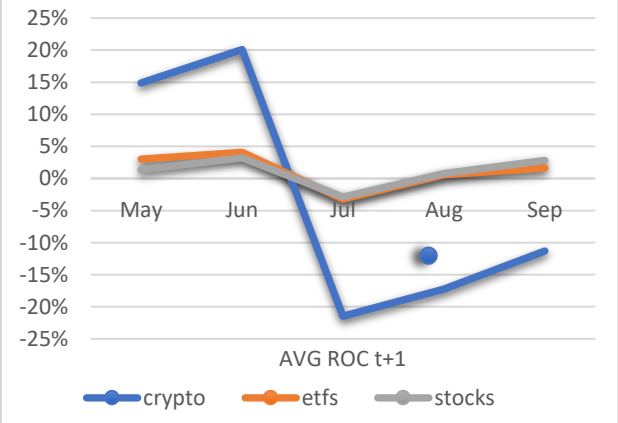
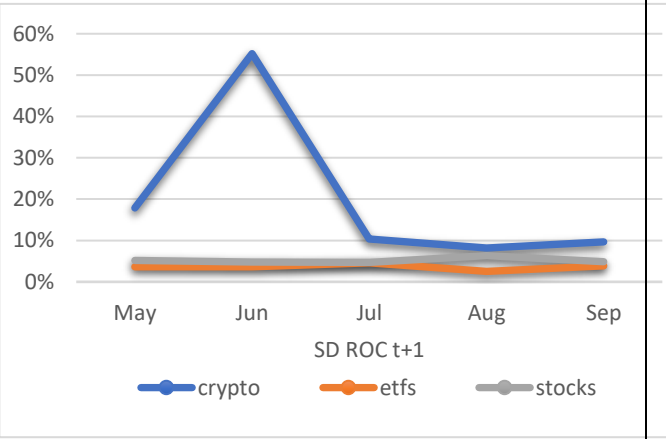
Big Data Alpha Model

1. Portfolio convention.

In this particular research we apply most common trading technics and assumptions to various types of assets. We will conduct 3 trading sessions: forming portfolio, 1st portfolio adjustment after 28 days, 2nd portfolio adjustment after 28 days. Despite some intentional simplification of the model we desire to evaluate 3 different strategies for three types of assets. So we will have 3 sessions and 3 different assets per session. Separately we select assents for a long position and for a short position

As a source we use data from yahoo finance via quantmod library from 1,5 years





Our research is focused on the following asset types:

- First one is a cryptocurrency. It is the hottest assets with high returns and high volatility. Our short list includes: BTC-USD, XRP-USD, ETH-USD, USDT-USD, LTC-USD, BNB-USD, LINK-USD, EOS-USD, XLM-USD, ADA-USD, TRX-USD. The main question more us is there any technique available that provide reliable prediction for 28 days period.

- S&P100 is our second short list. Those assets have individual characteristics, high correlation with fundamental data and lower volatility.

- And the last one - exchange-traded funds: SPYX, SPXV, YLCO, EWW, QLD, SPXN, HFXJ, ILF, RING, KBWP, IPKW, FBZ, XSD, CATH, PSI, FFTY, PSJ, KIE, EZA. This group has the lowest volatility and market driven returns.

As a result of this stage we get daily trading information for all our tickers from 18 months (from 01-01-2018 to 01-10-2019)

2. Factors identification.

Main purpose of this stage was to develop a model to aggregate different factors that could be used for further analysis. The main requirements are:

- Flexibility - be appropriate to different types of assets and different factors
- Scalability - able to scale up the model for different time periods and different number
- Simplicity - one-stage data processing that makes output feasible for further verification and analyses in Excel.

We use loop to aggregate time series object with daily trading information (from p.1) for all our tickers into one data matrix. This matrix we enrich with the following factors:

Fundamental factors from pdfetch library. Those factors will be useful to deliver middle term trend for assets price of stocks and ETF. We used the following factors: tenyTr_twoyrTR, ef.ffr , threemonthLibor, willshire5000, ecopolunct, DOW, CHINAUS, crudeoil, goldprice, coinbase.bitcoin, primar.credit.rate, Nasdqa, coinbaseethe.

Technical indicators from TTR were used for technical analysis of market trends. We used Momentum indicator RSI and Trend-following indicators MACD. Also we added such indicator as return on capital (daily) and 4 tips of moving average: MA3, MA12, EMA, SMA.

Performance measure from PerformanceAnalytics used mainly to control risk/return trade-off for all of our asset groups. We used Sharpe ratio, standard deviation and value at risk indicators.

As a result of this stage we deliver a “matrix” table with data across all tickers for the period.

+sentiments

3. Factors evaluations

From the prior stage we perform 4 different procedure based on truncated matrix (simulated available information on date): Alfa models, Beta models, Ranking factors and Measuring Performance.

Alfa model based on S&P 500

In the research we developed an alpha model for every asset. As a baseline for market performance we used S&P 500 index

Predicting beta models

We test the factors for stationary with kpss test. Based on the results we construct two models. One based on random forest and one linear model. We developed procedure for

comparison models with parameters (random forest model vs linear model, model for 30-days data, 90-days data and 365-days data)

Regarding random forest model, as anticipated, we faced with overfitting problem for the data set. So, predictive power of the model was unsatisfactory. Correction of time serials affect requires more time.

For the linear model we decided to use 90-days dataset as best trade-off between long term trends (fundamental factors) and sort term trends (market factors). Also, we increase impact of most significant factor using Bayesian information criterion optimization.

Despite some problems with endogeneity the model has satisfactory prediction power and use it as a main tool for returns prediction.

After this stage we deliver summary data frame for every session. This summary includes factors from p.2 and prediction of returns for t+1 and standard error for every linear model.

Ranking factors and assets selection.

We ranked all reasonable factors separately for each of the asset's types. The assets with the highest total rank we selected for further consideration. After this we tested some assumptions of the following assumptions:

- RSI low assumption that a stock price goes higher
- RSI high assumption that a stock price goes lower
- MACD >0 is a possible buy signal
- MACD <0 is a possible sell signal
- Alfa model – if alfa >0 – better is a possible buy signal
- Beta model – evaluation of individual stock in comparison to the unsystematic risk of the entire market (Fundamental factors from pdfetch library)
- Standard deviation – if the indicator is higher and stock is riskier
- Value at risk - a measure of the risk of loss for investments
- Sharpe ration - the greater the value of the Sharpe ratio, the more attractive the risk-adjusted return
- MA3/MA13>1 assumption that a stock price goes higher
- MA3/MA13<1 assumption that a stock price goes lower

Measuring Performance. We calculate investments returns and Shapiro ratio to evaluate our performance. Also, we compared our performance with a return on capital quartiles for every particular segment.

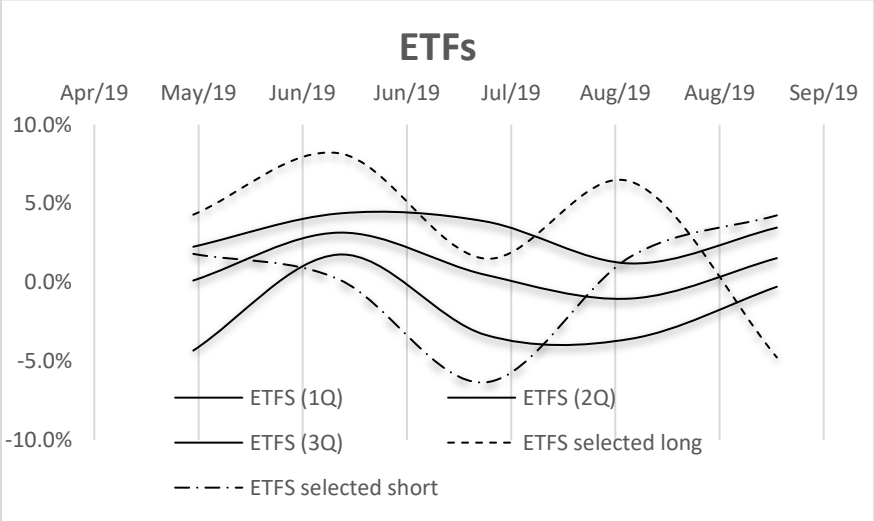
sentiments

4. Conclusions

In the research we tested 6 pools of assets ETFs(long and short), stock (long and short) and stock (long and short).

Due to different nature of the 3 groups securities in the research we apply different approaches to each of them. We review performance of our trading strategy in comparison with general market trend. For every group of assets, we draw charts with 1-3 quartiles ROC for our short list, performance of our asset for a long position and asset for a shot position. For every case we made a conclusion regarding opportunities for model enhancement.

1. ETFs



We assumed that ETFs have the high correlation with market trends, so we decided to select for long position such as: Momentum (RSI), Trend-following indicator (MACD) and beta model.

Our overall return for this asset is 102.9% for three months.

We had some short fall for ticker FFTY in SEP due to overestimation prediction from beta and we had not opportunity to recover from this loss.

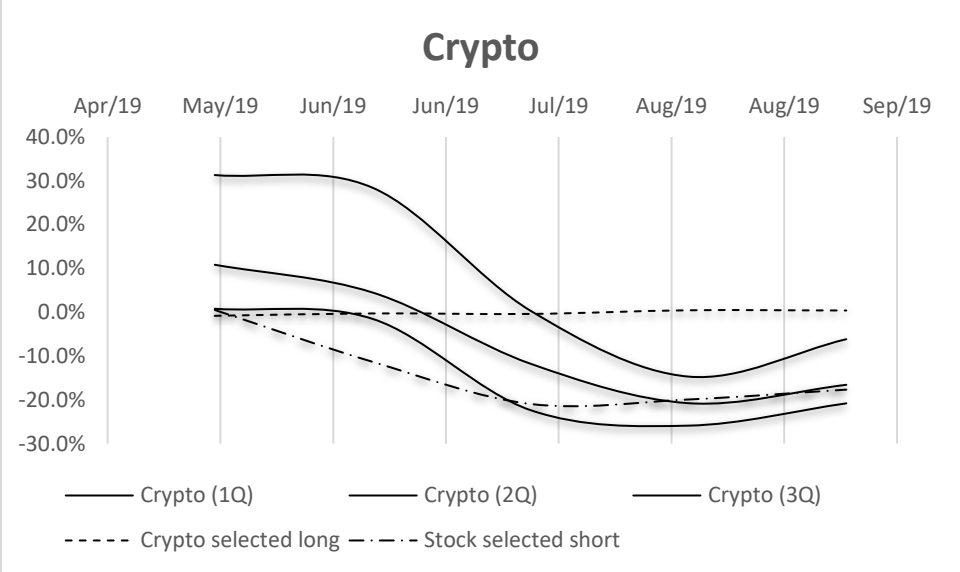
Period	Type	Ticker	ROC
September 2019	Etfs	FFTY	-4.8%
August 2019	Etfs	RING	6.4%
July 2019	Etfs	PSI	1.5%
June 2019	Etfs	QLD	8.2%
May 2019	Etfs	KBWP	4.3%

For the short position we applied different pool of factors: we decided to select under performance asset with low alpha but add to the scoring controls for risk: standard deviation and Sharpe ratio. Over result in this part was weak 100.8%, because we underestimate recovery trend for IPKW.

Period	Type	Ticker	ROC
September 2019	Etfs	IPKW	4.3%
August 2019	Etfs	IPKW	1.6%
July 2019	Etfs	IPKW	-6.3%
June 2019	Etfs	IPKW	0.2%
May 2019	Etfs	HFXJ	1.8%

Momentum (RSI), Trend-following indicator (MACD) performed well for this type od assets but we discover pure performance of our indicators in SEP session where we had ETFs market recovering.

2. Crypto



For the long position we try to be on a safe side and select Value at risk and Standard deviation for evaluate risk level, also we apply beta model. But we got a low return due to market shortfall. Our return on capital for the three months was 0.4%.

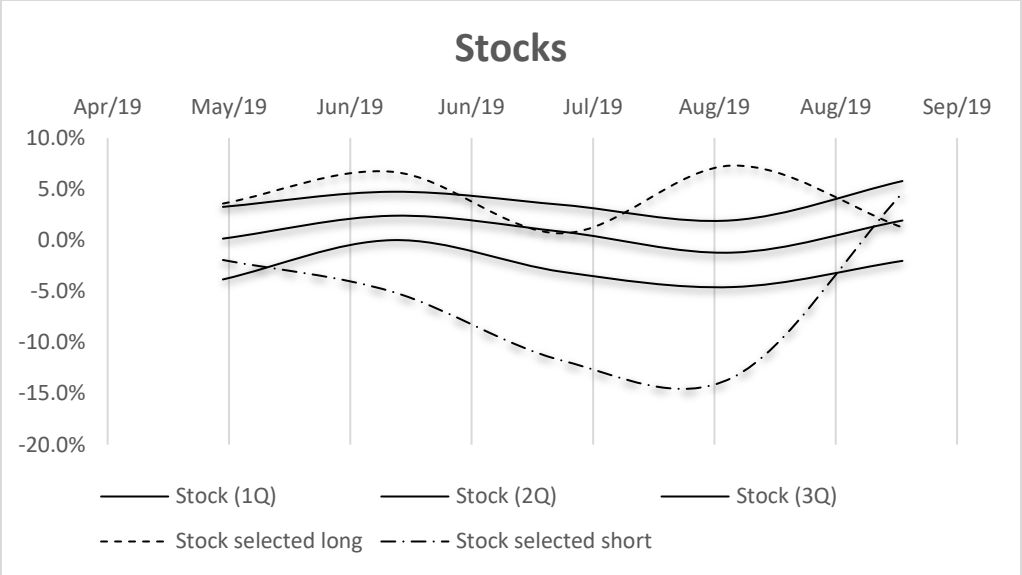
Period	Type	Ticker	ROC
September 2019	Crypto	USDT-USD	0.4%
August 2019	Crypto	USDT-USD	0.4%
July 2019	Crypto	USDT-USD	-0.4%
June 2019	Crypto	USDT-USD	-0.3%
May 2019	Crypto	BTC-USD	-0.8%

For the short position we make some modification of our scoring. We include Momentum indicator in the scoring. To control risk level, we used Sharpe ratio and value at risk. Also, we include MA3/MA13 because it indicates good performance for cryptocurrency. Due to high shortfall of the market we gain 91% for the three months. But this high performance is mainly driven by general market trend.

Period	Type	Ticker	ROC
September 2019	Crypto	ADA-USD	-17.7%
August 2019	Crypto	EOS-USD	-19.9%
July 2019	Crypto	LINK-USD	-21.0%
June 2019	Crypto	ADA-USD	-11.3%
May 2019	Crypto	USDT-USD	0.5%

All in all, this type on asset is very volatile and unpredictable, so it is not good asset for month-to-month trading. To gain some benefits we should reduce period between our sessions to have opportunity to react on short term trends. Technical indicators (essentially momentum) have the best predictive power for cryptocurrency, also it is vital to control a risk level.

3. Stocks



For the long position we used Trend-following indicator (MACD) to evaluate price reversing, Sharpe ration for risk control and Alpha model indicator to evaluate performance of the stocks compare to the market.
We had 9.4% return for the three months.

Period	Type	Ticker	ROC
September 2019	Stock	MSFT	1.2%
August 2019	Stock	NEE	7.3%
July 2019	Stock	SO	0.7%
June 2019	Stock	AMZN	6.7%
May 2019	Stock	PG	3.6%

For the sort position we used MACD Alfa model and value at risk. We change only risk control indicator compare to the long position.

Our return on capital was 25.5%. In September we underestimated NVDA performance due to low alpha indicator.

Period	Type	Ticker	ROC
September 2019	Stock	NVDA	4.6%
August 2019	Stock	KHC	-13.4%

July 2019	Stock	SLB	-11.8%
June 2019	Stock	DD	-5.1%
May 2019	Stock	GE	-1.9%

As a general conclusion, we tested 6 different cases in this research with the most common technical analysis tools. For every case described above factors predictive power changes very quickly and despite we include this changes in our beta model, our scoring procedure have a weakness in signals significance revaluation. We found out few reliable signals for July and August but in September the signal became exhausted.

The best performance of classical technical analysis tools demonstrated for stocks. Despite mediocre return, our model has a good predictive power for a short position as well as for a long position. I case of ETFs due to an economical nature of the assets Momentum and Trend-following indicator also performed well.

Cryptocurrency for middle frequency trading is not suitable. We have not found out any reliable mid-tern trend of signal. In addition, when we added control for the risk we get return near 0.