

RESEARCH ARTICLE

Loan Portfolio Dataset From MakerDAO Blockchain Project

YATIPA CHALEENUTTHAWUT¹, VYACHESLAV DAVYDOV^{2,3}, MICHAEL EVDOKIMOV⁴,
SUDARUT KASEMSUK¹, STANISLAV KRUGLIK⁵, (Member, IEEE),
GRIGORII MELNIKOV⁶, AND YURY YANOVICH^{1,4}, (Member, IEEE)

¹Skolkovo Institute of Science and Technology, 121205 Moscow, Russia

²HSE Tikhonov Moscow Institute of Electronics and Mathematics, HSE University, 123458 Moscow, Russia

³Quicktoken Tech Ltd., Dubai, United Arab Emirates

⁴Faculty of Computer Science, HSE University, 109028 Moscow, Russia

⁵School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371

⁶B4B.World, 85310 Budva, Montenegro

Corresponding author: Yury Yanovich (y.yanovich@skoltech.ru)


The work of Vyacheslav Davydov was supported by the Basic Research Program at HSE University.

ABSTRACT Decentralized finance (DeFi) offers a range of financial instruments and services that leverage the capabilities of web3 technology. Maker protocol, which enables users to obtain loans backed by cryptocurrencies, is one of them. Unlike traditional banks, Maker's data is transparently recorded on the Ethereum blockchain. In this research paper, we focus on analyzing the lending aspect of Maker from a traditional finance perspective. To achieve this, we create a unique dataset with loan portfolios from the MakerDAO project, making it the first dataset of its kind in the DeFi field. This publicly available dataset contains essential financial characteristics related to borrowing, including balance, loss given default, annual equivalent rate, and probability of default. Additionally, we develop a specialized mathematical model tailored specifically to this project. This model allows us to estimate the probability of default by considering the presence of crypto-collateral and utilizing Brownian motion passage levels. The results of this study provide valuable insights into lending practices in DeFi projects. They also help bridge the gap between traditional finance and blockchain-based financial services.

INDEX TERMS Blockchain, decentralized finance, knowledge discovery, dataset, smart contract, data mining, Brownian motion.

I. INTRODUCTION

Financial services refer to the range of activities and products provided by financial institutions, such as banks, insurance companies, and investment firms, to meet the financial needs of individuals and businesses. These services encompass various aspects of managing money, including lending, investing, insurance, and risk management [1], [2]. Financial services are highly regulated to maintain confidence in the financial system, provide a financial stability and protect consumers [3], [4]. For example, the Basel framework (BF) is the full set of standards of the Basel Committee on Banking

The associate editor coordinating the review of this manuscript and approving it for publication was Mueen Uddin .

Supervision, which is the primary global standard setter for the prudential regulation of banks [5]. As of early 2024, 28 jurisdictions covering a half of the humanity population use the BF. The key quantities of a loan in BF are an interest rate, a loss given default, and a probability of the default. But the real bank data is a part of bank confidential data and may contain sensitive personal information [6], hence it is not openly available to compute the parameters.

Decentralized finance (DeFi)—peer-to-peer financial services on public blockchains [7], [8]—not only bring new web3-based services but also provide analogs of traditional financial instruments [9], [10]. One of them is Maker, a blockchain protocol that facilitates crypto-backed loans [11]. The set of smart contracts implement Maker

protocol on the Ethereum blockchain, and a decentralized autonomous organization (DAO) named MakerDAO governs the project, including economic parameters assignment. Hereafter, we will use the terms MakerDAO and Maker interchangeably to refer to both the project and the protocol.

Maker's smart contracts are deployed on the Ethereum blockchain, making all related transactions visible to everyone. These transactions contain financial information, including user operation names and amounts. While it is technically possible to conceal this information using zero-knowledge proofs [12], [13], [14], doing so would greatly complicate and slow down the protocol, increase transaction fees, and reduce transparency for participants, ultimately making the project less reliable. Maker does not encrypt transaction data; the only hidden information relates to the real-world entities behind the protocol users identifiers. As a result, financial information can be extracted from the protocol. While user identifiers are visible, actual names are not. Hereafter, we will use the terms MakerDAO and Maker interchangeably to refer to both the project and the protocol.

Despite the transparency of transactions in blockchain-based projects, they lack regulation and standards as noted in previous studies [15], [16]. The goal of the current research is to analyze lending in the DeFi project Maker from a traditional finance perspective. The outcome is twofold: we provide a real lending portfolio dataset and equip it with standard banking numerical parameters.

The main contributions of our work are summarized below

- 1) We introduced a novel loan portfolio dataset obtained from the MakerDAO project, making it the first dataset of its kind in the field of DeFi. The dataset, along with utility functions for easy access, is publicly available at [17]. However, it should be noted that the dataset is currently limited to the ETH-A lending program.
- 2) We incorporated borrowing-driven financial characteristics, such as balance, loss given default, annual equivalent rate, and probability of default, into the aforementioned dataset.
- 3) We developed a project-specific mathematical model to estimate the probability of default. This model takes into account the presence of crypto-collateral and utilizes Brownian motion passage levels to provide a comprehensive understanding of both individual loan defaults and the correlation among different loans.

The rest of the paper is organized as follows. Section II provides an overview of the related papers. In Section III, we present a comprehensive perspective on the Maker protocol from the borrower's point of view. Section IV introduces mathematical models that describe the financial characteristics of loans. The structure of the dataset is outlined in Section V. In Section VI, a quantitative analysis is conducted to demonstrate the practicality of the collected data and the effectiveness of the proposed computational models. Finally, Section VII provides concluding remarks for the paper.

II. RELATED WORK

Banks use a variety of tools to maintain reasonable risk levels and increase efficiency, including regulator-required frameworks like the Basel framework [5] and machine learning models [18], [19]. Loan portfolio data has also attracted researchers' attention, with some studies accessing proprietary data that is not publicly available. For example, papers [20], [21] examine the impact of loan portfolio diversification on risk and capital efficiency based on German and Australian large banks, respectively. They have an access to more than thousand individual bank portfolios over seven years. Various machine learning techniques to predict non-performing loans using a portfolio dataset provided by a bank for four years, consisting of 181 thousand borrowers and hundreds of features, are compared in [22]. The paper [23] applies random forest to classify non-performing loans for Indonesia's bank loan dataset with 3300 borrowers and 12 features.

Several classic finance datasets are publicly available, such as Home Credit Default Risk on Kaggle, which challenges participants to predict the probability of default among 307 thousand debts using 239 features [24]. The UC Irvine Machine Learning repository contains several credit datasets, with the Taiwan credit card defaults dataset being the largest, containing 30 thousand debts and 24 features [25]. The peer-to-peer lending platform Lending Club provides a dataset containing 887 thousand debts collected from 2007 until 2015 with 79 features to predict the probability of the default [26].

In addition to traditional finance, there have been several studies on decentralized finance (DeFi) and its potential impact on traditional finance. For instance, the risk model of DeFi money lending was analyzed using tools from modern portfolio theory in papers [27], [28]. The challenges and opportunities of DeFi in the financial industry were analyzed in [29]. The risks and benefits of DeFi from the perspective of financial intermediation were analyzed in [30]. Also, the latter paper proposed the framework for analyzing DeFi projects.

While available data from traditional banking are limited due to trade secrecy and privacy concerns, DeFi loan data is openly available on public blockchains. Several studies have been conducted on DeFi lending platforms such as Maker, AAVE, Compound, and Spark Lend [11], [31], [32], [33], including their data collection, economic parameters estimation, and risk management. For example, the paper [34] analyzes data from the decentralized Ethereum protocol called Compound, using a relational database and providing statistical details to facilitate further analysis. The paper [31] assess the stability of the DAI stablecoin of the Maker project over the course of its first year deployment, including the cryptocurrency crisis in March 2020. The issue of high collateral requirements for blockchain-based loans using cryptocurrencies as collateral due to their high volatility is discussed in [35], and the authors propose a solution to make loans more accessible by offering

lower collateral requirements while keeping risk for lenders bound.

DeFi possesses unique operational characteristics. For instance, all computations are executed lazily and only upon a request. However, it is worth noting that the volume of centralized crypto finance surpasses that of the decentralized counterpart. Consequently, DeFi relies on oracles to incorporate centralized rates [36], [37], [38]. Oracles are compensated for their transactions and receive interest for their services. Typically, platforms obtain values from oracles at desired intervals. Nevertheless, in times of crisis [39], delays may occur, leading to deviations and unstable values as shown in [40]. Moreover, the unique features of DeFi pose challenges for new risk models. While models for the entire system exist, for example, the stochastic model for collateral-based stablecoins [41], models specifically addressing individual debts are yet to be presented.

However, there has been limited research on providing a real lending portfolio dataset for Maker and equipping it with standard banking numerical parameters. Such a dataset could be useful for both academic research and practical applications, such as risk management and portfolio optimization in Maker lending.

III. MAKER PROTOCOL FOR BORROWER

The Maker Protocol [11] operates using the native DAI token, which has a one-to-one soft peg to the United States dollar and is an ERC-20 token [42]. The protocol allows for collateral-secured DAI debts, with loan terms such as financial parameters. DAO mechanism to modify certain debts is included in corresponding smart contract. Financial parameters, such as the lending interest rate f (the multiplier applied to the loan balance over time) and the liquidation ratio r (the minimum allowed ratio of the locked collateral value to the debt value), are examples of these loan terms. Users can deposit Ethereum native cryptocurrency called Ether (ETH) or other tokens into their instance of a specific smart contract (Vault) and use them as collateral to mint DAI debt.

Let's consider a borrower's workflow in the Maker Protocol. The borrower starts by creating a Vault and depositing supporting collateral. The Maker Protocol's Oracle then evaluates the collateral, providing a real-time price feed for each asset. Based on the current market value of the collateral and the chosen borrowing program, the protocol calculates the maximum amount of DAI that can be borrowed. For example, ETH-A, ETH-B, and ETH-C are different borrowing programs with ETH as collateral, each with its own set of parameters and risk profiles. ETH-A is the original and most commonly used program, while ETH-B and ETH-C were introduced to offer additional options for users with different risk tolerances or preferences.

Once the borrower has minted DAI, they can use it for any purpose. The borrower is responsible for repaying the loan with interest, which is calculated based on the lending interest rate and duration of the loan. They can fully or partially repay the loan at any time, borrow more up to

the maximum permitted collateral program and size of the collateral amount, or increase or decrease the amount of collateral.

If the value of the collateral falls below a certain threshold, the Vault is at risk of being liquidated. In this case, the MakerDAO system will automatically initiate a liquidation process, which involves selling off a portion of the collateral to cover the outstanding debt. The borrower can not interact with the Vault under the liquidation process. The liquidation process is designed to be fast and efficient, with the goal of minimizing losses for both the user and the MakerDAO system. When a Vault is liquidated, the collateral is sold through an auction, allowing users to bid on the collateral using DAI. The auction is competitive, with bidders offering progressively lower prices until the collateral is sold.

If the auction is successful and the collateral is sold for a price that covers the outstanding debt, the remaining DAI is returned to the user. If the auction is unsuccessful and the collateral is not sold for a sufficient price, the MakerDAO system may take a loss on the liquidation. The resulting penalty for the borrower is flexible but usually ranges from 10% to 33%.

All actions involving the Vault and system parameters are recorded as plaintext Ethereum blockchain transactions. However, these transactions may be challenging for the general audience to understand due to Maker's use of technical terms such as `ilk`, `frob`, and `art`. To address this issue, we aim to present the loan portfolio dataset from the Maker project in a more accessible format.

IV. MATHEMATICAL MODELS

This study focuses on a single type of collateral. To obtain a loan in the Maker protocol, a user must have a Vault. Vaults can be associated with single user only and cannot be transferred. Ethereum addresses, which are 42-character hexadecimal strings, represent users. While an address does not reveal information about the real-world owner, some users may indirectly or directly disclose their identity. At the same time, a user can have multiple Vaults, and we keep track of which Vaults belongs to which user even in case of anonymous users.

We can determine if a Vault has an active loan at a given time by checking if its DAI debt is non-zero. Therefore, we define the beginning of a loan as when the DAI debt changes from zero to a positive number. We define the end of a loan as when the DAI debt becomes zero from any non-zero value. A loan can be active or ended, with the latter occurring either through successful repayment or liquidation. A single Vault can have multiple loans, and all loans in it have non-overlapping beginning-to-end time intervals. Now let's examine the financial characteristics of a loan.

A. BALANCE

When a user borrows DAI in the Maker project, they need to provide collateral. Without the loss of generality, we will

refer to the collateral asset as ETH. The loan starts at t_0 and lasts until T , which can be either the liquidation time, full repayment time, or maximum observed time if the loan is still active at the point T .

The amount of collateralization assets at any given time t is denoted by $a(t)$ (an example is shown in Figure 1). The blockchain records updates to the collateral balance as a piece-wise constant function, represented by update times τ and corresponding changes $\Delta a(\tau)$ due to collateral deposits or withdrawals, and liquidation processes.

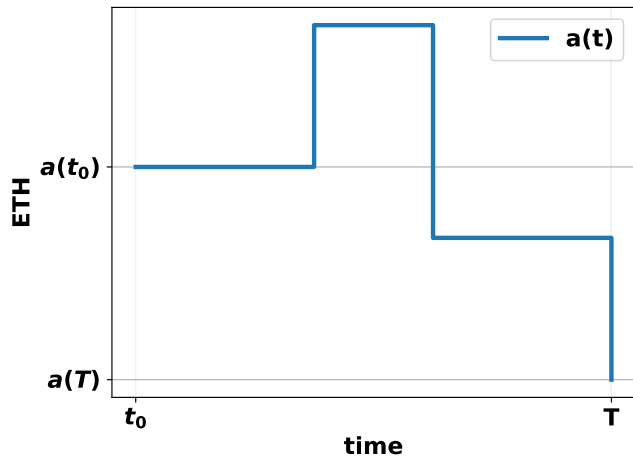


FIGURE 1. The amount of collateralization assets $a(t)$ as a function of time.

The maximum allowed debt is determined by the collateral price in DAI and the minimum allowed collateralization ratio $r_{min}(t)$. Oracles provide the ETH/DAI exchange rate $e(t)$, which is typically consistent with centralized exchange rates except in cases of extremely high transaction fees [31]. The minimum allowed collateralization ratio $r_{min}(t)$ is a piece-wise linear function with small slopes at non-constant intervals to ensure platform stability. Since debts in Maker project are over-collateralized we have that $r_{min}(t) > 1$.

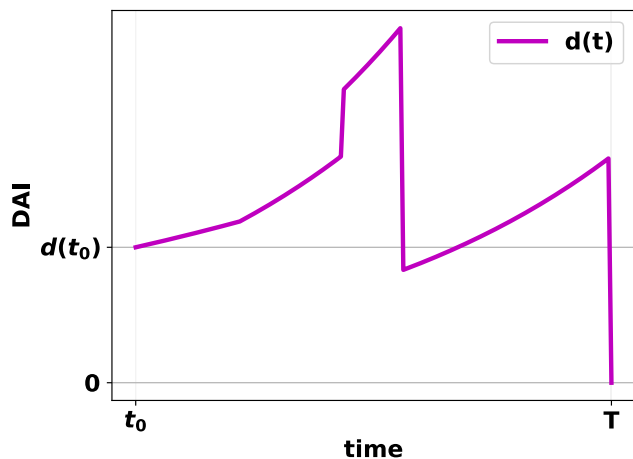


FIGURE 2. The debt $d(t)$ as a function of time.

Let $d(t)$ be the debt at time t (an example is shown in Figure 2). Interest is charged on the active debt, with the logarithm of the interest over time denoted by $f(t)$. If no actions are taken on the debt during an interval $(t_1, t_2]$, then the debt at time t_2 can be calculated as

$$d(t_2) = d(t_1) \cdot \exp\left(\int_{t_1}^{t_2} f(t)dt\right). \tag{1}$$

The log-interest $f(t)$ is piece-wise constant by design of the platform. If $f(t)$ is constant during $(t_1, t_2]$, then $d(t_2) = d(t_1) \cdot \exp(f(t_2) \cdot (t_2 - t_1))$. So the collateral balance is piece-wise exponential. The function breaks are due to the debt repayment, getting more or liquidation process. Changes in the log-interest cause derivative breaks without function breaks.

The current collateralization ratio $r(t)$ for $d(t) > 0$ equals to the following value (see Figure 3)

$$r(t) = \frac{e(t) \cdot a(t)}{d(t)}.$$

If $d(t) = 0$, we can set $r(t) = +\infty$ and if $r(t)$ drops below $r_{min}(t)$ at any point in time, the platform triggers the liquidation. The collateralization requirement check normally is near real-time. And the borrower is responsible for paying the interest during the liquidation period.

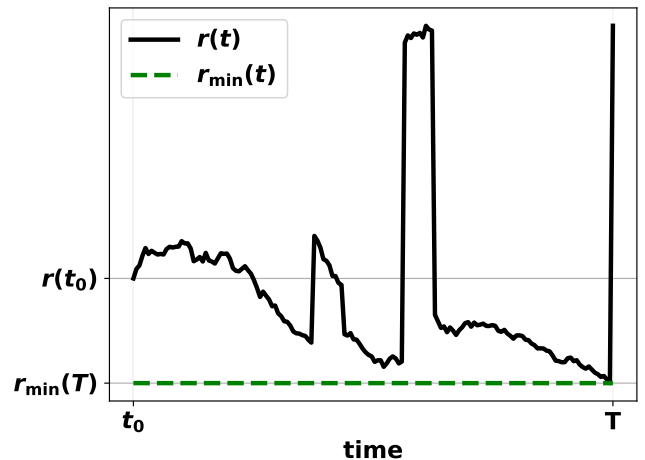


FIGURE 3. The collateralization ratio $r(t)$ and the minimum allowed collateralization ratio $r_{min}(t)$ as functions of time.

B. LOSS GIVEN DEFAULT

Loss given default (LGD) refers to the portion of an asset that is lost in the event of a borrower defaulting [5]. In the Maker protocol, debts are typically over-collateralized, resulting in losses for users in most cases. We can represent a user's balance at time t as

$$Bal(t) = a(t) \cdot e(t) - d(t), \tag{2}$$

and the left-side limit for any given function φ as $\varphi(t-) = \lim_{\tau \uparrow t} \varphi(\tau)$.

at time t , we use the following formula

$$\text{LGD}(t) = \frac{\text{Bal}(t-) - \text{Bal}(t)}{d(t-)} \quad (3)$$

A positive value for LGD indicates a loss for the user, while a negative value indicates a gain. To determine the average of D user defaults at times t_1, \dots, t_D , we use a weighted average:

$$\begin{aligned} \overline{\text{LGD}} &= \frac{\sum_{d=1}^D d(t_d-) \cdot \text{LGD}(t_d)}{\sum_{d=1}^D d(t_d-)} \\ &= \frac{\sum_{d=1}^D \text{Bal}(t_d-) - \text{Bal}(t_d)}{\sum_{d=1}^D d(t_d-)} \end{aligned} \quad (4)$$

C. ANNUAL EQUIVALENT RATE

The interest rate on the platform changes over time and is charged in second-wise intervals. If a loan is liquidated, the loss of collateral value during liquidation is calculated as $(a(T) - a(T-)) \cdot e(T)$. The log-equivalent rate (LER) is a constant log-interest rate that results in the same final debt, including any liquidation losses, for a debt from t_0 to T with debt changes $\Delta d_0, \dots, \Delta d_N$ at times t_0, \dots, t_N respectively.

To find the LER, we use the cumulative debt at time T with $\text{LER} = x$, denoted by $h(x)$, which is calculated as

$$h(x) = \sum_{n=1}^N \Delta d_n \cdot \exp(x(T - t_n)) \quad (5)$$

The LER is then determined by solving the following equation:

$$h(x) = d(T) + (a(T) - a(T-)) \cdot e(T), \quad (6)$$

where $h(x)$ equals the final debt plus any liquidation losses. The function $h(x)$ is monotonically increasing for $x > 0$ since $d(t) > 0$ for $t \in (t_0, T)$. Therefore, if there is no default, the LER falls within the range of $[\min_{t \in (t_0, T)} f(t), \max_{t \in (t_0, T)} f(t)]$. However, if there is a default, the LER can be large and the solution of (6) may be unstable. To avoid this issue, we only consider values of LER that are less than or equal to a fixed constant $f_{\max} > 0$.

To determine the average of D users at time t , we use a weighted average:

$$\overline{\text{LER}} = \frac{\sum_{i=1}^D d_i(t) \cdot \text{LER}_i}{\sum_{i=1}^D d_i(t)} \quad (7)$$

D. PROBABILITY OF DEFAULT

The probability of default (PD) is a risk assessment parameter commonly used by financial institutions. It is a financial term that describes the likelihood of default over a particular time horizon. Let us consider a dataset of loans from Maker platform in the format of time intervals as follows:

- N intervals of the time till default for default debts: t_1, \dots, t_N
- M intervals of the time during which there were no defaults: $\tau_{N+1}, \dots, \tau_{N+M}$. These intervals correspond

to either active debts or returned debts. τ s are intervals from the debt opening until the dataset generation for the active debts. τ s are intervals from the debt taking until debt repayment for the returned debts

and consider different debt models.

1) POISSON MODEL

The classic finance baseline model is a Poisson model that assumes all debts are independent and have an exponential distribution with an unknown parameter $\lambda > 0$ for time until default [43]. This simplification allows for the estimation of λ [44], [45]. However, this assumption does not hold for Maker data since all debts are based on the same collateral type but with different collateralized ratios.

Statement 1. Let λ be the parameter of the exponential distribution. Let X_1, \dots, X_{N+M} be independent and identically-distributed (i.i.d.) random variable from the exponential distribution with a parameter λ . Let x_1, \dots, x_{N+M} be realizations of X_1, \dots, X_{N+M} . Given x_1, \dots, x_N and deterministic parameters y_{N+1}, \dots, y_{N+M} such that $\forall n = N + 1, \dots, N + M : x_n > y_n$, the maximum likelihood estimator (MLE) $\hat{\lambda}$ of the parameter λ is

$$\hat{\lambda} = \frac{N + M}{\sum_{n=1}^N x_n + \sum_{m=N+1}^{N+M} y_m} \quad (8)$$

Proof: The likelihood defines as

$$\begin{aligned} L_{N,M}(\lambda) &= L(\lambda, x_{N+1}, \dots, x_{N+M} \\ &\quad |x_1, \dots, x_N, y_{N+1}, \dots, y_{N+M}) \\ &= \prod_{n=1}^{N+M} \lambda \cdot \exp(-\lambda x_n) \cdot \prod_{n=N+1}^{N+M} I(x_n \geq y_n), \end{aligned} \quad (9)$$

where $I(A)$ is an indicator function of the event A , i.e., $I(A) = 1$ if A is true and $I(A) = 0$ if A is not true. The likelihood is non-negative, and $L_{N,M}(\lambda)$ equals 0 if any $x_n < y_n$, $n = N + 1, \dots, N + M$. So the maximum of $L_{N,M}(\lambda)$ is for $x_{N+1}, \dots, x_{N+M} : x_n \geq y_n$, $n = N + 1, \dots, N + M$.

$L_{N,M}(\lambda)$ is a decreasing function of x_n , $n = N + 1, \dots, N + M$ for all λ in the region $x_n \geq y_n$, $n = N + 1, \dots, N + M$. So the maximum of $L_{N,M}$ is at $x_n = y_n$, $n = N + 1, \dots, N + M$, and the maximization of (9) is equivalent to the classic problem for the exponential distribution

$$\begin{aligned} L_{N+M}(\lambda) &= L(\lambda | x_1, \dots, x_N, y_{N+1}, \dots, y_{N+M}) \\ &= \prod_{n=1}^N \lambda \cdot \exp(-\lambda x_n) \cdot \prod_{n=N+1}^{N+M} \lambda \cdot \exp(-\lambda y_n) \rightarrow \max_{\lambda} \end{aligned} \quad (10)$$

with the log-likelihood expressed as

$$\begin{aligned} l_{N+M}(\lambda) &= (N + M) \ln \lambda \\ &\quad - \lambda \cdot \left(\sum_{n=1}^N x_n + \sum_{n=N+1}^{N+M} y_n \right). \end{aligned} \quad (11)$$

The observation that MLE $\hat{\lambda}$ of λ is given by (8) finishes the proof. ■

The probability of the default for a single debt during T , where X is an exponential random variable with parameter λ can be written as

$$PD(T) = P(X < T) = 1 - \exp(-\lambda T). \quad (12)$$

As the likelihood is functional equivariant [44], the MLE for PD is

$$\hat{PD}(T) = 1 - \exp(-\hat{\lambda}T), \quad (13)$$

where $\hat{\lambda}$ is given by (8). It is important to note that the model assumes independence between debts, so the covariance between different users is zero.

2) BROWNIAN MOTION MODEL

Another model considers the minimal allowed collateralization ratio $r_{\min}(t)$ in comparison to the actual user's collateralization $r(t)$. We assume that the logarithm of $e(t)/e_0$ follows a Brownian motion with zero mean and an unknown standard deviation $\sigma > 0$. Therefore, $\frac{1}{\sigma}(\ln \frac{e(t)}{e_0})$ is a Brownian motion B_t with zero mean and unit variance.

Let us denote

$$\psi(x) = \int_0^T \frac{|x + fs|}{\sqrt{2\pi s^3}} e^{-\frac{(x+fs)^2}{2s}} ds \quad (14)$$

for fixed parameters f and T .

Theorem 1. *If*

- 1) *the normalized exchange rate $\frac{1}{\sigma}(\ln \frac{e(t)}{e_0})$ for a given constant $\sigma > 0$ is a Brownian motion B_t with zero mean and unit variance*
- 2) *the borrower has a debt d_0 and collateral a_0 at time $t = 0$*
- 3) *the borrower has no actions with debt and collateral during $t \in (0, T]$*
- 4) *the platform's interest rate $f \geq 0$ and the minimum collateralization ratio $r_{\min} > 0$ are constant,*

then the probability of the borrower's default during the time interval $(0, T]$ and its variance are given by

$$PD = \psi(x_{\min}) \quad (15)$$

and

$$var = \psi(x_{\min}) \cdot (1 - \psi(x_{\min})), \quad (16)$$

respectively, where

$$x_{\min} = \frac{1}{\sigma} \ln \left(\frac{d_0 \cdot r_{\min}}{a_0 \cdot e_0} \right). \quad (17)$$

Proof: Firstly, let the stability fee $f = 0$. Then $a(t) = a(0) \equiv a_0$ and $d(t) = d(0) \equiv d_0$. Therefore, a debt default is equivalent to the existence of such $t > 0$ that

$$e(t) = \frac{d_0}{a_0} \cdot r_{\min} \equiv e_{\min}. \quad (18)$$

A debt default occurs when the Brownian motion B_t reaches the level x_{\min} (17). Let us denote $T_C = \inf\{t > 0 :$

$B_t = C\}$ for $C < 0$. Then for $T > 0$ [45] from the reflexion principle we have that

$$P(T_{x_{\min}} < T) = \int_0^T \frac{|x_{\min}|}{\sqrt{2\pi s^3}} e^{-\frac{x_{\min}^2}{2s}} ds, \quad (19)$$

therefore the distribution density $p_C(t) = \frac{C}{\sqrt{2\pi t^3}} e^{-\frac{C^2}{2t}}$.

Now, let stability fee be nonzero constant $f \geq 0$. Then $d(t) = d_0 \cdot e^{ft}$ and a default is equivalent to the existence of such $t > 0$ that

$$e(t) = \frac{d_0 \cdot e^{ft}}{a_0} \cdot r_{\min} \equiv e_{\min}(t).$$

I.e., a default is the passage of level

$$x_{\min}(t) = \frac{1}{\sigma} \ln \left(\frac{d_0 \cdot r_{\min}}{a_0 \cdot e_0} \right) + ft = x_{\min} + ft \quad (20)$$

by Brownian motion B_t (see Figure 4).

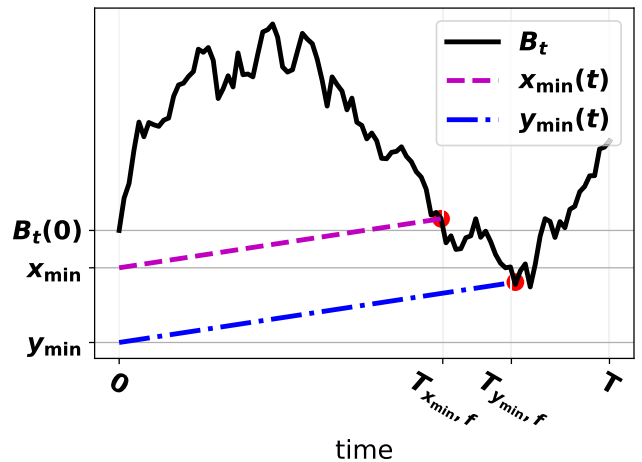


FIGURE 4. Borrower default can be described as a Brownian motion level passage. The black solid curve represents the normalized log-exchange rate of the collateral. The magenta dashed line indicates the minimum allowed rate before default for a user starting from x_{\min} . Similarly, the blue dash-dotted line represents the minimum allowed rate before default for a user starting from y_{\min} .

Let

$$T_{C,f} = \inf\{t > 0 : B_t = C + ft\}. \quad (21)$$

Then for $C < 0$ and $f > 0$

$$P(T_{x_{\min},f} < T) = \int_0^T \frac{|x_{\min} + fs|}{\sqrt{2\pi s^3}} e^{-\frac{(x_{\min}+fs)^2}{2s}} ds. \quad (22)$$

And from (14): $PD = \psi(x_{\min})$. As the default is a Bernoulli random variable, its variance is given by (16) and theorem statement follows. ■

Theorem 2. *If, in addition to the assumptions 1)-4) of Theorem 1,*

- 5) *the second borrower has a debt \tilde{d}_0 a collateral \tilde{a}_0 at time $t = 0$,*

then the covariance of two borrowers' defaults during time interval $(0, T]$ equals

$$cov = \psi(\min\{x_{\min}, y_{\min}\}) \cdot (1 - \psi(\max\{x_{\min}, y_{\min}\})) \quad (23)$$

where

$$y_{\min} = \frac{1}{\sigma} \ln\left(\frac{\tilde{d}_0 \cdot r_{\min}}{\tilde{a}_0 \cdot e_0}\right). \quad (24)$$

Proof: The second user has a single asset as a collateral with a default at level $y_{\min}(t) = y_{\min} + ft$, where y_{\min} is given by (24). Without loss of generality, $y_{\min} \leq x_{\min}$. Then the probability of the passage of both x_{\min} and y_{\min} is

$$P(T_{x_{\min},f} < T; T_{y_{\min},f} < T) = P(T_{y_{\min},f} < T) = \psi(y_{\min}). \quad (25)$$

Denote $I(A)$ the indicator of the event $T_{x_{\min},f} < T$, i.e., $I(A) = 1$, if $T_{x_{\min},f} < T$, and $I(A) = 0$, if $T_{x_{\min},f} \geq T$. Denote $I(B)$ the indicator of the event $T_{y_{\min},f} < T$. The mathematical expectations and the covariance can be written in (26) and (27), in turn.

$$\begin{aligned} EI(A) &= P(T_{x_{\min},f} < T) = \psi(x_{\min}) \\ EI(B) &= P(T_{y_{\min},f} < T) = \psi(y_{\min}) \end{aligned} \quad (26)$$

$$\begin{aligned} cov(I(A), I(B)) &= EI(A)EI(B) - EI(A)EI(B) \\ &= 1 \cdot P(T_{x_{\min},f} < T; T_{y_{\min},f} < T) - EI(A) \cdot EI(B) \\ &= \psi(y_{\min}) - \psi(x_{\min}) \cdot \psi(y_{\min}) \\ &= \psi(y_{\min}) \cdot (1 - \psi(x_{\min})) \\ &= \psi(\min\{x_{\min}, y_{\min}\}) \cdot (1 - \psi(\max\{x_{\min}, y_{\min}\})). \end{aligned} \quad (27)$$

The observation that (27) coincides with (23) finishes the proof. ■

As the standard deviation of the Brownian motion $\ln\left(\frac{e(t)}{e_0}\right)$ is unobservable, we can estimate it. More precisely, given a sample $\{(t_n, e(t_n))\}_{n=0}^N$, where $t_0 < t_1 < \dots < t_N$ the maximum likelihood estimator is

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\ln\left(\frac{e(t_n)}{e(t_{n-1})}\right) \right)^2 / (t_n - t_{n-1})}. \quad (28)$$

3) MODELS COMPARISON

Both the Poisson process and Brownian motion provide models for predicting probability of default PD. However, the true model cannot be observed directly. To test the accuracy of these models, we generated a dataset of daily defaults and utilized both models to predict defaults one day in advance. Initially, the models assume that the data is stationary. To verify this hypothesis, we employed the Augmented Dickey-Fuller ADF test [46]. The ADF test is based on the autoregressive model and assesses unit roots in time series, which causes trends. The null hypothesis of

the ADF test is the presence of a unit root in a time series. If the p -value (the probability that the null hypothesis is true) is less than a given significance level (usually 0.05), then the null hypothesis is rejected, indicating that the time series is stationary.

To compare the fit of models to data, we use four quantities [43], [44], [47]:

- Kullback-Leibler divergence (KL) measures the divergence between two probability distributions. It quantifies the amount of information lost when one distribution is used to approximate another.
- Total Variation (TV) measures the divergence between two probability distributions. It is defined as half the sum of the absolute differences between the corresponding probabilities in the two distributions.
- Relative Root Mean Squared Error (RRMSE) and Relative Mean Absolute Error (RMAE) are quadratic and linear aggregations of point-wise distances between two datasets, respectively. In our specific case, RMAE; is equivalent to TV, but we retain both terms since they are commonly used by machine learners and statisticians.

For each quantity, a smaller value indicates a better fit of the theoretical model to the empirical data.

V. DATASET STRUCTURE

Our focus is on the ETH-collateralized risk program A debts (ETH-A) within the MakerDAO protocol deployed on the Ethereum network. This program has the largest number of debts (137,441 out of 259,048) and debt volume (13.4 billion DAI out of 36.9 billion DAI). We utilized publicly available data from November 11th, 2019 (the first debt start in the considered asset) to July 31st, 2023, accessed via the Big Query project by Google. The collected raw data was decoded and further processed using Python. To verify specific information and ensure data correctness, we used a third-party API Ethereum provider, Infura.

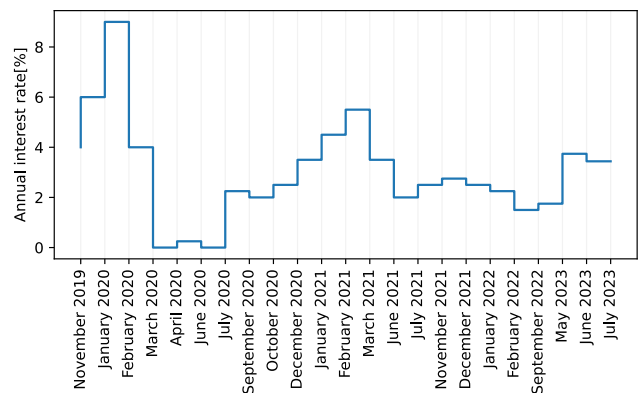


FIGURE 5. Annual Maker's interest rate for ETH-A.

After processing all the internal terms such as `frob` and `wad` and focusing solely on the borrower-related aspect of the Maker protocol, we collected a loan portfolio dataset. The dataset comprises two parts: *system* and *borrower* data.

The *system* data contains common parameters, such as the ETH/DAI exchange rate $e(t)$, which is used to estimate the collateralization ratio since the ETH-A program deals with ETH as collateral. The exchange rate is provided by oracles and typically corresponds to the centralized exchanges rate [31]. Maker’s loans are overcollateralized, and the minimum allowed collateralization ratio is defined as $r_{\min}(t)$. The platform receives interest rates from borrowers, and the dataset contains the log-interest $f(t)$.

The *borrower* data contains borrowers’ catalog and debt details. The borrower catalog lists all borrowers and their debts, each with start and end times, status, and loan actions (see Table 1). The possible statuses are

- *repaid*: the borrower returned the debt
- *liquidated*: the debt is fully repaid via liquidation process
- *restructured*: the debt is partially repaid via liquidation process, and a new debt started immediately after the liquidation
- *active*: the loan is active by the end of the observation period (July 31st, 2023).

The raw data only includes reference points for these parameters. Our utility functions allow obtaining their values over time and plotting system parameters for the entire period (see Figure 5) and loan characteristics for its lifetime (see Figure 6).

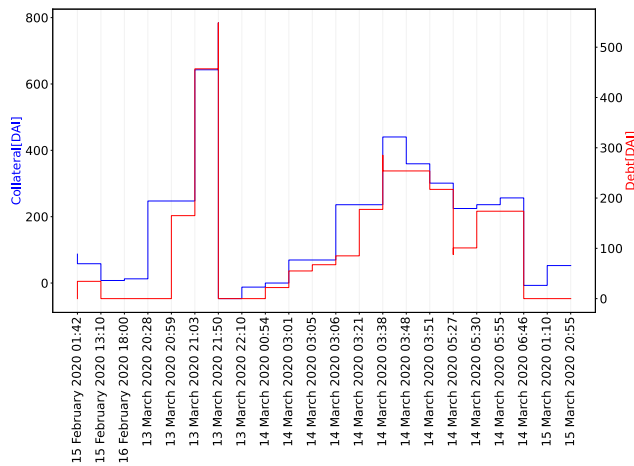


FIGURE 6. DAI debt over time for user 0 x 4032EE21404af045f6ba8022Cf4607950c87A39A.

The prepared dataset is publicly available on Gitlab [17].

VI. NUMERICAL EXPERIMENTS

To showcase the practicality of the gathered dataset (Section V) and viability of the suggested computational models (Section IV), we conducted a quantitative analysis. The code to reproduce the experiments is available on Gitlab [17]. We do note the possible presence of unexpected events that can significantly affect the market. Examples of the latter include the *Black Thursday* price crash on March 12th and 13th, 2020 [40], [41], and Maker’s announcement of their

TABLE 1. Loan actions of user 0 x 931dBd7001D14112D17304B78d305c4FE317E571.

Timestamp [yy-mm-dd hh-mm-ss]	Action	Collateral [ETH]	Debt [DAI]
20-02-15 01:42:39	lock asset	0.3000	0.0000
20-02-15 01:42:39	generate DAI	0.0000	34.3770
20-02-15 13:10:55	unlock asset	-0.1000	0.0000
20-02-16 18:00:01	liquidate	-0.2000	-34.3892
20-03-13 20:28:37	lock asset	0.0404	0.0000
20-03-13 20:59:40	lock asset	2.0000	0.0000
20-03-13 21:03:20	generate DAI	0.0000	165.0297
20-03-13 21:03:20	lock asset	3.1989	0.0000
20-03-13 21:03:20	generate DAI	0.0000	245.1908
20-03-13 21:50:37	generate DAI	0.0000	46.7635
20-03-13 21:50:37	lock asset	1.0593	0.0000
20-03-13 21:50:37	generate DAI	0.0000	91.3185
20-03-13 22:10:47	liquidate	-6.2986	-548.3069
20-03-14 00:54:08	lock asset	0.2500	0.0000
20-03-14 03:01:40	generate DAI	0.0000	22.0042
20-03-14 03:01:40	lock asset	0.4161	0.0000
20-03-14 03:01:40	generate DAI	0.0000	33.0255
20-03-14 03:05:13	lock asset	0.1000	0.0000
20-03-14 03:06:46	generate DAI	0.0000	12.3999
20-03-14 03:06:46	lock asset	0.2265	0.0000
20-03-14 03:06:46	generate DAI	0.0000	17.6159
20-03-14 03:21:27	lock asset	1.0000	0.0000
20-03-14 03:38:30	generate DAI	0.0000	92.3912
20-03-14 03:38:30	lock asset	1.5000	0.0000
20-03-14 03:38:30	generate DAI	0.0000	107.7055
20-03-14 03:48:09	unlock asset	-0.2760	0.0000
20-03-14 03:48:09	payback DAI	0.0000	-31.2139
20-03-14 03:51:13	unlock asset	-0.3179	0.0000
20-03-14 03:51:13	payback DAI	0.0000	-36.6915
20-03-14 05:27:38	unlock asset	-0.4331	0.0000
20-03-14 05:27:38	payback DAI	0.0000	-129.5872
20-03-14 05:27:38	unlock asset	-0.5674	0.0000
20-03-14 05:30:09	generate DAI	0.0000	13.1757
20-03-14 05:30:09	lock asset	0.0854	0.0000
20-03-14 05:55:06	generate DAI	0.0000	73.0000
20-03-14 06:46:34	lock asset	0.1500	0.0000
20-03-15 01:10:55	liquidate	-2.1336	-173.8455
20-03-15 20:55:55	lock asset	0.4748	0.0000
20-03-15 23:01:50	unlock asset	-0.4748	0.0000

upcoming Spark Lend project in 2023. In the following subsections, we will observe their effects.

A. BALANCE

The *borrower* data contains the details of all the debts. For our purposes we have to find the number of debts in Figure 7

and total collateral amount (total value locked, TVL) and debt in Figure 8, where the balance is their difference.

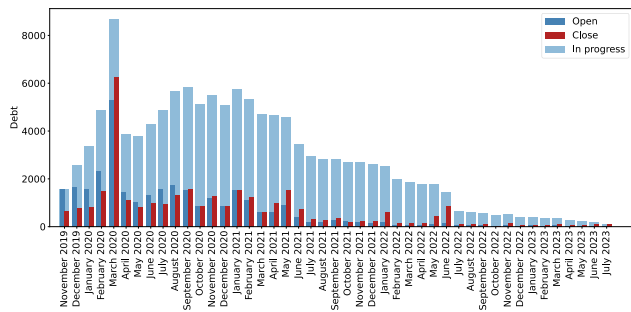


FIGURE 7. The number of ETH-A debts.

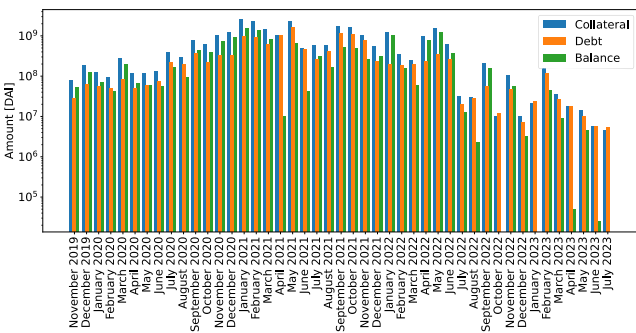


FIGURE 8. The total collateral amount and borrowers' balance.

The number of borrowers experienced a remarkable surge initially, but came to a halt during the unfortunate event of Black Thursday in 2020. The debt amount peaked in mid-2021, only to witness a decline due to the cryptocurrency market downturn known as the crypto winter in the same year. Currently, the total debt in ETH-A is steadily decreasing towards zero as Maker DAO transitions into the Spark Lend Protocol, marking a significant development.

B. LOSS GIVEN DEFAULT

The monthly average LGD (4) is depicted in Figure 9. Maker receives a fixed percentage of 13% from each auction (represented by the dashed black horizontal line), which indicates the typical loss level in an efficient market. The efficiency of the auction is a critical factor [40] and tends to decrease during periods of declining ETH prices.

C. ANNUAL EQUIVALENT RATE

The monthly average AER (7) is shown in Figure 10. Although the AER for returned debts aligns with the annual Maker's interest rate for ETH-A (Figure 5), the majority of liquidated debts have an AER that exceeds 100%.

D. PROBABILITY OF DEFAULT

The number of daily defaults has an ADF test statistic of -8.89 , which is significantly smaller than the 1% significance level of -3.45 . This statistic enables us to reject

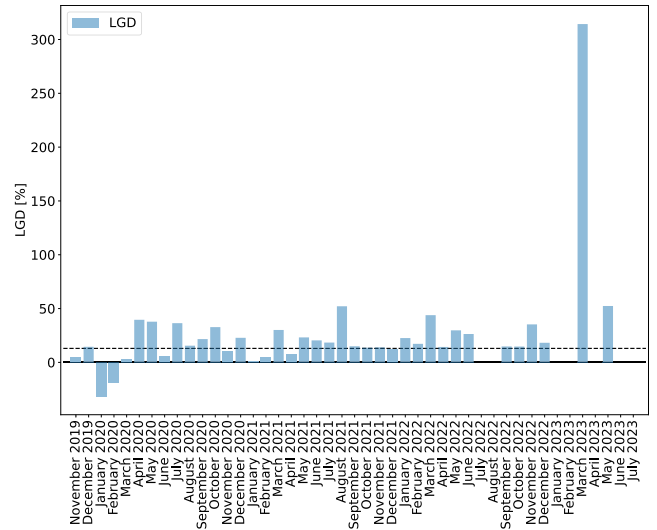


FIGURE 9. Monthly average LGD.

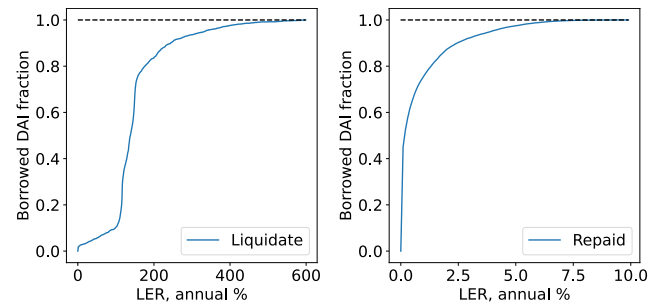


FIGURE 10. Monthly average AER for liquidated (left) and returned (right) debts.

the null hypothesis of data non-stationarity and conclude that the time series is stationary.

We have computed a day-ahead actual number of defaults together with the predictions by Poisson and Brownian motion models.

TABLE 2. Comparison of PD models. The lower the value in each column, the better the fit to the data. The best result in each column is highlighted.

	KL	TV	MSE	MAE
Poisson process	0.0061	0.00077	1.00	0.72
Brownian motion	0.0053	0.00078	1.08	0.72

The Brownian motion model is found to be superior to the baseline Poisson process model in describing and predicting debt defaults (see Table 2). This conclusion is supported by the KL; divergence, which measures the difference between the two probability distributions. The regression-specific RRMSE; is worse for the Brownian motion model compared to the Poisson process model. However, the RMAE; and TV; are comparable between the two models.

Both the RRMSE and RMAE values are close to one, indicating that both models have poor predictive power from

a regression perspective. This finding suggests that further analysis is needed.

Overall, the findings confirm the hypothesis that the Brownian motion model is better suited for capturing real-world data and demonstrates its superiority in fitting probabilistic distributions in collateral-based DeFi lending.

VII. CONCLUSION AND FUTURE WORK

This research focuses on analyzing the lending aspect of the Maker protocol in the DeFi space from a traditional finance perspective. The authors have gathered a unique dataset comprising loan portfolios sourced from the MakerDAO project, making it the first dataset of its kind in the DeFi field. This publicly available dataset contains essential financial characteristics related to borrowing, including balance, loss given default, annual equivalent rate, and probability of default. The current version of the dataset covers only the most popular Maker's borrowing program called ETH-A. However, the authors plan to expand the dataset to include other programs and new Spark Loan data in future work.

In addition to collecting this dataset, the authors have developed a specialized mathematical model tailored specifically to the Maker project. This model allows them to estimate the probability of default by considering the presence of crypto-collateral and utilizing Brownian motion passage levels. The proposed model outperformed the Poisson process baseline model on the loan portfolio dataset. By incorporating borrowing-driven financial characteristics into the dataset and developing this model, the authors provide a comprehensive understanding of both individual loan defaults and the correlation among different loans.

Expanding the analysis to include other borrowing programs beyond ETH-A presents challenges in finding the default correlation of level passage times between two correlated Brownian motions representing different collateral types. The authors acknowledge this as a future work. However, this also opens up opportunities to estimate the platform's risk, where simultaneous defaults of a significant portion of borrowers could pose a threat.

The findings of this study offer valuable insights into lending practices in DeFi projects and help bridge the gap between traditional finance and blockchain-based financial services. This research contributes to the understanding of how DeFi lending operates and offers a standardized approach to analyzing and evaluating loan portfolios in the DeFi space. Furthermore, the methodology can be extended to other DeFi lending platforms such as Compound and Aave.

REFERENCES

- [1] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, May 1970.
- [2] R. C. Merton, "Theory of rational option pricing," *Bell J. Econ. Manag. Sci.*, vol. 4, no. 1, p. 141, 1973.
- [3] C. Gorter and A. M. Bloem, "The treatment of nonperforming loans in macroeconomic statistics," *IMF Work. Papers*, vol. 1, no. 209, pp. 1–18, 2001.
- [4] L. Quaglia, "The 'old' and 'new' politics of financial services regulation in the European Union," *New Political Economy*, vol. 17, no. 4, pp. 515–535, Sep. 2012.
- [5] Basel Committee on Banking Supervision. (2022). *The Basel Framework. Bank for International Settlements*. [Online]. Available: https://www.bis.org/basel_framework/index.htm?export=pdf
- [6] P. Voigt and A. von dem Bussche, *The EU General Data Protection Regulation (GDPR)*. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-57959-7.
- [7] V. Buterin. (2014). *Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform*. [Online]. Available: <https://github.com/ethereum/wiki/wiki/White-Paper>
- [8] V. Buterin. (2015). *On Public and Private Blockchains—Ethereum Blog*. [Online]. Available: <https://blog.ethereum.org/2015/08/07/on-public-and-private-blockchains/>
- [9] F. Schär, "Decentralized finance: On blockchain- and smart contractbased financial markets," *SSRN Electron. J.*, pp. 3571335:1–3571335:24, Mar. 2020. [Online]. Available: <https://www.ssrn.com/abstract=3571335>
- [10] E. Meyer, I. M. Welpe, and P. Sandner, "Decentralized finance—A systematic literature review and research directions," in *ECIS Research Papers*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 1–25.
- [11] MakerDAO. (2020). *The Maker Protocol: MakerDAO's Multi-Collateral Dai (MCD) System*. [Online]. Available: <https://makerdao.com/en/whitepaper>
- [12] E. Ben Sasson, A. Chiesa, C. Garman, M. Green, I. Miers, E. Tromer, and M. Virza, "Zerocash: Decentralized anonymous payments from Bitcoin," in *Proc. IEEE Symp. Secur. Privacy*, May 2014, pp. 459–474.
- [13] B. Bunz, J. Bootle, D. Boneh, A. Poelstra, P. Wuille, and G. Maxwell, "Bulletproofs: Short proofs for confidential transactions and more," in *Proc. IEEE Symp. Secur. Privacy (SP)*, May 2018, pp. 315–334.
- [14] D. Korepanova, M. Nosyk, A. Ostrovsky, and Y. Yanovich, "Building a private currency service using exonum," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom)*, Jun. 2019, pp. 1–3. [Online]. Available: <https://ieeexplore.ieee.org/document/8812875/>
- [15] V. Rossikhin, M. Burdin, and O. Mykhalskiy, "Legal regulation issues of cryptocurrency circulation in Ukraine," *Baltic J. Econ. Stud.*, vol. 4, no. 3, pp. 254–258, 2018. [Online]. Available: <http://www.baltijapublishing.lv/index.php/issue/article/view/451/pdf>
- [16] B. D. Feinstein and K. Werbach, "The impact of cryptocurrency regulation on trading markets," *SSRN Electron. J.*, pp. 3649475:1–3649475:52, Mar. 2021. [Online]. Available: <https://papers.ssrn.com/abstract=3649475>
- [17] Y. Chaleenuthawut, M. Evdokimov, S. Kasemsuk, G. Melnikov, and Y. Yanovich. (2023). *MakerDAO Loan Portfolio Dataset*. [Online]. Available: <https://github.com/Sudarut-kas/Data-Mining-for-MakerDAO>
- [18] V. Makri, A. Tsagkanos, and A. Bellas, "Determinants of non-performing loans: The case of eurozone," *Panoeconomicus*, vol. 61, no. 2, pp. 193–206, 2014.
- [19] A. Khairi, B. Bahri, and B. Artha, "A literature review of non-performing loan," *J. Bus. Manage. Rev.*, vol. 2, no. 5, pp. 366–373, May 2021.
- [20] E. Hayden, D. Porath, and N. V. Westernhagen, "Does diversification improve the performance of German banks? Evidence from individual bank loan portfolios," *J. Financial Services Res.*, vol. 32, no. 3, pp. 123–140, Oct. 2007.
- [21] S. P. S. Rossi, M. S. Schwaiger, and G. Winkler, "How loan portfolio diversification affects risk, efficiency and capitalization: A managerial behavior model for Austrian banks," *J. Banking Finance*, vol. 33, no. 12, pp. 2218–2226, Dec. 2009.
- [22] S. I. Serengil, S. Imece, U. G. Tosun, E. B. Buyukbas, and B. Koroglu, "A comparative study of machine learning approaches for non performing loan prediction," in *Proc. 6th Int. Conf. Comput. Sci. Eng. (UBMK)*, Sep. 2021, pp. 326–331.
- [23] M. Annisa, "Prediction of non-performing loans for credit application analysis of rural bank using random forest," in *Proc. 9th Int. Conf. Electr. Eng., Comput. Sci. Informat. (EECSI)*, Oct. 2022, pp. 111–114.

- [24] Md. N. Alam and M. M. Ali, "Loan default risk prediction using knowledge graph," in *Proc. 14th Int. Conf. Knowl. Smart Technol. (KST)*, Jan. 2022, pp. 34–39.
- [25] S. R. Islam, W. Eberle, and S. K. Ghafoor, "Credit default mining using combined machine learning and heuristic approach," 2018, *arXiv:1807.01176*.
- [26] Y.-R. Chen, J.-S. Leu, S.-A. Huang, J.-T. Wang, and J.-I. Takada, "Predicting default risk on peer-to-peer lending imbalanced datasets," *IEEE Access*, vol. 9, pp. 73103–73109, 2021.
- [27] V. A. Davydov, S. A. Kruglik, and Y. A. Yanovich, "Comparison of banking and peer-to-peer lending risks," *Autom. Remote Control*, vol. 82, no. 12, pp. 2155–2168, Dec. 2021. [Online]. Available: <https://link.springer.com/article/10.1134/S0005117921120079>
- [28] V. A. Davydov, S. A. Kruglik, and Y. A. Yanovich, "Probability of the default-free state for token package from independent loans," *J. Commun. Technol. Electron.*, vol. 67, no. 6, pp. 778–786, 2022, doi: [10.1134/S1064226922060122](https://doi.org/10.1134/S1064226922060122).
- [29] D. A. Zetsche, D. W. Arner, and R. P. Buckley, "Decentralized Finance (DeFi)," *SSRN Electron. J.*, pp. 3539194:1–3539194:32, Sep. 2020.
- [30] F. Schär, "Decentralized finance: On blockchain- and smart contract-based financial markets," *Review*, vol. 103, no. 2, pp. 153–174, 2021.
- [31] M. Kjær, M. D. Angelo, and G. Salzer, "Empirical evaluation of MakerDAO's resilience," in *Proc. 3rd Conf. Blockchain Res. Appl. Innov. New. Services (BRAINS)*, Sep. 2021, pp. 193–200. [Online]. Available: <https://ieeexplore.ieee.org/document/9569811/>
- [32] AAVE. (2020). *AAVE Protocol Whitepaper V1.0*. [Online]. Available: https://github.com/aave/prhttps://github.com/aave/aave-protocol/blob/master/docs/Aave_Protocol_Whitepaper_v1_0.pdf
- [33] R. Leshner and G. Hayes. (2019). *Compound: The Money Market Protocol*. [Online]. Available: <https://compound.finance/documents/Compound.Whitepaper.pdf>
- [34] J. A. L. Escamilla and Y. Yanovich, "Data mining of compound DeFi project," in *Proc. 5th Int. Conf. Blockchain Technol. Appl.*, Dec. 2022, pp. 16–23, doi: [10.1145/3581971.3581974](https://doi.org/10.1145/3581971.3581974).
- [35] T. Azoulay, U. Carl, and O. Rottenstreich, "Allowing blockchain loans with low collateral," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2023, pp. 1–9.
- [36] S. Ellis, A. Juels, and S. Nazarov. (2017). *ChainLink: A Decentralized Oracle Network*. [Online]. Available: <https://link.smartcontract.com/whitepaper>
- [37] S. Kruglik, K. Nazirhanova, and Y. Yanovich, "Challenges beyond blockchain: Scaling, oracles and privacy preserving," in *Proc. 16th Int. Symp. 'Problems Redundancy Inf. Control Syst.' (REDUNDANCY)*, Oct. 2019, pp. 155–158. [Online]. Available: <https://ieeexplore.ieee.org/document/9003331/>
- [38] M. Bartholic, A. Laszka, G. Yamamoto, and E. W. Burger, "A taxonomy of blockchain oracles: The truth depends on the question," in *Proc. IEEE Int. Conf. Blockchain Cryptocurrency (ICBC)*, May 2022, pp. 1–15.
- [39] E. Bischof, A. Botezatu, S. Jakimov, I. Suharenko, A. Ostrovski, A. Verbitsky, Y. Yanovich, A. Zhavoronkov, and G. Zmudze, "Longevity foundation: Perspective on decentralized autonomous organization for special-purpose financing," *IEEE Access*, vol. 10, pp. 33048–33058, 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9739690/>
- [40] W. C. Gu, A. Raghuvanshi, and D. Boneh, "Empirical measurements on pricing oracles and decentralized governance for stablecoins," *Cryptoeconomic Syst.*, vol. 1, no. 2, pp. 1–29, Oct. 2021.
- [41] A. Klages-Mundt and A. Minca, "While stability lasts: A stochastic model of noncustodial stablecoins," *Math. Finance*, vol. 32, no. 4, pp. 943–981, Oct. 2022.
- [42] F. Vogelsteller and V. Buterin. (2015). *EIP-20: ERC-20 Token Standard*. [Online]. Available: <https://github.com/ethereum/EIPs/blob/master/EIPS/eip-20.md>
- [43] A. N. Shiryaev, *Probability* (Graduate Texts in Mathematics), vol. 95. New York, NY, USA: Springer, 1984.
- [44] L. Wasserman, *All of Statistics* (Springer Texts in Statistics). New York, NY, USA: Springer, 2004, doi: [10.1007/978-0-387-21736-9_24](https://doi.org/10.1007/978-0-387-21736-9_24).
- [45] A. N. Shiryaev, *Essentials of Stochastic Finance*. Singapore: World Scientific, 1999.
- [46] G. Elliott, T. J. Rothenberg, and J. H. Stock, "Efficient tests for an autoregressive unit root," *Econometrica*, vol. 64, no. 4, p. 813, Jul. 1996.
- [47] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning* (Lecture Notes in Computer Science), vol. 3176. Berlin, Germany: Springer, 2004, pp. 63–71, doi: [10.1007/978-3-540-28650-9_4](https://doi.org/10.1007/978-3-540-28650-9_4).



YATIPA CHALEENUTTHAWUT received the bachelor's degree in chemical engineering from Kasetsart University, Bangkok, Thailand, in 2019, and the master's degree in data science from the Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia, in 2022.

She is currently a Quantitative Trader. During the master's degree, her primary area of focus was on Decentralized Finance (DeFi) and Portfolio Optimization. Her research pursuits primarily revolve around quantitative trading strategies and time series analysis.



VYACHESLAV DAVYDOV received the Specialist degree (Hons.) in informatics and computer science and the Ph.D. degree in engineering from the Leningrad Institute of Aviation (now the Saint Petersburg State University of Aerospace Instrumentation), Saint Petersburg, Russia, in 1992 and 1995, respectively, and the Ph.D. degree in economics from Saint Petersburg State University, Saint Petersburg, in 2017.

Since 2017, he has been the Lead Research Scientist with the HSE Tikhonov Moscow Institute of Electronics and Mathematics, HSE University. Additionally, he is currently the CEO of Quicktoken Tech Ltd., Dubai, United Arab Emirates. He is an accomplished professional with more than 25 years of experience in the banking industry. Throughout his career, he has held various key positions, including the Chief Economist of the Stock Operations Department, Main Directorate of the Central Bank of the Russian Federation for Saint Petersburg, the First Deputy Chairman of the Board of Viking Commercial Bank, and the Director of the Department for Work with Problem Assets, North-West Sberbank. His research interests include risk assessment, portfolio management, tokenization, and cross-border payments.



MICHAEL EVDOKIMOV received the bachelor's degree (Hons.) in applied mathematics and information science from National Research University–Higher School of Economics (HSE), Moscow, Russia, in 2023.

Currently, he is a Middle Machine Learning Researcher with VK, Moscow. His research interests include the analysis and prediction of time series data and recommendation systems.



SUDARUT KASEMSUK received the bachelor's degree (Hons.) in applied mathematics from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2017. She is currently pursuing the master's degree with the Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia.

She was a Teaching Assistant of the Quick Success Blockchain Course in Innovation Workshop 2023 with Skoltech. Her primary research is Economic Data Mining and Analysis of the MakerDAO DeFi Project.



STANISLAV KRUGLIK (Member, IEEE) received the bachelor's and master's degrees (Hons.) in applied physics and mathematics from the Moscow Institute of Physics and Technology (MIPT), Moscow, Russia, in 2015 and 2017, respectively, the master's degree (Hons.) in data science from the Skolkovo Institute of Science and Technology (Skoltech), Moscow, in 2017, and the Ph.D. degree in computer science from MIPT, in 2021.

Currently, he is a Research Fellow with Nanyang Technological University, Singapore. His research interests include information theory and its applications, in particular to problems related to data storage and security. He was a recipient of the Potanin Foundation Scholarship, in 2017, the Russian President Scholarship, in 2016, and the Simons Foundation Scholarship, in 2015.



GRIGORII MELNIKOV received the bachelor's and master's degrees (Hons.) in applied physics and mathematics from the Moscow Institute of Physics and Technology (MIPT), Moscow, Russia, in 2018 and 2020, respectively, and the master's degree (Hons.) in data science from the Skolkovo Institute of Science and Technology (Skoltech), Moscow, in 2020.

He is currently the CTO of B4B.World. He has experience integrating private blockchains in enterprises and launching dApps on public chains. His research interests include consensus protocols, blockchain, and its applications.



YURY YANOVICH (Member, IEEE) received the bachelor's and master's degrees (Hons.) in applied physics and mathematics from the Moscow Institute of Physics and Technology, Moscow, Russia, in 2010 and 2012, respectively, and the Ph.D. degree in probability theory and mathematical statistics from The Institute for Information Transmission Problems, Moscow, in 2017.

He has been a Lecturer of the "Introduction to Blockchain" course with top Russian universities, since 2017. Currently, he is a Senior Research Scientist with the Skolkovo Institute of Science and Technology, Moscow. He is the author of the Exonum consensus protocol. His research interests include blockchain, consensus protocols, privacy, and applications.

• • •