Portfolio Sold-Out Problem in Numbers for DeFi Lending Protocols

A Research Proposal of the Master's thesis

submitted by

Waralak Pariwatphan

under the supervision of

Yury Yanovich

PhD, Senior Research Scientist

in partial fulfillment of

the requirements for the degree of

Master's Educational Program: Data Science

to

Skolkovo Institute of Science and Technology

Skolkovo Institute of Science and Technology, Moscow, Russia

December 2022

Contents

1 Background and Problem Statement

In quantitative finance, portfolio optimization based on Markowitz's framework is well recognized [7]. The fundamental concept is to diversify financial assets in a portfolio in order to decrease portfolio risk and choose the most efficient portfolio. Blockchain is a notable technology that allows for the distribution of digital ledgers without the requirement for a central entity to keep its state secured. Tokenization, which can fractionize assets into tokens and also transform ownership asset rights into that digital token, is the important characteristic.

Decentralized Finance (DeFi) is a financial technology based on blockchain that eliminates the need for intermediary and centralized financial organizations in financial transactions. MakerDao is the world's leading DeFi that enables the generate stable coin Dai and allows users to lock their cryptocurrencies as collateral and borrow Dai. According to these blockchain-based loan portfolios, we aim to investigate the optimal portfolio tokenization by implementing tokenization algorithms and evaluating the tokenized asset fraction in the portfolio package.

2 Purpose and Definitions

The aim of the research is to analyze the MakerDao dataset and purpose the suitable numerical optimization for tokenized asset on real loan portfolio dataset.

3 Literature Review

3.1 Geometric Brownian motion

Geometric Brownian motion is a stochastic process used to represent random motions. It is similar to a random walk. It is commonly applied in the study of financial markets, such as stock prices, foreign exchange, financial assets, and so on. A Wiener stochastic process is the standard Brownian motion [11]. Let $W(t)$ is the value in $\mathbf{R}$ when $t \in [0, \infty)$, the properties of Wiener process are

1. $W(0) = 0$.

2. for all $0 \leq t_1 < t_2, W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$.

3. $W(t)$ has independent increments. For all $0 \leq t_1 < t_2 < ... < t_n$ the random variables

$$W(t_2) - W(t_1), W(t_3) - W(t_2), ..., W(t_n) - W(t_{n-1})$$

are independent.

4. $W(t)$ paths are continuous.

First passage times of levels [12] are applied in a various applications. The time when the process hits an arbitrary threshold a for the first time is a random variable, according to Brownian Motion theory. It is possible to assume that the process begins at a certain position $x_0 > 0$ and is bounded to the positive semi-axis by an absorbing barrier x = 0. Assume that $a > 0$ and let $T_a = \inf \{t \geq 0 : B_t = a\}$. Then we will have

$$\mathrm{P}(T_a < t) = \mathrm{P}\left(\sup_{s \leq t} B_s > a\right).$$

Applying the reflection principle of D. Andre, it follows that

$$\mathrm{P}(T_a < t) = 2\mathrm{P}(B_t \geq a).$$

Since

$$\mathrm{P}(B_t \geq a) = \frac{1}{\sqrt{2\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}} dx,$$

So,

$$P(T_a < t) = \int_0^t \frac{a}{\sqrt{2\pi s^3}} e^{-\frac{a^2}{2s}} ds$$

3.2 Tokenization

The OECD has proposed two methods of asset tokenization as following [1]:

Tokenization of real assets that exist off-the-chain is the process of representing an existing asset on a distributed ledger in digital platform [8]. Real assets can be tokenized on blockchain using smart contracts and used separately as tokens and vaults. Tokens on the blockchain can represent economic value and asset rights.

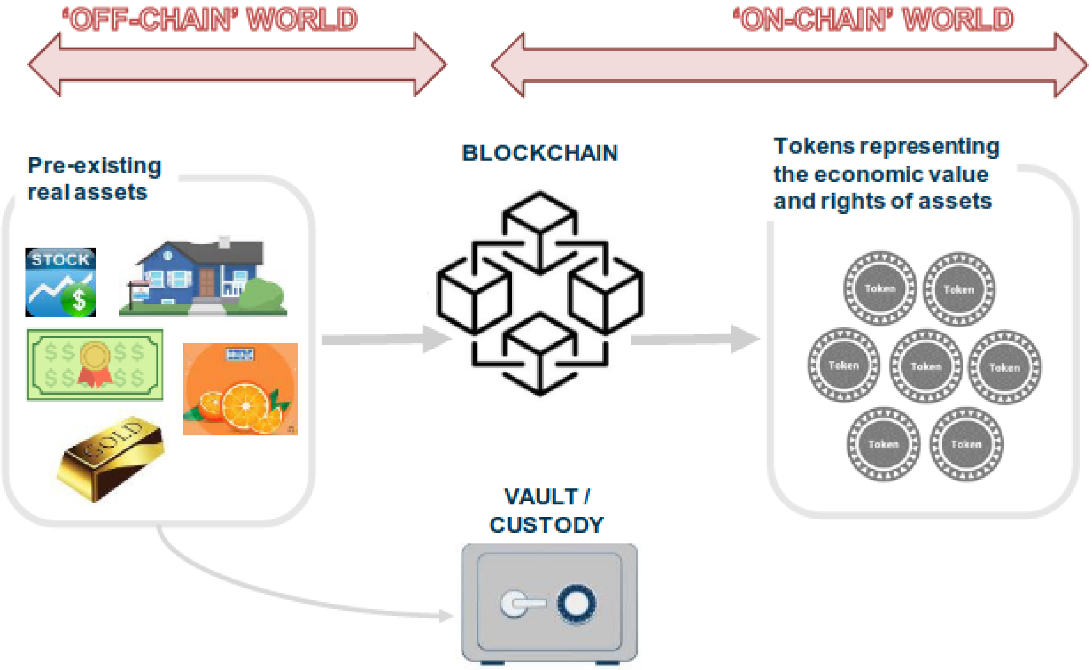

Figure 1: Tokenization of real assets that exist off-the-chain [1]

Tokenization of real assets that exist off-the-chain is the method of representing native tokens that are generated directly on-chain and distributed ledger, such as tokens issued in initial coin offerings (ICOs) that are created on the blockchain. ICOs are comprised of start-up enterprises creating digital tokens and distributing them to investors in return for funding and fundraising.

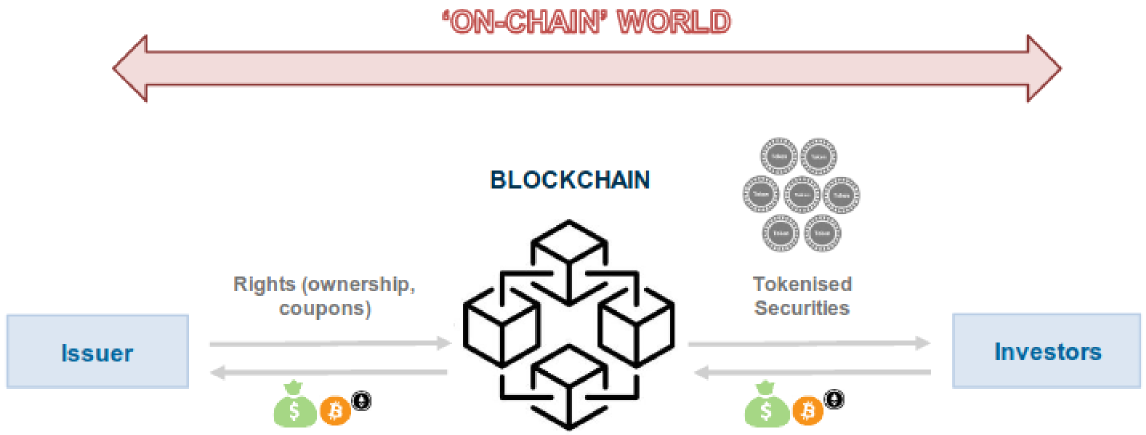

Figure 2: Tokenization of assets "native" to the blockchain [1]

3.3 Portfolio sold-out problem

According to the portfolio sold-out problem [4], the variance of the package $\vec{c}$ equals

$$\mathrm{V}(\overrightarrow{\mathrm{c}}) = \mathrm{Var}\ \overrightarrow{\mathrm{c}}^{T}\vec{\xi} = \overrightarrow{\mathrm{c}}^{\mathrm{T}}\mathbf{K}\overrightarrow{\mathrm{c}}.$$

A set of M packages $\mathbf{C}_M = (\vec{c}_1|\dots|\vec{c}_M) \in R^{N\times M}$ is the tokenization of the portfolio $(\vec{A},\vec{\xi})$ if $\sum_{m=1}^{M}\vec{c}_m \leq \vec{A}$.

The variance V of tokenization $\mathbf{C}_{\mathrm{M}}$ is the maximum variance of its packages:

$$\mathrm{V}\left(\mathbf{C}_{\mathrm{M}}\right) = \max_{\mathrm{m}\in\overline{\mathrm{M}}} \mathrm{V}\left(\overrightarrow{\mathrm{c}}_{\mathrm{m}}\right).$$

For a given portfolio $(\overrightarrow{\mathrm{A}},\vec{\xi})$ and a variance threshold $\sigma^2 > 0$, the portfolio sold-out problem is

$$\mathrm{M} \rightarrow \max_{\mathrm{M},\mathbf{C}_{\mathrm{M}}:\mathrm{V}(\mathbf{C}_{\mathrm{M}})\leq\sigma^2}.$$

Portfolio Sold-Out Problem: Special Cases

By assets:

- Homogeneous: $\mathbf{K} = \sigma_0^2\mathbf{I}_{\mathrm{N}}$
- Independent : $\mathrm{K}_{\mathrm{ij}} = 0$ for $\mathrm{i} \neq \mathrm{j}$
- General : any $\mathbf{K}$ is allowed

By packets:

- Discrete : $\mathbf{C}_{\mathrm{M}}$ is Boolean matrix
- Continuous: $\mathbf{C}_{\mathrm{M}}$ is real matrix

	Discrete	Continuous
Homogeneous	optimal solution	optimal solution
Independent	to be solved	to be solved
General	to be solved	numerical solution (for example, CVXPY [5])

3.4 Tokenization algorithms

Homogeneous systems have useful properties for analysis and design, such as stabilization, convergence rates, and trajectory scalability [3]. The algorithms' outputs are the number of created tokens M and the composition matrix $C_M = (\vec{c}_1|\dots|\vec{c}_M)$ with M columns and

N rows. The amount of assets that make up into the packages $\bar{a} = \sum_{m=1}^{M} \vec{c}_m$ [4].

3.4.1 Discrete algorithm

For discrete homogeneous tokenization, there are the set of M packages

$$C_M = (\vec{c}_1 | \dots | \vec{c}_M) \in \left\{0, \frac{1}{k}\right\}^{N \times M}$$

with the variance reduction parameter k of the portfolio $(\vec{A},\ \mathrm{K})$ if

$$\sum_{m=1}^{M} \vec{c}_m \leq \vec{A}$$

and

$$\forall m \in \bar{M} : \|\vec{c}_m\|_1 = 1.$$

3.4.2 Continuous algorithm

For continuous homogeneous tokenization, there are the set of M packages

$$C_M = (\vec{c}_1 | \dots | \vec{c}_M) \in \mathbf{R}^{N \times M}$$

with the variance reduction parameter k of the portfolio $(\vec{A},\ \mathrm{K})$ if

$$\forall m \in \bar{M} : \|\vec{c}_m\|_2 \leq \frac{1}{k} \wedge \|\vec{c}_m\|_1 = 1$$

3.4.3 Compact continuous algorithm

The portfolio sold-out problem formulation is

$$M \to \max_{M, \mathrm{C}_M} .$$

The constraints are:

$$\vec{c}^T \mathrm{K} \vec{c}_m \leq \sigma^2,$$

$$\sum_{m=1}^{M} \vec{c}_m \leq \vec{A},$$

$$\|\vec{c}\|_1 = 1.$$

There is a proof that an optimal matrix $\bar{C}_M$ with $\vec{c_1} = \vec{c_2} = ... = \vec{c_M} = constant$, where $\vec{\bar{c}} = \frac{1}{M}\sum_{m=1}^{M} \vec{c}_m$. Thus, original portfolio sold-out problem can be reformulated as

$$M \to \max_{\mathrm{C}_M}$$

with the constraints

$$\vec{a}^T K \vec{a} \leq \sigma^2 \cdot \|\vec{a}\|_1^2,$$

$$\vec{0} \leq \vec{a} \leq \vec{A}.$$

4 Methodology

The main milestones of this thesis research is consisted of:

4.1 Data Collection

The data will be collected using BigQuery public datasets. The blockchain transaction data was recorded in the *crypto-ethereum* dataset and this collection contains several tables relating to Ethereum footprints. BigQuery and SQL can access transaction data on smart contracts [2] and will be used to gather peer-to-peer lending data from MakerDAO in a variety of currencies, including ETH-A, BTC-A, ETH-B, SAI, BAT-A, and others. The challenge is to extract data from each smart contract since the blockchain platform is composed up of several contracts, each of which has a separate action and information.

4.2 Data Preparation

The BigQuery data is not in a human-readable format. Data transformation utilizing Web3 in Python and mapping with smart contract ABI [9] is a key step for converting raw data to human-readable data, extracting important information, and storing data in the format that can be used for the following step.

4.3 Computation of Financial Statistics

Financial statistics can be calculated in terms of Loss Given Default (LGD), Effective Interest Rate (EIR), and Probability of Default (PD). These statistics can be utilized to

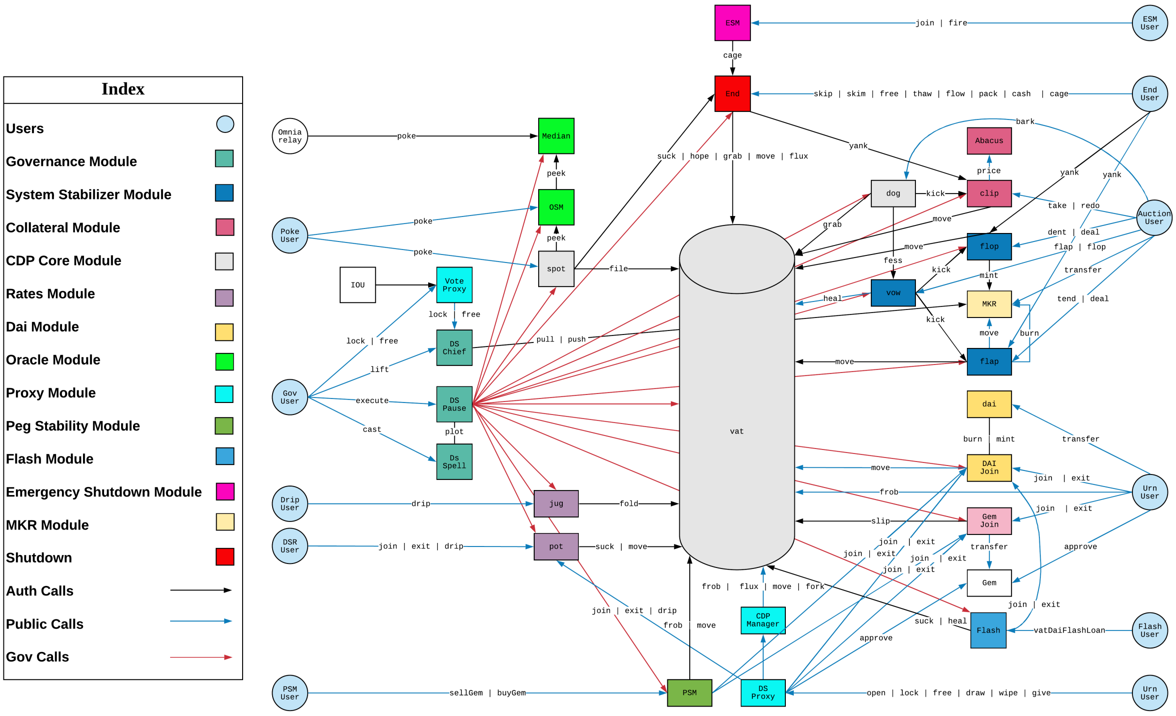

Figure 3: The Maker protocol smart contract modules [2]

represent an overview of transactions and as parameters for the following step.

4.4 Implementation Tokenization Algorithm

Tokenization Algorithms will be applied to the dataset to solve the portfolio sold-out problem and the solutions will be implemented. This research will concentrate on numerical optimization, for example, MEALPY [13], Second-order cone programming (SOCP) [6], Interior-Point Polynomial Algorithms in Convex Programming [10], which allow to specify the convex problem in intuitive syntax and other algorithms to determine the optimal portfolio.

5 Techniques and Tools

BigQuery is mainly used for collecting the data. **Python3**, especially Numpy, Web3, MEALPY, CVXPY, Matplotlib, and other implemented optimization algorithms are mainly used for implementing and presenting the results.

6 Work Plan

Plan of Work / Period	Year 2023									
	01	02	03	04	05	06	07	08	09	10
Data collection from BigQuery through Web3 and transformation of raw data to human-readable data										
Data preparation and financial statistics computation (LGD, EIR, and PD)										
Using optimization methods and doing a thorough evaluation of several ways to determine the optimal portfolio										
Using tokenization methods, summarizing results, and drafting a report										
Pre-defense										
Defense										

7 Potential Impacts: innovation, research and education

Decentralized finance is gaining traction and plays an important role in traditional finance. Furthermore, tokenization is growing in popularity in terms of generating tokens that will be the foundation of the most of digital transactions in the near future. This research may be used as a resource for further research in the tokenized method as well as portfolio optimization, which will benefit financial institutions and boost innovation in the blockchain approach. Furthermore, this research can serve as a foundation for learning about portfolio optimization.

References

[1] *The tokenisation of assets and potential implications for financial markets, oecd blockchain policy series*, Tech. report, OECD, 2020.

[2] N. N. Author, *Vat - detailed documentation, the maker protocol's core accounting system*, 2022.

[3] Yatipa Chaleenutthawut, Viacheslav Davydov, Aleksandr Kuzmin, and Yury Yanovich, *Practical blockchain-based financial assets tokenization*, ICBTA 2021: 2021 4th International Conference on Blockchain Technology and Applications (2021), 51–57.

[4] Viacheslav Davydov and Yury Yanovich, *Optimal portfolio sold-out via blockchain tokenization*, IECC 2020: 2020 2nd International Electronics Communication Conference (2020).

[5] Steven Diamond and Stephen Boyd, *CVXPY: A Python-embedded modeling language for convex optimization*, Journal of Machine Learning Research **17** (2016), no. 83, 1–5.

[6] Alizadeh F. and Goldfarb D., *Second-order cone programming*, Math. Program., Ser B **95** (2003), 3–51.

[7] Jules C. Mba Herve M. Tenkam and Sutene M. Mwambi, *Optimization and diversification of cryptocurrency portfolios: A composite copula-based approach*, Appl. Sci. 2022, 12, 6408 (2022).

[8] Garrick Hileman and Michel Rauchs, *Global blockchain benchmarking study*, Tech. report, Cambridge Centre for Alternative Finance, Cambridge, UK, 2017.

[9] Yifei Huang, *Decoding ethereum smart contract data*, 2021.

[10] Yurii Nesterov and Arkadii Nemirovskii (eds.), *Interior-point polynomial algorithms in convex programming*, Society for Industrial and Applied Mathematics, 1994.

[11] Hossein Pishro-Nik (ed.), *Introduction to probability, statistics, and random processes*, Kappa Research, 2014.

[12] Albert N. Shiryaev (ed.), *Essentials of stochastic finance: Facts, models, theory*, World Scientific, 1999.

[13] Nguyen Van Thieu and Seyedali Mirjalili, *MEALPY: a Framework of The State-of-The-Art Meta-Heuristic Algorithms in Python*, June 2022.