DeFi Survival Analysis: Insights Into the Emerging Decentralized Financial Ecosystem

AARON M. GREEN, MICHAEL P. GIANNATTASIO, JOHN S. ERICKSON, OSHANI SENEVIRATNE, and KRISTIN P. BENNETT, Rensselaer Polytechnic Institute, USA

We propose a survival analysis approach for discovering and characterizing user behavior and risks for lending protocols in decentralized finance (DeFi). We demonstrate how to gather and prepare DeFi transaction data for survival analysis. We illustrate our approach using transactions in Aave, one of the largest lending protocols. We develop a DeFi survival analysis pipeline that 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 modified for competing risks when appropriate, such as Kaplan–Meier survival curves, cumulative incidence functions, Cox hazard regression, and Fine-Gray models for sub-distribution hazards to gain insights into usage patterns and risks within the protocol. We show how, by varying the index and outcome events as well as covariates, we can use DeFi survival analysis to answer questions like "How does loan size affect the repayment schedule of the loan?"; "How does loan size affect the likelihood that an account gets liquidated?"; "How does user behavior vary between Aave markets?"; "How has user behavior in Aave varied from quarter to quarter?" 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.

CCS Concepts: • **Mathematics of computing** → **Mathematical analysis**; • **Applied computing** → **Electronic commerce;**

Additional Key Words and Phrases: Decentralized finance, aave, survival analysis, competing risks, financial technologies, fintech, economics, data analytics

ACM Reference Format:
Aaron M. Green, Michael P. Giannattasio, John S. Erickson, Oshani Seneviratne, and Kristin P. Bennett. 2024. DeFi Survival Analysis: Insights Into the Emerging Decentralized Financial Ecosystem. *Distrib. Ledger Technol.* 3, 1, Article 4 (March 2024), 23 pages. https://doi.org/10.1145/3638064

1 INTRODUCTION

The rise in popularity of Bitcoin in the last decade has brought with it the novel study of blockchain technologies in both academic and industrial environments. An accompanying new financial ecosystem has started to emerge, called **Decentralized Finance** (**DeFi**). Built using smart contracts (code deployed onto blockchains), the infancy of this novel financial market has seen developers attempting to recreate the functions of traditional, centralized

The authors acknowledge the support from NSF IUCRC CRAFT center research grant (CRAFT Grant #22003) for this research. The opinions expressed in this publication do not necessarily represent the views of NSF IUCRC CRAFT. This work was supported by the Rensselaer Institute for Data Exploration and Applications (IDEA).
Authors' address: A. M. Green, M. P. Giannattasio, J. S. Erickson, O. Seneviratne, and K. P. Bennett, Rensselaer Polytechnic Institute, 110 8th St, Troy, New York 12180; e-mails: {greena12, giannm, erickj4, senevo, bennek}@rpi.edu.

ACM 2769-6472/2024/03-ART4
https://doi.org/10.1145/3638064

financial services in a decentralized way. Banking has been remade into **lending protocols**, currency exchanges have been recreated as **decentralized exchanges**, and so on. Each DeFi protocol comes with its own set of novelties, attempting to innovate on traditional financial infrastructure in a unique way. As DeFi grows and matures, creating tools that help track and interpret changes in the DeFi ecosystem should prove interesting and useful, as the richness and volume of data publicly available in DeFi is unprecedented. We propose survival analysis tools which help address two big-picture questions about DeFi: "How are users behaving within DeFi protocols?" and "How are user-level behaviors within DeFi protocols changing through time?"

For this work, we demonstrate our approach on Aave, one of the largest lending protocols [4][1]. Our general survival analysis approach, however, can be readily generalized to the many existing and new DeFi protocols. Aave [4] has a market size of over $8.14 billion as of November 1, 2022, and is deployed on seven different networks, including Ethereum, Polygon, and Avalanche, to name a few. Lending protocols serve a similar role in DeFi to banks in traditional finance; they allow users to deposit cryptocurrencies into a lending pool, granting them interest on their deposited assets. Users can withdraw the deposits as desired. Users can also borrow funds from the lending pool so long as they provide sufficient collateral for the asset type and quantity they seek to borrow. Users can repay loans or keep them as long as desired as long as they maintain sufficient collateral. Should the value of their collateral drop too much or they allow too much interest to accrue on their loans, their account is at risk of being liquidated. Liquidations in Aave are one of the methods implemented by the developers to allow for the maintenance of the health of the lending pool. The protocol is configured to pay out a small bonus to any user (called a keeper) who is willing to repay another user's unhealthy loan balance in exchange for the collateral. When the keeper repays a liquidatee's loan, they receive the appropriate proportion of the liquidatee's collateral (their deposited assets), and also receive a small, percentage-based reward from Aave in order to incentivize liquidation.

Since these transactions are the building blocks of DeFi data streams, we applied a method of data-driven time-to-event analysis called "survival analysis" to the transactions streams. Survival analysis in the context of DeFi proves a versatile technique for derivation of myriad insights into macro-level user behaviors and how these behaviors have changed through time. We show here just a few of the many compelling results from applying survival analysis methods to Aave's transactions. This article is organized as follows: in the methods section, we describe the source and structure of Aave data used, the survival analysis methods employed, and an overview of the application created in conjunction with this article. In the results section, we demonstrate the use provided by survival analysis to answer various questions. We conclude with a discussion of the contributions of this work and promising directions for future work.

2 METHODS

2.1 Transaction-Level Data

The data used for this analysis is from The Graph[2]. Aave pushes its own data to The Graph, and each network Aave is deployed on has its own subgraph. These subgraphs are structured identically, at least with respect to the transaction-level data with which our analyses are concerned. We have collected all of the transaction data from nine subgraphs maintained by Aave, which include the data from the following Aave markets: Ethereum [6] (both Mainnet and Automated Market Maker versions), Polygon [15], Avalanche [30], Optimism [37], Harmony [34], Fantom [7], and Arbitrum [16]. There are currently two versions of Aave in deployment, referred to simply as V2 [5] and V3 [11] (V1 went by a different name and is, for all intents and purposes, no longer in use). The Ethereum markets only operate on AaveV2, the Polygon and Avalanche markets have both AaveV2 and AaveV3 versions in deployment, and the rest are exclusive to AaveV3. We illustrate survival analyses of a single market

[1] https://aave.com/
[2] https://thegraph.com/en/

Table 1. Summary of Transaction Types from Aave's Ethereum Market Collected from November 30, 2020 to October 01, 2022 Providing the Average Values of Each Transaction Type

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	191,103	$350,886.90	$11,048.17	NA	NA
Deposit	363,282	$653,439.90	$11,100.46	NA	NA
Withdraw	260,079	$871,974.2	$25,044.53	NA	NA
Repay	132,567	$481,908.4	$22,352.36	NA	NA
Liquidation	26,040	NA	NA	$39,026.36	$41,571.71

Note that for liquidation transactions there are two currencies involved: the principal currency that the keeper is paying back to Aave, and the collateral currency that the keeper is receiving from the liquidatee's account.

Table 2. Snapshot of Transaction-Level Data from Aave's Ethereum Market that was Used as the Dataset for this Analysis

Date and Time	Type	User	Coin	Amount	Amount (in USD)	···
Dec. 1, 2020 05:15:00	Deposit	<ID>	USDT	100.00	100.00	···
Dec. 1, 2020 05:15:30	Borrow	<ID>	XSUSHI	15.52	100.00	···
⋮	⋮	⋮	⋮	⋮	⋮	⋱
Sept. 30, 2022 23:50:00	Repay	<ID>	DAI	25,000.667	24,978.34	···
Sept. 30, 2022 23:50:45	Withdraw	<ID>	WETH	3.652	8,976.09	···

and across different markets. Since, the various Aave markets come with the advantages and disadvantages of the networks (blockchains) on which they are deployed, examining differences in user behaviors across markets can naturally lead to interesting insights regarding how user behaviors change across markets. The data used here includes all transactions made on Aave's V2 deployment on the Ethereum blockchain from its creation on November 30, 2020 through October 1, 2022.

The majority of the analyses presented in this article focus solely on the Aave Ethereum market. This is for a few reasons: Aave is still fairly new, and the Ethereum market is the oldest of the deployments and, thus, its data spans the longest duration; the value locked on the Ethereum market is, as of November 1, 2022, approximately $6.16 billion, accounting for nearly 75% of the value locked across all of Aave's markets; and the number of transactions from this market comprises only roughly 5% of the total transactions across all markets combined, which makes computation of its associated survival data easier to test. For the data used in this article, we include five transaction types from Aave: deposits, withdraws, borrows, repays, and liquidations. There are some other transaction types that Aave collects, but they vary from market to market and the quantity of these transactions are so low that we do not feel they are important for the sake of this analysis; e.g., swaps (a user swapping from stable to variable borrow rates, or vice versa) comprise just 3,382 transactions out of the 1,475,175 in Aave's Ethereum market, and is not tracked in other markets. A summary of the data used from the Ethereum market is given in Table 1. Descriptions of each transaction type are provided in Section 2.1.1 below, and Table 2 shows a sample of the raw transaction data.

Note that Aave allows the usage of 92 different coins across all markets. For insightful analysis, we divide the coins into two types: stablecoins and non-stablecoins. A stablecoin is from a class of cryptocurrencies that attempts to offer price stability, typically in terms of USD. The other types of coins in the dataset are non-stablecoins. A breakdown of the coins available in Aave's Ethereum market and whether they are classified as stable or non-stable can be found in Appendix B. There are 11 different stablecoins in the Aave's Ethereum market, and 28 other coins which we call non-stable.

2.1.1 Transaction Type Descriptions. The most fundamental transaction type in Aave is the "deposit". Like opening a savings account with a bank, deposited assets will slowly accrue interest, providing users some incentive to use the platform even passively. Naturally, users can withdraw their deposited assets from the protocol using the "withdraw" transaction type.

An important characteristic of an account in Aave is how much collateral a user has posted. When a user deposits a currency into Aave, in most cases, they are able to select whether they want that currency to be used as collateral in their account (certain currencies are disallowed as collateral by Aave). Having one or more assets enabled as collateral in an account will allow a user to "borrow" assets from the Aave lending pool. The amount of cryptocurrency a user wants to borrow is capped by a percentage of the total value of the assets they have enabled as collateral in their account. These assets are all valued relative to their conversion rate with the Ethereum cryptocurrency. When a loan has been taken out through a borrow transaction, the user can "repay" the loan over time.

The most complex transaction type in Aave is the "liquidation." When someone borrows funds in Aave, they borrow them against the value of their collateral assets. As the loan accrues interest, and as the values of the collateral and borrowed currencies fluctuate with the market, a loan that was originally healthy can become unhealthy. If the health of a user's account (calculated using the relative value of the user's collateral assets and borrowed assets) deteriorates too much, the user's account can be liquidated. This means another user (called a keeper) is allowed to pay off a portion of the unhealthy loan and claim the appropriate portion of the user's collateral, and gain a small bonus from the Aave protocol for having done so. Liquidations are an important method for keeping the protocol healthy overall, since they provide an incentive for someone to repay a loan that was otherwise losing value for the lending pool.

2.2 Creating Survival Analysis Data

Survival analysis is a collection of statistical procedures for data analysis in which the variable of interest is the time from an index event until the outcome event [9]. Basically, survival analysis allows for time-to-event analysis. There are four primary choices to make when using survival analysis: an "index event," an "outcome event," a "covariate" of interest, and an "observation period." Index events trigger the beginning of a single observation. After some time has elapsed following an index event, an outcome event may occur. Outcome events trigger the end of an observation that began with an associated index event and the appropriate elapsed time is calculated. It is possible as well that the period of observation ends prior to the occurrence of an outcome event for a specific index event. In this case, the event is considered to be "right-censored," and a note of this censoring is made. The elapsed time is the time between the index event and the end of the observation period. The survival data will be segmented by each possible value of the chosen covariate, allowing us to see the effects the covariate has on the outcomes. Also, in some cases there are multiple kinds of transactions competing to cause the same outcome. When relevant, we keep track of what event/transaction type caused the outcome in order to account for competing risks.

As an example, consider the question "How long do users take to repay stablecoins versus non-stablecoins after a borrow?" We can use survival analysis to answer this question by choosing borrow transactions as index events, repay transactions of the same coin (i.e., the principal reserve coin) as outcome events, the coin type (stable or nonstable) as the covariate, and an observation period of interest. For the sake of this example, let the observation period be the time from Aave's inception until the end of our data collection (Oct. 1, 2022). From these choices, the survival-style data we need to create will look like Table 3.

Survival analysis with respect to covariates can allows us to address different questions. Covariates can be any factor associated with an index event. For instance, in the example in Table 3, we added the covariates such as which currency was borrowed (i.e., coin), or whether the currency is a stablecoin (i.e., coin type.) New covariates can easily be defined and added as desired to achieve analysis goals.

Table 3. Sample Survival-Style Data with **Index Events** being Borrows and **Outcome Events** being Repayments, with Additional Covariates

Elapsed Time (days)	Censored?	User	Coin	Coin Type	Aave Market	⋯
12.6	False	<ID>	WETH	Non-Stable	Ethereum	⋯
68.3	False	<ID>	USDC	Stable	PolygonV2	⋯
54.1	True	<ID>	DAI	Stable	Optimism	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱
155.0	False	<ID>	AMPL	Non-Stable	AvalancheV2	⋯

2.2.1 Computing Survival Data with Different Observation Periods. The choice of observation period allows for additional versatility in what questions we can answer with survival analysis, and brings with it slightly different steps for computing the survival data. For the work presented here, we use two different observation periods. The default observation period we use is **since inception**. For survival data computed since inception, we consider data from the start of Aave through the end of our collected data (October 1, 2022). This observation period only requires consideration of right-censored events, where the index event is observed but no outcome event is observed. In this case, we mark that an index event occurred and that a duration from the time of the index event until the end of the observation period has passed, but that the event is censored. Unless otherwise specified, this is the observation period used in our analysis.

In Section 3.4, we change how we handle the observation period to **quarterly**. Instead of having just one observation period for the entire history of Aave, we break the data into quarterly observation periods (01-Jan through 31-Mar, 01-Apr through 30-Jun, 01-Jul through 30-Sep, and 01-Oct through 31-Dec). We only include quarters that we have complete data for, which includes 2021 Q1 through 2022 Q3. We treat each quarter as a separate observation period, which means that it is possible for an outcome event to occur with no associated index event having taken place in the observation period. This leads to "left-censored" events. If an outcome event occurs, say, 30 days after the start of the observation period and it was not preceded by an associated index event, we record in the survival data that an observation occurred with 30 days of elapsed time, and that the event was censored.

2.3 Survival Analysis Methods

Our primary method for visualizing survival analysis is with Kaplan–Meier survival curves [17]. To quantify the characteristics of the survival data we use a couple of different metrics: the **hazard ratios** (**HR**) [10] with corresponding p-values; and the **restricted mean survival time** (**RMST**) [28]. In cases where multiple event types compete to cause one outcome, we follow standard practices for presenting competing risk analyses. We use the **Cumulative Incidence Function** (**CIF**) to visualize competing risks. We also compare the **cause-specific hazard ratios** (**CSHR**) from multi-state Cox models with the sub-distribution hazard ratios (SDHR) from the Fine and Gray model to more accurately compare the effects that competing risks have on an outcome. We briefly explain how to read and interpret these below.

To create the Kaplan–Meier curves, compute the RMST, and calculate the hazard ratios, we employ the `survminer` package [18] and the `survival` package [35, 36] in R [26]. For Kaplan–Meier curves we use the `ggsurvplot` function from `survminer` package. To compute the RMSTs we use the `survival` package's `survmean` function. For calculating the (cause-specific) hazard ratios (and p-values), we use the `coxph` function from the `survminer` package. For plotting CIFs, we use the `ggcuminc` function from the `tidycmprsk` package [33]. We use the the `crr` function from the same package for creating the Fine and Gray models and calculating the SDHRs.

2.3.1 Kaplan–Meier Survival Curves. Kaplan–Meier curves show the probability of the outcome event having *not yet occurred* after an increasing amount of time has elapsed. Interpreting a Kaplan–Meier curve requires the

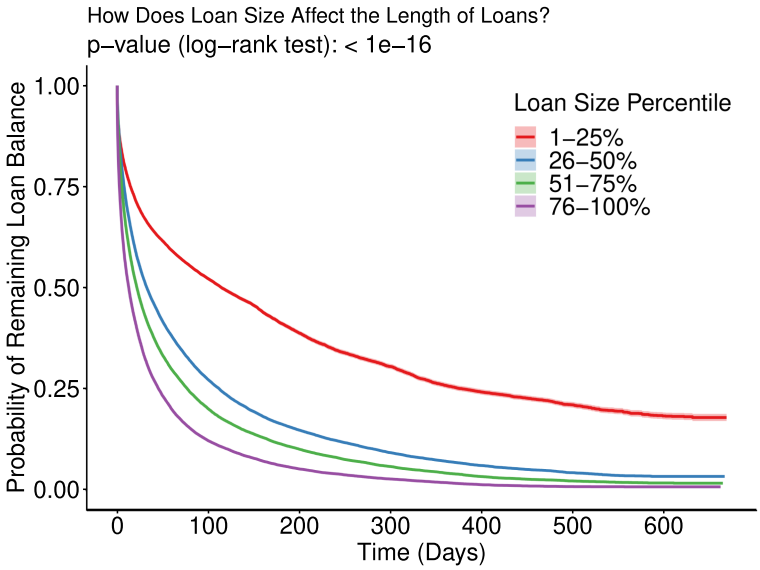


Fig. 1. Kaplan–Meier plot using borrows as index events, the full repayment of the borrow as outcome events, and the loan size quartile as a covariate.

understanding of three main components: the index and outcome events used to create the underlying survival data, and the covariate selected as a strata. Numerous Kaplan–Meier plots can be created from the same survival data based on which covariate is selected for the strata, and for each unique value of the selected covariate, a separate Kaplan–Meier curve will be drawn on the same plot. To read the Kaplan–Meier plots, then, we will use Figure 1 as an example.

In Figure 1, the index events are "when a user borrows from Aave", the outcome events are "when the balance of that user's loan has been fully repaid", and the covariate selected as the strata is the quartile of the amount in USD that was initially borrowed. Since we have segmented the amount borrowed into quartiles, there are four Kaplan–Meier curves included in this plot. The individual curves show how likely it is for the outcome event to have not yet occurred after time passes. In this example, if we focus on the red curve (1–25% of loans by size), we see that it takes nearly 150 days after a borrow has been made for the loan to have been repaid in full. Considering the curves for the other loan sizes, we see a clear gradient in how quickly loans are repaid based on how much was initially borrowed. The larger the loan was that a user took out, the faster they tend to repay that loan.

The statistical significance of the difference between any Kaplan–Meier curves in a plot are calculated and given as a p-value using the log-rank test. The p-values have been included atop the figures, and due to the magnitude of data present in the creation of these plots, these p-values tend to be very small. If the p-value is less than $1e-16$, then we round it to 0. Lastly, each Kaplan–Meier curve is plotted with a 95% confidence interval included. Similar to the p-values rounding to 0, the confidence intervals tend to be so closely fit to the curves themselves that they often are not visible, but nonetheless they are included in each Kaplan–Meier curve drawn.

2.3.2 Cox Proportional Hazards Regression. The Cox proportional-hazards model [10] is a model for quantifying the effects of covariates on the survival time of events. For a categorical covariate, Cox regression will use one of the values of that covariate as a reference point against which to compare the likelihood of survival through time. We compute and give the hazard ratios for the non-reference values of the covariates. Hazard ratios that are greater than 1 indicate that the value of the covariate "reduces survival likelihood" (the outcome

event is more likely to occur) relative to the reference. Hazard ratios less than 1 indicate the value of the covariate "increases survival time" (the outcome is less likely to occur). We also give the p-value provided for each Hazard Ratio to indicate the statistical significance of these results. The p-values are computed by a Wald test [38].

2.3.3 Restricted Mean Survival Times. We use the restricted mean survival time to calculate the mean survival time for types of events stratified by covariates across the entire observation period. For instance, in Table 6, we calculate the RMST of borrow-to-full-repay survival data stratified by loan size. For the smallest loans (1st Quartile) the RMST is 225.08 days, meaning that over the course of the entire observation period the average time it took for the smallest loans to be repaid was 225.08 days. Likewise, for the largest loans the RMST is significantly lower at 45.81 days. When competing risks are involved, we still report the RMST, treating separate causes as censoring. In this context, the RMST indicates the mean time until the event of interest without the occurrence of competing risks.

2.3.4 Competing Risks Analysis. In certain cases, the application of pure survival analysis can lead to incomplete or deceptive results due to multiple types of events competing as possible outcomes for the same index event. These are called "competing risks", and arise when the occurrence of one event type precludes the occurrence of another.

The standard method for analyzing competing risks data involves presenting both the "cause specific hazard ratios" (CSHR) and the "subdistribution hazard ratios" (SDHR) for a particular analysis. The CSHR is just the hazard ratio from the Cox model, quantifying a user's immediate risk of experiencing a specific outcome type. The SDHR is computed via the Fine and Gray model [12]. This model generalizes the Cox proportional-hazards model to accommodate competing risks. The SDHR quantifies the effect of covariates on the marginal probability of the event occurring, considering the presence of competing risks. The difference between the two hazard ratios is subtle. CSHRs show what happens to those who are still at risk for the event, whereas SDHRs tell you about the cumulative incidence in the whole population. Using both ratios can help confirm the robustness of our findings.

To perform this analysis, we first convert the boolean "Censored?" column to instead store the specific outcome event (i.e., liquidation, repay, censor). This data is then used to create the respective Fine and Gray Model and to plot the CIFs. Interpreting CIFs is similar to interpreting Kaplan–Meier curves, but the curves are inverted. Considering Figure 2, there are four CIFs plotted here. Each curve represents the likelihood over time that the outcome in question has taken place. The legend tells us what proportion of the events were using stable coins and non-stable coins, as well as what proportion of the total borrow events experienced each type of outcome. Note that the outcome percentages do not sum to 100%. This is because of censored events not experiencing either outcome. In the accompanying table (Table 5), we give the percentage of the combination of covariate and outcome.

3 RESULTS

3.1 Insights Into User Behavior After Depositing Currency

We show that survival analysis can provide a useful picture of how users behave after depositing money into their accounts. Deposits are the natural first transaction for a user to make in the Aave lending protocol, since before borrowing they must deposit collateral. Thus, looking at how users behave after making deposits seems a natural place to begin our analysis. Using deposits as index events and the first subsequent transaction of any type as outcome events, we get Figure 3. We are interested in what outcome events users are actually making, and so we must consider each transaction type as a competing risk. Therefore, we plot the CIFs for each transaction type. Since we are not factoring in any covariates to this analysis, there is no need to report any HRs. Of the 357,485 deposits, there were 28,058 that were not followed by another transaction and thus are not represented in this analysis. This analysis is not limited to users' initial deposits, and thus it is possible for these deposits to be followed by events like repays. This would indicate that the account has had previous activity.

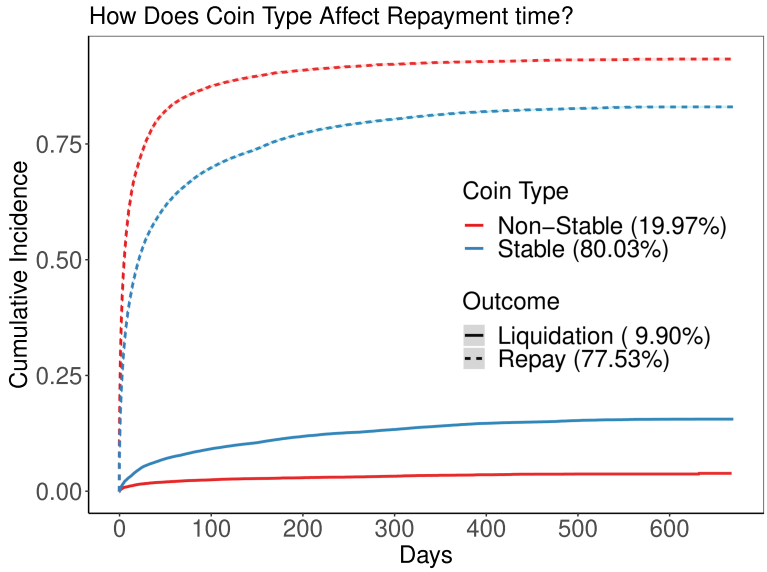


Fig. 2. CIFs with borrows as index events, first repayments or first liquidations as competing outcome events, and using the coin type as a covariate. The red curves, which focus on borrows made with non-stable coins, are repaid faster and liquidated slower than borrows made with stable coins.

Figure 3 provides us insights into what users are doing with their deposits. Users frequently make deposits and then immediately use them as collateral in borrows (36.52% of next transactions after a deposit are borrows). They tend to do so more quickly than any other transaction. However, based on the percentages of each outcome event, users are slightly more likely to make an additional deposit (37.61%) as their next transaction following a deposit fairly rapidly. Users depositing and then using the funds to repay loans occurs slightly less rapidly and are less common (6.40%). Deposits followed by withdraws occurs 18.58% of the time, but these tend to take much longer than the time to all other outcome types except liquidations. Unsurprisingly, since users have to borrow before making a liquidation, deposit-to-liquidations occur less frequently and tend to take the most time.

We can use Table 4 to better understand this analysis. From the RMSTs, we see that when a user borrows after a deposit, the mean time to do so is only 2.18 days. This is the action that users are the fastest to make by a significant margin. The next fastest action is making another deposit, for which the mean time to make is 6.89 days. Users are also more quick to make repayments after deposits than they are to withdraw their deposits. This seems intuitive enough, since a user will need to make a deposit of whatever currency they had borrowed in the past in order to repay it. The HRs and significant p-values further support these conclusions.

3.2 Insights Into Loans and Their Outcomes

Next we consider user behavior with respect to loans. Enabling users to take out loans without the need for explicit approval by another party is one of the more unique features of DeFi lending protocols. A user is able to borrow currency based on two factors: the amount of other currencies they have deposited into the lending pool; and which of their deposits they are willing to post as collateral for their loan. This smart-contract enabled lending is interesting, because it has no direct analog in traditional banking and also because the reasons for which someone would take out a loan in this setting are not immediately obvious. The requisite over-collateralization for loans makes motivation for loans less obvious, since the ability to take out a loan requires prior ownership of a relatively greater amount of crypto-assets in the first place. One possible motivation for borrowing assets

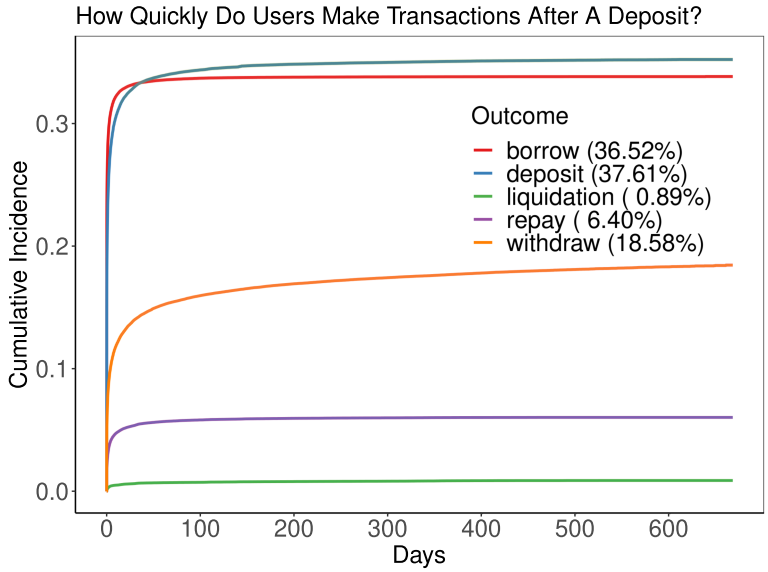


Fig. 3. CIFs for deposits as index events, any subsequent transaction as an outcome event, and separated by the type of outcome event. Legend includes percentage of each transaction type after deposit.

Table 4. Deposit-to-Next-Transaction: Quantifiers of the Survival Data Corresponding to Figure 3

Outcome Type	Percentage	RMST (days)
borrow	36.52	2.18
deposit	37.61	6.89
liquidation	0.89	36.62
withdraw	18.58	21.66
repay	6.40	10.64

Using Deposits as Index Events and Next Transaction as Outcome Events, with Each Transaction Type being a Competing Event.

in this way could be as a way to short a cryptocurrency while holding onto one or more assets that you believe will maintain or increase in value. Since the allowance of lending is the primary reason for lending protocols to exist, we feel that emphasizing behavior related to loans and their outcomes is the most natural and interesting place for survival analysis.

In Aave, loans are indefinite until some combination of two potential outcomes occur: repayment of the loan and liquidation of the account. Loans can be repaid in as many or as few repay transactions as the debtor desires, and there is no required timeline to pay back loans; the loan can last for as long as the user's account remains healthy. If the user's account becomes unhealthy, they can be liquidated. When a user is getting liquidated, the keeper can choose one of the unhealthy user's collateral coins and one of their principal reserve coins and pay off up to 50% of that principal coin the user has borrowed in order to claim an equal proportion of the collateral coin. The liquidator is paid a small bonus by Aave to incentivize this action. This is the other way that a loan can be partially repaid.

Table 5. Quantifiers of the Survival Data Corresponding to Figure 2, Using Borrows as Index Events, Repays or Liquidations as Outcome Events, and the Coin Type as Covariates

Coin Type	Outcome	RMST	CSHR	CSHR P-Val.	SDHR	SDHR P-Val.	Percentage
Non-Stable	Liquidation	585.79	Ref.	Ref.	Ref.	Ref.	2.07
Stable	Liquidation	496.17	2.43	<0.001 (***)	3.58	<0.001 (***)	8.23
Non-Stable	Repay	58.73	Ref.	Ref.	Ref.	Ref.	15.48
Stable	Repay	117.97	0.62	<0.001 (***)	0.59	<0.001 (***)	61.39

In either case, we will be using borrow transactions as index events and either repay transactions or liquidation transactions as the outcome events. We use competing risk analysis for some of these analyses, since both liquidations and repay transactions serve as ways to partially repay a loan. We note, however, that while these methods of repayment are competing, they are not mutually exclusive. It is possible for a loan to be partially repaid by the borrower, and to still be the subject of a liquidation event in the future. Likewise, it is possible for a loan to be partially liquidated and for the remaining balance to still be repaid by the borrower at a future time. Note that these outcome events only mark the first repayment or the first liquidation following the borrow, not the time for the loan to be fully repaid or fully liquidated.

3.2.1 How does Coin Type Affect the Outcomes of Loans? One covariate we hypothesize to have a significant effect on how users behave with their loans is whether the loan is of a stable or non-stable coin. We compute survival data using borrows as index events and either repays as outcome events, using the coin type as covariates (Figure 2). For repayments, we see users repaying non-stable coins significantly quicker than stable coins. From Table 5, we get that the RMST for repaying non-stable coins is 58.73 days, which is just about half as long as the RMST for repaying stable coins (117.97 days). In contrast, non-stable coins are liquidated much less frequently than stable coins. The mean time it takes for stablecoin loans to be liquidated is 496.17 days, which is almost 100 days sooner than the mean time of 585.79 days for non-stablecoin loans. From the CSHR and SDHRs, we see that borrowing a stable coin significantly reduces the incidence of repayments while simultaneously increasing the incidence of liquidations.

3.2.2 How does Loan Size Affect the Length of Loans? Next, we examine the effect of loan size on the time it takes for users to pay off the loan. In Aave, loans do not have defined lengths, the collateral and loan principal are typically in different coins, and the interest rates can vary drastically throughout the course of a loan if the borrower chooses to use variable-rate interest. Loans can be paid back incrementally through repay transactions or through liquidations. This analysis considers how long it takes for a loan to be repaid entirely, regardless of whether the loan was partially liquidated. As stated earlier, the reasons why someone would take out an over-collateralized loan in DeFi are not completely clear. In traditional finance, someone might take out a loan to buy a house, slowly paying back the loan over time, but it seems unlikely that many people are using these loans for that purpose. Even if they were, we would not have lengthy enough data to know that the loan had been paid back. Also, in traditional finance, we might expect to see shorter-term loans by businesses for the sake of corporate restructuring. These loans could be much larger than what one person might take out for a mortgage, and could be repaid much more quickly. Of course, there are plenty of other reasons a person or business might take out a loan as well. The hope is that we can use survival analysis to present a clear pattern in loan repayment times in Aave in order to glean some insight into how these loans might be being used.

Fortunately, a clear pattern is exactly what we see. In Aave's history, the larger a loan is, the faster the loan is likely to be fully repaid. We see this result in Figure 1, and the fact that the differences are significant is clearly shown by the large separation between the curves. The p-value for the log-rank test confirms this significance. In Table 6, we give the RMST and HRs for the different quartiles of loan amounts. The smallest quarter of loans have a mean survival time of 225.08 days, meaning that users are averaging over 225 days to fully repay the

Table 6. Borrow-to-Full-Repayment: Quantifiers of the Survival Data Corresponding to Figure 1, Using Borrows as Index Events, Full Loan Repayments as Outcomes, and Loan Size as a Covariate

USD Amount Quartile	Percentage	RMST (days)	Hazard Ratio	*P*-value
1st Quartile (1–25%)	25.00	225.08	Reference	Reference
2nd Quartile (26–50%)	25.00	96.93	1.8801	<0.001 (***)
3rd Quartile (51–75%)	25.00	71.30	2.2991	<0.001 (***)
4th Quartile (75–100%)	25.00	45.81	3.0552	<0.001 (***)

Table 7. Borrow-to-Liquidation: Quantifiers of the Survival Data Corresponding to Figure 4, Using Borrows as Index Events, Account Liquidations as Outcomes, and the Combination of Principal and Collateral Coin Types as Covariates

Principal:Collateral Combination	Percentage	RMST (days)	Hazard Ratio	*P*-Value
Stable:Stable	1.18	113.63	Reference	Reference
Non-Stable:Non-Stable	6.84	128.28	0.88801	<0.001 (***)
Non-Stable:Mixed	0.05	117.73	0.99592	0.97
Non-Stable:Stable	0.62	64.56	1.66800	<0.001 (***)
Stable:Non-Stable	90.79	102.61	1.09799	<0.001 (***)
Stable:Mixed	0.51	90.54	1.22647	<0.001 (***)

smallest loans in Aave. In contrast, the largest loans are being repaid in just 45.81 days on average. The HRs confirm that the larger a loan is, the higher the risk is for the loan to be fully repaid through time.

3.2.3 How do Coin Types Influence Risk of Liquidation? We hypothesize that the combination of stable and non-stable coins of the principal reserve and the collateral may lead to further insight into the risk of borrows. As shown in Figure 4, we stratify the borrow-to-liquidation data by factoring in what collateral was purchased and what principal types were specifically paid off by a liquidator. 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 3. 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. We aggregated user's liquidation events to gain more information as to which coins the users have used 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, e.g., one user on October 30, 2021 was the subject of 82 liquidations in just over an hour. 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 (a 24 hour period) 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." As such, the competing risks approach is not suited to this analysis, since the covariate of focus is not applicable to the outcome of repayment.

Using the stable:stable combination of principal and collateral coins as the benchmark, we get the quantification of risk is seen in Table 7.

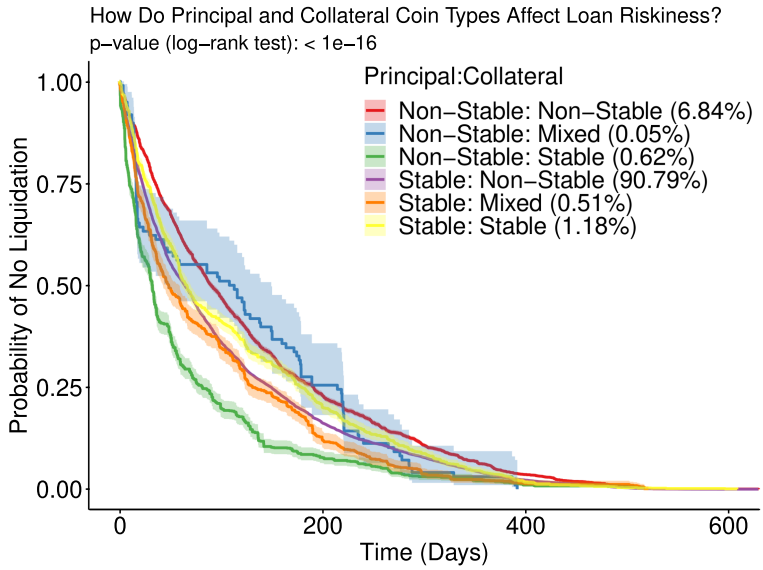


Fig. 4. Kaplan–Meier survival curves using borrows as index events and liquidations as outcomes using different combinations of principal and collateral coin types as covariates.

The results show that the combination of non-stablecoin principal with non-stablecoin collateral tends to be safer than that of stablecoin principal and collateral. This is a little surprising, since the relative value of an account's principal and collateral are what causes its ability to be liquidated. When functioning properly, stablecoins should always have the same relative value and the only reason an account should liquidate with stablecoins as both the principal and the collateral would be when enough interest has accrued and gone unpaid on the account. Non-stablecoins' values are much more volatile, and if the relative value of the principal coin is not in sync with the value of the collateral coin, the account could be liquidated much more quickly. However, this is not what we see. This likely indicates that users who borrow non-stablecoins using non-stablecoins as collateral are using collateral assets that are less likely to drop in value than their principal asset(s).

We also see that loans are liquidated more quickly when they have non-stablecoin principal and stable collateral. This behavior makes sense logically during a bull market when many coins will be gaining value relative to the USD which most stablecoins are pegged to. If the value of the principal assets rise relative to the value of the collateral, the loan can become unhealthy very quickly. If an account consists of stablecoin principal and non-stablecoin collateral, we see a slight increase in risk relative to stable:stable accounts. This combination also accounts for the vast majority (90.79%) of liquidated accounts.

3.3 Insights Into Differences Between Markets

The analysis so far has focused on Aave's Ethereum market. We now examine how user behavior compares in Aave's other markets of AaveV3. The transaction volumes for the other markets are given in Appendix A. The six V3 markets were launched on March 12, 2022 and the data used here runs through October 01, 2022. Because we only have 203 days of data for the V3 markets, we restrict the time window when computing the RMSTs to the first 203 days of observation.

In Table 8, we compute survival data using borrows as index events, liquidations as outcome events, and the Aave market as covariates. The RMSTs and HRs paint an interesting picture here. Using the Ethereum market as a benchmark, we see all but one of the V3 markets showing less risk of liquidation following a borrow. The Fantom

Table 8. Borrow-to-Liquidation: Quantifiers of Survival Data using Borrows as Index Events, Liquidations as Outcome Events, and the Aave Market Where the Transactions Took Place as Covariates

Market	Percentage	RMST (days)	Hazard Ratio	*P*-Value
Ethereum	30.86	168.74	Reference	Reference
Arbitrum	7.42	193.63	0.25331	<0.001 (***)
Avalanche	11.99	185.93	0.46450	<0.001 (***)
Fantom	1.87	145.72	1.75018	<0.001 (***)
Harmony	0.66	179.06	0.62611	<0.001 (***)
Optimism	14.33	191.90	0.28715	<0.001 (***)
Polygon	32.86	184.03	0.53000	<0.001 (***)

The RMSTs were computed with a cutoff of 203 days to account for the shorter span of data from the non-Ethereum markets.

market, which accounts for just 1.87% of the borrow events in these transactions, has a RMST of 145.72 days. This indicates that in the first 203 days of these markets, the mean time it takes for a loan to be liquidated following a borrow is 23.02 days less than the mean time for borrows to be liquidated in the Ethereum market. In contrast, the other V3 markets show significantly longer times to liquidation. For instance, the Arbitrum market shows a mean survival time of 193.63 days, which is almost a 25 day increase over the Ethereum market's survival time.

In Table 9, we see a different side of the markets. Using deposits as index events, borrows as outcome events, and again using the Aave market as covariates, we can see how quickly users will borrow funds after making a deposit in Aave. In the Ethereum market, the mean time for users to borrow after a deposit is 85.76 days, and again, we see the biggest contrast in the Fantom market, where the mean time is only 44.55 days. The Polygon market, which is the market with the second-biggest share of deposits in this data at 23.76% of the deposits, also shows users being much more likely to borrow funds following a deposit. The mean survival time for the Polygon market is just 58.63 days.

One factor that likely causes differences in user behavior across markets is the size of transaction fees in the market. The transaction fees are just flat fees, not scaled by transaction size, so in markets with higher transaction fees the use of smaller transactions is more heavily penalized. Additionally, making lots of transactions is more heavily penalized. While our data source does not tell us how much users paid in transaction fees, we do know that the Ethereum market has non-negligible fees, usually at least $10. The other most popular markets of Polygon and Optimism have very low transaction fees, usually costing less than $0.01. As a result, we suspect that these markets attract retail users who are looking to experiment more with the protocol and engage in behaviors such as yield farming to make shorter term profits, whereas the Ethereum market likely attracts much larger, institutional users. This would likely account for the large discrepancy between the value share of the Ethereum market, which stands at about 75%, and its share of the total number of transactions within Aave, which is only around 5%.

3.4 Insights Into Evolving Behavior

Since DeFi is still in its infancy, we would expect to see changes in macro-level user behaviors as people get excited about the new technology and start using it, but perhaps do not have a well-established understanding for how the technology should be used. In this section, we look at a couple of results of how user behaviors have changed from quarter to quarter. We again compute how long it takes for loans to see their first repayment after a user borrows money (Figure 5(a)), and we also look at behavior relating to how long users keep all their funds in their accounts (Figure 5(b)).

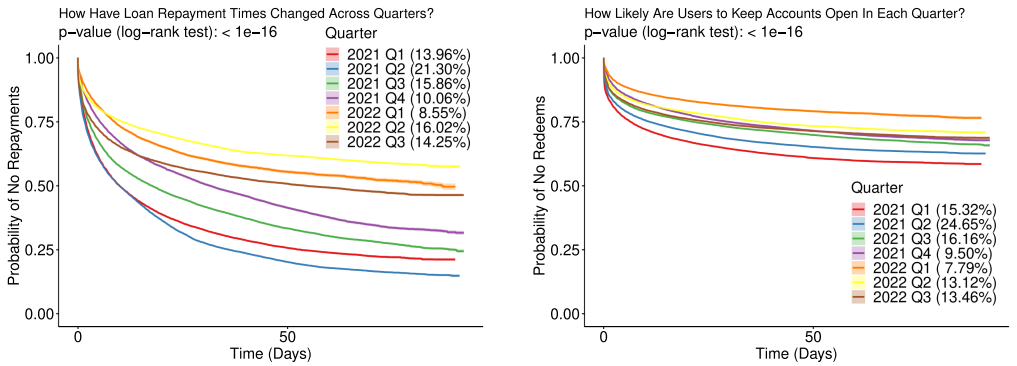
Table 9. Deposit-to-Borrow: Quantifiers of Survival Data Using Deposits as Index Events, Borrows as Outcome Events, and the Aave Market Where the Transactions Took Place as Covariates

Market	Percentage	RMST (days)	Hazard Ratio	*P*-Value
Ethereum	29.97	85.76	Reference	Reference
Arbitrum	9.86	89.05	0.98123	<0.001 (***)
Avalanche	14.52	52.72	1.55080	<0.001 (***)
Fantom	2.12	44.55	1.78117	<0.001 (***)
Harmony	1.52	94.88	0.87052	<0.001 (***)
Optimism	18.26	75.31	1.26596	<0.001 (***)
Polygon	23.76	58.63	1.48999	<0.001 (***)

The RMSTs were computed with a cutoff of 203 days to account for the shorter span of data from the non-Ethereum markets.

In Figure 5(a), we look at Kaplan–Meier curves for borrow-to-first-repay events with quarters as covariates. Note that first-repay factors in the first outcome of either repayment or liquidation, so competing risks is not suited to this outcome event. Each distinct curve represents a different quarter of transactions. We note that, while the shapes of the curves are consistent from quarter to quarter, there are drastic differences between how quickly users are repaying loans depending on the quarter. For instance, we can see that in 2022 Q2, in the entire quarter only about 50% of the loans had their first repayment. This is in contrast to 2021 Q2 where over 75% of the loans saw their first repayment by the end of the quarter. The HRs for these curves are in Table 10, along with the average amount per borrow in each quarter. What we can observe from these coefficients is that the tendency through time has been for users to take longer to have their first repayment on their loans. 2021 Q2 showed a slightly increased risk of loans having their first repayment during the quarter, but the subsequent four quarters each show reduced likelihood of loans having their first repayment. This trend was broken in 2022 Q3, but the fact remains that users have been tending to slow down their repayment schedules as Aave has matured. Though not perfect, the average amount per borrow has followed a similar trend to first-repayment times, increasing steadily between 2021 Q1 and 2022 Q1, then decreasing drastically in the subsequent two quarters. This makes some logical sense with the increasing first-repayment times, as it stands to reason that larger loans on average would take more time to be fully repaid. However, this does contrast with the findings from 3.2.2 where we showed that larger loans have tended to be repaid more quickly, so a more in-depth analysis is required to fully understand the effects of loan size on their duration in Aave.

In Figure 5(b), we change our focus to how long users are leaving money in their Aave accounts. As a result, the only possible events are redeem or continuing to leave the money (censored), so competing risks analysis is not applicable. We hypothesize that users who leave money in their accounts longer are more invested in both the protocol of Aave and in the emerging DeFi ecosystem, and thus, it is interesting to see how the duration that users keep money in their accounts has changed through time. Similar to the behavioral changes seen in Figure 5(a), the time it takes for users to withdraw funds after depositing them changes drastically from quarter to quarter. Generally speaking, users are leaving funds in their accounts much longer than the terms of loans, but the relative difference from quarter to quarter for how quickly users are withdrawing funds is significant. We see a similar trend in the deposit-to-withdraw analysis as in the borrow-to-repay analysis, where the first quarter in the data shows users being the most likely to withdraw funds from their accounts, and the general trend has been that as time has passed, the risk of users withdrawing funds has decreased. Again, these observations are confirmed by the hazard ratios in Table 11 along with the average amount per deposit in each quarter. Similarly to the average borrow values, the average deposit values per quarter increase steadily from 2021 Q1 through 2022 Q1, and then decline drastically through 2022 Q3.



(a) Kaplan–Meier plot for borrow-to-repayment (b) Kaplan–Meier plot for deposit-to-withdraw

Fig. 5. Kaplan–Meier Survival Curves showing the differences in repayment schedules and withdrawal schedules throughout individual quarters since Aave V2 launched. 2020 Q4 has been removed due to the small portion of the quarter for which Aave existed (November 30-December 31, 2020).

Table 10. Borrow-to-Repayment: Quantifiers of the Survival Data Corresponding to Figure 5(a), Using Borrows as Index Events, First Repayments as Outcome Events, and the Quarter in which the Events Took Place as Covariates

Quarter	Percentage	RMST (days)	Hazard Ratio	*P*-Value	Mean Loan Amount
2021 Q1	13.96	30.91	Reference	Reference	$105,312.75
2021 Q2	21.30	26.64	1.10977	<0.001 (***)	$418,012.49
2021 Q3	15.86	36.93	0.82244	<0.001 (***)	$405,710.39
2021 Q4	10.06	44.36	0.64165	<0.001 (***)	$521,328.58
2022 Q1	8.55	52.06	0.50875	<0.001 (***)	$582,755.78
2022 Q2	16.02	55.77	0.45300	<0.001 (***)	$424,048.96
2022 Q3	14.25	48.61	0.59914	<0.001 (***)	$236,371.63

This data was computed using each quarter as a separate observation period. The mean value (in USD) of borrows during each quarter has also been included.

Table 11. Deposit-to-Withdraw: Quantifiers of the Survival Data Corresponding to Figure 5(b), Using Deposits as Index Events, First Withdraws as Outcome Events, and the Quarter in which the Events Took Place as Covariates

Quarter	Percentage	RMST (days)	Hazard Ratio	*P*-value	Mean Deposit Amount
2021 Q1	15.32	58.70	Reference	Reference	$162,796.81
2021 Q2	24.65	62.36	0.85933	<0.001 (***)	$629,919.20
2021 Q3	16.16	66.07	0.72749	<0.001 (***)	$724,694.00
2021 Q4	9.50	67.94	0.65520	<0.001 (***)	$1,897,922.48
2022 Q1	7.79	74.23	0.46595	<0.001 (***)	$1,038,654.83
2022 Q2	13.12	69.43	0.61763	<0.001 (***)	$667,335.69
2022 Q3	13.46	67.43	0.68514	<0.001 (***)	$358,047.14

This data was computed using each quarter as a separate observation period. The mean value (in USD) of deposits during each quarter has also been included.

3.5 Survival Analysis App and GitHub Repository

Since there are so many choices for index events, outcome events, and covariates, we thought it appropriate to help create these analyses and allow other people to explore our survival data. The app can be accessed at this link: https://inciteprojects.idea.rpi.edu/defitoolkit/app/defitoolkit/. This app allows the user to create Kaplan–Meier survival curves and tables of related data similar to what has been presented in this article. In addition to the app, the code used to create the analysis presented in this article can be found on GitHub[3] along with some instructions to help acquire the necessary data to run the code (the datasets are far too large to be stored in a GitHub repository). We hope that anyone interested in this work can use the app and our code to further explore and expand on the insights we have shown in this article.

4 RELATED WORK

Survival analysis as a tool in economic and financial analysis is growing in popularity, but is still largely overshadowed by discriminant and logistic analyses. That said, there have been some significant articles applying survival analysis to financial questions. Some classic examples of questions to which survival analysis has been applied include questions related to bank failures [20], credit default risks [29] [14], and business failure prediction [13], to name a few. These articles demonstrate that survival analysis can be used effectively in financial theory and applications as a model for understanding and predicting events related to the financial wellbeing of businesses both small and large.

Some work has tried to incorporate loan prepayments as competing risks with loan defaults, either in the context of mortgages [8], or more recently in the context of online loans [21]. The argument is that the end goal of providing someone with a loan is to accrue additional money through interest payments. Both loan prepayments and defaulting on the loan act as outcomes that cause less interest to be paid over the life of the loan. This is similar to our analysis in Section 3.2, but is not directly applicable to loan analysis in the DeFi space because liquidations and defaults operate differently, and because a loan in DeFi is constantly accruing interest, so the notion of making a "prepayment" before an interest payment is due does not make sense. However, there are still similarities in the analysis and the application of competing risks is necessary in both cases.

A key feature of these survival-analysis studies is the kind of data to which they have access. For instance, in Ref. [13], a study is made that includes data from an 18-year span detailing 27 financial variables of 189 businesses during that span. These data are well-labeled and understood, and the data spans a long duration. This is in stark contrast to the data available in DeFi, where the transactions are all public, but the big picture of what might be happening with these transactions is unknown. Additionally, the time span that DeFi data can bridge right now is limited. There might be a lot of data, but it is all from a relatively short period of time. For this reason, we believe it is more appropriate to use survival analysis as an expository tool for user-level behavior in DeFi rather than as a predictive tool for how various entities in the DeFi space might perform.

With an over-collateralized loan, a borrower must post collateral which exceeds the value of the debt. This way, collateralization simultaneously ensures that the lender (likely a smart contract) can recover their loaned value. The "health factor" (HF) is a custom threshold in DeFi 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. [25] 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 [22], Aave, Compound [27], and dYdX [1]) and

[3] https://github.com/aaronmicahgreen/DeFi-Survival-Analysis-ACM-DLT

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 Ethereum 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. [25] 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 take a loan with a stablecoin as collateral with the intent of holding the loan indefinitely. If the stablecoin collateral is accruing higher interest than the borrowed principal coin, this can lead to a form of passive income. Early empirical evidence on the stability of crypto-backed loans with stablecoins has been studied by Kozhan and Viswanath-Natraj [19]. 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., 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. [24], 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. [2] 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.

Most of the related works approach the issue of liquidation at a conceptual level or rely on aggregate flow data. In contrast, our article uses transaction-level blockchain data to provide a more "microscopic" view of liquidations, 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 transaction data in the scenarios above, we have shown these survival analysis methods to be versatile tools for answering all kinds of questions in the DeFi sphere.

The possibilities for what questions to ask and answer with survival analysis are myriad. What we have presented in the results here were limited to just a handful of selections for index and outcome events, and using a few different variables to stratify the results. We showed that, counter to basic intuition, DeFi loans tend to be repaid more quickly when the loan is larger. We have shown how the combinations of stable and non-stable coins as both principal and collateral assets affect the riskiness of an account ending up being liquidated. We have also applied considered loan repayments and liquidations as competing risks, applying competing risks

[4] https://coinmarketcap.com/historical/20200313
[5] https://compound.finance

analysis to more accurately understand the interactions between coin types and the different outcomes for loans in DeFi. Expanding the scope of the data briefly, we showed that depending on the Aave market, behaviors can change drastically. We hypothesize that these behavioral changes have mostly to do with the magnitude of the transaction fees in the market, and show that there are large differences in risk of liquidation over time between markets with high transaction fees and markets with low transaction fees. Lastly, we show that the data can be segmented into separate observation periods for each quarter, and that we can use these quarterly observation periods to find differences in user behavior through time.

This work represents just a few steps in the usage of survival analysis techniques to help form a complete picture of behavioral trends in the DeFi ecosystem. 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 well. In some analyses, consideration of repeated events could provide additional insights into user behavior. For instance, in the process of repaying a loan, a user can make any number of repay transactions. They could repay it all at once, or they could repay it in small installments over a long period of time. There are some existing methods for handling repeated events, like multi-state models [23]. Furthermore, survival analysis models can be used to for making predictions, like in Ref. [21], where the competing risks models are used to predict how much profit will be made from specific loans. In addition to this article, we have also been preparing a toolkit for analysis of DeFi protocols through many different lenses. Survival analysis techniques are just one part of the toolkit, and eventually the results of these survival analysis applications will be integrated into other results to help form more robust analyses of DeFi protocols.

This research only partially addresses 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 [3, 32] 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 [31]. 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.

Moving forward, we plan to continue using the survival analysis methods to help with ongoing work identifying and quantifying characteristics of the emerging DeFi ecosystem. We are working towards solidifying some methods for clustering the users of DeFi protocols and explaining what kinds of users each cluster represents. When the clusters are completed we plan to add user cluster as an additional covariate in our survival analysis, and we hope to see big differences in behavior across clusters. To improve the results relating to differences in user behavior over time, we plan to quantify the "bullishness" of the market through time and try to identify different behavioral trends based on how bullish the market is at the time. We also hope to incorporate flash loan data into our analyses as soon as we can correlate users' transactions between protocols, in order to get the true picture of how flash loans are being used.

All of this analysis will eventually be incorporated into our open-source DeFi Toolkit.[6] Currently, the toolkit allows for the dynamic creation of survival analysis plots like what have been presented here. As we continue to refine certain techniques, they will be added to the toolkit to make for a more robust application.

APPENDICES

A TRANSACTION SUMMARIES BY MARKET

In Section 2.1, we presented a table summarizing the transaction data that was used in this article from the Aave Ethereum market. Since the data from other markets was also used for parts in this article (e.g., Section 3.3), we

[6] https://inciteprojects.idea.rpi.edu/defitoolkit/app/defitoolkit

feel some might be interested in seeing the data summaries from other Aave markets. These are included below, with the market name, Aave version, and date-range of data included in the caption for each table.

Table 12. Summary of Transaction Types from Aave's **Avalanche V2** Market Collected from March 31, 2021 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	318,227	$582,645.30	$1,200.00	NA	NA
Deposit	1,027,789	$383,204.90	$3,233.96	NA	NA
Withdraw	783,107	$501,630.90	$5,660.20	NA	NA
Repay	199,300	$929,626.30	$5,650.52	NA	NA
Liquidation	10,900	NA	NA	$6,452.82	$6,873.20

Table 13. Summary of Transaction Types from Aave's **Polygon V2** Market Collected from March 31, 2021 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	1,704,749	$29,383.27	$499.93	NA	NA
Deposit	13,338,917	$15,494.35	$7.22	NA	NA
Withdraw	6,030,779	$14,106.76	$407.73	NA	NA
Repay	1,297,373	$31,252.15	$777.32	NA	NA
Liquidation	64,762	NA	NA	$7,296.67	$7,828.70

Table 14. Summary of Transaction Types from Aave's **Arbitrum V3** Market Collected from March 12, 2022 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	45,866	$2,572.83	$5.99	NA	NA
Deposit	119,543	$2,424.03	$11.07	NA	NA
Withdraw	60,371	$3,838.99	$66.45	NA	NA
Repay	26,120	$3,966.26	$69.90	NA	NA
Liquidation	1,046	NA	NA	$1,502.62	$1,571.00

Table 15. Summary of Transaction Types from Aave's **Avalanche V3** Market Collected from March 12, 2022 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	74,243	$208,964.30	$711.14	NA	NA
Deposit	175,990	$129,160.30	$228.43	NA	NA
Withdraw	85,420	$198,728.40	$1,759.38	NA	NA
Repay	50,476	$294,342.80	$1,792.30	NA	NA
Liquidation	1,999	NA	NA	$7,197.06	$7,618.99

Table 16. Summary of Transaction Types from Aave's **Fantom V3** Market Collected from March 12, 2022 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	11,591	$3,269.36	$200.02	NA	NA
Deposit	25,654	$4,598.34	$102.23	NA	NA
Withdraw	17,794	$6,120.27	$213.73	NA	NA
Repay	9,580	$3,689.34	$299.99	NA	NA
Liquidation	475	NA	NA	$1,553.61	$1,655.37

Table 17. Summary of Transaction Types from Aave's **Harmony V3** Market Collected from March 12, 2022 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	4,064	$5,496.71	$199.98	NA	NA
Deposit	18,379	$3,302.44	$20.87	NA	NA
Withdraw	7,791	$5,708.04	$104.40	NA	NA
Repay	3,602	$4,994.46	$86.14	NA	NA
Liquidation	269	NA	NA	$197.23	$210.21

Table 18. Summary of Transaction Types from Aave's **Optimism V3** Market Collected from March 12, 2022 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	88,759	$44,458.98	$1.00	NA	NA
Deposit	221,326	$23,822.92	$8.35	NA	NA
Withdraw	97,036	$24,175.12	$23.14	NA	NA
Repay	40,934	$81,033.41	$19.43	NA	NA
Liquidation	824	NA	NA	$1,834.86	$1,961.56

Table 19. Summary of Transaction Types from Aave's **Polygon V3** Market Collected from March 12, 2022 to October 01, 2022

Transaction	Occurrences	Mean Value	Median Value	Mean Principal	Mean Collateral
Borrow	203,353	$27,890.93	$1,174.48	NA	NA
Deposit	287,995	$3,305.92	$50.00	NA	NA
Withdraw	157,448	$5,062.66	$141.45	NA	NA
Repay	215,159	$26,196.11	$780.54	NA	NA
Liquidation	3,815	NA	NA	$1,302.76	$1,371.30

B COIN TYPES IN AAVE'S ETHEREUM MARKET

Coin Symbol	Coin Type
DAI	Stable
LINK	Non-Stable
Aave	Non-Stable
WBTC	Non-Stable
SNX	Non-Stable
USDC	Stable
TUSD	Stable
USDT	Stable
SUSD	Stable
BUSD	Stable
WETH	Non-Stable
YFI	Non-Stable
UNI	Non-Stable
BAT	Non-Stable
REN	Non-Stable
ENJ	Non-Stable
KNC	Non-Stable
MANA	Non-Stable
MKR	Non-Stable
ZRX	Non-Stable
CRV	Non-Stable
GUSD	Stable
BAL	Non-Stable
XSUSHI	Non-Stable
RENFIL	Non-Stable
RAI	Stable (Not pegged to USD)
AMPL	Non-Stable
PAX	Non-Stable
DPI	Non-Stable
FRAX	Stable
FEI	Stable
ENS	Non-Stable
UST	Non-Stable
CVX	Non-Stable
1INCH	Non-Stable
LUSD	Stable
STETH	Non-Stable

ACKNOWLEDGMENTS

The authors acknowledge the support from NSF IUCRC CRAFT center research grant (CRAFT Grant #22003) for this research. The opinions expressed in this publication do not necessarily represent the views of NSF IUCRC

CRAFT. This work was supported by the Rensselaer Institute for Data Exploration and Applications (IDEA)7. We would like to thank Leyi Lin and Shuyu Meng for helping create and interpret results related to this work.

REFERENCES

[1] Antonio Juliano. 2017. *dYdX: A Standard for Decentralized Margin Trading and Derivatives*. Technical Report. Retrieved from https://whitepaper.dydx.exchange/

[2] Massimo Bartoletti, James Hsin-yu Chiang, and Alberto Lluch Lafuente. 2021. SoK: Lending pools in decentralized finance. In *International Conference on Financial Cryptography and Data Security*. Springer, 553–578.

[3] Kristin P. Bennett. 2019. Artificial intelligence for public health. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 1–2. DOI: https://doi.org/10.1109/BIBM47256.2019.8983112

[4] Boado, Ernesto. 2020. *AAVE Protocol Whitepaper*. Technical Report. Retrieved from https://www.cryptocompare.com/media/38553941/aave_protocol_whitepaper_v1_0.pdf

[5] Boado, Ernesto. 2020. *AAVE Protocol Whitepaper V2.0*. Technical Report. Retrieved from https://cryptorating.eu/whitepapers/Aave/aave-v2-whitepaper.pdf

[6] Vitalik Buterin. 2014. *Ethereum: A Next-generation Smart Contract and Decentralized Application Platform*. Retrieved from https://github.com/ethereum/wiki/wiki/White-Paper

[7] Sang-Min Choi, Jiho Park, Quan Nguyen, and Andre Cronje. 2018. Fantom: A scalable framework for asynchronous distributed systems. DOI: https://doi.org/10.48550/ARXIV.1810.10360

[8] Brian Ciochetti, Yongheng Deng, Bin Gao, and Rui Yao. 2002. The termination of commercial mortgage contracts through prepayment and default: A proportional hazard approach with competing risks. *Real Estate Economics* 30, 4 (02 2002), 595–633. DOI: https://doi.org/10.1111/1540-6229.t01-1-00053

[9] Taane G. Clark, Michael J. Bradburn, Sharon B. Love, and Douglas G. Altman. 2003. Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer* 89, 2 (2003), 232–238.

[10] D. R. Cox. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34, 2 (1972), 187–220. Retrieved from http://www.jstor.org/stable/2985181

[11] Emilio Frangella, Lasse Herskind. 2022. *AAVE V3 Technical Paper*. Technical Report. Retrieved from https://github.com/aave/aave-v3-core/blob/master/techpaper/Aave_V3_Technical_Paper.pdf

[12] Jason P. Fine and Robert J. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* 94, 446 (1999), 496–509. DOI: https://doi.org/10.1080/01621459.1999.10474144

[13] Adrian Gepp and Kuldeep Kumar. 2008. The role of survival analysis in financial distress prediction. *International Research Journal of Finance and Economics* 16, 16 (2008), 13–34.

[14] Dennis Glennon and Peter Nigro. 2005. Measuring the default risk of small business loans: A survival analysis approach. *Journal of Money, Credit and Banking* 37, 5 (2005), 923–947. Retrieved from http://www.jstor.org/stable/3839153

[15] Anurag Arjun Jaynti Kanani, Sandeep Nailwal. 2020. Matic Network Whitepaper. Retrieved from https://github.com/maticnetwork/whitepaper

[16] Harry A. Kalodner, Steven Goldfeder, Xiaoqi Chen, S. Matthew Weinberg, and Edward W. Felten. 2018. Arbitrum: Scalable, private smart contracts. In *USENIX Security Symposium*.

[17] E. L. Kaplan and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53, 282 (1958), 457–481. Retrieved from http://www.jstor.org/stable/2281868

[18] Alboukadel Kassambara, Marcin Kosinski, and Przemyslaw Biecek. 2021. *survminer: Drawing Survival Curves using 'ggplot2'*. Retrieved from https://CRAN.R-project.org/package=survminer R package version 0.4.9.

[19] Roman Kozhan and Ganesh Viswanath-Natraj. 2021. Decentralized stablecoins and collateral risk. *WBS Finance Group Research Paper*, Available at SSRN: https://ssrn.com/abstract=3866975 or http://dx.doi.org/10.2139/ssrn.3866975

[20] William R. Lane, Stephen W. Looney, and James W. Wansley. 1986. An application of the cox proportional hazards model to bank failure. *Journal of Banking and Finance* 10, 4 (1986), 511–531. DOI: https://doi.org/10.1016/S0378-4266(86)80003-6

[21] Zhiyong Li, Aimin Li, Anthony Bellotti, and Xiao Yao. 2023. The profitability of online loans: A competing risks analysis on default and prepayment. *European Journal of Operational Research* 306, 2 (2023), 968–985. DOI: https://doi.org/10.1016/j.ejor.2022.08.013

[22] MakerDAO. 2017. *The Maker Protocol: MakerDAO's Multi-Collateral Dai (MCD) System*. Technical Report. Retrieved from https://makerdao.com/en/whitepaper/#abstract

[23] Luís Meira-Machado, Jacobo de Uña-Álvarez, Carmen Cadarso-Suárez, and Per K. Andersen. 2009. Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research* 18, 2 (2009), 195–222. DOI: https://doi.org/10.1177/0962280208092301.

[24] Daniel Perez, Sam M. Werner, Jiahua Xu, and Benjamin Livshits. 2021. Liquidations: DeFi on a knife-edge. In *International Conference on Financial Cryptography and Data Security*. Springer, 457–476.

[25] Kaihua Qin, Liyi Zhou, Pablo Gamito, Philipp Jovanovic, and Arthur Gervais. 2021. An empirical study of defi liquidations: Incentives, risks, and instabilities. In *Proceedings of the 21st ACM Internet Measurement Conference*. 336–350.

[26] R. Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/
[27] Robert Leshner, Geoffrey Hayes. 2019. *Compound: The Money Market Protocol*. Technical Report. Retrieved from https://compound.finance/documents/Compound.Whitepaper.pdf
[28] Patrick Royston and Mahesh K. B. Parmar. 2013. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Medical Research Methodology* 13, 1 (Dec. 2013). DOI: https://doi.org/10.1186/1471-2288-13-152
[29] Michal Rychnovský. 2018. Survival analysis as a tool for better probability of default prediction. *Acta Oeconomica Pragensia* 2018, 1 (2018), 34–46. DOI: https://doi.org/10.18267/j.aop.594
[30] Kevin Sekniqi, Daniel Laine, Stephen Buttolph, and Emin Gün Sirer. 2020. Avalanche Platform. Retrieved from https://assets.website-files.com/5d80307810123f5ffbb34d6e/6008d7bbf8b10d1eb01e7e16_Avalanche%20Platform%20Whitepaper.pdf
[31] Xiao Shou, Tian Gao, Dharmashankar Subramanian, and Kristin P. Bennett. 2023. Multi-Label Event Prediction in Continuous Time. To appear.
[32] Xiao Shou, Georgios Mavroudeas, Alexander New, Kofi Arhin, Jason N. Kuruzovich, Malik Magdon-Ismail, and Kristin P. Bennett. 2019. Supervised mixture models for population health. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 1057–1064. DOI: https://doi.org/10.1109/BIBM47256.2019.8983339
[33] Daniel D. Sjoberg and Teng Fei. 2022. *tidycmprsk: Competing Risks Estimation*. Retrieved from https://CRAN.R-project.org/package=tidycmprsk R package version 0.2.0.
[34] Harmony Team. 2020. Harmony Technical Whitepaper. Retrieved from https://harmony.one/whitepaper.pdf
[35] Terry M. Therneau and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. Springer, New York.
[36] Terry M. Therneau. 2022. *A Package for Survival Analysis in R*. Retrieved from https://CRAN.R-project.org/package=survival R package version 3.4-0.
[37] Mark Tyneway. 2020. Optimism. Retrieved from https://github.com/ethereum-optimism/optimism
[38] Abraham Wald. 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 54, 3 (1943), 426–482. Retrieved from http://www.jstor.org/stable/1990256

Received 17 January 2023; revised 5 November 2023; accepted 7 December 2023