Reinforcement Learning Portfolio Optimization for FX Trading

Krzysztof Wojdalski

2018-07-20

Contents

Abstract

The master's thesis is about the Reinforcement Learning application in the foreign exchange market. The author starts with describing the FX market, analyzing market organization, participants, and changes in the last years. He tries to explain current trends and the possible directions. The next part consists of theoretical pattern for the research - description of financial models, and the AI algorithms. Implementation of the RL-based approach in the third chapter, based on Q-learning, gives spurious results.

Introduction

Trading floors are usually perceived as places with noisy shouting people, frenzy, and a lot of bizarre situations. This is justified since the reality was pretty much as above, but 30 years ago. Financial markets have been constantly interested in inventions from computer science world. It is part of the game - having a competitive edge could result in abnormal return. Hence, finance was often the first industry that was adopting state-of-the-art technologies. In its essence, the discipline was always extremely focused on increasing efficiency. Even now, blockchain, one of the most sought-after technologies, is expected to succeed in finance first. For trading entities the infinite goal is to maximize profits. There are many ways to achieve it, e.g. by passive methods such as buy-and-hold, however one of the most promising and emerging categories are AI-based strategies. This shift of replacing human with robots for the decision process is more than likely to take place over the next years or decades. It contradicts with the cliches mentioned at the beginning. Even though machine learning as a discipline is nothing new - the fundamentals come from the 50s, the industry still has not embraced it very widely. For

instance, in FX market only a few biggest names managed to prepare machine learning-based trading systems.

The majority of systems described in the literature aim to maximize trading profits or risk-adjusted measures, such as the Sharpe ratio. Many attempts have been made to come up with a consistently profitable system and inspiration has come from different fields ranging from fundamental analysis, econometric modelling of financial markets, to machine learning. A few attempts were successful and those that seemed most promising often could not be used to trade actual markets due to associated practical disadvantages. Among others these included large drawdowns in profits and excessive switching behaviour resulting in high transaction costs. Professional traders have generally regarded those automated systems as risky in comparison to the returns they were able to deliver by themselves. Even if a trading model was shown to produce an acceptable risk-return profile on historical data there was no guarantee that the system would keep on working in the future. It could cease working precisely at the moment it became unable to adapt to changing market conditions.

aims to deal with the above problems to obtain a usable, fully automated and intelligent trading system. To accomplish this, a risk management layer and a dynamic optimization layer are added to known machine learning (RL) algorithms. The middle layer manages risk in an intelligent manner so that it protects past gains and avoids losses by limiting or even shutting down trading activity in times of high uncertainty. The top layer dynamically optimizes the global trading performance in terms of a trader's risk preferences by automatically tuning the system's hyper-parameters.

While the machine learning system is designed to learn from its past trading experiences, the optimization overlay is an attempt to adapt the evolutionary behaviour of the system and its perception of risk to the evolution of the market itself.

This research departs from a similar principle by developing a fully layered system where risk management, automatic parameter tuning and dynamic utility optimization are combined. The machine learning algorithm combined with the dynamic optimization is termed adaptive reinforcement learning.

The first part consists of the introduction to the problem. It outlines the whole concept of the AI-related fields in finance. It brought up historical background of finance and computer sciences, and its interdependency.

This chapter starts with the outline of selected papers from quantitative finance. It includes both classic models, such as CAPM, the gold standard in equity research, and modern ones. The part is descriptive as it regards implicit pros and cons of financial models. The literature review is specifically about algorithmic trading and the methodology of other similar researches, e.g. Sakowski et al. (2013). Second chapter also includes the explanation of machine-learning algorithms underlying (or inspiring) the trading system.

The third chapter starts with detailed explanation of the research. The main hypotheses are as follows:

- Algorithms based on artificial intelligence can be fruitful for investors by outperforming benchmarks in both risk and return;
- Better performance turns out to be true in high-frequency trading and on longer period intervals;
- Algorithms can learn how to spot overreacting on markets and choose the most under/overpriced security by exploiting time series analysis tools.

It contains description of the methodology - all formulas and steps that directed to final results. It looks at each layer of the trading system. In 3.1 the modifications to the standard algorithm

are set out and in 3.2 and 3.3 the risk management and optimization layers are explained.

The used algorithms are based on dynamic optimization approach. Besides value function based on Differential Sharpe Ratio, there will be several indicators, e.g. RSI, which serve as a base for decision taking of the algorithm. The methodology will include transactional costs, so that the optimization is going to be implemented in a real-like environment

The value function will be based by several statistics, such as the Sharpe and the Differental Sharpe Ratio to capture both risk and return. The output of my algorithm will be a set of the agents' actions in the form of $-1, 0, 1$. Moreover, I will enclose all elements of Reinforcement Learning-model, i.e.:

- environment
- states - cumulated returns, and risk measures, such as the Sharpe ratio, MD, MDD, the Sortino ratio.
- actions -
- rewards

In the last part of the chapter the performance of the trading system (RL-based agents) is demonstrated and examined against several benchmarks: * Buy-and-hold strategy what means holding long-position in selected currency pairs. * Random actions - this part of the algorithm will generate random values in a domain of $\{-1, 0, 1\}$. These values will serve as a position in the underlying pairs. The benchmark will not include any transactional costs as this obvious that this extreme case would have an enormous cumulated transactional cost (position would change in $\frac{2}{3}$ of states).

The final section outlines conclusions. It compares the results with similar works and suggests possible directions to extend the research. It addresses such questions as: * What can be additionally implemented? * What were limitations and what must be done to avoid them in the future?

Data

Datasets used for the purpose of this workpaper are from the following databases:

- Thomson Reuters Tick Database for FX market
- Tick database from a vendor aggregating quotations from liquidity providers (TradAir, yet to be discussed)

Formulas

FX Market Organization

Explaining the institutional structure of FX market requires introducing formal definitions of market organization. According to Lyons Lyons (2002), these are:

- Auction market - a participant can place a market and a limit order. The first action is aimed at buying X units at the best price. Alternatively, limit orders set a threshold, i.e. they are executed only if the market quotations reach a certain price. Limit orders are aggregated into an order book
- Single dealer market - in this kind of market organization, there is just one dealer. It is obliged to quote an asset, i.e. to match demand and supply. Its quotations are always

the best bid and the best ask. The main task is to manage the risk to make profit off his spread.
- Multiple dealer market - it is extension of single dealer market. There is more than one dealer and they compete against each other. It might be centralized or decentralized. In the first version, all dealers are put into the same location while in the second it is not the case. When the market is decentralized, it is possible for price takers to gain profits by arbitrage transactions.

The FX market is a kind of decentralized multiple-dealer market. There is no single indicator that would show the best bid and the best ask. Hence, the market transparency is low. It is especially important at tail events. It is hard to determine when the market was at a given time and findings are usually spurious. The foreign exchange market is perceived as the largest and most liquid one, with a year-on-year turnover of €69 trillion.

The FX market is an over the counter, global (OTC) market, i.e. participants can trade currencies with relatively low level of legal obstacles. The market core is built up by the biggest banks in the world. Hence, the FX organization is often referred as an inter-bank market. The participants of the FX market differ by access, spreads, impact, turnover they generate, order size, and purpose. They can be divided into five main groups:

- Central Banks - Central banks have the biggest market impact - they control money supply, interest and reserve rates. Through their set of tools, they can strengthen or weaken local currency. In the developed markets, their turnover is rather small due to the fact that intervenings in the open market happen rarely, but order size is usually bigger than for other four groups due to the effect they want to achieve.
- Commercial Banks - Most of the flow in the market belongs to commercial banks. Although the environment in which FX trading occurs is highly dispersed in terms of location, over 85% of flow is generated by top 15 banks, as seen in 1. It can be observed even for currencies that the banks do not have real interest in. It means that in fact banks stay with flat position. Their role is to gain a difference between quotations from their own liquidity pool and their quotations for a counterparty. Hence, over the years, the market have changed dramatically. Even though turnovers are higher than ten years, market practitioners tend to claim that liquidity is much worse. It is mostly due to the fact that new regulation, internal and external (The Basels), have been introduced. Banks are required to stay with rather small positions, especially in non-G10 pairs. Their approach to risk is much more conservative than it used to be.
- Non-bank Financial Institutions - In the state of new regulations, their market significance is on the rise, e.g. hedge funds might serve as liquidity providers for banks. The non-bank financial institutions category is very broad and entities in it are very heterogenous
- Commercial Companies - their conditions as price takers are significantly worse than commercial banks due to the fact they trade bigger size and mainly hedge their main business.
- Retail Traders - their main purpose is to speculate. The conditions they receive from financial institutions are generally worse but it might not be always be the case.

% latex table generated in R 3.4.0 by xtable 1.8-2 package % Fri Jul 20 16:55:36 2018

In last years, there have been observed shifting towards eFX. Commercial banks, as mentioned in the previous subsection, are subject to new regulations. Therefore, right now they are more concerned about increasing their turnover than benefiting off a good view (speculation). eFX helps in this goal. It requires more technology while a number of traditional dealers is effectively reduced. The activity require quantitative analysts, “quants”, who can manage pricing engines in order to maximize profit while staying in the risk threshold. Over 4 years, eFX gained 13 percent point and in 2015 for the first time surpassed voice trading, with 53.2% of client flow

	Rank	Bank	MarketShare
1	1	Citi	0.16
2	2	Deutsche Bank	0.15
3	3	Barclays	0.08
4	4	JPMorgan	0.08
5	5	UBS	0.07
6	6	Bank of America Merrill Lynch	0.06
7	7	HSBC	0.05
8	8	BNP Paribas	0.04
9	9	Goldman Sachs	0.03
10	10	RBS	0.03
11	11	Societe Generale	0.02
12	12	Standard Chartered	0.02
13	13	Morgan Stanley	0.02
14	14	Credit Suisse	0.02
15	15	State Street	0.02

Table 1: Market share of top financial institutions in FX trading in 2014

share JeffPatterson2015 Chung2015.

The following chapter introduces articles that correspond with the subject of the current thesis and are considered as fundamentals of modern finance. Specifically, the beginning contains financial market models. The next subchapter includes basic investment effectiveness indicators that implicitly or explicitly result from the fundamental formulas from the first subchapter.

Selected financial market models and theory

Works considered as a fundament of quantitative finance and investments are Sharpe Sharpe (1964), Lintner Lintner (1965), and Mossin Mossin (1966). All these authors, almost simultaneously, formulated Capital Asset Pricing Model (CAPM) that describes dependability between rate of return and its risk, risk of the market portfolio, and risk premium. Assumptions in the model are as follows:

- Decisions in the model regard only one period,
- Market participants has risk aversion, i.e. their utility function is related with plus sign to rate of return, and negatively to variance of portfolio rate of return,
- Risk-free rate exists,
- Asymmetry of information non-existent,
- Lack of speculative transactions,
- Lack of transactional costs, taxes included,
- Market participants can buy a fraction of the asset,
- Both sides are price takers,
- Short selling exists,

Described by the following model formula is as follows:

$$E(R_P) = R_F + \frac{\sigma_P}{\sigma_M} \times [E(R_M) - R_F]$$

where:

- $E(R_P)$ – the expected portfolio rate of return,
- $E(R_M)$ – the expected market rate of return,

- R_F – risk-free rate,
- σ_P – the standard deviation of the rate of return on the portfolio,
- σ_M – the standard deviation of the rate of return on the market portfolio.

$E(R_P)$ function is also known as Capital Market Line (CML). Any portfolio lies on that line is effective, i.e. its rate of return corresponds to embedded risk. The next formula includes all portfolios, single assets included. It is also known as Security Market Line (SML) and is given by the following equation:

$$E(R_i) = R_F + \beta_i \times [E(R_M) - R_F]$$

where:

- $E(R_i)$ – the expected i-th portfolio rate of return,
- $E(R_M)$ – the expected market rate of return,
- R_F – risk-free rate,
- β_i – Beta factor of the i-th portfolio.

The Modern Portfolio Theory

The following section discuss the Modern Portfolio Theory developed by Henry Markowitz Stulz (1995). The author introduced the model in which the goal (investment criteria) is not only to maximize the return but also to minimize the variance. He claimed that by combining assets in different composition it is possible to obtain the portfolios with the same return but different levels of risk. The risk reduction is possible by diversification, i.e. giving proper weights for each asset in the portfolio. Variance of portfolio value can be effectively reduced by analyzing mutual relations between returns on assets with use of methods in statistics (correlation and covariance matrices). It is important to say that any additional asset in portfolio reduces minimal variance for a given portfolio but it is the correlation what really impacts the magnitude. The Markowitz theory implies that for any assumed expected return there is the only one portfolio that minimizes risk. Alternatively, there is only one portfolio that maximizes return for the assumed risk level. The important term, which is brought in literature, is the effective portfolio, i.e. the one that meets conditions above. The combination of optimal portfolios on the bullet.

Bullet figure

The Markowitz concept is determined by the assumption that investors are risk-averse. This observation is described by the following formula:

$$E(U) < U(E(X))$$

where:

- $E(U)$ – the expected value of utility from payoff;
- $U(E(X))$ – utility of the expected value of payoff.

The expected value of payoff is given by the following formula:

$$E(U) = \sum_{i=1}^{n} \pi_i U(c_i)$$

where:

- π_i – probability of the c_i payoff,

- $U(c_i)$ – utility from the c_i payoff.

One of the MPT biggest flaws is the fact that it is used for ex post analysis. Correlation between assets changes overtime so results must be recalculated. Real portfolio risk may be underestimated. Also, time window can influence the results.

The efficient market hypothesis

In 1965, Eugene Fama introduced the efficient market term (**???**). Fama claimed that an efficient market is the one that instanteneously discounts the new information arrival in market price of a given asset. Because this definition applies to financial markets, it had determined the further belief that it is not possible to beat the market because assets are perfectly priced. Also, if this hypothesis would be true, market participants cannot be better or worse. Their portfolio return would be a function of new, unpredictable information. In that respect, the only role of an investor is to manage his assets so that the risk is acceptable.

Selected investment performance measures

Introduced articles does not include any indicator that would explicitly measure portfolio management effectiveness. Equations that result from the authors' work are important because some of further developed measures are CAPM-based. The most known are the Sharpe ratio, the Treynor ratio, and the Jensen's alpha. Popularity of these indicator comes from the fact that they are easy to understand for the average investor. (**???**) In (**???**), the author introduced the $\frac{R}{V}$ indicator, also known as the Sharpe Ratio (S), which is given by the following formula:

$$S_i = \frac{E(R_i - R_F)}{\sigma_i}$$

where:

- R_i – the i-th portfolio rate of return,
- R_F – risk-free rate
- σ_i – the standard deviation of the rate of return on the i-th portfolio.

Treynor (Treynor1965) proposed other approach in which denominator includes β_i instead of σ_i. The discussed formula is given by:

$$T_i = \frac{R_i - R_F}{\beta_i}$$

where:

- R_i – the i-th portfolio rate of return,
- R_F – Risk-free rate
- β_i – Beta factor of the i-th portfolio.

Both indicators, i.e. S and T are relative measures. Their value should be compared with a benchmark to determine if a given portfolio is well-managed. If they are higher (lower), it means that analyzed portfolios were better (worse) than a benchmark. The last measure, very popular among market participants, is the Jensen's alpha. It is given as follows:

where:

- R_i – the i-th portfolio rate of return,
- R_F – Risk-free rate
- β_i – Beta factor of the i-th portfolio.

The Jensen's alpha is an absolute measure and is calculated as the difference between actual and CAPM model-implied rate of return. The greater the value is, the better for the i-th observation.

The differential Sharpe ratio - this measure is a dynamic extension of Sharpe ratio. By using the indicator, it can be possible to capture a marginal impact of return at time t on the Sharpe Ratio. The procedure of computing it starts with the following two formulas:

$$A_n = \frac{1}{n}R_n + \frac{n-1}{n}A_{n-1}$$

$$B_n = \frac{1}{n}R_n^2 + \frac{n-1}{n}B_{n-1}$$

At $t = 0$ both values equal to 0. They serve as the base for calculating the actual measure - an exponentially moving Sharpe ratio on η time scale.

$$S_t = \frac{A_t}{K_\eta\sqrt{B_t - A_t^2}}$$

where:

- $A_t = \eta R_t + (1-\eta)A_{t_1}$
- $B_t = \eta R_t^2 + (1-\eta)B_{t_1}$
- $K_\eta = (\frac{1-\frac{\eta}{2}}{1-\eta})$

Using of the differential Sharpe ratio in algorithmic systems is highly desirable due to the following facts (**???**):

- Recursive updating - it is not needed to recompute the mean and standard deviation of returns every time the measure value is evaluated. Formula for A_t (B_t) enables to very straightforward calculation of the exponential moving Sharpe ratio, just by updating for R_t (R_t^2)
- Efficient on-line optimization - the way the formula is provided directs to very fast computation of the whole statistic with just updating the most recent values
- Interpretability - the differential Sharpe ratio can be easily explained, i.e. it measures how the most recent return affect the Sharpe ratio (risk and reward).

The drawdown is the measure of the decline from a historical peak in an asset. The formula is given as follows:

$$D(T) = \max\{max_{0,t\in(0,T)}X(t) - X(\tau)\}$$

The Sterling ratio (SR)

The maximum drawdown (MDD) at time T is the maximum of the Drawdown over the asset history. The formula is given as follows:

$$MDD(T) = \max_{\tau\in(0,T)}[\max_{t\in(0,\tau)} X(t) - X(\tau)]$$

In this chapter term **machine learning** and its subfields are explained. Discussion also contains possible applications for trading financial instruments.

Machine Learning

There are many definitions of machine learning sources provide as the field evolves. In this subchapter, the author has selected arbitrarly definitions that accurately captures the spirit of the discipline. What is machine learning then? The most accepted and widely used definitions are as follows:

- "Field of study that gives computers the ability to learn without being explicitly programmed." - Arthur Samuel, a pioneer in machine learning and computer gaming (1959)
- "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." - Tom Mitchell, a computer scientist and E. Fredkin University Professor at the Carnegie Mellon University (CMU)

Especially the latter is considered as an elegant and modern definition. Less formal, but also relevant remarks, comes from two renown authors of textbooks from the discipline:

- "Pattern recognition has its origins in engineering, whereas machine learning grew out of computer science. However, these activities can be viewed as two facets of the same field…" - Christopher Bishop
- "One of the most interesting features of machine learning is that it lies on the boundary of several different academic disciplines, principally computer science, statistics, mathematics, and engineering. …machine learning is usually studied as part of artificial intelligence, which puts it firmly into computer science …understanding why these algorithms work requires a certain amount of statistical and mathematical sophistication that is often missing from computer science undergraduates." - Stephen Marsland, https://www.amazon.com/dp/1420067184?tag=inspiredalgor-20

https://machinelearningmastery.com/what-is-machine-learning/

Altough there are many more concepts, ideas, and comments as to what exactly machine learning is, the general goal is the same: Machine learning is about building such models that resemble the reality to a sufficient extent, are optimal in terms of a value function and can be later used for predictions on new data.

Why is machine learning important?

Machine learning helps in solving problems that are difficult or even impossible to solve in a determinisic way. Sometimes variables can be missing or observed values can contain an embedded error. Traditional models are often prone to be under- or overdetemined, i.e. they might not generalize well or are too general. An appropriate machine learning model should contain approximate solution containing only relevant parts.

Classification of machine learning algorithms

In machine learning (ML), tasks are classified into broader categories based on how learning/feedback (P) is received and/or what kind of problem they solve. One can distinct the following ones:

- Supervised Learning - the whole set $(Y_t; X_{t,1}, ..., X_{t,n})$ is available. The goal is to model the special variable Y_t using a subset of X_t variables, i.e. find a functional relationship $Y_t = f(\mathbb{X}_{\approx})$ between the input variables and the output variables which minimizes a predefined

loss function $g(f(\mathbb{X}_t); Yt)$. The structural form of this relationship is constrained by the class of functions considered. For example we can assume that there is a linear relationship between input and output variables and a square loss function, then the problem becomes:

$$\min_{b1...bn} \mathbb{E}[(Y_t - (b_1 X_{t,1} + \cdots + b_n X_{t,n}))^2]$$

The utilized estimation method above is called least squares method for linear regression. Even though it is considered a simple one, it sometimes provides sufficient results. Other popular methods for supervised learning are:

- K-nearest neighboors, Neural Networks,
- SVM - Support Vector Machines,
- Random Forests
- Unsupervised learning - it is the category that deals with only $\mathbb{X}_{\approx}$ set. In other words, The goal is to find patterns among the dataset and categorize observations. The most popular methods are:
- Clustering - based on finding groups of instances which are similar as possible to observations from the same groups while as different as possible to observations from other ones
- Feature extraction - this subcategory of unsupervised learning consists of methods for extracting relevant variables from a set of variables $\mathbb{X}_t$. Often, a subset of a dataset can contain a similar amount of information as the original one while reducing dimensionality so that a model computation is much faster and efficient. improves the model in Occam's Razor sense.
- Anomaly detection - this type helps in identification of observations that are outliers and should be carefully investigated. Sometimes the whole variable needs to be transformed or spotted observations must be removed due to their invalidity.
- Reinforcement Learning - an algorithms in reinforcement learning approach tries to maximize long term cumulated reward and interacts with the environment. Reinforcement learning and its variations are very useful when the solved problem is not stationary.Here the algorithm does not just try to structure in environment but it interacts with it in order to maximize long-term cumulative reward. Especially suited when the problem is not stationary and relationships between variables change over time. Y_t is a value function that is aimed to help in solving only specific types of problems. More on reinforcement learning is in the next section.

Reinforcement Learning (RL) is a subfield of machine learning that consists of an agent which learns how to act in an unknown or not fully known environment. It is probably the most intuitive category of ML in terms of what people implicitly believe to be artificial intelligence. According to David Silver [ref], it captures influences from disciplines such as engineering, economics, mathematics, neuroscience, psychology and computer science.

The only feedback an agent receives is a scalar reward. The goal of it is to maximize long-run value function which consists of summed up discounted rewards in subsequent states. The goal of the agent is to learn by trial-and-error which actions maximize his long-run rewards. The environment changes stochastically and in some cases interacts with the agent. The agent must choose such a policy that optimizes amount of rewards it receives. The design must capture this fact by adjusting the agent so that it does not act greedily, i.e. it should explore new actions instead of exploiting existing optimal (possibly suboptimal) solutions.

Reinforcement learning algorithms can be classified into three general subcategories:

- Model Based - they are based on the idea that an model of the environment is known. Actions are chosen by searching and planning in this model.
- Value Based - modelfree - it uses experience to learn in a direct way from state-action values or policies. They can achieve the same behaviour, but without any knowledge on the world model an agent acts in. Given a policy, a state has some value, which is defined as cumulated utility (reward) starting from the state. Model-free methods are generally less efficient than model-based ones because information about the environment is combined with possibly incorrect estimates about state values. Moreover, the values for states are # to do methods, the transition matrix is not stationary
- Policy Based

https://www.princeton.edu/~yael/Publications/DayanNiv2008.pdf

Components of an reinforcement learning system

Reinforcement learning systems are developed to solve sequential decision making problems, to select such actions that eventually maximize cumulative discounted future rewards. In the following section the author explained components of reinforcement learning on the example of game of chess and trading

- Environment (E) - it defines what states and actions are possible. In the game of chess it is the whole set of rules and possible combination of figures on the chessboard. It must be stated that some states are not available and will be never reached. In trading such rules might constitute that for instance the only position an agent can take is 0 or 1, or that weights of assets in a portfolio must sum up to 1.

- State (s) - can be seen as a snapshot of the environment. It contains a set of information in time t that a RL agent uses to pick the next action. States can be terminal, i.e. the agent will no longer be able to choose any action. In such scenario they end an episode (epoch), a sequence of state-action pairs from the start to the end of the game. For a trading application, a state in time t can be a vector of different financial measures, such as rate of return, implied/realized volatility, moving averages, economics measures, technical indicators, market sentiment measures, etc.

- Action (a) - givn a current state the agent chooses an action which directs him into a new state, either deterministically or stochastically. The action choice process itself may also be deterministic or based on probability distributions. In the game of chess analogy, an action is to move a figure in accordance to the game's rules. In trading it could be for instance going long, short, staying flat, outweighing.

- Reward (r)

- Policy (π) - a policy is a mapping from state to action. It determines agent's choices and may be stochastic. Policies do not imply deterministic nature of the mapping. Even after countless number of episodes and states, there is a chance that an efficient RL algorithm will explore other states rather than by exploiting the then-optimal action

- Value Function - it is a prediction of future, usually discounted rewards. Value functions are used for determining how much a state should be desired by the agent. They depend on initial states (S_0), and a policy that is picked up by the agent. Every state should have an associated value, even if the path it's part of was never explored - in such cases they usually equal to zero. The general formula for value function is as follows:

$$V^{\pi} = \mathbb{E}_{\pi}[\sum_{k=1}^{\infty} \gamma^k r_{t+k} | s_t = s]$$

where γ is a discount factor from the range $[0; 1]$. It measures how much more instant rewards are valued. The smaller it is the more immediate values are relatively more relevant and cause algorithm to be more greedy. Sometimes γ is equal to 1 if it is justified by the design of the whole agent.

- Model (m) - a model shows the dynamics of environment, how it will evolve from S_{t-1} to S_t. Formally, it's a set of transition matrices:

$$\mathbb{P}^a_{ss'} = \mathbb{P}[s^{'}|s, a]$$
$$\mathbb{R}^a_s = \mathbb{E}[r|s, a]$$

where:

$\mathbb{P}_s s^{'} a$ is a matrix of probability of transitions from state s to state $s^{'}$ when taking action a. Analogously, $\mathbb{R}^a_s$ is an expected value of reward when an agent is in state s and taking action a

Research Objective

The primary research goal was to evaluate the Reinforcement Learning-based algorithm for multiasset trading. The main idea behind the algorithm deployment is that it can systematically outperform benchmarks in terms of selected risk and return measures. Designed trading system was aimed to spot non-trivial patterns in data, more efficiently than human, and exploit them accordingly.

In the project author wanted to assess the possibility of using a reinforcement learning agent for trading domain. The objectives are as follows:

- Implementation - consisting of creating reinforcement learning-based agent that are capable of trading financial instruments basing on time series tables
- Evaluation - testing if trading agents for out-of-sample periods can outperform benchmarks in the measures provided by the author
- Conclusion - answering the question if such approach might help in generating abnormal positive results and determining if the method can be feasible and efficient in real-like environment

Design of the research

The whole system can be divided into three main parts:

- Data preprocessing - taking FX data from Bloomberg with use of the dedicated API, parsing the data and adjusting it for the further analysis. The system is dedicated for currency trading, however with little adjustments it could fit in other asset classes as well.
- Variable extraction - not all preprocessed currency pairs are relevant and worth adding. For instance, if USD/CNH is highly correlated with USD/CNY it is senseless to add the latter to the portfolio. #TO DO
- State-action space - the extracted variables, based on time series for currency pairs, are merged into state space.

Lintner, John. 1965. "Security prices, risk, and maximal gains from diversifivation." *Jf* 20 (4): 587–615. doi:10.1111/j.1540-6261.1965.tb02930.x.

Lyons, Richard K. 2002. "The Microstructure Approach to Exchange Rates (a review)." *Financial Analysts Journal* 58 (5): 101–3. doi:10.2469/faj.v58.n5.2475.

Mossin, Jan. 1966. "Equilibrium in a capital asset market." *Econometrica* 34 (4): 768–83. doi:10.2307/1910098.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *The Journal of Finance* 19 (3): 425–42. doi:10.2307/2329297.

Stulz, Rene M. 1995. "American Finance Association, Report of the Managing Editor of the Journal of Finance for the Year 1994." *The Journal of Finance* 50 (3): 1013. doi:10.2307/2329297.