# On-Chain Credit Risk Scores: A Catalyst for DeFi Lending Product Development

Daniel Stephens
*Bird.Money*
daniel@bird.money

Alex Kirby, PhD
*Bird.Money*
alex.kirby@bird.money

Luciano De Monte
*Bird.Money*
luciano@bird.money

May 17, 2021

## Abstract

Decentralized cryptocurrency lending protocols have experienced exponential growth in the past 12 months, despite catering primarily to a narrow clientele of traders that seek hedging and leveraging solutions. At the same time, significant frictions and deficiencies persist within traditional lending markets, both in developed and developing nations. This paper establishes a framework for expanding the potential customer base of cryptocurrency lending by relaxing collateral requirements primarily through the development of an individualized risk rating assigned at a public key level. It also contrasts the proposed framework with the core deficiency of the predominant approach adopted by existing firms. Recommendations are made for enhancements to both product and protocol design to ensure the long-term viability of a decentralized lending platform in the absence of collateral protections.

# Introduction: The State of DeFi

In the span of less than one year, the overcollateralized debt market in the decentralized finance (DeFI) sector has grown from virtual non-existence to over $30B USD[1]. The overcollateralized debt position (hereinafter called simply CDP) instrument is used largely by investors to achieve esoteric financial objectives, for example hedging and leveraging. However, CDPs have become colloquially termed "loans". When applied in this context, it is a terminology that is inconsistent with the lay understanding of a loan because a borrower must be in possession of greater capital than they wish to borrow. This misnomer serves to mask not only linguistic meaning, but a tremendous market opportunity. Given the likelihood that DeFi will serve as a "tipping point" innovation that enables greater global mainstream adoption of cryptocurrency, the market potential associated with a viable approach to undercollateralized or uncollateralized debt vehicles in DeFi is inestimably large.

# Business Problem: Individualized Risk

The fundamental mechanics of lending as a for-profit enterprise have changed little in millenia: one entity (i.e., the lender) is incentivized to transfer value (i.e, loan principal) to a second entity (i.e., the borrower) in exchange for a promise to repay the original loan amount plus a fee that accounts for both default risk and operating costs (i.e., interest). The DeFi CDP represents a breakthrough in operating cost reduction as the marginal cost of loan issuance via an autonomous protocol is near-zero. However, in its current state, the DeFi CDP also represents significant atavism in its pricing of risk.

Currently, the DeFi CDP effectively treats all borrowers as equal risks. The regressive effect of this feature is quite alarming when one considers the fact that borrower-agnostic interest rates first emerged in 1754 BC Mesopotamia[2] and, although they were derived subjectively, borrower-specific interest rates first emerged in 4th century Athens[3]. In essence, the lack of risk discrimination used in the formulation of a DeFi CDP has led to all default risk being naively captured within collateral requirements and not by interest rates. The core problem is therefore, how does one begin to shift the capture of risk premiums away from collateral requirements and towards interest rates?

# Constraints: Decentralized Challenges

Distributed ledger technology (DLT) and cryptocurrency asset markets possess features that simultaneously enable and also complicate the methods by which default risk can be estimated and subsequently incorporated into protocol-derived loan terms. The factors that uniquely facilitate the innovations that this paper calls for shall be evident in a later discussion of a potential framework. Therefore, let us first explore the challenges posed by certain features of cryptocurrency markets

---

[1] See https://defipulse.com/ for Total Value Locked of major lending protocols

[2] Garfinkle, Steven. "Shepherds, merchants, and credit: some observations on lending practices in Ur III Mesopotamia." *Journal of the Economic and Social History of the Orient* 47.1 (2004): 1-30.

[3] Millett, P. (2002). Lending and borrowing in ancient Athens. Cambridge, England: Cambridge University Press.

and DLT as they exist today.

To be clear, the extent to which the following problems are resolved suggests the proportion of estimable risk that can be re-captured within an interest rate, that can lead to lowered collateral requirements, or a combination of the two. Because analytical capabilities evolve gradually, this discussion should not be misunderstood as an implication of general viability in binary terms. As solutions to these challenges evolve, so too will the DeFi lending protocol's ability to develop new products that capture individualized default risk in more sophisticated ways.

## Anonymity

Central to the estimation of individual risk, whether financial or otherwise, is the knowledge of individual characteristics. However, the trustless nature of DLT has facilitated unprecedented levels of user privacy. This is an oft-cited criticism of the cryptocurrency industry, though there do in fact exist legitimate, non-illicit reasons for this desire.

It should come as no surprise that financial discrimination is rampant in developing countries, with women, minorities and other less privileged classes being denied access to capital based on factors that bear no relationship to default risk[4]. Given the proportion of the global population that may ascribe to these classes, in conjunction with the consequences of inadequate access to capital, the scale of this problem is exceedingly large. Research has even shown that in developed economies, preferential loan terms are offered to entities that may share a mutual interest with the lender that is unrelated to default risk, which often conflicts with greater economic welfare[5]. This is fairly described as corruption and can be mitigated, in part, by allowing credit seekers to remain anonymous.

For these reasons, the most fundamental challenge associated with undercollateralized DeFi lending is the tension between the collation of individual risk factors and the legitimate desire for anonymity.

## Public Key as Identity Proxy

Traditional "Know Your Customer" verification routines may eventually provide an explicit solution to the anonymity challenge mentioned above. However, because this is both not yet available and also justifiably undesirable for certain borrowers (i.e., to prevent discrimination), a borrower's unique public key must be used as a proxy for individual identity.

A public key, or wallet address, being nothing more than the output from a particular cryptographic function, is necessarily costless to create. Therefore, its use as a proxy for identity creates a likelihood for two scenarios to emerge: first, instantiation of a new public key in response to the occurrence of a

---

[4] Demirgüç-Kunt, Asli and Klapper, Leora F. and Singer, Dorothe, Financial Inclusion and Legal Discrimination Against Women: Evidence from Developing Countries (April 1, 2013). World Bank Policy Research Working Paper No. 6416, Available at SSRN: https://ssrn.com/abstract=2254240

[5] Ying Ge, Jiaping Qiu "Financial development, bank discrimination and trade credit" Journal of Banking & Finance Volume 31, Issue 2 (2007)

negative, risk-related event and second, a one-to-many relationship between an individual and their public keys. The implications of the former are perhaps obvious, however, in the case of the latter, certain analytic ramifications exist that are far less clear and even knowable.

As will be detailed in later sections, the network structure of DLT transactions enables the analysis of all interactions a single wallet has with all other wallets in the network[6]. However, if this analysis is performed as a means of inferring the default risk of a given address (i.e., a borrower), misspecification of which secondary addresses are also owned by the borrower will in turn lead to misspecification of that borrower's default risk.

Generally speaking, the tractability of this problem hinges on the "costs" that can be imposed on public key instantiation in terms of credit offerings. Despite the fact that in traditional finance there exists a large cost of identity creation (i.e., it is an unlawful act), length of credit history has still emerged as a powerful determinant of creditworthiness. This implies that duration-based factors are innately good predictors of risk. As such, it's likely that similar effects will also be uncovered in DeFi lending and shall begin to serve as a cost imposed on the abandonment of an existing public key in exchange for one that has been newly created.

## Asset Price Volatility

Regardless of the degree to which a loan is collateralized, the extreme price volatility endemic to cryptocurrency asset markets presents unique challenges for portfolio management, loan structuring and individual default events. In the event that stringent collateral ratios are imposed, fluctuations in both the collateral or principal assets will affect the risk profile of that account. The value of the collateral asset could depreciate, the value of the principal asset could appreciate, or both could occur, ultimately impairing the ability for collateral to sufficiently mitigate default risk.

However, in the event that interest rates serve as the primary means of risk management, changes in the principal asset price can dramatically affect a borrower's ability or incentive to repay their loan. This is particularly so when principal asset prices rise sharply. In fact, the relationship between inflation and consumer loan defaults in volatile, emerging markets is well studied[7,8]. So it follows that if interest payments are denominated in the principal asset, effects seen in traditional financial markets during exceptionally high inflationary conditions can be expected to manifest in DeFi as well.

Increased price volatility is a feature of all nascent markets. And while improvement should be expected as the DeFi market matures, immediately available solutions will be necessary for the introduction of

---

[6] Somin, Shahar, Goren Gordon, and Yaniv Altshuler. "Network analysis of erc20 tokens trading on ethereum blockchain." International Conference on Complex Systems. Springer, Cham, 2018.
[7] Badar, Munib, Atiya Yasmin Javid, and S. Zulfiquar. "Impact of macroeconomic forces on nonperforming loans: An empirical study of commercial banks in Pakistan." wseas Transactions on Business and Economics 10.1 (2013): 40-48.
[8] Malik, Madhur, and Lyn C. Thomas. "Modelling credit risk of portfolio of consumer loans." Journal of the Operational Research Society 61.3 (2010): 411-420.

undercollateralized loan structures. These solutions will likely vary depending on which particular concern is to be addressed. When attempting to limit the effect of price volatility, derivative protocols may be useful for mitigating portfolio risk; advanced decentralized governance mechanisms may be useful for loan structuring risks; and novel predictive technologies may allow protocols to instigate preemptive user behaviors to lessen individual default risk.

## Information Security

Over the past two decades, the number of significant data breaches has grown at an alarming pace, with estimates of their cost to society reaching well over $100B per year[9]. Financial firms have not been exempt from such attacks. Victims have included the U.S. credit bureau Experian and the U.S. credit card company Capital One, which received fines from regulators of $575M and $80M respectively following significant security failures. It therefore should go without saying that as a necessary custodian of sensitive financial data, any DLT-based credit risk estimation product must be aware of security risks. Thorough vulnerability assessments leading to adequate cybersecurity measures must be implemented and continually updated.

Due to the public nature of DLT protocols, these concerns exist in a further heightened state. While the breach of personal information is not common within the DLT space, exploits leading to the loss of user's funds are. In 2018, the Japanese firm Coincheck suffered a hack leading to a loss of $534M of user funds and the Italian firm BitGrail lost $195M.

Fortunately, decentralized encrypted data storage networks and mathematical frameworks such as zero-knowledge proofs are widely used in DLT. These tools may be useful in efforts to mitigate cybersecurity risks described above.

# Individualized Risk Estimation: A Framework for DeFi Loans

Up to this point, the language and perspectives deployed in this paper are borne out of observations made solely through a lens engineered by traditional financial markets, a line of sight which can be followed from ancient mesopotamia through to present day Wall Street. One must be dispossessed of these notions to begin realizing the potential that exists owing to a confluence of technological factors, of which DLT is but one. Advancements in machine learning and computational power will enable dramatic improvements in the capabilities of analytic product development. In conjunction with methodological advancements, the ubiquity and richness of data, both on- and off-chain, will empower these analytical products to draw truly novel inference. Entirely new models for incentivizing ethical user data capture that are a response to the surveillance capitalism boom times initiated by the founding of the likes of Google and Facebook will serve to rebalance the inequities mass scale analytic products can create. And finally, the democratization of technology that was once the domain of only the most developed societies

[9] Benjamin Edwards, Steven Hofmeyr, Stephanie Forrest, Hype and heavy tails: A closer look at data breaches, Journal of Cybersecurity, Volume 2, Issue 1, December 2016, 3–14

will facilitate deeper and more varied applications of the innovations described below.

It is therefore important to conceptualize these innovations more broadly as risk estimation rather than a mere “credit score”. While the most direct, and therefore first, applications of inventions from this arena shall be financial in nature, adaptation is inevitable. It is through adaptation that we will uncover as yet unknowable utility of these tools in other aspects of modern life. The construct of a “credit score” therefore serves to limit one’s imagination of this future potential.

## Pseudonymous Identity

As previously discussed, the challenge laying at the core of this problem is the tension between the fundamental need for a stable digital identity and the justifiable desire for a certain degree of anonymity, as represented by a user’s public key address. For example, allowances for anonymity raise the likelihood of exploits such as a Sybil attack[10]. Due to the nature of public key instantiation, and in the absence of traditional “Know Your Customer” (KYC) verification, this exploit vector cannot be entirely eliminated; potential attackers must instead be disincentivized by limiting the potential rewards of a successful attack. It’s likely that this disincentivization can be achieved via two mechanisms: setting a low and variable maximum value on each loan and overweighting certain elements during risk modeling. Relating to the former, unknown risks (i.e., “empty” wallets) should be eligible for loan amounts that do not pose systemic risk to the lending protocol, which would increase with time and modeling information. Relating to the latter, elements less vulnerable to manipulation can be over-weighted during the risk modeling process, for example length of transaction or borrowing histories.

It’s likely that a certain subset of users will consent to a KYC verification process akin to that which exists in traditional loan applications; however, this will not apply to users who can not, or simply prefer not to, undertake such an exercise. Fortunately this challenge extends well past the boundaries of DeFi. Multiple teams are at varying stages of development with products that seek to solve this problem by different means. As it stands today, each suffers from often significant limitations, however. The following is therefore offered solely to affirm that resources are currently being expended in this research area, and not to claim a suitable solution presently exists.

Upala is a DLT project that seeks to establish a mechanism for assigning economic value to one’s digital identity as means of disincentivizing fraud, forgery and other abuses[11]. Because the value assigned is only redeemable should the holder delete the ID, one can infer the value the holder places on that representation of their identity from the ID’s economic value. Following over four years of R&D, Microsoft has launched its ION Decentralized Identifier (DIDs) product on its hyperledger blockchain[12]. Its intended uses relate almost exclusively to enterprise-level centralized network and resource

[10] Douceur, John R. "The sybil attack." International workshop on peer-to-peer systems. Springer, Berlin, Heidelberg, 2002.

[11] See https://docs.upala.id/en/latest/whitepaper.html

[12] See https://techcommunity.microsoft.com/t5/identity-standards-blog/ion-we-have-liftoff/ba-p/1441555

management, but the learnings acquired will likely translate to other end-uses. There is also extensive literature available on the topic of decentralized KYC[13,14,15].

The absence of a complete, or even partial, solution to this problem doesn't preclude the creation and use of a decentralized default risk estimate; rather, it merely limits the degree to which this product can facilitate the shift of risk premium from collateral requirements to individualized interest rates. As decentralized identity solutions become more reliable and acceptable, increasingly more complex and useful individualized DeFi loans will become viable.

## Data Capture and Storage

With respect to data capture, two avenues of discussion are warranted to fully explore the innovative potential being proposed: collection incentives and collection infrastructure. Novel incentives are required to correct inequities that exist in contemporary analytical product development life cycles, which are perhaps best evidenced by the digital advertising practices responsible for creating the world's largest technology companies[16]. Data-derived inequities also extend deep into the traditional financial industry, often in an even more secretive and detrimental manner. The very same clandestine, monopolistic practices that have led to a concentration of wealth derived from data, have also created a concentration of infrastructure designed to collect, store and monetize data.

Several teams have begun to demonstrate the possibilities of a participatory data capture and monetization model that ensure profits generated from user data are also enjoyed by those from which the data were derived. The Measurable Data Token project has developed applications that allow consumers to recapture economic value from their purchasing data that would otherwise be expropriated by companies such as Amazon[17]. Streamr has generalized this model by creating utilities that allow app developers to deploy financially-ethical, turnkey data marketplaces seamlessly within the products they build[18].

Similar models can be deployed for the collection of financial data that will be vital to the production of a well-specified DeFi risk estimation framework. Such an approach is not motivated by ethics alone; certain classes of user data will not be publicly available on-chain and their reliable submission must therefore be incentivized. DLT has enabled these models to spring from the pages of economics text books into everyday life because the same technological system that can capture and store data can convey value to its creators in return (i.e., in the form of cryptocurrency tokens). This feature, along with

---

[13] Kapsoulis, Nikolaos, et al."Know Your Customer (KYC) Implementation with Smart Contracts on a Privacy-Oriented Decentralized Architecture." Future Internet 12.2 (2020): 41.
[14] Sinha, Prince & Kaul, Ayush. (2018). Decentralized KYC System. 10.13140/RG.2.2.35562.16323.
[15] Moyano, José Parra, and Omri Ross. "KYC optimization using distributed ledger technology." Business & Information Systems Engineering 59.6 (2017): 411-423.
[16] Zuboff, Shoshana. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs, 2019.
[17] See https://mdt.io/
[18] See https://streamr.network/

additional efficiencies gleaned from autonomously-controlled protocols, dramatically lower the overhead costs to administer a system as described here.

The advent of decentralized encrypted data storage networks like IPFS[19] and Swarm[20] extend these efficiencies to include mitigation of various barriers typically associated with operating large data storage infrastructure. Significant start-up capital requirements, ongoing hardware optimization costs and challenges arising from long distance geo-distribution create economies of scale that prevent firms from operating their own data centers, thus leading to only a small number of cloud providers[21]. This oligopolistic market structure in turn raises concerns relating to information security, fair market pricing and other anti-competitive behaviors, each of which serves as a threat to a firm seeking to challenge existing credit risk modeling infrastructure. By delivering comparable costs relative to performance and eliminating anti-competitive threats, decentralized encrypted data storage networks have become an important development catalyst in the DLT industry.

These advancements in the collection and storage of rich user data have begun to enable the next generation of analytic product development, such as will be necessary for a blockchain based credit risk estimate. Novel incentive structures will elicit participation in data collection efforts and distributed storage architecture will lessen both economic and accounting costs associated with managing these data.

## Risk Attribute Discovery

All analytic products are only as capable of producing utility as their inputs are reflective of the real-world phenomena the product intends to model. The importance of curating a rich dataset of relevant, timely and comprehensive attributes is therefore self-evident. However, the nature of these attributes and the methods by which they are captured may always remain a closely guarded trade-secret. For this reason, the following is offered merely to classify various data into similar groups, for example by source or by inferential utility, and to discuss their potential benefits and costs in analytical terms.

Summarily, attributes can be sourced in one of two ways: on-chain observation and off-chain collection. On-chain observation is computationally expensive, but highly reliable, whereas off-chain collection is computationally simple *per se,* but carries with it verification challenges. As a result, initial products can be expected to rely heavily on on-chain observation, where the descriptor “on-chain” refers to data that was generated by any function of a given DLT protocol and not simply its presence in a distributed ledger block. This is to distinguish between such data and the activity of a DLT oracle, which is a tool that simply exposes off-chain data within an on-chain structure.

Borrowing from graph theory, on-chain observations can be categorized as being descriptive of the target node in the first order or that target’s vertices (i.e., other nodes and the edges that connect them) in the

[19] See https://ipfs.io/

[20] See https://swarm.ethereum.org/

[21] Albert Greenberg, James Hamilton, David A. Maltz, and Parveen Patel. 2009. The cost of a cloud: research problems in data center networks. SIGCOMM Comput. Commun. Rev. 39, 1 (January 2009), 68–73.

second order[22]. In this construct, public key addresses are nodes and transactions are edges. Given that the objective is to characterize a single wallet, "target" denotes the address the risk estimate is to be computed for. Examples of first order attributes may include asset balances, transaction frequency, address age and essentially any derivation that can be computed without the need to access another node in its network. Examples of second order attributes may include prior loan performance, interactions with smart contracts or first order attributes of other connected nodes.

Due to their potential for exploitation, attributes sourced from off-chain resources are best characterized by their placement on a two-dimensional continuum describing expected inferential value in a predictive model versus their verification costs. This model should be used to prioritize development resources during construction of off-chain attribute collection architectures. For a KYC-identified user, an example of a high value-low cost attribute may be their bank account balance as it reliably describes the ability to repay a loan and can likely be verified by a trusted third-party (i.e., a banking institution). For a pseudonymous user, an example of a high value-low cost attribute may be their web browsing history as recorded by a trusted application, which has been shown to be predictive in credit scoring risk models[23]. Attributes that fall into this high value-low cost quadrant would logically be deemed the highest priority when allocating scarce development resources. Because there is a general tendency for real-world data to be correlated, low value-low cost attributes would be the next best suitable priority once the former has been exhausted. This strategy emphasizes scarce development resources and speed-to-market imperatives, while producing a greater number of potentially orthogonal risk factors.

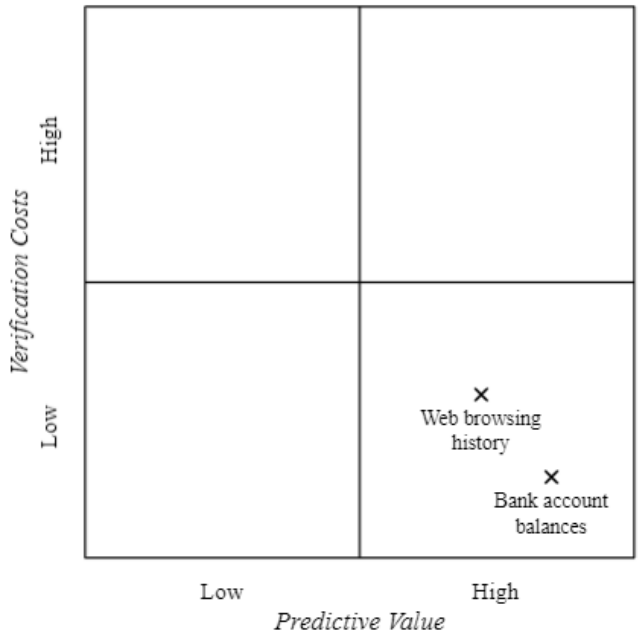


***Figure 1: Prioritization via predictive value vs. verification cost***

[22] Diestel, Reinhard (2017), Graph Theory (5th ed.), Berlin, New York: Springer-Verlag

[23] Berg, Tobias, et al. "On the rise of fintechs: Credit scoring using digital footprints." The Review of Financial Studies 33.7 (2020): 2845-2897.

A thorough discussion of verification methods for self-reported, off-chain data is out of scope for this paper. However, certain innate advantages exist for a verification system deployed on or alongside DLT. First, DLT systems are tamper proof; once data has been recorded, and assuming certain network conditions are maintained, encryption methods make retroactive modification impossible. This feature can be extended to maintain verification records of centralized, legacy data systems using hashing methods[24]. Their immutable nature yields tremendous verification cost savings because the activity must only be performed once. Second, the collocation of network structures and economic value (i.e., tokens) creates an environment naturally suited to peer-to-peer verification schemas; for example, third parties can be incentivized to attest to the authenticity of certificates, claims or other self-reported data[25].

# Analytic Methodology

At present, several teams are positioned to deliver a default risk product for use in DeFi lending, though they are each using suboptimal analytic methods. The following will demonstrate that these deficiencies may in fact be catastrophic at scale. The predominant approach is best described as a "heuristic factor weighting" model. Translated to a statistical regression framework, this method entails devising a series of variables and their associated coefficients using intuitive human judgements of their relationships to credit risk. The shortcomings of such a method are easily understood through the lens of adverse risk selection.

Adverse selection occurs in any marketplace characterized by information asymmetries between buyers and sellers; in our example buyers are credit seekers and sellers are decentralized lending protocols[26]. In this case, credit seekers possess knowledge of their creditworthiness that lending protocols necessarily cannot when no risk factors are taken into account during the origination of a loan. It is well understood that these risk factors must therefore be observed in order to eliminate the effects of asymmetric information and the resulting potential for adverse selection[27]. In both macroeconomic and microeconomic contexts, a consequence of pervasive adverse selection is the collapse of the market or the firm, respectively. Particularly so for firms that issue risk-related products, adverse selection produces a tendency for high risk borrowers to favor those products that most inadequately price the severe risks that they pose. This is best understood by example: if a prior loan default is a reliable indicator of a future loan default, and firm A adequately incorporates this factor into its price but firm B does not, all rational

[24] Kalis, Rosco, and Adam Belloum. "Validating data integrity with blockchain." 2018 IEEE International Conference on Cloud Computing Technology and Science (CloudCom). IEEE, 2018.
[25] Aydar, Mehmet, and Serkan Ayvaz. "Towards a Blockchain based digital identity verification, record attestation and record sharing system." arXiv preprint arXiv:1906.09791 (2019).
[26] Akerlof, George A. "The market for "lemons": Quality uncertainty and the market mechanism." Uncertainty in economics. Academic Press, 1978.
[27] Weiss, Gregor NF, Katharina Pelger, and Andreas Horsch. "Mitigating adverse selection in p2p lending–Empirical evidence from prosper. com." (2010).

consumers that have previously defaulted shall seek loans from firm B, leading to firm B's insolvency if the effect persists.

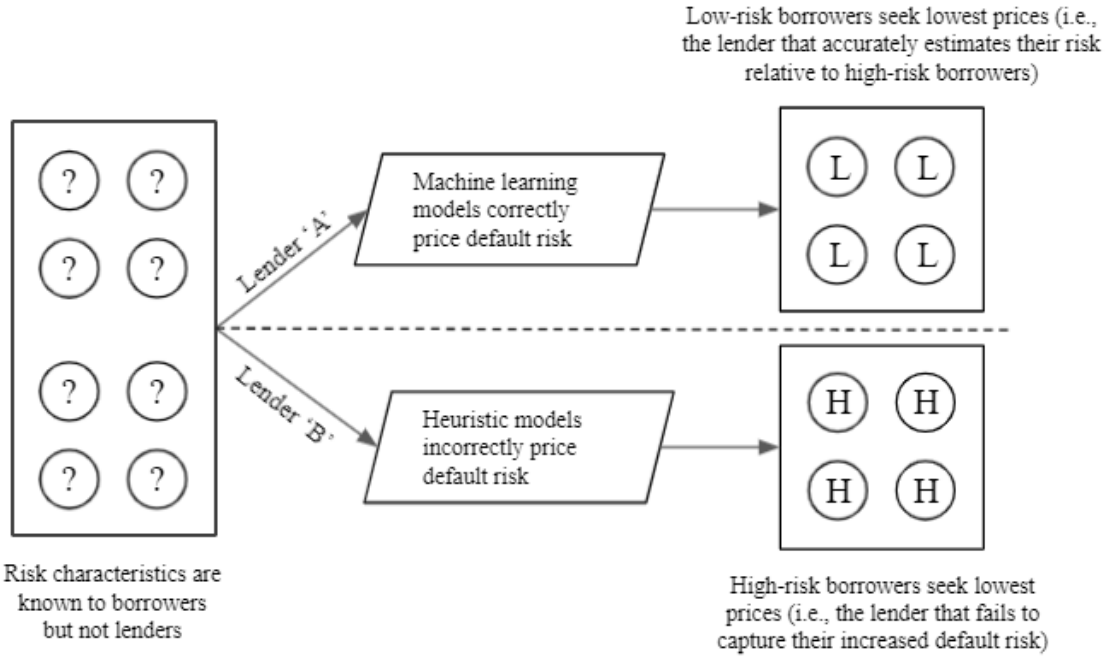


***Figure 2: Adverse selection due to risk misspecification***

The field of behavioral economics has perhaps exceeded all other disciplines in successfully demonstrating the severity of judgement errors that the human mind makes in its assessment of risk[28]. Even supposed "subject matter experts" have been consistently shown to be victims of similar fallacies and biases as lay individuals. This fact, combined with an understanding of adverse selection, leads to a simple, yet important, assertion: when issuing undercollateralized DeFi loans, the firms that price risk using methods that avoid the shortcomings of human judgement shall outperform all others.

Statistics have been used to price risk going as far back as 1693 with Edmond Halley's rudimentary actuarial analysis of mortality rates and the costs these rates implied for life annuity premiums[29]. In contemporary finance, statistical models are used almost exclusively in the pricing of risk-based products, particularly those dealing with consumer credit. Given the availability of suitable statistical tools, the abundance of pertinent risk data and size of the potential DeFi lending market, statistical modeling must therefore be the optimal method for the development of trust estimates sufficient to safely issue undercollateralized debt.

When competing solely against heuristic models, even the most simple statistical models can be expected to produce significant performance advantages. Initially, Generalized Linear Models (GLM), which have been widely used in insurance since the 1970s, offer certain attractive features relative to more modern

[28] Kahneman, Daniel, and Amos Tversky. "Prospect theory: An analysis of decision under risk." Handbook of the fundamentals of financial decision making: Part I. 2013.
[29] Halley, Edmund. "An estimate of the degrees of the mortality of mankind." Mathematical Demography. Springer, Berlin, Heidelberg, 1977. 21-26.

tools[30]. They are relatively inexpensive to develop, both in terms of statistical expertise and computational costs, and easy to understand. This accessibility makes them an excellent framework to facilitate early-stage breakthrough market adoption.

On the other hand, more advanced statistical methods have recently become widely available thanks to the percolation of open source software from the likes of Google or Facebook. Among these are Artificial Neural Networks (ANNs), which have been demonstrated to outperform traditional credit scoring methods in a commercial setting, such as the GLM[31]. ANNs are a statistical technique inspired by a simplified reconstruction of organic neural pathways. Genetic Programming (GP) has been demonstrated to further outperform ANNs in the development of credit scoring models[32]. GP is an extension of an ANN that recursively spawns subsequent models that each select only the best performing models from which further spawning is permitted. This technique is similarly inspired by our understanding of evolutionary biology. Both ANNs and GP are frameworks well within reach for a moderately funded firm that seeks to advance data-driven credit risk estimates when the potential returns are that which can be expected from the introduction of such a product.

## Product Delivery

With respect to distribution, traditional credit reporting bureaus suffer from at least two critical flaws that should be resolved by a decentralized alternative: a lack of discriminative exchange of the information contained in its products and a lack of market pricing for these products. Relating to the former, borrowers are typically required to consent to complete disclosure to a lender of all information contained in their credit file. Given the increasingly varied use of credit scores, examples abound where this degree of disclosure is unnecessary. Relating to the latter, the presence of a small number of credit reporting bureaus in a given jurisdiction affords agencies near-monopolistic status, including the pricing freedoms associated with this market structure[33]. Moreover, the absence of traditional competitive forces may act to disincentivize innovation. Cryptographic tools such as zero-knowledge proofs, oracle protocols, and token smart contracts offer revolutionary new possibilities for addressing these issues.

Zero-knowledge proofs (ZKP) are field of mathematics that allow one party to prove to another party that they are in possession of certain information without needing to convey anything other than the truthfulness of that assertion[34]. ZKPs have been utilized as a primary protocol for privacy-oriented DLT projects like Monero and ZCash, though they can also be used as subprotocols without limiting their

---

[30] Graybill, Franklin A. Theory and application of the linear model. Vol. 183. North Scituate, MA: Duxbury press, 1976.

[31] West, David. "Neural network credit scoring models." Computers & Operations Research 27.11-12 (2000)

[32] Ong, Chorng-Shyong, Jih-Jeng Huang, and Gwo-Hshiung Tzeng. "Building credit scoring models using genetic programming." Expert Systems with Applications 29.1 (2005): 41-47.

[33] Pagano, Marco, and Tullio Jappelli. "Information sharing in credit markets." The journal of finance 48.5 (1993): 1693-1718.

[34] Goldwasser, Shafi, Silvio Micali, and Charles Rackoff. "The knowledge complexity of interactive proof systems." SIAM Journal on computing 18.1 (1989): 186-208.

privacy-preserving features[35]. This shall enable a protocol to maintain, for example, transparent governance practices, while also restricting access to sensitive user information on a "need to know" basis. In the context of credit risk scoring, ZKPs enable a potential lender to merely stipulate that a borrower's score be above a certain threshold to grant him credit, without requiring disclosure of other information in the borrower's file. This example can be easily extended to other issuance criteria, for example score-dependent interest rates, by simply reformatting the proof to reflect whether a score falls within each range that would necessitate a different interest rate.

While many operational challenges still exist, an oracle protocol can in theory serve as an autonomous delivery interface for borrowers and lenders that wish to retrieve credit risk analytics. In combination with token smart contracts, the introduction of open market pricing for these products is remarkably simple. An oracle can be designed to grant metered, time-bound access to a user on the condition that an access token is transferred to the oracle protocol. The access token can then be permanently removed from circulation, or "burned. If the access token's smart contract is designed such that new tokens are emitted at a fixed frequency, it's market price will equilibrate freely based on the available supply of the token and the demand for the information available within the oracle.

As will be discussed further in the next section, the market pricing dynamics of a token intended to meter access to a scarce resource, such as an analytic product delivered on a computer network, are necessarily quite different from that of, for example, a protocol governance token. Economic theory suggests that equilibrium price is achieved when the quantity supplied by a producer equals the quantity demanded by consumers. It also suggests that additional production should cease when the marginal cost equals marginal revenue. When applied to pricing models for access tokens versus governance tokens, these theories must therefore imply divergent pricing functions because, on an accounting basis, the marginal cost of governance is always zero while the marginal cost of access is always a positive, non-zero, real number. Perhaps in more practical terms, access and governance are simply very different economic goods which in all likelihood implies equally different market prices. So by decoupling the value assigned to governance from the value assigned to access, each price discovery market is able to reach equilibrium without being constrained or distorted by the other.

## Analytic Governance

Even more so than pricing, the greatest detriment to consumer welfare generated by the monopolistic market structure of credit reporting agencies is the lack of incentive for ethical and transparent operational governance. This is made evident by the enactment of sweeping legislation such as the Fair Credit Reporting Act of 1970[36] in the United States and the General Data Protection Regulation of 2018 in the European Union[37]. Despite such laws, credit bureaus account for as much as 75% of all complaints

[35] Goldreich, Oded, and Yair Oren. "Definitions and properties of zero-knowledge proof systems." Journal of Cryptology 7.1 (1994): 1-32.
[36] Frenzel, Lawrence D. "Fair credit reporting act: the case for revision." Loy. LAL Rev. 10 (1976): 409.
[37] Lopez Rojas, Edgar Alonso, Dincer Gultemen, and Erjon Zoto. "On the GDPR introduction in EU and its impact on financial fraud research." European Modeling and Simulation Symposium, EMSS 2018. Cal-tek Srl, 2018.

received by the U.S. Consumer Financial Protection Bureau[38] and 1 in 20 U.S. consumers are denied or over-charged for credit due to errors in their reports[39]. These and other similar statistics demonstrate the ineffectiveness of legislation as a solely adequate solution to ensure credit reporting operational governance practices are well balanced between lender and borrower welfare. Other mechanisms must therefore be needed to supplement legislative efforts.

One such innovation that has emerged from DLT is a Decentralized Autonomous Organization (DAO), which is an entity whose operations are defined by smart contracts and decided by a voting mechanism. Candidly, algorithmic governance as it exists today suffers from many known, and likely many more unknown, limitations necessitating significant future research on the topic[40]. However, a DAO represents great potential in an effort to address the shortcomings mentioned above, assuming certain underlying conditions are met.

Perhaps the greatest challenge of DAO smart contract construction is the fact that undeclared future constraints must be accounted for in the design of its codebase. Rather than explore direct technical solutions to this problem, a more general one is proposed: a firm seeking to employ a DAO in the context of DLT credit risk reporting must assume a high probability of exploit and balance the proportion of operational control ceded to a DAO against the systemic threat posed by such an exploit being successfully executed. Stated plainly, operational control must be slowly shifted from a centralized process to a decentralized one, the pace of which being dictated by our understanding of the strengths and weaknesses of the DAO's design.

Risk model development and implementation offers an excellent example to demonstrate this approach of incremental DAO implementation, while at the same time resolving a shortcoming of traditional credit reporting agency governance. None of the major agencies publish verifiable data describing the accuracy of its consumer credit models to a source available to the general public. Similarly, the rating variables and associated rating factors used in these models are typically considered trade secrets. Initially, control of only major new model implementations could be ceded to a DAO and voters informed only by model validation reports that are generated without reference to individual rating variables. Because one can expect significant improvement between major updates, and individual rating variables are not revealed, the DAOs behavior can be more reliably predicted and the potential for exploit lessened. As the firm's and the research community's understanding of DAOs advances, more control and information can be shifted from traditional managerial structures to a DAO framework.

Voting eligibility is another topic that requires careful consideration if a DAO is to be successfully applied in this context. Typically a vote is enabled by possession of a governance token. Because this token has a financial value, and a cost is therefore incurred by its acquisition, there's a tendency for income

---

[38] 6 Consumer Financial Protection Bureau, Consumer Response Annual Report January 1 – December 31, 2017 (March 2018)

[39] Federal Trade Comm'n Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003 (Dec. 2012)

[40] DuPont, Quinn. "Experiments in algorithmic governance: A history and ethnography of "The DAO," a failed decentralized autonomous organization." Bitcoin and beyond (2017): 157-177.

discrimination to arise as the price of the governance token increases. On the other hand, an attempt to artificially suppress the price of a governance token allows voting power to be more easily concentrated within a small faction whose interests may not be aligned with the broader community. For this reason, novel governance token emission and acquisition mechanisms must be explored. Mechanisms that encourage participatory behaviors in support of the protocol, for example validating self-reported data or acting as a guarantor (a topic discussed in the next section), are preferable to a purely free-market purchase model.

# Undercollateralized DeFi Loans: Novel Products and Platforms

Given the simplicity of current CDP loan structures, the introduction of undercollateralized loans will require enhancements to both loan terms as well as the protocols from which they are originated. With respect to loan terms structure, much can be learned from traditional finance thanks to their extensive and long standing use across a variety of markets. However, traditional financial institutions rely heavily on labor to administrate loan products, which limits the potential for these experiences to inform DeFi lending protocol development. Particularly as it relates to the latter, the following should therefore be interpreted as an exploration of possible features, rather than an exhaustive specification proposal.

## Loan Structure

The stringent collateral requirements common to all current DeFi lending products allow individual loans to be issued without time-based repayment terms of any kind. However, as collateral requirements are relaxed, pre-specified repayment terms will become increasingly critical to manage protocol liquidity, and ultimately, protocol solvency. This fact can be observed easily in developing nations as increases in economic and population growth lead to more frequent loan origination and loan default, requiring regulators and institutions to modify institutional capital requirements[41].

Initially, it's likely that a maturity date shall be the only necessary time-based repayment term added to an undercollateralized DeFi loan, though this assumes many of the requirements previously outlined such as a gradual relaxation of collateral ratios and prudent maximum loan amounts. At this stage, borrower-specific default risk as measured by a model estimate can still be incorporated solely within collateral requirements, leaving the interest rate as a vehicle for conveying time-dependent default risk. This structure allows for the simplest possible technical implementation of an individualized, undercollateralized loan product, which shall accelerate speed-to-market.

As collateral requirements for a loan approach zero, additional terms will be needed to maintain sufficient cash flow planning for appropriate liquidity levels to be preserved. These terms can likely be drawn almost entirely from traditional finance. For example, an incremental repayment schedule can be

[41] Sharma, Divesh S., and Errol R. Iselin. "The Relative Relevance of Cash Flow and Accrual Information for Solvency Assessments: A Multi‐Method Approach." Journal of Business Finance & Accounting 30.7‐8 (2003): 1115-1140.

employed to ensure funds are being continually redeposited to liquidity pools and for early identification of delinquent accounts. While the latter cannot be expected to aid in loss recoupment, it can allow the protocol to more quickly recompute future loan terms in a manner that preserves target liquidity levels, a topic that will be explored further in the following section.

Presently, DeFi loans exist in only one of three states: open active, closed positive or closed liquidated (i.e, negative). The absence of any continuum between these states precludes late fee assessment as a precursor step to default (i.e., liquidation) and severely hinders the information content available to post-hoc risk modeling exercises. What's more, assessment of a late fee can be expected to elicit behavioral change in many borrowers leading to the eventual positive closure of an account. As a risk class, this borrower is inherently different from one who adheres to all loan terms and another who fully defaults. Capturing more granular borrower performance data, such as that which would be generated by this example, enables improved risk model specification.

## Protocol Structure

As stated above, existing information describing the implications of undercollateralized loan products on protocol design is virtually non-existent. Not only are autonomous lending protocols a relatively new invention, but undercollateralized loan products issued by such a protocol are exceedingly rare. Within the following, certain protocol design enhancements shall be clearly and immediately necessary, while others may only emerge as useful only once protocols evolve to a more advanced state.

For a protocol to maintain target liquidity levels, a feedback loop must be established between portfolio performance and new loan issuance. Stated plainly, if default rates rise, credit granting parameters must be constrained to maintain solvency. If default rates fall, these parameters can be relaxed to maximize liquidity and operational profit. The precise targets that are chosen and the degree to which actual performance deviates from targets will likely also be a feature upon which protocols compete for borrowers in an environment where multiple platforms exist, contrary to the oligopoly seen in traditional credit scoring environments.

It's possible that protocols can be redesigned such that third-party guarantors are incentivized to replace some portion of borrower-supplied collateral with their own. If this is to be tenable for potential guarantors, adequate portfolio management tools must be available, for example the credit risk products described above. Sufficient financial incentives must also be offered to guarantors, which can be made possible by the increased market interest rates undercollateralized or uncollateralized loan products will support. This may prove to be an ideal mechanism for issuing governance rights in the form of tokens, which would provide further incentive to guarantors. The proposed guarantor-supplied collateral model also gives rise to a secondary market for consumer loans that can be tranched to offer fixed risk and return products, which is an area of significant interest for venture capitalists at present. Individual guarantors can set their own risk and return parameters in a manner that generates a portfolio of interest-bearing assets pursuant to their specific financial goals. This is not unlike the secondary market for mortgage-backed securities in traditional finance, except that it is inherently more inclusive thanks to DeFi's dramatically lower barriers to entry.

Features that mitigate borrower default risks associated with price volatility can also be implemented at a protocol level. A research area that shows particular promise is the deployment of "two-way collateral functions", which involves deploying a portion of the borrower's collateral to derivatives that serve as upside and downside hedges[42]. The ratio of native and derivative collateral can be set algorithmically by the protocol or set manually by the borrower, and in either case updated periodically based on market conditions. All else equal, this structure also enables greater individual loan amounts while maintaining the same level of systemic collateral risk to the protocol, promoting greater scalability.

Perhaps the most significant unresolved feature gap of an undercollateralized lending protocol is the lack of an equivalent monitoring incentive to what currently exists as a liquidation market in existing protocols. Potential liquidators maintain a protocol's financial health by closing loans that are in violation of collateral ratio requirements in exchange for the right to seize the borrower's collateral. Thus, in the absence of collateral, a new monitoring incentive must be devised. Issuing governance tokens, conferring the liquidation right directly to guarantors, conveying ownership of future interest revenues, and other incentives can be explored. Because the immediate objective is merely to relax collateral requirements, and not to eliminate them, the absence of a complete solution is not prohibitive.

# Conclusions: Research Beyond Credit Risk

Out of the foregoing should have emerged one central theme: that product development in this space is best cast as a series of research problems. Creation of a sufficiently performant DeFi default risk score will require application of cutting edge network analysis for dataset curation and advanced statistical methods for risk estimation. Mitigation against attack vectors that are enabled by pseudonymous identity may call for novel adjustments to risk modeling or creation of new marketplaces for identity price discovery. Existing schemas for DAO governance models must be overhauled to securely and reliably address socioeconomic and access disparities that pervade all financial markets, decentralized or otherwise. Advanced cryptographic methods must be deployed to preserve privacy in some areas, while leaving others capable of much needed transparency. Perhaps most importantly, deployment of each of these solutions must be conducted in a manner that balances prudence against the inevitable pressures of capitalist forces, as an existential threat can just as easily come from within (e.g., security exploits) or without (e.g., competition).

One particularly exciting upshot arises from the potential to deploy the architecture built for credit risk products to other forms of risk estimation in the cryptocurrency industry. Just as a traditional credit score has been adapted to become a powerful determinant in auto insurance and other traditional financial products, this framework will prove useful beyond DeFi lending, for example new project capital raises, token supply control measures such as staking and liquidity farming, or more targeted token issuance tactics (i.e., airdrops). The technologies and philosophies proposed above represent a broader leap forward in scalable, ethical individualization more so than simply an advancement of DeFi loan product offerings.

---

[42] Kim, Sungil. "New Crypto-Secured Lending System with a Two-Way Collateral Function." Ledger 6 (2021).