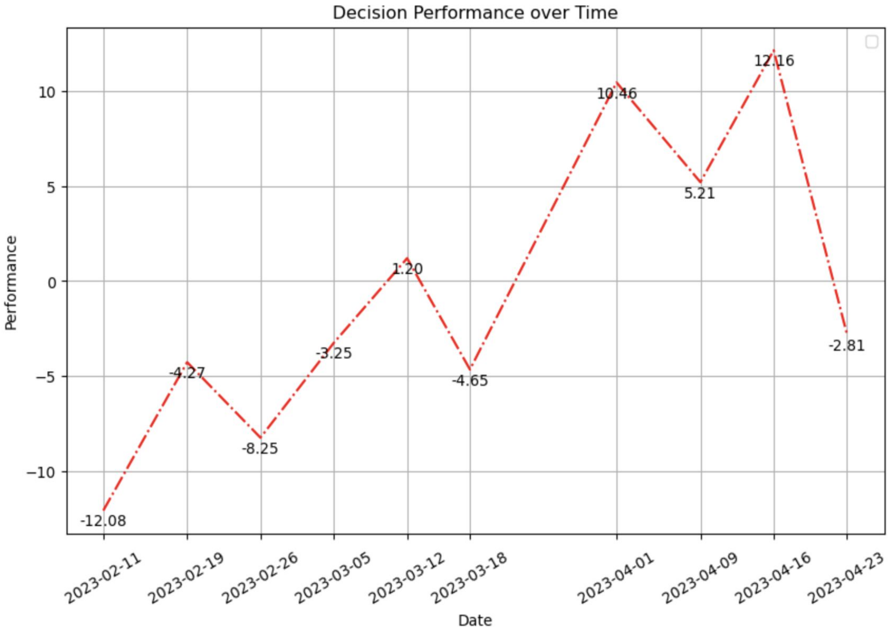
To reduce risk, we added a constraint (less than **10%** for each stock)



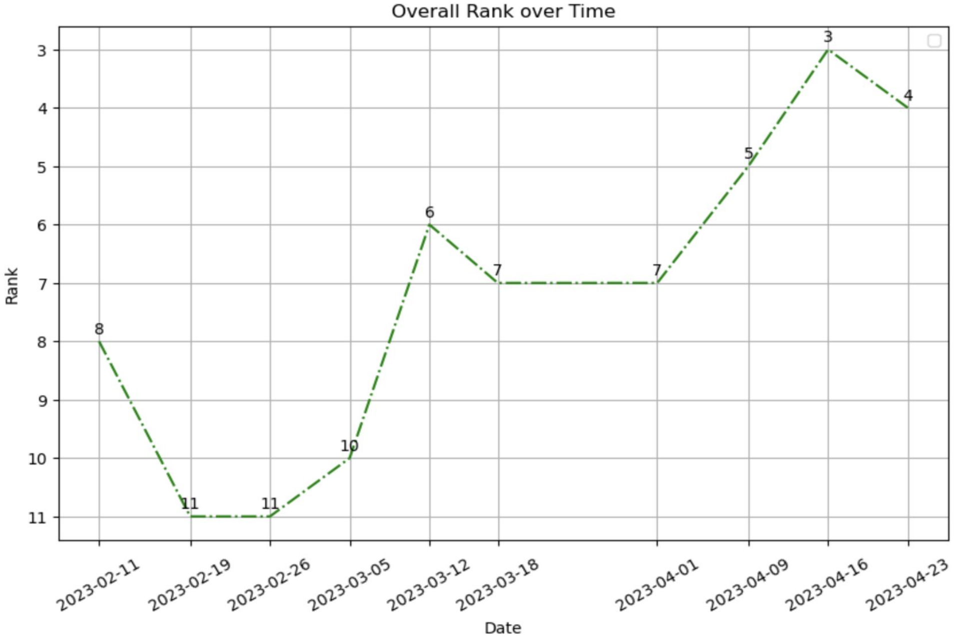
Covariance Matrix of the different stocks

Results – Overall Progress in Performance





Results – A Big Comeback ?



Key Takeaways:

- Increased finance **knowledge** and better finance **intuition**
- Implemented alternative methods to make **short-term predictions**, as factor modeling works better on the long run.
- But making accurate predictions is hard → we can try to predict the **direction** of the price
- Technical analysis is not enough (EMH)
- **Quantity and quality of the data** is important to beat the market → difficult/expensive to garner
- Importance of human psychology