Adjusted Pair Trading Strategy with GBM to Altcoin Market

Becky Bian
Advised by Prof. Leon Tatevossian

Agenda

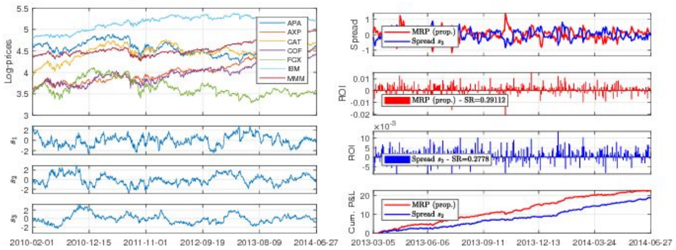
- Preliminary and Motivation
- Literature and Concept Review
- Modified Pair Trading Strategy with Multidimensional GBM
- Implementation & Results – Altcoin trading experiment
- Conclusion and Future Work

Preliminary and Motivation — About Altcoin

- Altcoins, recently introduced decentralized digital currency, were aimed to address the loopholes and weaknesses of Bitcoin application. It includes Ethereum (ETH), Litecoin (LTC), Bitcoin Cash (BCH), Tether (USDT), USD Coin (USDC) and etc.
- Overall, altcoin market has the features of extremely high volatility and great trading return. Quantitatively, compared to the return rate of S&P 500 of 26.89% in 2021, Ethereum returned 399.2% and Binance Cash even generated an astonishing amount of 1268.9% return within 2021. Suggesting the necessity of market-neutral trading strategy.

Preliminary and Motivation — About Statistical Arbitrage

- Statistical Arbitrage: A market-neutral trading strategy which is widely used to build portfolio across the Hedge Fund Industry in the Quantitative space. Currently, it has been applied widely for the establishment of trading baskets in traditional fixed income Wagner market (2005), equities and ETF market (Kang and Leung, 2017).
- Pair Trading: A popular type of statistical arbitrage and it involves matching a long position with a short position in two stocks with a high correlation based on their price spreads. Usually, the two stocks have to pass cointegration test and price spread stationarity test to prove their fitness for the mean-reverting trading strategy.



Preliminary and Motivation — About GBM

- Geometric Brown Motion: GBM is is a continuous stochastic process in which the logarithm of a random variables follows a Brownian motion/Wiener process with drift. And it is an important instance among the stochastic processes which satisfies a stochastic differential equation (SDE).

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$$

- Its application were very prevalent in the traditional financial market, such as the establishment of Black–Scholes model in the option contract market (1973), the stock price forecasting model based on Geometric Brownian Motion by Reddy, and Clinton (2016). However, most of its use on price forecasting are specific to the single market with little attention paid to portofolio establishment under cointegration strategy.

Literature and Concept Review — Pair Trading

1. Stock selection: conduct **cointegration tests** on selected stocks and test the stationarity of the spread in pair
2. Trading threshold determination: conduct hypothesis testing and estimate z-scored spread mean with the use of **training data**
3. Trading signal determination: compute actual daily spread using real-time data and trade according to the principle:
    a. Take short when actual spread is **higher than threshold**
    b. Unwind the position when the spread reverts back to mean

What's the cointegration relationships among different altcoins and which cointegration test is most efficient?

Can we expand the training data in real time and adjust trading threshold to adapt the market changes?

Can pricing forecasting algorithms be introduced here to determine the transaction point where highest profit can be made?

Literature and Concept Review

- **Definition 1** Cointegration:

  y_{1t} *and* y_{2t} *are cointegrated, if* y_{1t} *and* y_{2t} *are* $I(1)$ *series and* $\exists \beta$ *such that* $\beta Y_t = y_{1t} - \beta_1 y_{2t} \sim I(0)$.

  $y_{1t}, y_{2t},$ *and* y_{3t} *are cointegrated, if* $y_{1t}, y_{2t},$ *and* y_{3t} *are* $I(1)$ *series and* $\exists \beta$ *such that* $\beta Y_t = y_{1t} - \beta_1 y_{2t} - \beta_2 y_{3t} \sim I(0)$.

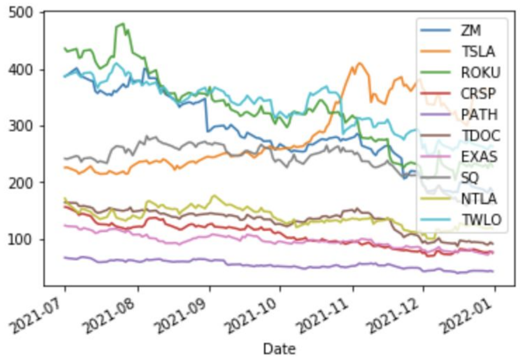
- **Definition 2** Vector Error Correction Model (VECM):

$$\Delta X_t = \mu + \Phi D_t - \Gamma_{p-1}\Delta X_{t-p+1} - \cdots - \Gamma_1 \Delta X_{t-1} + \Pi X_{t-1} + \varepsilon_t, \quad t = 1, \cdots, T$$

$$Long-run\ impact\ matrix: \quad \Gamma_i = (\Pi_{i+1} + \cdots + \Pi_p), \quad i = 1, \ldots, p-1.$$

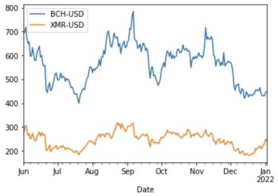
$$Short-run\ impact\ matrix: \quad \Pi = \Pi_1 + \cdots + \Pi_p - I.$$

Literature and Concept Review — ARKK Example




```
correlation= 0.9020319539285295
cointegration= (-4.012055652282486, 0.006902110133543827, array([-3.94810437, -3.36469667, -3.06423535]))

<matplotlib.axes._subplots.AxesSubplot at 0x7fb039dbe590>
```

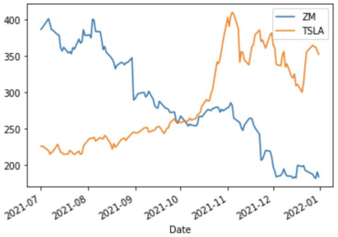



```
correlation= 0.8841877659256762
cointegration= (-2.0924854603269383, 0.48029583694003747, array([-3.98475411, -3.38466405, -3.07801392]))

<matplotlib.axes._subplots.AxesSubplot at 0x7ff751fdf590>
```


```
correlation= -0.7531881932546062
cointegration= (-1.7684109373249726, 0.6449352363803815, array([-3.98475411, -3.38466405, -3.07801392]))

<matplotlib.axes._subplots.AxesSubplot at 0x7ff751f78390>
```



Literature and Concept Review

- **Definition 3** Engle-Granger Cointegration Test:

$$If\ \widehat{\mu}_t = y_{1t} - \hat{c} - \widehat{\beta}_2 y_{2t}\ \ is\ stationary,\ y_{1t}\ and\ y_{2t}\ are\ cointegrated.$$

- **Definition 4** Johansen Cointegration Test:
  - Hypothesis

$$H_0:\ r = r^* < k \qquad r := number\ of\ \ cointegration\ vectors$$

$$H_1:\ r = k$$

  - Steps
    - Setting k as different values, estimate the VECM model using maximum likelihood
    - Compare the models using likelihood ratio tests

Literature and Concept Review — Pair Trading

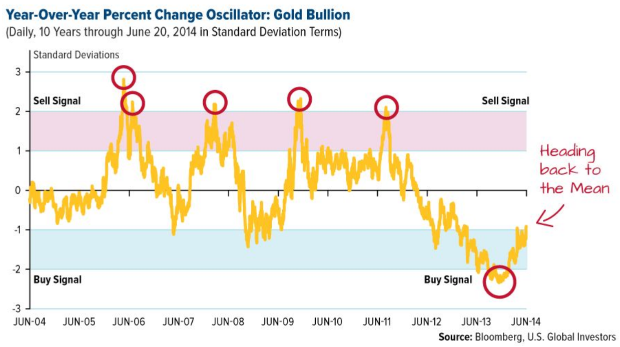
1. Stock selection: conduct **cointegration tests** on selected stocks and test the stationarity of the spread in pair
2. Trading threshold determination: conduct hypothesis testing and estimate z-scored spread mean with the use of **training data**
3. Trading signal determination: compute actual daily spread using real-time data and trade according to the principle:
    a. Take short when actual spread is **higher than threshold**
    b. Unwind the position when the spread reverts back to mean

What's the cointegration relationships among different altcoins and which cointegration test is most efficient?

Can we expand the training data in real time and adjust trading threshold to adapt the market changes?

Can pricing forecasting algorithms be introduced here to determine the transaction point where highest profit can be made?

Literature and Concept Review



- **Definition 5** Pair Trading Spread (mean-reverting):

$$a.\ \ y_{1t}/y_{2t} \sim N\left(\mu_{y_{1t}/y_{2t}}, \sigma^2_{y_{1t}/y_{2t}}\right)$$

$$b.\ \ \log(y_{1t})/\log(y_{2t}) \sim N\left(\mu_{\log(y_{1t})/\log(y_{2t})}, \sigma^2_{\log(y_{1t})/\log(y_{2t})}\right)$$

$$c.\ \ \log(y_{1t}) - hedge\ ratio \cdot \log(y_{2t}) \sim N\left(\mu_{\log(y_{1t}) - hedge\ ratio \cdot \log(y_{2t})}, \sigma^2_{\log(y_{1t}) - hedge\ ratio \cdot \log(y_{2t})}\right)$$

- **Definition 6** Pair Trading Principle:
  - If observations of spread are larger than the predicted value, take short position for stock1 immediately
  - If the observations of spread reverts, unwind the position and profit is generated

Literature and Concept Review

- **Theorem 1** Two-dimensional Ito's Lemma:

$$\textit{Let } \left\{\left(X_1(t), X_2(t)\right)\right\} \textit{ be a solution to } \ln X(t) = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma\left(W(t) - W(0)\right) \textit{ and } g\left(t, x_1, x_2\right)$$

be a function which is continuously differentiable in t and continuously twice differentiable jointly with respect to x_1 *and* x_2*. Then,* $g\left(t, x_1, x_2\right)$ *is a solution of :*

$$\mathrm{d}g\left(t, X_1, X_2\right) = \frac{\partial g\left(t, X_1, X_2\right)}{\partial t}\mathrm{d}t + \frac{\partial g\left(t, X_1, X_2\right)}{\partial x_1}\mathrm{d}X_1 + \frac{\partial g\left(t, X_1, X_2\right)}{\partial x_2}\mathrm{d}X_2$$

$$+ \frac{1}{2}\left[\left(\left(\sigma_{11}{}^2 + \sigma_{12}{}^2\right)\frac{\partial^2 g\left(t, X_1, X_2\right)}{\partial x_1{}^2} + \left(\sigma_{21}{}^2 + \sigma_{22}{}^2\right)\frac{\partial^2 g\left(t, X_1, X_2\right)}{\partial x_2{}^2}\right) + 2\left(\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}\right)\frac{\partial^2 g\left(t, X_1, X_2\right)}{\partial x_1 \partial x_2}\right]\mathrm{d}t$$

Literature and Concept Review

- **Definition 7** Two-dimensional GBM model:

$$dX_1 = \mu_1(t, X_1, X_2)\,dt + \sigma_{11}(t, X_1, X_2)\,dW_1 + \sigma_{12}(t, X_1, X_2)\,dW_2$$

$$dX_2 = \mu_2(t, X_1, X_2)\,dt + \sigma_{21}(t, X_1, X_2)\,dW_1 + \sigma_{22}(t, X_1, X_2)\,dW_2$$

To solve the above SDE system, suppose $S(X) = \log X$ *and* $X(0)$ *as the initial value of* X.

By using Two – dimensional Ito's Lemma, we are able to obtain solutions:

$$X_1(t) = X_1(0)\, exp\left(\mu_1 - \frac{1}{2}\sum_{j=1}^{2}\sigma_{1j}{}^2\right)t + \sum_{j=1}^{2}\sigma_{1j}(W(t) - W(0))$$

$$X_2(t) = X_2(0)\, exp\left(\mu_2 - \frac{1}{2}\sum_{j=1}^{2}\sigma_{2j}{}^2\right)t + \sum_{j=1}^{2}\sigma_{2j}(W(t) - W(0))$$

Based on the correlated price forecasting solution, we are able to compute price spread accordingly then.

Pair Trading Strategy with Multidimensional GBM

1. Stock selection: conduct **cointegration tests** on selected stocks and test the stationarity of the spread in pair
2. Trading threshold determination: conduct hypothesis testing and estimate z-scored spread mean with the use of **training data**
3. Trading signal determination: compute actual daily spread using real-time data and trade according to the principle:
    a. Take short when actual spread is **higher than threshold**
    b. Unwind the position when the spread reverts back to mean

Engle-Granger Cointegration Test and Johansen Cointegration Test are both applied to calculate all possible stationary spreads for the further testing.

Expand the training data to include forecasted price spread daily obtained from multidimensional GBM as well and reconduct hypothesis testing during each trading day to compute for the up-to-date trading threshold.

Take short when:

1. Today's the only day above threshold compared to the next few consecutive days' forecasted price (forecasted daily)
2. Today has highest price spread compared to the above threshold forecasted price spread for the consecutive following days

Pair Trading Strategy with Multidimensional GBM

1. Stock selection: conduct **cointegration tests** on selected stocks and test the stationarity of the spread in pair
2. Trading threshold determination: conduct hypothesis testing and estimate z-scored spread mean with the use of **training data**
3. Trading signal determination: compute actual daily spread using real-time data and trade according to the principle:
   a. Take short when actual spread is **higher than threshold**
   b. Unwind the position when the spread reverts back to mean

On a daily basis:

1. Expand the training data to forecast all future price spread value with multidimensional GBM
2. Reconduct hypothesis testing on the spread mean and variability to compute the adjusted trading threshold
3. Determine if today is the point where the transaction is the most profitable compared to following days

Implementation — Altcoin Trading Experiment

Data Set-up:

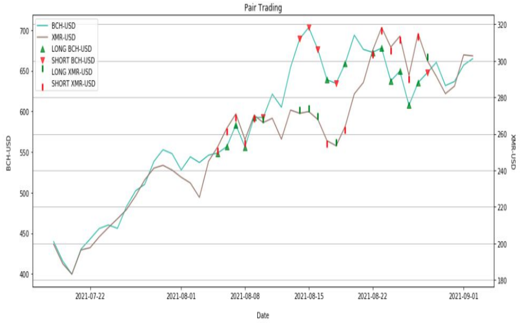
- 2-year daily close price data for 11 altcoins from yahoo finance (2020.8.20-2022.8.20)
- BNB-USD, SOL-USD, ADA-USD, ETH-USD, LTC-USD, BCH-USD, DOT-USD, DOGE-USD, USDT-USD, MATIC-USD, XMR-USD

Date	BNB-USD	SOL-USD	ADA-USD	ETH-USD	LTC-USD	BCH-USD	DOT-USD	DOGE-USD	USDT-USD	MATIC-USD	XMR-USD
2020-08-20	23.092575	3.208267	0.133642	416.439789	62.966755	294.791626	2.900080	0.003459	1.000409	0.026709	101.311401
2020-08-21	22.128994	2.943668	0.123760	389.126343	59.379021	284.298248	2.875028	0.003393	1.004549	0.024932	92.569595
2020-08-22	22.239820	3.020886	0.125276	395.835144	60.311157	287.362427	4.484690	0.003442	1.000956	0.026151	94.739082
2020-08-23	21.870842	3.322431	0.121595	391.384491	60.623260	284.957062	3.967066	0.003414	1.002345	0.025308	91.312607
2020-08-24	22.605852	3.344986	0.124488	408.144196	62.199425	291.325958	4.602614	0.003418	1.001976	0.027981	93.939880
2020-08-25	21.475367	3.264850	0.113292	384.001038	58.580666	276.799194	5.523393	0.003305	1.002365	0.025861	88.828781
2020-08-26	22.277306	3.558375	0.115285	386.466125	58.358173	275.571014	6.081716	0.003303	1.001853	0.026905	89.123993
2020-08-27	23.080753	3.445060	0.106824	382.632629	56.041016	264.584320	5.628337	0.003203	1.002206	0.025737	89.539352
2020-08-28	23.133812	4.011393	0.109453	395.874664	57.354404	268.914703	6.159955	0.003268	1.002036	0.026855	94.359978

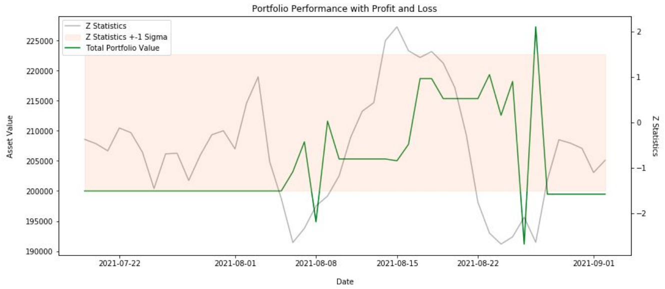
Objective:

- Determine if the modified strategy fits the altcoin market better than traditional pair trading strategy
- Investigate the best time frame and price spread for the real strategy implementation

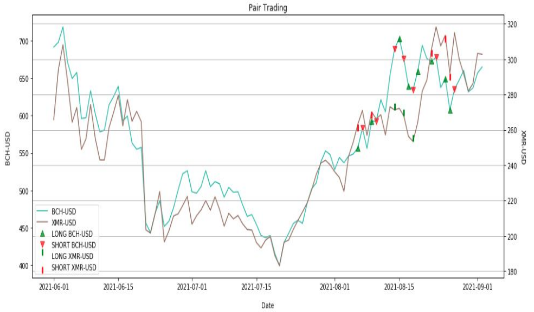
Results — Altcoin Trading Experiment



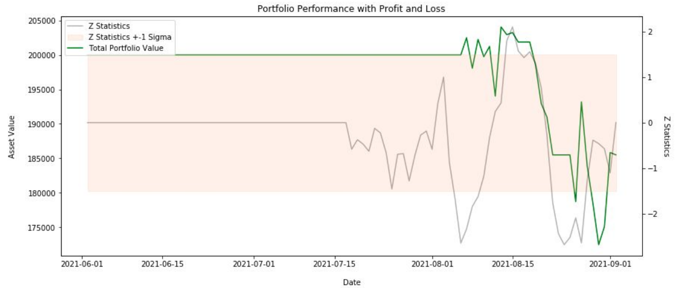
CAGR = 26965.417%



Traditional Pair Trading result with **log price ratio** spread

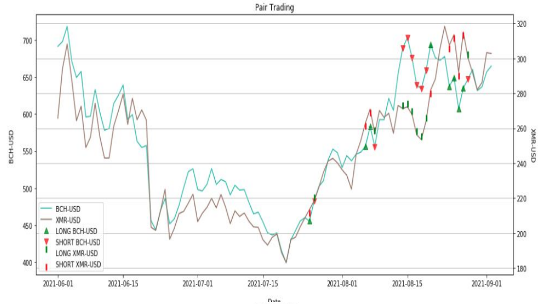


CAGR = 13370.187%

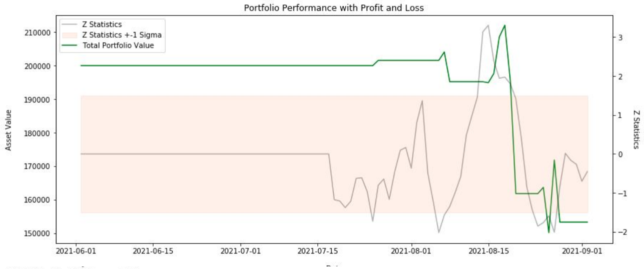


Modified Pair Trading based on GBM result with **log price ratio** spread

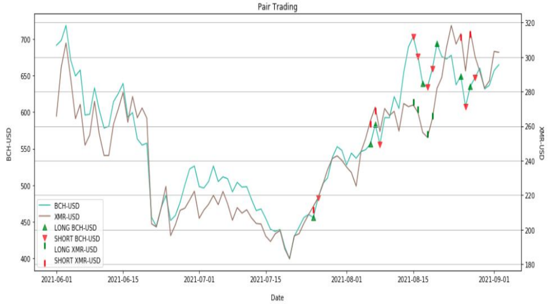
Results — Altcoin Trading Experiment



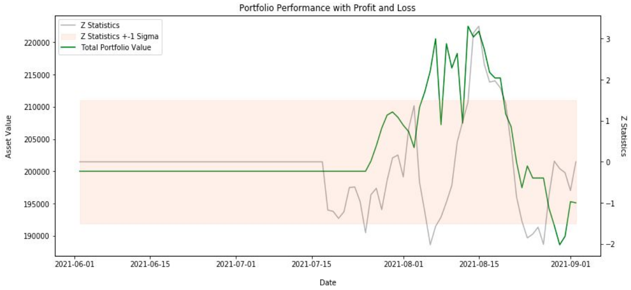
CAGR = 444.175%



Traditional Pair Trading result with **price ratio** spread



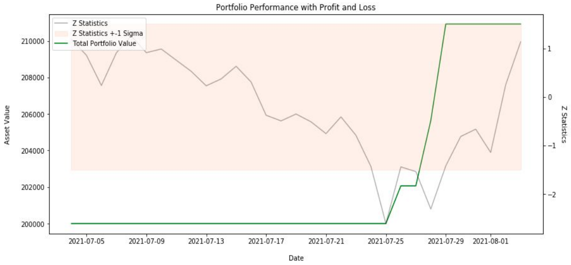
CAGR = 20005.058%



Modified Pair Trading based on GBM result with **price ratio** spread

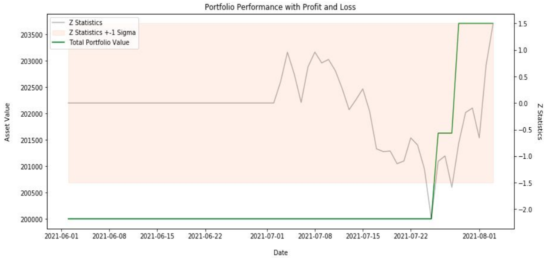
Results — Altcoin Trading Experiment

Number of days = 30
CAGR = 878140.967%



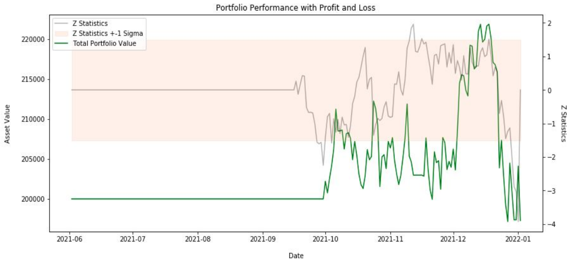
Trading days=30

Number of days = 62
CAGR = 6494.966%



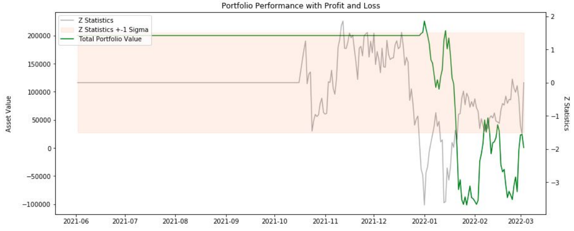
Trading days=62

Number of days = 107.0
CAGR = 915.534%



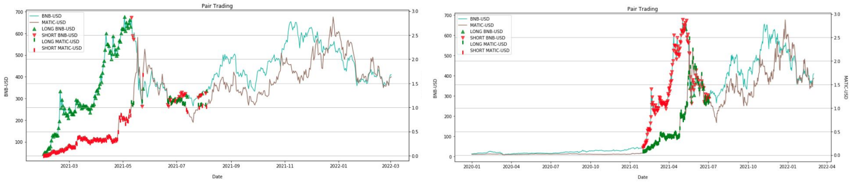
Trading days=107

Number of days = 137.0
CAGR = -100.000%



Trading days=**137**

Conclusion and Future Work



Trading frequency comparison

Findings:

- For this dataset, both Engle-Granger cointegration test and Johansen test works efficiently.
- For the pairs passing cointegrated tests in this dataset, only log price ratio and price ratio spread pass adf test
- GBM modified strategies work better than traditional one only for price ratio spread
- Both strategies work poor as trading time period lengthens

Conclusions:

- Overall, traditional pair trading strategy is much more likely to generate higher CAGR than the modified strategy
- Both strategies are most powerful within 30-day trading period whereas its effectiveness shrinks gradually

Future Work:

- Further optimize the price spread forecasting accuracy to resolve the trading period influence
- Discover the multiple-stock pair trading strategy modification possibility

References

- Maruddani, Di & Trimono, Trimono. (2018). Modeling stock prices in a portfolio using multidimensional geometric brownian motion. *Journal of Physics: Conference Series*. 1025. 012122. 10.1088/1742-6596/1025/1/012122.
- Kloeden PE, Platen E 1992 *Numerical Solution of Stochastic Differential Equation*, New York: Springer-Verlag Berlin Heidelberg
- Platen E, Rendek R 2009 Exact Scenario Simulation for Selected Multi-Dimensional Stochastic Processes, *Communications on Stochastic Analysis* **3** 3 443-465

Thank You!

Many Thanks to my advisor — Prof. Leon Tatevossian!!!