DeFi Survival Analysis: Insights into Risks and User Behaviors

Aaron Green[1], Christopher Cammilleri[1], John S. Erickson[1], Oshani Seneviratne[1], and Kristin P. Bennett[1]

The Rensselaer Institute for Data Exploration and Applications
Rensselaer Polytechnic Institute
http://idea.rpi.edu
{greena12,cammic,erickj4,senevo,bennek}@rpi.edu

Abstract. We propose a decentralized finance (DeFi) survival analysis approach for discovering and characterizing user behavior and risks in lending protocols. We demonstrate how to gather and prepare DeFi transaction data for survival analysis. We demonstrate our approach using transactions in AAVE, one of the largest lending protocols. We develop a DeFi survival analysis pipeline which first prepares transaction data for survival analysis through the selection of different index events (or transactions) and associated outcome events. Then we apply survival analysis statistical and visualization methods such as median survival times, Kaplan–Meier survival curves, and Cox hazard regression to gain insights into usage patterns and risks within the protocol. We show how by varying the index and outcome events, we can utilize DeFi survival analysis to answer three different questions. What do users do after a deposit? How long until borrows are first repaid or liquidated? How does coin type influence liquidation risk? The proposed DeFi survival analysis can easily be generalized to other DeFi lending protocols. By defining appropriate index and outcome events, DeFi survival analysis can be applied to any cryptocurrency protocol with transactions.

1 Introduction

The rapid growth in popularity of blockchain-based products like cryptocurrencies has brought with it a growth in the complexity of the blockchain ecosystem. Myriad new products are being developed and deployed on the numerous blockchains that now exist. Following the invention and adoption of these products by an increasing number of users, a new financial ecosystem has emerged: Decentralized Finance (DeFi). The world of DeFi, though young, is already comprised of hundreds, if not thousands, of products and services. One such product is known as a lending protocol. DeFi lending protocols offer a similar set of services as banks offer to consumers in the world of traditional finance. Users utilize the protocols to conduct *transactions*. For instance, a user of a DeFi lending protocol can take some cryptocurrency they own and deposit it into the protocol, then accruing interest on their account balance. These users can also take out loans through the protocol, much like a person can take out a loan from a bank.

Since these lending protocols are growing in size along with the cryptocurrency market as a whole, an obvious first question might be, "how and why are people using DeFi lending protocols?" There are multiple popular lending protocols, and their data streams are varied. But these data streams share a common structure; entities are conducting different types of transactions through time, each involving varying amounts and types of cryptocurrency. The fact that these entities (we will call them users, but they may be smart contracts) conduct transactions at irregular intervals and the many types of transactions makes DeFi lending streams challenging to understand. However, this complexity also exists in domains such as healthcare and commerce, so there are myriad tools available to analyze temporal data streams. In this paper, we demonstrate how to use one such tool, "survival analysis," to gain insight into DeFi transactions.

Survival analysis models time-to-event data. Survival analysis is widely used in healthcare to understand the risk of death (or other events of interest) after treatment, but can be used more generally to analyze the time between any two events [5]. For example, it can be used to analyze the time between a user's borrow transaction and a transaction to repay that coin. As in healthcare, the data is frequently *right-censored*: users are likely to have an outstanding loan at any given time when we stop observing the data stream. In this analysis, we demonstrate how to gather and prepare DeFi transaction data for survival analysis and apply survival analysis statistical methods such as Kaplan–Meier survival curves and Cox Hazard regression to understand usage patterns within a protocol. Survival analysis has been previously used to understand loan defaults in Centralized Finance (CeFi) [2, 6, 10], but applying this analysis directly to DeFi is not straightforward since DeFi varies significantly from CeFi.

In order to interpret the results of this first use of survival analysis in DeFi, we focus our analysis on looking solely at one lending protocol, developing generic survival analysis tools for transaction analysis, and examining the results of these tools for that protocol. These same tools can eventually be applied to other lending protocols. The protocol chosen was AAVE[1]. At the time of this writing (March 14, 2022), AAVE is the second-largest DeFi lending protocol, with approximately $8.42 billion worth of crypto-assets locked in the protocol according to DeFi Pulse[2]. We note, however, that survival analysis tools for DeFi transactions can be used to analyze and gain insight into any DeFi protocol that consists of transactions through time, including other lending protocols and exchanges.

This paper is organized as follows: in the methods section, we describe the AAVE data and the survival analysis methods used to study it. In the results section, we demonstrate the use provided by survival analysis to answer three different questions. We conclude with a discussion of the contributions of this work and promising directions for future work.

[1] https://aave.com
[2] https://defipulse.com

2 Methods

2.1 Data

The data used in this analysis comes primarily from The Graph[3], a service that indexes data from blockchains and allows for the querying of the indexed data. For the work presented here, we sought to combine data from the primary seven transaction types that AAVE records in order to give a comprehensive view of all transactions that have taken place in AAVEv2 [1] since its deployment on November 30, 2020. The data used in this analysis starts with the first transactions of the AAVEv2 protocol on November 30, 2020, and ends on January 6, 2022. The seven transaction types we've pulled from The Graph include deposits, redeems, borrows, repays, liquidations, interest-rate swaps, and reserve collateral usage toggling.

Combining these seven transaction types, we get a table that includes one transaction per row, totaling 847,798 transactions. Table 1 summarizes the number of transactions of each type and their mean and median values. There are some common features for each transaction, such as the user involved and the time the transaction was made. Aside from liquidations, the transactions also have one specific coin involved. Liquidations have two coins: a principal coin and a collateral coin. AAVE transactions in our dataset have used 54 different coins. We divide these coins into two types: stablecoin and non-stablecoin. A stablecoin is from a class of cryptocurrencies that attempt to offer price stability (typically in terms of USD), and that is backed by a reserve asset. The other types of coins in the dataset are non-stablecoins.

Deposits and redeems in AAVE function as one might expect deposits and withdrawals to function at a bank. A user can deposit a currency into the AAVE protocol, accruing interest on their deposit through time. Upon depositing a currency, AAVE mints the user some corresponding interest-bearing `aTokens`, which represents how much of a reserve they've deposited into the lending pool. These `aTokens` can be redeemed through the protocol to functionally withdraw their previously-deposited currency from the lending pool.

Borrows and repays function as their names would suggest. It is important to understand that borrowing a currency in AAVE is governed by smart contracts. Anyone can borrow any amount of currency from the AAVE lending pool as long as they follow the criteria specified by the appropriate smart contracts. Users who borrow in AAVE use the currency they have deposited into the protocol as collateral. Not all currencies that can be deposited in AAVE are allowable as collateral. Additionally, users can choose which of their deposited assets they want to allow for usage as collateral. In order to qualify to borrow some asset, a user must have enough deposited assets in the system that are usable as collateral so that the loan would be over-collateralized. The extent of over-collateralization required depends on the specific currencies being used as collateral. More specific details about the requirements for borrowing in AAVE can be read about in the AAVE whitepaper [1].

[3] https://www.thegraph.com

Liquidations, the most complicated of the transactions in our data, have a lot of unique information in each transaction. When a user performs a liquidation, they are always liquidating the account of another user. This means there are two users recorded for each liquidation transaction: the user being liquidated (the *liquidatee*), and the user performing the liquidation (the *liquidator*). Additionally, whereas other transaction types only interact with a single currency at a time, liquidations involve two currencies. There is the principal currency that the liquidator is paying off and the collateral currency that the liquidator is buying.

Collateral and swap transactions are the simplest transactions. Each one is functionally just the toggling of a setting in a user's account. Collateral transactions are made when a user wants to toggle whether a deposited currency can be used as collateral for loans they've made or plan to make. Swap transactions allow users to switch loans between stable interest rates and variable interest rates for an individual currency.

Table 1: Summary of transaction types in AAVEv2 data collected from November 30, 2020 to January 06, 2022

Transaction Type	Occurrences	Mean Value (USD)	Median Value (USD)
Borrow	124,899	$337,019.10	$14,983.18
Collateral	220,046	NA	NA
Deposit	239,836	$482,453.00	$4783.75
Redeem	170,516	$843,124.80	$26,633.09
Repay	81,650	$448,525.00	$25,314.20
Swap	2937	NA	NA

Transaction Type	Occurrences	Mean Principal (USD)	Mean Collateral (USD)
Liquidation	7,914	$74,798.06	$79,682.95

2.2 Survival analysis for DeFi

Survival analysis is a collection of statistical procedures for data analysis in which the outcome variable of interest is the time from an index event until the outcome event [5]. To apply survival analysis to AAVE, we must pick two events: the index transaction and the outcome transaction. The survival time is the elapsed time between the index and outcome transactions. If, for example, we want to understand how soon users borrow money after making a deposit, then the index transaction is the user's deposit, and the outcome transaction is the first borrow the user makes thereafter. The data is right-censored since we can only analyze data until the final date in our data. If a deposit transaction has no matching borrow, it could be because the user never borrowed or because the user borrow had not occurred during the period of analysis. Survival analysis correctly analyzes right-censored data.

The power of DeFi results from the fact that it can be performed on the time duration between any desired pair of transactions analyzed by any of the widely used statistical survival analysis techniques. In Section 3, we demonstrate how we can address many different questions by changing the definition of the index and outcome transactions. For each analysis below, we define precisely the index and outcome transactions. Survival analysis estimates the survival function for each of these pairs of transactions. The survival function captures the probability of the outcome transaction *not* occurring through time. We utilize the `ggsurvplot` function from the `survminer` package in R to produce Kaplan–Meier curves, the most popular way to both estimate and visualize survival functions. For example, Figure 1 shows survival curves for each transactions type. They represent the time from the deposit to the first transaction of that type. If we are interested in how variables affect time to the outcome event, we utilize Cox regression (or proportional hazards regression). In Section 3.2, we utilize Cox regression to see if the coin type of borrows (stable or non-stable) is associated with faster liquidation rates.

3 Results

We use survival analysis to dissect the relationships between certain pairs of transaction types. In doing so, we demonstrate the effectiveness of survival analysis in visualizing and quantifying user behavior in AAVE. We explain how to transform the raw transaction data into forms suitable for survival analysis. Then we show different ways to use this data and survival analysis tools to uncover patterns of user behavior through the selection of different index events, the narrowing down of outcome events, and the separation of the data by other relevant features.

3.1 What do users do after a deposit?

We show that a Kaplan–Meier survival curve can provide a useful picture of how users behave after they deposit money into their accounts. Deposits are the natural first transaction for a user to make in a lending protocol, since before depositing any currency into an account, there aren't really other possible actions one can take. Thus, looking at how users behave after making deposits seems a natural place to begin our analysis.

To convert transaction data into survival data, we treat each deposit present in our data as an index event. This means we have 239,836 index events in the survival analysis. The outcome of each event occurs when the user makes their next transaction. The time difference between the deposit and the next transaction serves as the "survival" time for this analysis, and so if a user just makes a deposit and performs no subsequent transactions, that would manifest in our data as a deposit that has "survived" so far, and would be censored by time. For each deposit that is eventually followed by a subsequent transaction,

Table 2: Survival Data from Deposits to Next Transaction

Time From Index Event (in hours)	Censored?	Next Transaction Type
0.0295	False	Deposit
16.373	False	Deposit
0.2765	False	Borrow
2.58	True	NA
⋮	⋮	⋮

we also record the type of transaction that follows. This gives data in the form given in Table 2.

Starting with the simplest survival analysis method, we used the R function `surv_median` to calculate the median time to the first transaction after a deposit. We computed the median time to any transaction and for each transaction type. We show these values in Table 3 ordered by median survival time. The percentage of the time this transaction type followed a borrow is provided as well. The median time between any two transactions is .017 hours. That means 50% of the time, a deposit by a user is followed by a transaction by that same user in less than 1.02 minutes. Note that for this specific analysis, we do not consider the type of coin used in the second transaction (outcome event.)

From the median times to each transaction type following deposits, we can more clearly compare the magnitudes of elapsed time for different outcome transaction types. The speed at which users tend to engage in redeem and collateral transactions after a deposit is less than a second. The median survival to the next deposit and borrows is on the order of minutes. In contrast, many users take more than two days to repay or be liquidated after a borrow.

Table 3: Median time from deposit to next transaction in hours with percentage occurring for each transaction type

Next Transaction Type	Median Survival Time in hours	Percentage
Any	0.017	100%
Collateral	0.000 (<1 second)	32.62
Redeem	0.000 (<1 second)	25.69
Deposit	0.103	27.11
Borrow	0.231	10.74
Swap	15.108	0.13
Repay	25.488	2.68
Liquidation	61.053	0.22

We can gain further analysis by plotting survival curves for this data split by the type of subsequent transaction to gain a more nuanced understanding of how quickly users make each type of transaction following a deposit. When we separate the survival curve by the next transaction type, the usage of survival analysis deviates slightly from its original intent. This is because when we separate the curve by subsequent transaction type, we are effectively separating the curve by a variable that only exists because the index event has "not survived," i.e., the existence of a "next transaction type" means the outcome event has occurred, and thus we've cut out the censored transactions. However, we believe the value of the visualization remains intact.

Figure 1 shows seven different survival curves. As is the case for any survival curves, each curve begins at time 0 and has a probability of survival of 1. Then, as each curve progresses through time, the probability of "survival" drops in proportion to how many of the cases represented by each curve have reached their corresponding outcome event. If a curve drops quickly to a low survival probability, that means users tend to make that transaction type quickly after making a deposit.

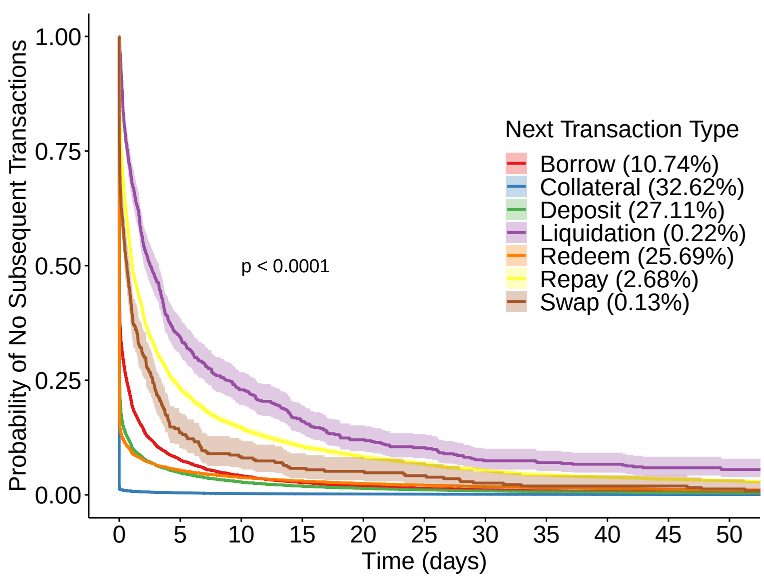


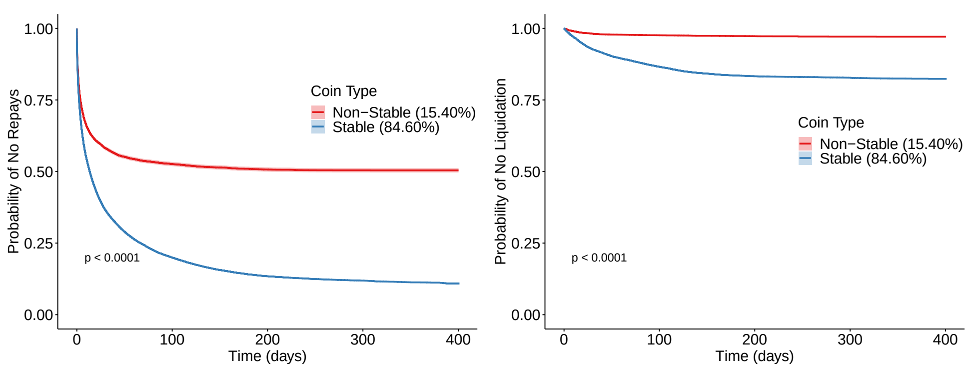
Fig. 1: Kaplan–Meier survival curves from a user's deposits to their next transaction shows noticeable differences between which transactions users tend to make after a deposit.

Looking at Figure 1, it is apparent that the most common transaction users make after a deposit is the collateral transaction. This makes sense, as one of the requirements in AAVE for a user to take out a loan is that the loan is over-collateralized; according to the requirements of the protocol, the user cannot take out a loan without having any currency in their account that is marked as collateral. Still, the "survival" of a deposit until a collateral transaction is brief. The survival curve for collateral dips straight down and hits zero almost immediately, indicating that users don't wait long before making collateral transactions.

3.2 How long until borrows are repaid or liquidated?

Since survival analysis allows for flexible choices of index and outcome events, we turn our focus now to using borrows as the index events and the relevant transactions of repays and liquidations as the outcome events. Using only transaction-level data to analyze borrows makes it difficult to track a loan in its entirety. There is no end date to the loan. The user can make several borrows of a coin and then maintain the loan that accrues interest until they repay it in one or more repay transactions for that coin or until part or all of the loan is liquidated. When converting the transaction data into a form usable for survival analysis with borrows as the index events, we had to decide what was an appropriate outcome event. To avoid making assumptions about when a loan is totally repaid (either through liquidations or repay transactions), we define the outcome events as just the first repayment that a user makes or the first liquidation that is made for the borrowed currency. Thus, to be clear, the following analysis in Figure 2 does not show how long it takes for loans to be totally repaid through repays or liquidations, but just how long it takes for them to *start* being repaid through either means.

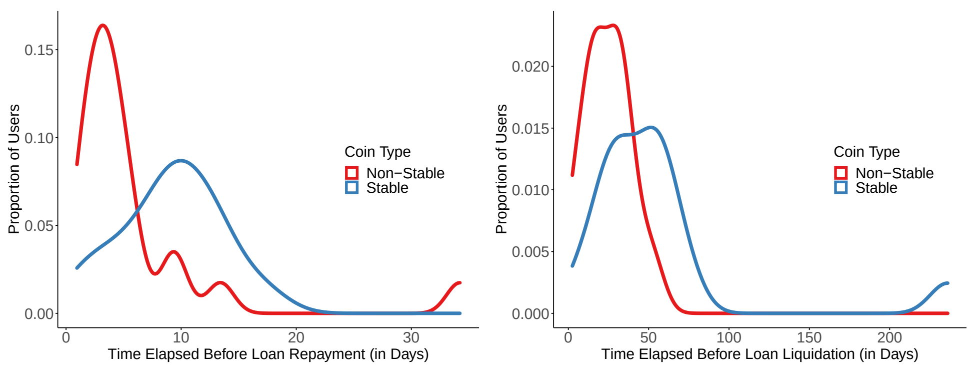
We also choose to split the two curves by whether the borrowed coin is a stablecoin. The use cases for borrowing stablecoins versus non-stablecoins are quite different, so we hoped to see a drastic difference in the repayment schedules and liquidation tendencies for these different coin types, and indeed this is what we see in Figure 2.



(a) Borrow to First Repay (b) Borrow to First Liquidation

Fig. 2: Borrow to first repay and first liquidation analysis shows significant differences between stablecoins (red) and non-stablecoins (blue). Note that the time ranges for repay is smaller than that of liquidation.

Clearly, from these survival curves, users tend to repay loans of stablecoins much more often than non-stablecoin loans. Almost all loans of stablecoins end up being repaid at least in part by the 402-day cutoff, which is as much data as we have. This is in stark contrast with the non-stablecoin borrows, for which only about 50% of all loans have seen even a single repayment. We see similarly contrasting behavior for the frequency that loans of each coin type are liquidated. Loans of stablecoins are liquidated significantly more often than loans of non-stablecoins. Unsurprisingly, liquidations as a whole occur far less frequently than repayments, so the survival probability of loans relative to liquidations is much higher than loans to repayments, but the same patterns are present with respect to the type of coin being borrowed.



(a) Borrow to First Repay (b) Borrow to First Liquidation

Fig. 3: Borrow to first repay and first liquidation analysis both show differences between stable and non-stablecoins.

We can also use the survival data to create density plots for the median time to repayment or liquidation for each coin type, which provides additional clarity on exactly how long it takes for repayments or liquidations to occur. These median repayment and liquidation times can be seen in Figure 3. From these, we can see that, despite fewer repayments and liquidations of non-stablecoins, the median times to these events are quicker than they tend to be for stablecoins. Most people make their first repayment of a borrowed non-stablecoin in five days or less, whereas for borrowed stablecoins, the median time until the first repayment is about ten days. The timelines for liquidations to occur are longer, with the highest proportion of liquidated borrows of non-stablecoins occurring between 20 and 30 days following the borrow, and between 25 and 75 days for borrows of stablecoins. Still, the contrasting behavior between stablecoin and non-stablecoin borrows are similar for the time taken to repays and time taken to liquidations.

3.3 How does coin type influence liquidations?

In the previous section, we saw the dramatic impact of coin-type of the principal on time to the first liquidation of a borrow. We hypothesize that the combination of the principal and the collateral may lead to further insight into the risk of borrows. Thus we further separate the borrow-to-liquidation data by factoring in what collateral was purchased and what principal types were specifically paid off by the liquidator. We perform the same index and outcome events as the prior liquidation analysis; only now do we analyze borrows associated with liquidation. This gives the curves seen in Figure 4. Since we are splitting the curves by what principal and collateral were paid off and purchased at the time of the liquidation, all the curves do end up with a 0% probability of survival, similar to the curves in Figure 1. Again though, we can still use the curves to gain insight into the relative riskiness of the principal:collateral combinations that people can have in their accounts. According to the log-rank test, the differences in the curves are statistically significant.

The definition of the outcome event in this analysis is quite different. To gain a more accurate picture of the liquidated user's account, we aggregated liquidation events to gain more information as to which coins the users have as collateral in their account. Even though each liquidation transaction only records one principal type and one collateral type, sometimes a user will be the subject of multiple liquidations in quick succession. It would be inaccurate to consider these liquidations as separate events; they really are all part of one bigger liquidation event. Thus, in our transaction data, if a user is liquidated multiple times in quick succession with no intermittent non-liquidation transactions, we aggregate them into one bigger liquidation transaction. The outcome event is the combined liquidation transaction, with the time being the first liquidation transaction. This lets us see whether there were multiple types of collateral and principal coins involved in the event. Thus, if a user has both stablecoins and non-stablecoins in their account as collateral, or if they've taken out loans of both stablecoins and non-stablecoins, we mark the collateral or principal, respectively, as "Mixed."

Table 4 shows an example of where we can use a Cox proportional hazard model to more effectively quantify the differences between survival probabilities of each type of principal:collateral combination leading to liquidations. For the Cox proportional hazard model, we need to choose one of the combinations of principal:collateral to which to compare the others, which tells us proportionally how risky the other types of principal:collateral combinations are relative to the benchmark combination. If we select the stable:stable combination as the benchmark, we get the quantification of risk via the `coxph` function from the `survminer` package as seen in Table 4.

The "Coefficient" column of this table indicates the relative riskiness of the loan types as compared to the benchmark type of stable principal and stable collateral. Negative coefficients are indicative of lesser risk, meaning that any of the principal:collateral combinations that have a negative coefficient are less likely to be liquidated than the stable:stable combination. The p-value tells the statistical significance of these coefficients, with lower p-values indicating that

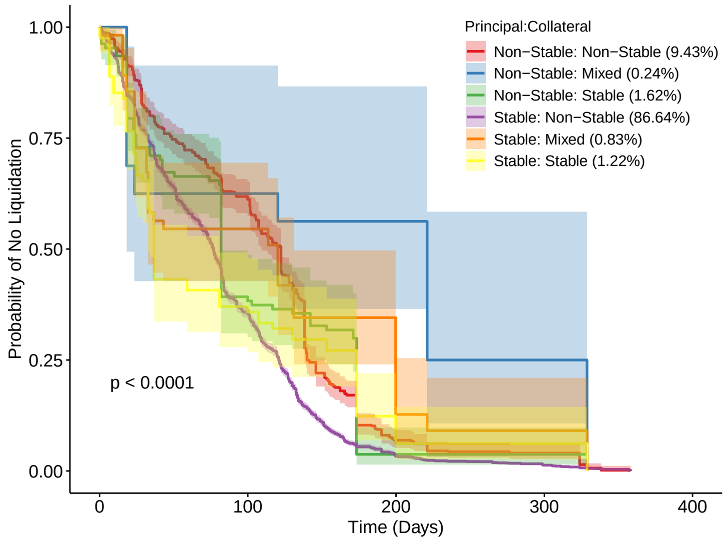


Fig. 4: Survival curves for different combinations of principal and collateral coin types.

the corresponding coefficients were less likely to be generated by chance. The results show that the combinations of stablecoin principal with non-stablecoin collateral tend to be liquidated significantly sooner than stablecoin principal with stablecoin collateral loans. Borrows with stablecoin principal with non-stablecoin collateral are much riskier since the price of the cryptocurrency in dollars is highly volatile. We see that liquidations for borrows with mixed collaterals reduce the risk for both types of principals. There were no significant differences in time to liquidation compared with stable:stable combinations for the non-stable:stable and non-stable:non-stable combinations.

Table 4: Cox proportional hazards coefficients quantifying risk of liquidation. Bolded principal-collateral combinations have significant differences in risk of liquidation relative to stable:stable. Percentage of liquidation events indicates number of events of each type. Stable:stable constitutes 1.22% of liquidations.

Principal:Collateral Combination	Coefficient	p-value	Percentage
Stable:Non-Stable	0.24499	0.02903	86.64
Non-Stable:Stable	-0.08756	0.55238	1.62
Non-Stable:Non-Stable	-0.17044	0.14967	9.43
Stable:Mixed	-0.32461	0.06328	0.83
Non-Stable:Mixed	-0.74976	0.00617	0.24

4 Related Work

With an over-collateralized loan, a borrower must post collateral, i.e., provide something of value as security to cover the value of the debt, where the value of the collateral posted exceeds the value of the debt. This way, collateralization simultaneously ensures that the lender (likely a smart contract) can recover their loaned value and provides the borrower with an incentive to repay the loan. The "health factor" (HF) is a custom threshold in lending systems. If the debt collateral falls below the HF (typically below 1), the debt position may be opened for liquidation. Then the liquidators can purchase the locked collateral at a discount and close the borrower's debt position. Thus, leveraged positions are subject to liquidation when the debt becomes unhealthy, and a liquidator can repay the debt and benefit from a liquidation spread.

Given this novel form of automatic lending, a growing body of literature has studied liquidations on borrowing and lending platforms in DeFi. Qin et al. [9] have analyzed risk management provided by liquidators, acting on the protocol's user accounts. They have measured various risks that liquidation participants are exposed to on four major Ethereum lending pools (i.e., MakerDAO, AAVE, Compound, and dYdX) and quantified the instabilities of existing lending protocols. They have illustrated that the commonly used incentive mechanisms tend to favor liquidators over borrowers, causing the problem of so-called over-liquidation, leading to unnecessary high losses for borrowers. The only recourse the borrowers have to avoid such liquidations is to monitor their loan-to-value ratio when the market changes quickly because even a random drop in market prices can result in a cascade of liquidations. If there are any drops in the market, it can lead to self-accelerating pressure to sell, which further causes more problems for a blockchain-based DeFi, such as network congestion that leads to steep gas costs. We witnessed such an event in the ETH market collapse of March 13, 2020[4] that left some borrowers unable to react, despite imminent liquidations. It can be particularly bad for borrowers who get liquidated if market prices recover after a dip again, leaving them deprived of subsequent upward price participation. In general, regardless of market conditions, liquidations in DeFi are widely practiced, and related works such as Qin et al. [9] have quantified that over the years 2020 and 2021, liquidators realized a financial profit of over 800M USD while performing liquidations.

Stablecoins play a significant role in liquidations, as they have several characteristics that are directly tied to liquidation mechanics. For example, a user may not want to sell the token collateral, which is usually in the form of a stablecoin, but instead hold it indefinitely as a means of passive income, which might exceed the cost of borrowing, making the transaction profitable. Early empirical evidence on the stability of crypto-backed loans with stablecoins has been studied by Kozhan and Viswanath-Natraj [7]. They specifically focused on the price volatility in the MakerDao protocol, which introduced the world's first decentralized stablecoin called Dai that is soft-pegged to the US Dollar, i.e.,

[4] https://coinmarketcap.com/historical/20200313

it uses a collateralized debt position mechanism to keep the price stable with respect to the US Dollar. They have analyzed how collateral stability increases peg stability and found a positive relationship between collateral risk and the price volatility of the stablecoin Dai.

The efficiency of lending pool liquidations has been studied by Perez et al. [8], in which they introduced a lending pool state model that is instantiated with historical user transactions observable in the Compound[5] implementation deployed on Ethereum. Their model abstraction facilitates the observation of state effects of each interaction and investigates the latency of user liquidations following the under-collateralization of borrowing accounts. Similarly, Bartoletti et al. [3] provide an abstract formal state transition model of lending pools and prove fundamental behavioral properties, which had previously only been presented informally in the literature. Additionally, the authors examine attack vectors and risks, such as utilization attacks and interest-bearing derivative token risk.

As the demand for loans in crypto-assets grows, the borrowing interest rate goes up. In a bullish crypto market, speculators may be keen to borrow funds even if there is a high interest rate, in expectation of an appreciation in the assets of their leveraged long position as demonstrated by Xu et al. [13]. Such an environment is advantageous for lenders, resulting in higher yields to them. Compound and AAVE, two major DeFi lending protocols, have witnessed the borrow APY of the stablecoin USDC increasing from a low of 2-3% in May 2020 to as high as 10% in April 2021 (as of this paper writing in March 2022, the APY is hovering at 2% in Compound and AAVE, but other protocols such as Celcius, BlockFi and Nexo offer upwards of 8% APY) [6]. In a bullish market, the yield generated is incorporated in interest-bearing tokens, such as `aTokens` from AAVE analyzed in this paper. However, as was already noted, the wild fluctuations in the market result in unexpected liquidation events, as evidenced from this paper's results.

Most of the related works approach the issue of liquidation at a conceptual level or rely on aggregate flow data. In contrast, our paper uses transaction-level blockchain data to provide a more microscopic view on the issue combined with survival analysis techniques.

5 Discussion and Future Work

This work defines a pipeline for survival analysis of DeFi lending protocols which includes data aggregation, cleaning, converting to a data abstraction model, and performing powerful survival statistical analyses and visualizations to gain insights. Using AAVE lending data in three different scenarios, we demonstrate how to gain insights into user behaviors and loan risks by defining appropriate index and outcome events and then applying survival analysis.

Each scenario is characterized by distinct definitions of the index and outcome transactions in the survival analysis. We characterize user behaviors using

[5] https://compound.finance
[6] https://defirate.com/usdc/?amount=100&symbol=USDC&term=365&rate_type=lend

survival analysis of AAVE users' next transaction. Our analysis of borrowed coin types shows the value of survival analysis for discovering factors contributing to events. Our survival analysis of borrows to repays and liquidations showed that borrows of stablecoins versus non-stablecoins exhibit very different characteristics. Users hold non-stable loans longer before the first repayment. But if they do repay, they tend to repay more quickly. We could get a deeper understanding of AAVE user behavior by taking a more refined look at the types of coins and transaction volumes, incorporating external factors such as coin prices in the survival analysis, and defining alternative index and outcome events. Using machine learning to create clusters that capture different behaviors (e.g., retail versus institutional investors), and then doing survival analysis could also be very illuminating. We leave these to future work.

This work represents just the first step in the use of survival analysis in DeFi. We note that these DeFi survival analysis techniques could be generalized to other DeFi lending protocols. DeFi survival analysis can be applied to any cryptocurrency protocol with transactions. Hazard analysis and other types of survival analysis and visualization methods could be used. As future work, we will prepare a toolkit for DeFi survival analysis with associated dashboards and demonstrate it on other DeFi Protocols.

One limitation of this research is that it does not address the rich DeFi ecosystem, which has many interacting protocols and coin prices. We are already exploring the use of more advanced Artificial Intelligence (AI) methods for the analysis of transaction data developed for commerce and health [4, 12] that could incorporate more aspects of the DeFi ecosystem. These could be used for segmenting users and predicting behaviors and prices. Early results analyzing AAVE transactions using Neural Temporal Point Processes are very promising [11]. DeFi represents an exciting new domain for AI research in transaction modeling since DeFi is a compelling use case, and all the datasets are public by definition.

The code used to generate the figures in this paper is available in a public GitHub repository[7].

6 Acknowledgements

This work was partially supported by the Rensselaer Institute for Data Exploration and Applications. We would like to thank Roman Vakhrushev for his contributions to the liquidation analysis, Soumya Mishra, and Jaimin Vyas for contributions to early versions of the survival analysis.

References

1. AAVE Developers: AAVE Protocol Whitepaper V2.0. Tech. rep. (12 2020)

[7] https://github.com/aaronmicahgreen/DeFi-Survival-Analysis-Insights-into-Risks-and-User-Behaviors

2. Allen, L.N., Rose, L.C.: Financial survival analysis of defaulted debtors. Journal of the Operational Research society **57**(6), 630–636 (2006)
3. Bartoletti, M., Chiang, J.H.y., Lafuente, A.L.: Sok: lending pools in decentralized finance. In: International Conference on Financial Cryptography and Data Security. pp. 553–578. Springer (2021)
4. Bennett, K.P.: Artificial intelligence for public health. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1–2 (2019). https://doi.org/10.1109/BIBM47256.2019.8983112
5. Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G.: Survival analysis part i: basic concepts and first analyses. British journal of cancer **89**(2), 232–238 (2003)
6. Dirick, L., Claeskens, G., Baesens, B.: Time to default in credit scoring using survival analysis: a benchmark study. Journal of the Operational Research Society **68**(6), 652–665 (2017)
7. Kozhan, R., Viswanath-Natraj, G.: Decentralized stablecoins and collateral risk. WBS Finance Group Research Paper Forthcoming (2021)
8. Perez, D., Werner, S.M., Xu, J., Livshits, B.: Liquidations: Defi on a knife-edge. In: International Conference on Financial Cryptography and Data Security. pp. 457–476. Springer (2021)
9. Qin, K., Zhou, L., Gamito, P., Jovanovic, P., Gervais, A.: An empirical study of defi liquidations: Incentives, risks, and instabilities. In: Proceedings of the 21st ACM Internet Measurement Conference. pp. 336–350 (2021)
10. Roszkowska, P., Prorokowski, Ł.: Model of financial crisis contagion: A survey-based simulation by means of the modified kaplan-meier survival plots. Folia Oeconomica Stetinensia **13**(1), 22–55 (2013)
11. Shou, X., Gao, T., Subramanian, D., Bennett, K.P.: Multi-label event prediction in continuous time. Under review (2022)
12. Shou, X., Mavroudeas, G., New, A., Arhin, K., Kuruzovich, J.N., Magdon-Ismail, M., Bennett, K.P.: Supervised mixture models for population health. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1057–1064 (2019). https://doi.org/10.1109/BIBM47256.2019.8983339
13. Xu, J., Vadgama, N.: From banks to defi: the evolution of the lending market. In: Enabling the Internet of Value, pp. 53–66. Springer (2022)