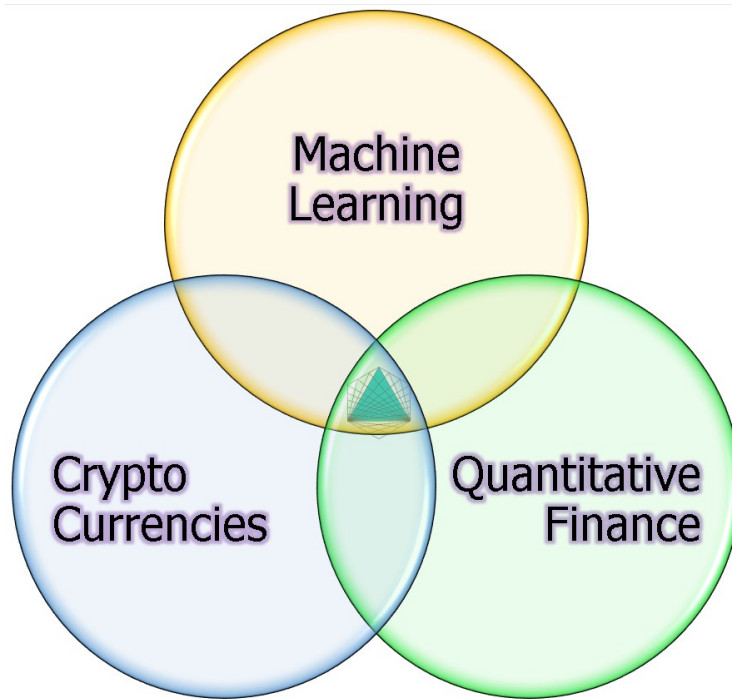


# CASE: THE A.I. CRYPTO TRADER

## 1. INTRODUCTION

In attempt to make a case which is to be somewhat universally understandable by various types of students, the case is financial time-series prediction, while making it more engaging with the hot topic of cryptocurrencies.

The case integrates knowledge from various sources - Crypto Currencies, Quantitative Finance and Machine learning. At the same time the case is stratified as the teams solving it could complete various levels - as far as they could solve it.



## 2. BUSINESS PROBLEM FORMULATION

The goal is to build a successful investing/trading model on the cryptocurrency markets. The data consists of time-series of various cryptocurrencies prices and volumes. The data could also be enriched by the teams.

Cryptocurrencies' portals: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#) ; [Link 4...](#)

Crypto guides & glossaries: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#) ; [Link 4...](#)

Cryptocurrencies data: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#) ; [Link 4...](#) ; [Link 5...](#)

The case has two levels of difficulty - in order to go to start solving the next level, the students would probably need to complete the previous:

- **LEVEL 1: Make a prediction model of the major cryptocurrencies' prices**

The process of investment could be analyzed in several consecutive phases, which are processes of transforming information. First are the input processes - from the environment, from the current state of the investment and from the investors' goals. Using these inputs, the next phase is a predictor for forecasting/estimating the expected values for the significant variables and the external factors. Statistical analysis of past portfolio structure is also necessary.

The prediction model should be able to forecast the prices of the major cryptocurrencies throughout the most volatile periods with some acceptable accuracy, utilizing all available information (including additional enrichment of data) for one period ahead.

- **LEVEL 2: Make an autonomous A.I. decision-maker for trading/investing.**

Autonomous investment is a concept which implies that the process of investment management should be conducted by and self-maintained by non-human entity. The autonomous investment corresponds to the ever-faster changing investment environment accelerated by the rapid development of the communicational and computational abilities of the modern computer.

So as an effort to keep up with the technological advancements in the field as well as trying to keep the playing field level, there is now a new tendency among researchers to propose autonomous solutions for trading (while of course making it harder and harder for human entities to participate). The general approach preached by the authors of the current case is based on the idea of machine learning. The A.I system should be able to make use of the available data (including the predictions from level 1) to form and simulate a viable investment strategy, working on its own for the whole period of supplied data.

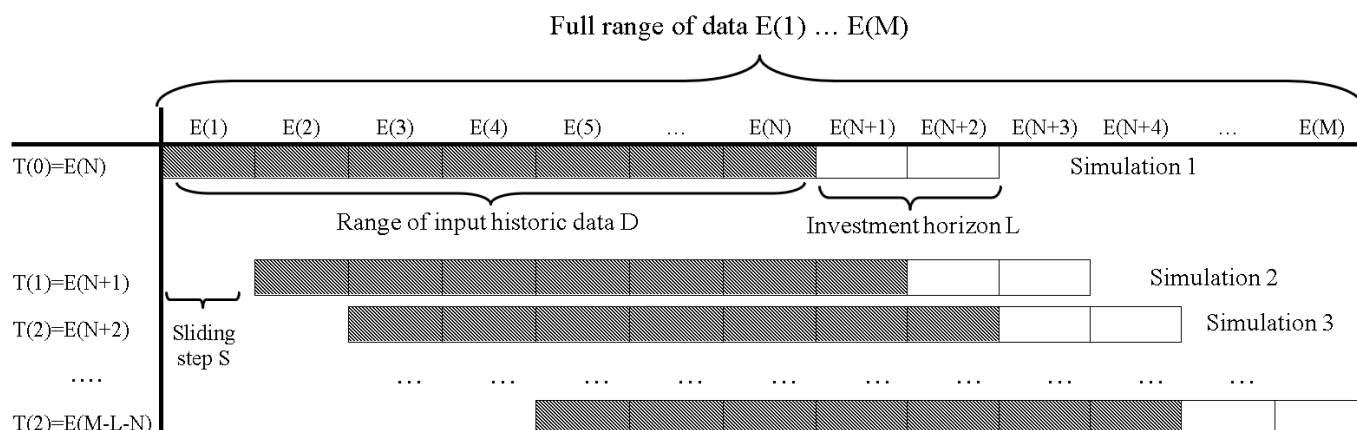
### 3. RESEARCH PROBLEM SPECIFICATION

This section consists of problem description from data science point of view. It is meant to assist the teams to what general directions should they take in their solution.

#### The concept of the competition

Generally the case requires series of historical simulation experiments over the crypto currencies market, using the supplied pricing data. Every experiment uses prehistory ex-post data to predict ex-ante several points into the future. These predictions are then compared with the real data for these same several points and this concludes one experiment. To begin the next experiment the historical ex-post data window slides to the future and starts over again.

All of the “models of investors” are to be back-tested on the unified competition data track. Such competition track consists of complete time series of all possible assets. The back-testing is done for every data-point (historical trading time moment) with all possibilities (all assets available for trading in that time moment). The direct result of such systematic approach would be a ranking list of the most successful (according to given criteria) A. I. investors.



Every A. I. investor is to be tested on each data point with the full variety of possible modifications. There is a minimal incremental step of one observation. The historical data window, which is used for estimation, reflects the period of previous data that the model uses is maximum 2000

observations. With each advance to the next data point, the window “slides” forward as well. The horizon of prediction (minimum one observation) also slides forward.

For sake of equivalency among the teams we will assume that the earliest period for prediction (level 1) or investment decision (level 2) would be the first time data point after 00:00 on 01.01.2021. The prediction/investment horizon would be 1 hour ahead, but the teams are free to implement methods working for more steps ahead. The sliding step of the prediction window would be 1 hour, but again the teams may implement systems which re-evaluate at longer time periods. The last simulation experiment should finish with the last data point of the supplied data.

During the historical simulation all necessary computations are calculated according to the business logic or the algorithms of the experimented model of investor based on the all the data points in the data set. As a result, there is a new formed data array with selected solutions for each observation. So that this new data array has dimensions of  $k$  columns and  $m-1-D$  rows, where  $k$  is the number of all assets (some/most/all of them could be equal to 0), included in the data set,  $m$  is the number of all observations, and  $D$  is the size of the necessary pre-history data for the given model of investor. The A. I. investor reconsiders his investment portfolio every  $L$  periods, where  $L$  is the investment horizon.

### **General Assumptions for the case**

These assumptions are necessary for the methodically correct run of the historical simulation:

- the investor is independent, private, small entity investing own funds;
- the transactions (frequency and amounts) do not influence the market prices;
- all other investment markets except for the presented in the data set are disallowed;
- minimum discretization step is 1 hour;
- the moment “Now” is notated by  $t=0$  or  $t(0)$
- all necessary predictions are based on ex-post data before  $t(0)$ ;
- technical limitations - the once trained models should be able to work on regular computer (1 cpu 2 GhZ, 4 GB RAM will be assumed)

### **Level 1**

The main aim of time series modeling is to carefully collect and rigorously study the past observations of a time series to develop an appropriate model which describes the inherent structure of the series. This model is then used to generate future values for the series, i.e. to make forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past.

There are many techniques which could be used to forecast the cryptocurrencies’ prices, and the choice would be left to the discretion of the participants. Some of the main classes of forecasting models are: Regression analysis; Autoregression & moving averages; Support vector machines, Artificial neural networks, etc.

Introduction to Time series forecasting: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#) ; [Link 4...](#)

Examples of Time series forecasting: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#) ; [Link 4...](#)

[Have a look at the article, written especially for this case by one of our mentors...](#)

### **Prediction scoring**

The prediction models should be tested on the prices of the chosen 20 cryptocurrencies (see below) and every metric would be averaged among all currencies in the final score. For the sake of scoring the accuracy of the prediction models, before the end of the deadline you would be provided with several various periods for testing out of the supplied data, with each testing period consisting of 96

hours. These periods would be unknown to the participants and would be chosen specifically for their contradictory movements and volatility. The prediction models should be able to perform within certain limits and will be competed among each other on the following metrics:

- 1) The **mean absolute percentage error (MAPE)** is a measure of prediction accuracy of a forecasting. It is measuring how close the prediction is to the actual data for the each whole testing period. It expresses accuracy as a percentage, with the smallest possible value being most desirable:

$$M(y_t, f_t) = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - f_t}{y_t} \right|$$

Where:

$y_t$  - actual data point at period  $t$

$f_t$  - forecasted data point for period  $t$

$n$  - total number of periods considered for evaluation

- 2) The **directional symmetry (DS)** statistic gives the percentage of occurrences in which the sign of the change in value from one time period to the next is the same for both the actual and predicted time series. The DS statistic is a measure of the performance of a model in predicting the direction of value changes. The case  $DS = 100\%$  would indicate that a model perfectly predicts the direction of change of a time series from one time period to the next. The desirable score here should be higher than 50%.

$$D(y_t, f_t) = \frac{100}{n-1} \sum_{t=2}^n d_t,$$

$$d_t = \begin{cases} 1, & \text{if } (y_t - y_{t-1})(f_t - f_{t-1}) > 0 \\ 0, & \text{if } (y_t - y_{t-1})(f_t - f_{t-1}) \leq 0 \end{cases}$$

Where:

$y_t$  - actual data point at period  $t$

$f_t$  - forecasted data point for period  $t$

$n$  - total number of periods considered for evaluation

- 3) As an additional and necessary characteristic of the prediction model will be calculated stability over the testing sets of periods. The stability will be measured by **coefficient of variation of the mean absolute percentage error**. Coefficient of variation is a relative measure of dispersion that corresponds to standard deviation, but it is expressed in percentage terms, which means it could be compared across different prediction models. The coefficient would be computed among the values of MAPE for the various testing sets of periods. The desired value should be less than 1.

$$S(M_i) = \sqrt{\frac{\sum_{i=1}^m (M_i - \bar{M})^2}{m-1}},$$

$$\bar{M} = \frac{\sum_{i=1}^m M_i}{m},$$

$$R = \frac{S(M_i)}{\bar{M}}$$

Where:

$M_i$  - mean average absolute error for testing set  $i$

$R$  - generalized coefficient of variation

m - total number of testing sets

- 4) In order for the prediction models to have any practical value **Computational efficiency** will be computed as a relative time to calculate the next prediction compared to the minimal time step of the next actual observation. The model should be able to be executed on a regular computer (1 cpu 2 GhZ, 4 GB RAM will be assumed) but there are no restrictions for the computer on which the model is trained. Desired value is minimum possible.

$$U = \frac{\sum_{c=1}^p T_c}{T_g}$$

Where:

$T_c$  - time for calculating prediction for currency c in seconds

p - total number of predicted currencies

$T_g$  - time interval for next actual data point in seconds

U - relative time for execution of all predictions in % of time interval

- 5) The combined score for each model will be computed as a linear combination of the above specified metrics averaged over the studied crypto currencies. The highest possible value of the score is most desirable. Please, have in mind that this score may be a negative number.

$$Z = \overline{D} - \overline{M} - \overline{R} - U$$

## **Level 2**

Over the years a significant number of portfolio models, methods, procedures, and strategies ("investors") have been proposed by theoreticians and practitioners in the field. Application of each of them should be considered as a systematic process consisting of several phases (i.e. goal setting, data collection, data structuring, statistical testing, enforcing limitations, forecasting, generating and developing feasible solutions, selection of optimal solution, realizing the investment solution, retrieving feedback of the significant outcomes, etc).

You should build an autonomous trading/investment system which presents the best chance for profiting on the cryptocurrencies market. This system (A.I.) should run throughout the whole period of the data set, starting on the first time data point after 00:00 on 01.01.2021 and finishing on the last time step. The A.I should be able to decide trades when to buy and sell the cryptocurrencies on its own without human interaction. And the task at hand is to simulate what would be the results of such A.I. for the whole data set period.

The participants have the choice to which approach to use to build their autonomous trading system. Possible (out of many) approaches are Trading robots (typically directed towards trading with one or two assets), Statistical arbitrage (utilizing pricing inefficiencies among several assets), Autonomous portfolio management (using dynamic optimization techniques to allocate capital among many assets).

Trading robots introduction: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#)

Trading robot example: [Link 1...](#)

Trading robot builders: [Link 1...](#) ; [Link 2...](#)

Statistical arbitrage introduction: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#) ; [Link 4...](#)

Statistical arbitrage examples: [Link 1...](#) ; [Link 2...](#)

Statistical arbitrage advanced examples: [Link 1...](#) ; [Link 2...](#)

Investment portfolio introduction: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#)

Investment portfolio further reads: [Link 1...](#) ; [Link 2...](#) ; [Link 3...](#)

Investment portfolio advanced example: [Link 1...](#)

## Additional assumptions for level 2

In order to define the case for Level 2, additional assumptions about the trading procedures are needed.

Assumptions on simulating the investment value:

- the initial invested amount is 10000 USD, simulating a small investor;
- being small investor also means, that the transactions do not influence the market prices;
- there is no new capital inflow, nor capital outflow (no new funds added, no funds withdrawn) for the full period of the data;
- the initial capital could be invested no earlier than the first time data point after 00:00 on 01.01.2021;
- if the investment value reaches 0 USD (or real return of -100%), the investor is insolvent/bankrupt (none-the-less you should keep the results even from these tryouts as they are usually useful experiments);

$$Mv(t) = \sum_{i=1}^k Nu_i(t-1).P_i(t)$$

Where:

Mv (t) - market value of investment at the moment t

Nu<sub>i</sub> (t) - equivalent number of units of asset i at moment t

P<sub>i</sub> (t) - market price of asset i at moment t

k - total number of assets

- the value of the USD positions is unchanging (regardless of the real world trade possibilities for USD), which means zero return - the case is about cryptocurrencies, not FIAT currencies;
- If the investor runs into non-computable values by the A.I.'s procedures (the model could not present a realizable result and there is not possible solution for the current time moment), the states of investments in all assets are kept as the values from previous time moment;
- No fundamental events influence the value of the investment (e.g. do not account for effects from forks, airdrops, burns, etc.).

Assumptions on simulation the market:

- there is perfect liquidity on the market, so the investor could always make a trade at the given price;
- all transactions in time moment t (and after moment t-1) are made at price P(t);
- there is no short selling;
- there is no margin trading.

Assumptions on simulation the market frictions:

- all trades are charged 0.25% of the turnover of the trade;

$$St = Nu_i(t).P_i.Bf$$

Nu<sub>i</sub> (t) - equivalent number of units of asset i at moment t

$P_i(t)$  - purchase price of asset  $i$  at moment  $t$  in USD

$B_f$  - exchange fee (0.0025)

- the trading fee includes all blockchain transaction fees (since there are no withdrawals allowed);
- no income taxes are owed to any fiscal entity;
- no market spread - all trades are simulated on the given market prices;
- The minimum trade value for orders within same time frame and with the same cryptocurrency is 0.001 BTC (100,000 Satoshis);
- The maximum trade value for orders within same time frame and with the same cryptocurrency is 0.5 BTC;
- Rationalizing the transactions – due to the various market frictions it is rational that before realizing a transaction the new desired investment structure (computed at period  $t$ ) is compared to the old investment structure (computed at period  $t-1$ ), so that the new market order trades only the differences.

$$MOr_i(t) = Nu_i(t) - Nu_i(t-1)$$

Where:

$MOr_i(t)$  - market order in number of units of asset  $i$  at moment  $t$

$Nu_i(t)$  - equivalent number of units of asset  $i$  at moment  $t$

- for any other market frictions, you should refer to the rules of Bittrex ([here...](#))

## Investment scoring

**1) Risk-adjusted return measure.** In order measure performance of the investment A.I. it is essential to implement a viable indicator. It has to be risk adjusted return due to:

- Firstly – since all possible benefits from investing in a given asset are already reflected in the individual return of the asset, then the return becomes general measure of the benefit.
- Secondly – the return alone is not a good measure as the same value of return may be achieved at various values of risk. Ergo risk adjusted return is a better measurement for risky investments.
- Thirdly – using risk adjusted return effectively compresses two-dimensional problem into one-dimensional and among other advantages it also relaxes the multi-criteria optimization problem to single-criterion optimization problem.

The risk adjusted return should be computed by using an asymmetric measure modified from Sortino ratio (Frank A. Sortino and Robert van der Meer. "Downside Risk." Journal of Portfolio Management. Vol. 17, No. 5, Spring 1991, pp. 27–31). The asymmetric measures calculate the risk far more precisely but require at least twice as many data points for computing a robust estimation. Desired value is maximum possible.

$$G = \frac{1 + R_p}{\sqrt[2]{v_0(R_p)}}$$

$$R_p = \frac{Mv(M)}{Mv(1)} - 1$$

$$v_0(R_p) > 0$$

$$v_0(R_p) = \frac{\sum_{t=1}^M \max\{0, -R_p(t)\}^2}{M}$$

Where:

$G$  - Generalised asymmetric risk adjusted return measure  
 $R_p$  - Return of the investment A.I. for the whole data set  
 $R_p(t)$  - Compound return of the investment A.I. for time data point  $t$   
 $Mv(M)$  - Market value of the investment at data point  $M$  (endpoint)  
 $MV(1)$  - Market value at the start of the competition (10000 USD)  
 $vo(R_p)$  - Lower partial moment of return with minimum acceptable return = 0

- 2) In order for the A.I. investor to have any practical value **Computational efficiency** will be computed as a relative time to calculate the next solution compared to the minimal time step of the next actual observation. The A.I. should be able to be executed on a regular computer (1 cpu 2 GhZ, 4 GB RAM will be assumed) but there are no restrictions for the computer on which the model is trained. Critical value is  $< 1$ , (in other words time for calculation should be less than 300 seconds).

$$U(t) = \frac{T_s(t)}{T_g(t)}$$

$$\forall U(t) < 1$$

Where:

$T_s(t)$  - time for calculating solution  $t$  in seconds  
 $T_g(t)$  - time interval for actual data point  $t$  in seconds  
 $U(t)$  - relative time for execution of solution  $t$  in % of time interval

#### 4. DATA DESCRIPTION

The used dataset is the same for all participants but could be enriched with any publicly available data (of course any enrichment should be diligently described). The dataset is considered a test track for tryout, evaluation, competition, and comparative analysis of the A.I. investors.

##### Data transcript

See the full data transcript here:

<https://www.kaggle.com/bizzyvinci/coinmarketcap-historical-data>

##### Selected cryptocurrencies

List of the currencies to be used in level 1 (prediction modeling):

Currency	ticker	Coin ID
Bitcoin	BTC	1442
Bitcoin Cash	BCH	1445
Bitcoin Gold	BTG	1456
Cardano	ADA	1446
Dash	DASH	1453
Dogecoin	DOGE	1447
Eos	EOS	1452
Ethereum	ETH	1443
Ethereum Classic	ETC	1457
Iota	MIOTA	1451



Lisk	LSK	1460
Litecoin	LTC	1448
Monero	XMR	1454
NEMcoin	XEM	1447
Neo	NEO	1449
Ripple	XRP	1444
Stellar	XLM	1450
Tether	USDT	1474
Tron	TRX	1455
Zcash	ZEC	1465

### **Possible data challenges and solutions**

Data challenge: there are price time series which do not have enough data points to conduct reasonable analysis or to apply a certain method.

Solution: making a well-grounded selection of the assets in the research database, but still leaving as many assets as possible in the data set.

Data challenge: in the initial trading sessions of some assets there are relatively large periods of trading inactivity or strange, uncharacteristic behavior.

Solution: eliminating the boundary effects by removing certain number of initial observations.

Data challenge: in price time series there is missing data.

Solution: data imputation in the time series. There are many approaches to data imputation of financial time series. According to Lazarov, D., "Missing data evaluation in financial time series", these are the most promising ones:

LVCF - Last value carried forward imputation

LINT - Linear interpolation

R1 - Multivariate regression imputation

NP100 - Non-parametric multivariate regression imputation using a moving window of length 100

MARX1 - Multivariate autoregression imputation of lag 1

AR5X - Univariate autoregression imputation of lag1 – lag5

MLP - Univariate multi-layer perceptron imputation

Data challenge: not enough data for a given A.I. trader.

Solution: collect additional data and enrich the data set. This may include economic/business data, data from various markets, etc. Some popular data sources for cryptocurrencies are following (but any data source is acceptable, as long as it is publicly accessible): [Link1...](#), [Link2...](#), [Link3...](#), [Link4...](#), [Link5...](#), [Link6...](#), [Link7...](#)

## **5. EXPECTED OUTPUTS**

**The teams will have to prepare and upload the following:**

- a co-authored article / notebook (see below for more instructions)
- with source codes, including all necessary libraries/used environments or workflows, so that the code works with normal effort.
- Prediction datasets for the solution of Level 1.
- Calculated metrics, as required by the case for Level 1 and Level 2

**Questions:****1) Level 1: Prediction modelling**

What necessary data preparation did you need?

Which is the most suitable method for forecasting cryptocurrencies?

Which cryptocurrencies were hardest to predict?

What anomalies in behavior of the cryptocurrencies did you detect?

**2) Level 2: A.I. Trading**

How many assets did you include in your A.I. Trading? How did you make the selection?

Which is the best Risk adjusted return have you achieved? Which model? Would this be the winning model if we ranked them on Return only?

What is the optimal time for reevaluating the investment (optimal investment horizon)? What is the optimal size of the training data set?

(Bonus) If you use the calendar for fundamental events such as airdrops, burns, forks ([here...](#)), how would you change your A.I. Trader to be more profitable?

## READER ON FINANCE TERMS

### 1) Glossary

#### Investment

- The investment process is a chain of considerations and actions for an individual, from thinking about investing to placing buy/sell orders for investment assets. (Investments, Bodie, Kane, Marcus, pp. 858)
- The sacrifice of certain present value for (possibly uncertain) future value. (Investments, Sharpe, Alexander, Bailey, pp. 1013)
- Investing is a process of consciously sacrificing own resources in the pursuit of future reward or goal. Important notion is that uncertainty of future events may clash with some current actions. None of the future outcomes is guaranteed to offset the undertaken restriction of investor's degrees of freedom (Alexander et al, 1993, pp. 840).
- Although generally it is accepted that "investing" differs from "trading", based on the term of the strategy (longer terms for investing), for the current needs trading will be assumed as a special case of investing, as it fits perfectly well with the above definitions (presuming "future" to be both short and long term).

#### Investor

- The investor is an entity (physical or legal), purposefully using financial (and other) resources for investment and pursuing future rewards. It is assumed that such entity acts rationally as a real Homo Economicus (Mill, 1836).

#### Investment asset/asset

- asset is an instrument that can be traded freely and easily on a well-developed market. (Investment Science, Luenberger, pp. 40)
- asset (a.k.a. investment instrument, investment asset, position) according to Marchev (2012) is investment opportunity, traded freely on a transparent market which transmits publicly enough relevant information.

#### Organized market

- A central location where trading of assets is done under a set of rules and regulations.
- A mechanism designed to facilitate the exchange of financial assets by bringing buyers and sellers of assets together. (Investments, Sharpe, Alexander, Bailey, pp. 1010 -1017)
- Organized exchange are centralized auction-type markets, while the over-the-counter market is an intricate network of asset dealers that take position in various assets and buy and sell from their own portfolios. (Modern Investment Theory, Haugen, pp. 26)

#### Market price

- A market price is the last price at which an asset traded, meaning the most recent price on which a buyer and seller have agreed and at which some amount of the asset was transacted.
- Closing price is the final price at which an asset was traded for a given trading period. The closing price represents the most current valuation of an asset until the end of the next trading period. Note that the trading period could be virtual (for the sake of research) and not officially regulated.

#### Return

- A rate of return expresses the percentage change in your wealth from one period to another. (Modern Investment Theory, Haugen, pp. 389)

- Returns from investing are crucial to investors; An assessment of return is the only rational way (after allowing for risk) for investors to compare alternative investments that differ in what they promise. The measurement of actual (historical) returns is necessary for investors to assess how well they have done or how well investment managers have done on their behalf. (Investments Analysis and Management , Jones, pp. 114)

## Risk

- Risk means uncertainty about future rates of return. (Investments, Bodie, Kane, Marcus, pp. 123)
- The chance that the actual outcome from an investment will differ from the expected outcome. (Investments Analysis and Management , Jones, pp. 11)
- A measure of uncertainty about the outcome from a given event. The greater the variability of possible outcomes, on both the high side and the low side, the greater the risk. (Foundations of Financial Management, Block, Hirt)
- Risk is identified with the dispersion of returns, i.e. with possible deviations (both positive and negative) from the expected return. (Capital investment & financial decisions, Levy, Sarnat, pp. 237)
- The standard deviation around the expected return. (asset Analysis and Portfolio Management, Fischer, pp. 560)

## Liquidity

- Liquidity is the ease (speed) at which an asset can be sold and still fetch a fair price. (Investments, Bodie, Kane, Marcus, pp. 864)
- Liquidity refers to the ease of converting an asset into money quickly, conveniently, and at little exchange cost. (asset Analysis and Portfolio Management, Fischer, pp. 6)
- The ability to sell an asset quickly without having to make a substantial price concession. (Investments, Sharpe, Alexander, Bailey, pp. 1014)

## Arbitrage

- Arbitrage is the simultaneous purchase and sale of an asset to profit from a difference in the price. It is a trade that profits by exploiting the price differences of identical or similar financial instruments on different markets or in different forms. Arbitrage exists as a result of market inefficiencies and would therefore not exist if all markets were perfectly efficient.
- Statistical arbitrage comprises a set of quantitatively driven trading strategies, looking to exploit the relative price movements across thousands of financial instruments by analyzing the price patterns and the price differences between financial instruments.
- Possible scenarios for exploiting arbitrage conditions:
  - Market Neutral Arbitrage - involves taking a long position in an undervalued asset and shorting an overvalued asset simultaneously. The assets should be selected to have similar volatilities and thus, an increase in the market will cause the long position to appreciate in value and the short position to depreciate by a roughly the same amount.
  - Cross Market Arbitrage - exploiting the price discrepancy of the same asset across markets, buying the asset in the lower-valuing market and selling it in the highly valuing market simultaneously.
  - Cross Asset Arbitrage - using the price discrepancy between a financial derivative and its underlying asset.

## Investment portfolio

- Collection of assets. (Investments, Sharpe, Alexander, Bailey, pp. 167)
- The assets held by an investor taken as a group. (Investments Analysis and Management, Jones, pp. 7)
- assets that have return and risk characteristics of their own, in combination, make up a portfolio. (asset Analysis and Portfolio Management, Fischer, pp. 2)
- Portfolio management is the process of combining assets in to a portfolio tailored to the investor's preferences and needs, monitoring that portfolio, and evaluating its performance. (pp. 2) Portfolio theory deals with the big picture - the risk and return attributes of the investor's overall portfolio of assets. (pp. 810) (Investments, Bodie, Kane, Marcus)
- The purpose of using a portfolio approach is to improve the conditions of the investment process by obtaining such properties (values of significant variables) of the combined assets that are not obtainable by any single asset. The most often (but not the only) considered significant variables are risk and return. A certain configuration of risk and return is only possible within a given combination of assets. Improving risk and return conditions through portfolio management is diversification (Marchev, 2012a).

### Diversification

- Another means to control portfolio risk is diversification, by which we mean that investments are made in a wide variety of assets so that the exposure to the risk of any particular asset is limited. (Investments, Bodie, Kane, Marcus, pp. 810)
- The process of adding assets to a portfolio in order to reduce the portfolio's unique risk and, thereby, the portfolio's total risk. (Investments, Sharpe, Alexander, Bailey, pp. 1007)
- The variance of the return of a portfolio can be reduced by including additional assets in the portfolio, a process referred to as diversification. (Investment Science, Luenberger, pp. 151)
- Diversification is related to the Central Limit Theorem, which states that the sum of identical and independent random variables with bounded variance is asymptotically Gaussian (the notion of diversification can be extended to more general random variables by the concept of mixing). (Robust Portfolio Optimization and Management, Fabozzi, pp. 19)
- The more traditional forms of diversification have concentrated upon holding a number of asset types (stock, bonds) across industry lines (utility, mining manufacturing groups). (asset Analysis and Portfolio Management, Fischer, pp. 560)
- An investor can construct a diversified portfolio and eliminate part of the total risk, the diversifiable or non-market part (unsystematic part). (Investments Analysis and Management, Jones, pp. 254)

## 2) Computation considerations on portfolio significant variables

The **mathematical description of a portfolio** is in the form of vector consisting of  $k+1$  positions each with respective weights, where  $k$  is the total number of assets and the additional position is the cash position  $c(t)$ . For unwanted positions the values are set to 0. Not invested funds are always assumed to be in the cash position. The sum of all weights is equal to 1.00. In brief the mathematical description of an investment portfolio is a  $k+1$ -dimensional vector of weights summing to 1.00 (i.e. a singular simplex).

$$W(t) = [w_1(t) \quad w_2(t) \quad \dots \quad w_i(t) \quad \dots \quad w_k(t) \quad c(t)]$$

where:

$W(t)$  - weight structure of the portfolio at discrete time (t)

$c(t)$  - relative weight of the cash position at discrete time (t)

$w_i(t)$  - relative weight of  $i$ -th asset at discrete time (t), subject to (no short positions):

$$w_i(t) \in [0,1] \wedge \sum_{i=1}^{k+1} w_i(t) = 1$$

To calculate **return of each asset** at the discrete time  $t$  the particular case is considered, where:

- Short selling is not possible transaction, meaning that the discrete times  $s$  (time of selling) follows  $b$  (time of buying).
- No income taxes are assumed.
- Besides the return derived from price change, there are other forms of return of a asset arising during the time of investing. These include airdrops, forks, rewards, burns and etc. All these must be estimated as financial inflow or outflow per one unit of currency (e.g. if while holding a position there is an airdrop of  $z$  amount per unit, then this is a positive return of  $z$ ).

$$R_i(t) = \frac{(P_i(s) - P_i(b) - K(s) - K(b)) + \sum_{t=b}^s B_i(t)}{P_i(b)}$$

where:

$P_i(s)$  - sell price at the discrete time  $s$  of asset  $i$

$P_i(b)$  - buy price at the discrete time  $b$  of asset  $i$

$B_i(t)$  - quantified complimentary benefits of asset  $i$  for the period between discrete time  $s$   $b$  and  $s$

$K(s)$  - brokerage at the discrete time  $s$

$K(b)$  - brokerage at the discrete time  $b$

$t$  - discrete time of return.

**Return of an investment portfolio** is calculated as a weighted average of the returns of all included assets. The weights correspond to the configuration of the portfolio – the allocated investment in each position. The sum of all weights (including cash position) is always equal to 1. The return of a cash position is normally assumed 0.

$$R_p(t) = \sum_{i=1}^k w_i(t) \cdot R_i(t)$$

where:

$R_p(t)$  - return of the portfolio  $P$  at the discrete time  $t$

$R_i(t)$  - return of the asset  $i$  at the discrete time  $t$

$w_i(t)$  - relative weight of position  $i$  at the discrete time  $t$

There are several approaches to calculating **asset risk**. The dominant concept is to use variance and/or standard deviation and/or volatility as a measure of risk. A good case could be built around using information entropy as a risk measure of a portfolio. So measuring the risk of an individual asset may be formulated as function of historical data of the return of asset:

$$V_i(t) = F\left(R_i\left(\overline{t-d}\right), R_i(t)\right)$$

where:

$V_i(t)$  - risk of the asset  $i$  at the discrete time  $t$ ;

$F$  - function for measuring the risk of asset  $i$ ;

$d$  - number (depth) of historical data considered for calculation of risk.

No matter what measure is used for **portfolio risk**, there is a strong agreement among authors that “the risk of a portfolio is not a weighted average of the risks of all included assets” (Jones, 1994, p. 573). The risk of a portfolio depends not only on the risks of every included asset, but also on the mutual dependence (interdependence) between and among the assets. Cash position is assumed to have a risk of 0. An example approach to measure portfolio risk is described below, where there are two additive terms – one for weighted average of the risks of included assets and the other for calculating pair by pair the interdependence of the assets.

$$V_p(t) = \sum_{i=1}^k w_i(t)^2 \cdot V_i(t) + \sum_{i=1}^k \sum_{j=1}^k w_i(t) \cdot w_j(t) \cdot \rho(V_i(t), V_j(t))$$

where:

$V_p(t)$  - risk of the portfolio  $p$  at the discrete time  $t$

$V_i(t)$  - risk of the asset  $i$  at the discrete time  $t$

$\rho(V_i(t), V_j(t))$  - measure for interdependence of the assets  $i$  and  $j$  at the discrete time  $t$