Journal of Corporate Finance 96 (2026) 102908

Contents lists available at ScienceDirect

Journal of Corporate Finance

journal homepage: www.elsevier.com/locate/jcorpfin

Blockchain price oracles: Accuracy and violation recovery☆

Matthias Nadler, Katrin Schuler *, Fabian Schär

Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland

ARTICLE INFO

Editor: E. Lyandres

JEL classification:
G29
G32

Keywords:
Asset prices
Blockchain
Decentralized finance
Financial risk
Oracles

ABSTRACT

Reliable asset price data are critical for the functioning of decentralized finance (DeFi) protocols, particularly those involving collateralized lending. The accuracy of blockchain-based price oracles directly affects key processes such as collateral valuation, liquidation, and risk management. This paper presents a comprehensive empirical analysis of Chainlink Price Feeds (CPFs), the dominant oracle infrastructure in DeFi. We compile a novel dataset of over 150 million observations from 40 CPFs on Ethereum over an 18-month period, matched to benchmark prices from a centralized exchange. To identify the determinants of oracle inaccuracy, we estimate pooled OLS and fixed effects regressions, relating price deviations to design parameters, reporter dynamics, and market conditions. We then introduce a Markov-like state transition framework to model the resolution of target corridor violations, using multinomial logistic regression to estimate transition probabilities. Finally, we exploit position-level data from one of the largest decentralized lending markets and apply entity fixed effects regressions to examine how users adjust collateralization in response to oracle design. Our findings highlight economically significant deviations that are systematically related to oracle accuracy configurations and market stress, and show that users internalize these risks in their financial decisions. The results offer new insights for the design of resilient oracle systems and the management of risk in decentralized financial markets.

1. Introduction

Accurate and timely information lies at the heart of financial markets, shaping asset valuation, contract enforcement, and risk management. The reliable provision of price data is therefore a fundamental prerequisite for market efficiency and the proper functioning of financial contracts. Contracting parties rely on timely and accurate asset price information to determine contract terms, assess collateral adequacy, and manage risk exposures. In conventional financial markets, this information is typically supplied by centralized intermediaries such as exchanges, rating agencies, trade repositories, or commercial data vendors, whose data quality, incentive structures, and operational risks have been extensively examined in the finance literature (e.g., Brennan et al., 1993; Bessembinder et al., 2006; D'Souza et al., 2010; Duffie and Stein, 2015).

With the emergence of blockchain-based financial applications, often referred to as Decentralized Finance (DeFi), the mechanisms for information provision differ fundamentally. Many smart contract-based financial arrangements require external data to be enforced. Whenever users cannot be trusted to provide this information truthfully, DeFi applications rely on specialized entities known as *oracles*, which transmit off-chain information (e.g., asset prices, macroeconomic indicators, or event outcomes) onto the blockchain. However, unlike centralized data vendors, oracles operate in a partially decentralized and often economically incentivized framework that may exhibit different forms of risk, including manipulation, latency, and misreporting.

☆ This article is part of a Special issue entitled: 'Blockchain-based finance' published in Journal of Corporate Finance.

* Corresponding author.
E-mail address: katrin.schuler@unibas.ch (K. Schuler).

https://doi.org/10.1016/j.jcorpfin.2025.102908
Received 15 December 2023; Received in revised form 13 October 2025; Accepted 14 October 2025
Available online 21 October 2025

The reliability of asset price oracles is particularly important for DeFi applications, commonly referred to as protocols, that contain elements of collateralized lending. Inaccurate or delayed price updates can result in mispriced collateral, flawed liquidations, or unanticipated losses, potentially amplifying systemic risk within the broader DeFi ecosystem (e.g., Perez et al., 2021; Lehar and Parlour, 2022; Schuler et al., 2023). Given that these protocols currently secure assets worth several tens of billions of U.S. dollars,[1] understanding the accuracy and reliability of oracles is critical for DeFi market participants, as well as for regulators and supervisory bodies.

The critical role of asset price oracles is widely acknowledged and has been qualitatively explored in the literature on decentralized financial ecosystems (e.g., Eskandari et al., 2021; Schär, 2021; Werner et al., 2022; Duley et al., 2023). In contrast, empirical research on asset price oracles remains relatively scarce. Notable exceptions include Cong et al. (2025), who aggregate data from various sources across multiple DeFi ecosystems to examine the effects of oracle availability on economic activity metrics. A broader analysis of dependency structures within the Ethereum ecosystem is provided by Kaleem and Shi (2021), who compile a novel usage data panel for 88 asset pairs supported by the most widely used oracle service at the time.

Oracle accuracy has thus far been examined primarily in studies with a limited asset scope. Liu et al. (2021) investigate the accuracy of Ether-USD price feeds used by four leading DeFi protocols, while Gu et al. (2021) analyze the proprietary oracle of a major on-chain collateralized stablecoin protocol, likewise focusing on the Ether-USD pair over time.

This study contributes to the literature by providing a detailed empirical analysis of oracle accuracy across a broad range of assets supported by the dominant oracle service in decentralized finance: Chainlink Price Feeds (CPFs) on Ethereum. Using block-level data for 40 CPFs over an 18-month period as well as 146,620 user capitalization choices in a leading DeFi lending market, we examine three aspects of oracle performance: (i) the magnitude and determinants of price deviations relative to benchmark market prices; (ii) the dynamics of recovery following deviation violations; and (iii) the impact of oracle accuracy configurations on DeFi users' collateralization choices and overall capital efficiency.

Our empirical approach leverages two novel datasets that combine on-chain oracle activity, update mechanics, and protocol participation metrics with high-frequency reference prices from Binance, the most liquid centralized exchange during the observation period.

We document several important findings. First, CPFs exhibit an average deviation of 57 basis points relative to the reference price, with tighter update parameters (e.g., narrower target corridors and faster heartbeats) significantly reducing deviations. Second, we observe that elevated asset price volatility and trading volume are associated with larger deviations, while the transition to proof of stake is linked to a marginally improved accuracy. Network congestion, as measured by gas prices, has a negligible effect on oracle deviations. Third, the effects of reporter participation and reporter uncertainty, measured as the variance in the reported prices at a given point in time, are only persistent prior to the Ethereum Merge and when controlling for CPF fixed effects. In those cases the participation is linked to a decrease in deviation, while the uncertainty among the various reporters is linked to an increase in deviation. Fourth, based on a Markov transition model, we estimate that accuracy target corridor violations occur in approximately 2.5% of all observations and are typically resolved within 10 blocks (roughly two minutes), although larger violations tend to persist for longer periods. Our analysis further shows that both the length of the violation sequence and the magnitude of the deviation significantly influence the subsequent transition path, with more severe or prolonged deviations increasing the likelihood of correction via an oracle update rather than a random recovery. Finally, we find evidence that both oracle accuracy and user-specific factors influence capitalization decisions. More sophisticated users appear better equipped to operate with tighter collateral margins, while less sophisticated users maintain larger buffers as a precaution. Moreover, the composition of asset pairs also matters, with cross-category positions (i.e., crypto collateral with stablecoin debt or vice versa) typically requiring higher collateral buffers due to greater relative price risk. These findings suggest that improved oracle accuracy enables all users to be more capital-efficient while remaining within protocol risk thresholds, though the extent of this effect varies across user types and asset configurations.

Our findings carry relevant insights for risk management and capital efficiency concerns in blockchain-based financial markets. Protocol designers must account for the probabilistic nature of oracle deviations when setting liquidation thresholds and collateral requirements. At the same time, users need to be aware of the potential risks arising from oracle imperfections, particularly during periods of elevated volatility. Our results are supported by an extensive set of robustness checks, including alternative model specifications, subsample analyses, and adjustments for outliers, which confirm the stability of the main conclusions.

To the best of our knowledge, this study is the first to provide comprehensive empirical evidence on the determinants and consequences of oracle accuracy across a wide variety of assets within a large-scale DeFi market. As such, our results contribute to the growing literature at the intersection of financial technology, information provision, and decentralized market design.

The remainder of the paper is structured as follows. Section 2 provides an overview of relevant literature, different categories of oracles, the design of CPFs, and the role of oracles in decentralized finance, offering the foundation for the subsequent analysis. Section 3 formalizes the research questions and outlines the key hypotheses to be tested. Section 4 describes the data collection process, variable construction, and sample selection. Section 5 introduces the empirical methodologies employed and presents the main findings, focusing on price deviations, violation recovery and the effect of relevant oracle configurations on users' collateralization choices. Section 6 conducts a series of robustness checks to assess the sensitivity and reliability of the results, including treatments for outliers and approaches addressing potential endogeneity concerns. Finally, Section 7 summarizes the main contributions and discusses implications for DeFi market design, regulatory oversight, and future research.

[1] Approximately USD 40 billion according to DeFiLlama.com (retrieved August 8, 2024).

2. Background

This section provides the necessary conceptual and technical background on blockchain price oracles, with a particular focus on their role in DeFi. Given the novelty of this topic in the finance literature, we aim to offer a concise overview of the oracle problem, its implications for financial contracts on blockchains, and the design choices that shape oracle reliability. By outlining the fundamental mechanisms and challenges of price oracle systems, we establish the context for our subsequent empirical analyses and highlight why understanding oracle accuracy is essential for assessing risk management and capital efficiency in DeFi protocols.

2.1. Oracle classification and literature

The challenge of providing accurate and timely external data to blockchain environments, commonly referred to as the *Oracle Problem*, has been extensively studied in the computer science literature (see e.g., Al-Breiki et al., 2020; Albizri and Appelbaum, 2021; Pasdar et al., 2023). Specifically in the context of DeFi ecosystems, the particular relevance of reliable asset price oracles is well understood (see e.g., Bartoletti et al., 2021; Werner et al., 2022; Zhou et al., 2023). With the exception of decentralized exchanges that offer endogenous price discovery by design (e.g., Lin et al., 2019; Buterin, 2017), most DeFi protocols rely on external price information to function as intended.

Eskandari et al. (2021) provide a comprehensive overview of blockchain oracle architectures, offering a classification framework and an assessment of key attack and corruption vectors. These vulnerabilities are further elaborated upon in recent surveys (Liu et al., 2021; Zhao et al., 2022; Heimbach and Wattenhofer, 2022).

Reliance on off-chain price oracles is implied whenever tokens are tied to real-world assets. However, also in the case of blockchain-native assets, there are several reasons why on-chain price sources are unsuitable as DeFi oracles. First, the vast majority of trading, and consequently, the formation of ground truth prices, occurs off-chain on centralized exchanges (Aspris et al., 2021). Second, while Angeris and Chitra (2020) formally show that decentralized exchanges can serve as robust on-chain oracles under efficient market conditions, such mechanisms remain vulnerable to manipulation in low-liquidity environments or through tools like flash loans (Qin et al., 2021; Wang et al., 2021; Lehar and Parlour, 2025). Time-Weighted Average Price (TWAP) mechanisms have been proposed to enhance manipulation resistance of on-chain price sources but come at a cost in terms of timeliness and precision and still exhibit inherent weaknesses (Mackinga et al., 2022; Aspembitova and Bentley, 2023). In practice, today's quality and resilience of on-chain price sources do not meet the requirements of DeFi protocols. In consequence, they are rarely used in a capacity beyond quality checks for off-chain reported prices.

Off-chain oracles introduce reliance on designated agents (so-called reporters) that are responsible for transmitting price data to the blockchain. This dependency presents specific vulnerabilities tied to both the update mechanism as well as the reliability and trustworthiness of the reporting agents.

In *pull-based* oracle architectures, an off-chain client monitors an on-chain oracle component for information requests. The client retrieves data from a specified source (e.g., a trusted API) and feeds the cryptographically signed response back to the oracle contract through a transaction (e.g., Zhang et al., 2016; Woo et al., 2020). To access price information, this design requires multiple transactions across several blocks. It thereby breaks composability and raises scalability concerns, which is generally unsuitable for the purpose of DeFi protocols. Consequently, most asset price oracles are designed to have prices continuously available and regularly updated, independent of the current demand.

A pragmatic, though highly centralized example are oracles where prices are pushed by a single trusted reporter at regular intervals. In this case, both data availability and accuracy depend entirely on that reporter, creating a single point of failure and critical concentration of influence over depending protocols. A partial mitigation strategy involves allowing any party to submit price updates, provided those values are cryptographically signed by a trusted source. However, such open systems often struggle to ensure timely and reliable updates due to weak incentive structures.

The dominant off-chain oracle design in DeFi today are *Decentralized Oracle Networks* (DONs). Although implementations vary, they generally involve a group of reporters that independently monitor the ground truth price and collectively submit updates on-chain. This decentralizes both the power and responsibility for updating prices. Reporter communication and coordination occur within a separate network. Whenever an update is due, a new price is independently determined by each reporter and subsequently processed by a predefined logic. This logic may include statistical aggregation, weighing mechanisms depending on the economic stake reporters put in the accuracy of their data, or their reputation based on past performance. Zhao et al. (2022) analyze popular DON setups and propose a framework for evaluating their trustworthiness. More recently, Cong et al. (2025) examine the impact of DON oracle adoption and availability on DeFi key metrics of protocols and blockchains respectively, and conclude that there are symbiotic network effects.

2.2. Chainlink Price Feeds

Chainlink is the most widely adopted price oracle infrastructure in DeFi and, during our observation period, served as the primary source of asset price data for the majority of Ethereum-based financial protocols. Despite the emergence of alternative oracle solutions, Chainlink Price Feeds (CPF) continue to be the dominant source of third-party asset price data on the Ethereum blockchain. A more technical description of the operational mechanics of CPFs is provided in Appendix B.

From the perspective of a dependent protocol and its users, the trigger criteria of a CPF can be interpreted as an accuracy and minimal timeliness commitment. They are specified in the form of a *target corridor* and a *heartbeat*. The former is defined as the

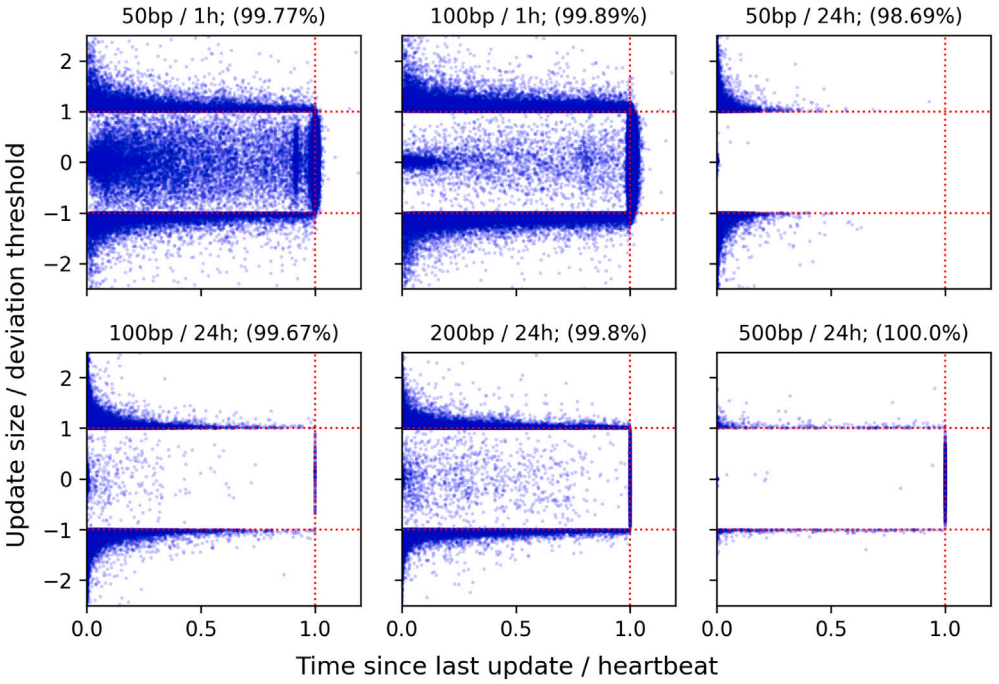


Fig. 1. Distribution of Chainlink Price Feed (CPF) updates over an 18-month period (March 2022–August 2023), grouped by trigger criteria (heartbeat and target corridor). Each observation corresponds to a successful on-chain price update. The vertical axis measures the normalized magnitude of changes relative to the preceding update, scaled by the CPF's parameters. The red rectangles denote the expected region of updates if all triggers were binding, providing a benchmark for responsiveness. Data are based on 40 verified CPFs deployed on Ethereum and percentages in brackets refer to the proportion of observations within range of the plot.

relative deviation of the reporters' current aggregated view of the ground truth price compared to the latest CPF price on-chain. The latter is defined as the maximum time passed since the last successfully transmitted price update.

Fig. 1 presents a descriptive analysis of the update behavior of 40 Chainlink Price Feeds (CPFs) over an 18-month period, grouped according to their trigger criteria. Each observation represents a successful price update and displays the magnitude of the normalized change relative to the preceding update, scaled by the CPF's respective trigger criteria. Presuming fast response time to CPF criteria being triggered, we would expect all updates to be in close proximity of the red rectangles.

This expectation appears to be broadly satisfied in the case of the heartbeat mechanism. However, a number of updates fall well outside the expected range defined by the deviation threshold. Observations beyond this corridor indicate that, in certain instances, the difference between the reported on-chain price and the aggregated off-chain reference price exceeded the stated threshold by a substantial margin. Such deviations warrant a more formal and systematic analysis of the conditions under which updates occur, as well as the extent to which trigger criteria are reliably adhered to in practice.

From a financial risk management perspective, the deviation threshold represents the more critical commitment. While a higher heartbeat frequency may improve average oracle accuracy, most agents can accommodate slower update intervals, provided the reported prices remain within a predictable range that is consistent with their expectations and reflected in their risk parameters. If the actual accuracy corridor is wider than anticipated, however, this may result in outcomes where a protocol no longer functions as intended. Accordingly, this study focuses on evaluating the performance of CPFs primarily in terms of their price accuracy.

2.3. Oracle accuracy and capital efficiency

Asset price oracles play a crucial role in many DeFi applications, primarily in collateral valuation. The accuracy of these oracles affects the risk parameter choices of dependent protocols and their users, directly influencing capital efficiency. This section provides a stylized formalization of this relationship, and serves as a motivation for the study.

Recall that the permissionless and pseudonymous nature of DeFi protocols preclude reputation-based lending. As a result, collateral is a fundamental tool in the economic design of these protocols. Any liability l of a position must be covered by collateral c, which is locked within the protocol. Generally, protocols enforce a minimum equity requirement $\lambda \in (0, 1)$ and rely on oracles to provide a relative price p_o for the collateral asset in terms of the debt asset, in order to monitor the relative value of both c and l and determine a position's *health*:

$$\frac{l}{c \cdot p_o} \leq (1 - \lambda) \tag{1}$$

Here, λ represents the maintenance margin in derivative protocols or the complement of the maximum loan to value in lending applications. If a position does not meet this health condition, its liabilities cannot be increased and it becomes open for liquidation. In such cases, the collateral is sold to cover the outstanding liabilities, typically at a discounted price $p_o \cdot (1-k)$, where $k \in [0,1)$ serves as an incentive for liquidators.

Thus, the accuracy of an asset price oracle is vital for the integrity of financial risk processes in dependent protocols. First, accurate oracle prices ensure timely identification of unhealthy positions. Second, they enable correct pricing of collateral in liquidations. Deviations in p_o can result in liabilities remaining uncovered or the collateral of healthy positions being sold below value. The former constitutes bad debt for the protocol, potentially leading to default on user deposits if not absorbed by protocol reserves. The latter primarily affects the equity of the position owner, with losses from underpriced collateral sales further exacerbated by any $k > 0$.

Ideally, the oracle price would always be in perfect sync with the underlying ground truth or market price and unhealthy positions would be liquidated (or topped up) immediately. However, some deviation is inevitable since the ground truth is moving quasi-continuously and the blockchain follows a discrete update interval. As a result, p_o will likely diverge from the ground truth price p_g between updates. This deviation can be expressed as:

$$\upsilon = \frac{p_o}{p_g} - 1 \tag{2}$$

Note that there are two distinct cases to consider: When $\upsilon < 0$, the collateral is undervalued relative to the position's liabilities and when $\upsilon > 0$, the asset price oracle overvalues the collateral. Assuming the oracle reports truthfully, we expect $p_o \approx p_g$ with each update and consequently $\upsilon \approx 0$. This allows us to approximate the relative update size by $\frac{p_o'}{p_o} \approx \frac{1}{1+\upsilon}$.

Next, we assume that agents, including protocols and users, have formed expectations regarding the distribution of υ. Based on this distribution and confidence levels determined by their risk preferences, agents derive threshold values $\tilde{\upsilon}_{min}$ and $\tilde{\upsilon}_{max}$. Intuitively, these thresholds represent the range of oracle price deviations that agents account for in setting their risk parameters to mitigate potential financial losses due to collateral under- or overvaluation by the oracle. Through these risk parameters, $\tilde{\upsilon}$ directly affects the capital efficiency of depending protocols.

Protocols consider $\tilde{\upsilon}_{max}$ in choosing λ in order to avoid losses from delayed liquidation of unhealthy positions. Specifically, the minimum equity requirement must prevent situations where positions transition from healthy to being liquidated for less than their liabilities in a single oracle update. Based on our assumptions as well as Eqs. (1) and (2), this can be formalized as:

$$\lambda \geq 1 - \frac{1-k}{1+\tilde{\upsilon}_{max}} \tag{3}$$

With $(1-k) > 0$, the derivative of inequality (3) confirms a strictly positive relationship between λ and $\tilde{\upsilon}_{max}$ for $\tilde{\upsilon}_{max} > -1$. Hence, increasing asset price oracle deviations, i.e., $\Delta\tilde{\upsilon}_{max} > 0$, lead to higher minimum equity margins, reducing capital efficiency.

$$\frac{\partial\lambda}{\partial\tilde{\upsilon}_{max}} \geq \frac{1-k}{(1+\tilde{\upsilon}_{max})^2}$$

Additionally, users may further increase their collateral in order to avoid equity loss due to undue liquidations. They account for possible undervaluation of their collateral in form of $\tilde{\upsilon}_{min}$ and maintain collateral in excess of the protocol's minimum equity requirements. Based on Eqs. (1) and (2), the impact on rational users' choice of collateralization ratio $\varphi = \frac{c \cdot p_g}{l}$ can be formalized as:

$$\varphi \geq \frac{1}{(1-\lambda)\cdot(1+\tilde{\upsilon}_{min})} \tag{4}$$

With $(1-\lambda) > 0$, the derivative of inequality (4) confirms a strictly negative relationship between φ and $\tilde{\upsilon}_{min}$ for $\tilde{\upsilon}_{min} > -1$. Therefore, an increase in asset price oracle deviations, i.e., $\Delta\tilde{\upsilon}_{min} < 0$, implies increasing collateral levels chosen by users, further reducing capital efficiency of protocol usage.

$$\frac{\partial\varphi}{\partial\tilde{\upsilon}_{min}} \leq -\frac{1}{(1-\lambda)\cdot(1+\tilde{\upsilon}_{min})^2}$$

In summary, asset price oracle accuracy affects capital efficiency by shaping the risk parameter choices of both protocols and users. First, protocols take $\tilde{\upsilon}_{max}$ into account when specifying λ in order to mitigate the risk of adverse conditions following periods of collateral overvaluation by p_o. Second, rational users will bolster collateral per unit of liability relative to $\tilde{\upsilon}_{min}$ to avoid undue liquidation losses in periods of collateral undervaluation by p_o.

The stylized framework developed above underscores how oracle price deviations shape the economic trade-off between risk management and capital efficiency in decentralized lending markets. By affecting both protocol-level parameters and user-level collateralization strategies, oracle accuracy emerges as a key determinant of system dynamics. These theoretical considerations motivate our empirical investigation, which aims to (i) quantify the drivers and magnitude of oracle price deviations, (ii) analyze the mechanisms through which these deviations are corrected, and (iii) evaluate how oracle accuracy parameters influence user risk-taking and capital efficiency. The following section formalizes these ideas as our central research questions.

3. Research questions

We address three central research questions, each linked to a set of testable hypotheses. Research Question 1 (RQ1) investigates the factors that drive deviations between Chainlink Price Feeds (CPFs) and benchmark market prices. Research Question 2 (RQ2) focuses on instances where deviations fall outside the target corridor, that is, beyond the bounds implied by the oracle's stated accuracy parameters. Conditional on such violations, we analyze the recovery dynamics, including the likelihood of returning to the acceptable range and the associated implications for price behavior and oracle responsiveness. Research Question 3 (RQ3) examines the extent to which oracle accuracy configurations influence user behavior, with particular attention to collateralization decisions and capital efficiency within DeFi lending protocols.

The three research questions, together with their corresponding hypotheses, are outlined in detail below.

RQ1: What are the driving factors of price deviation?

The CPF trigger criteria, specifically the heartbeat and the target corridor, should have clear effects on how well the CPF tracks the underlying ground truth price. Hence, we formalize the following two hypotheses:

H1.1 Heartbeat: *A faster heartbeat increases the average update frequency of the price feed and thereby decreases the deviation.*

H1.2 Target Corridor: *A wider target corridor decreases the responsiveness to small price changes and therefore increases deviation.*

Each CPF update consists of individual price observations from each participating reporter. These individual observations are recorded and aggregated on-chain with each CPF update. The set of authorized reporter addresses in the aggregator contract configuration restricts who can submit price observations in a report. While this set can be changed, it is usually constant for long periods of time. The participation rate among the authorized set of reporters and the variance of their submitted price observations could help explain deviation in the CPFs. We acknowledge that reporter uncertainty might be correlated with market volatility variables and will discuss this in the results Section 5.1. Thus, we formalize the following two hypotheses:

H1.3a Reporter Participation: *The number of participating reporters affects the deviation.*

The direction is unclear, as there might be two competing sub-effects at work. On the one hand, a higher participation may decrease timeliness of the report and thereby increase deviation. On the other hand, a higher participation may increase the price feeds robustness to outliers.

H1.3b Reporter Uncertainty: *The uncertainty among the reporters increases deviation.*

Congestion on a blockchain network occurs when there is an increased demand for limited block space. This can lead to slower confirmation times and higher transaction fees. It is very likely that this interferes with timely CPF updates. We formalize this in the following hypothesis:

H1.4 Congestion: *A higher network congestion can delay oracle updates and therefore increases deviation.*

Market dynamics likely also have an effect on CPF deviation. We expect recent market volatility to be an indicator for current uncertainty and unpredictability in price trends, making it more difficult for the CPF to be accurate. Further, Epps and Epps (1976) find evidence of stochastic dependence between trading volume and price changes. The direction of this effect, however, is unclear for our analysis of CPF deviation. For these reasons we opt to treat trading volume as a control variable and formalize the following hypothesis for market volatility:

H1.5 Market Volatility: *A higher past volatility leads to less robust price signals and therefore increases present deviation.*

Lastly, we expect a strong autoregressive component in our model. Past deviations are likely to be carried forward and serve as an indicator for high current deviations. However, if the CPF updates were perfectly timely, a past deviation that exceeds the target corridor would always lead to a decrease in the current deviation through a CPF update. To better capture this dynamic and explore which of the two competing effects is dominant, we use a dummy variable that indicates whether there was a target corridor violation in the last block. The hypothesis is formalized as follows:

H1.6 TC Violation: *A past target corridor violation affects the present deviation.*

The direction is unclear, as there might be two competing sub-effects at work. On the one hand, there is an autoregressive component where part of the deviation might be carried forward and thereby increase the deviation. On the other hand, there is an urgency component, where the violation itself should trigger an update and therefore decrease the deviation.

Fig. 2 summarizes the hypotheses and the anticipated effects on deviations. The strategies employed to address potential endogeneity concerns will be detailed in Section 6. In addition, we will examine the potential correlations among the independent variables.

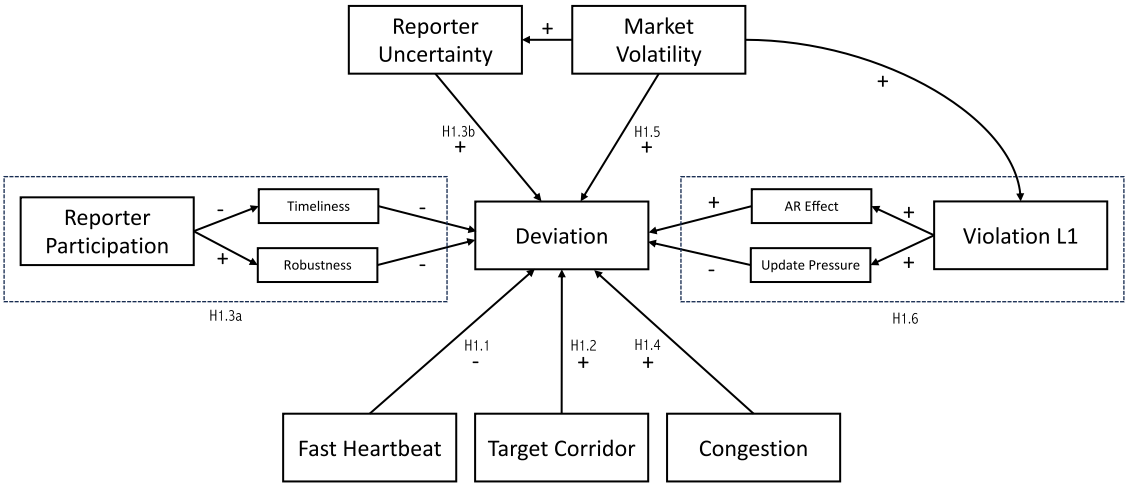


Fig. 2. Conceptual framework of the hypothesized determinants of oracle price deviations. The diagram summarizes the expected direction of effects for six hypotheses: heartbeat frequency (H1.1), target corridor width (H1.2), reporter participation (H1.3a), reporter uncertainty (H1.3b), network congestion (H1.4), market volatility (H1.5), and lagged violations (H1.6). Positive and negative signs indicate whether the factor is expected to increase or decrease deviations from the ground truth price.

RQ2: How does an oracle recover after a target corridor violation?

This research question is a follow-up to **H1.6** and explores the transitions between a state where the oracle price is outside the target corridor (target corridor violation) and a state where the oracle price is within the target corridor. We investigate whether the duration and magnitude of the violation has an impact on the transition probability and transition path. The hypotheses can be formalized as follows:

H2.1: *The duration of the violation has a significant effect on the state transition path.*

Specifically, the likelihood of returning to the target corridor through a CPF update rather than random price movement increases with larger consecutive violation sequences.

H2.2: *The magnitude of the violation has a significant effect on state transition path.*

In particular, the probability to transition back to a state where reported prices are within the target corridor via a CPF update relative to a random price movement increases with higher magnitudes of violations.

RQ3: How does the quality of asset price oracles influence users' collateralization decisions?

This research question examines the practical implications of asset price oracle accuracy on capital efficiency. In particular, we explore the excess capitalization levels chosen by users, i.e., position collateralization beyond the minimum requirements imposed by protocol risk parameters, dependent on the underlying asset price oracle. The relationship between oracle accuracy and position capitalization choice of rational users described in Section 2 in combination with **H1.1**, **H1.2** and **H1.5** are the basis for the following hypotheses:

H3.1 Heartbeat: *Users may choose lower capitalization levels for positions depending on oracles with faster heartbeats.*

According to H.1.1, faster heartbeats lower the maximum time between oracle updates and are hence expected to decrease average oracle deviations as well as the probability of extreme deviations. Depending upon the proportion of updates triggered by the heartbeat, the negative impact on capitalization levels may be more or less significant.

H3.2 Target Corridor: *Users choose higher capitalization levels for positions depending upon oracles with wider target corridors.*

H1.2 presumes that a wider target corridor increases the range of oracle deviations. Consequently, we expect users to increase their capitalization levels to account for higher maximum deviations.

H3.3 User Sophistication: *More sophisticated users choose more aggressive capitalization levels for their position.*

Larger positions are expected to be monitored more closely and adjusted more frequently, due to lower relative adjustment costs.[2] In consequence, these positions are expected to have a lower excess capitalization.

H3.4 Volatility: *Price volatility has a significant positive effect on position capitalization choices.*

Debt and collateral asset combinations with high ground truth price volatility are expected to have higher average oracle deviations and increased probability of extreme deviations. Users are expected to account for this with increased excess capitalization of their positions.

4. Data

To test our hypotheses, we compiled two novel datasets. The first dataset is used in our analyses of **RQ1** and **RQ2**, capturing the on-chain performance of 40 Chainlink Price Feeds (CPFs) at block-level granularity over an 18-month period (March 2022 to August 2023). The resulting 150 million observations allow for a detailed examination of price deviations and update behavior across CPFs. The second dataset is used in our analysis of **RQ3** and is constructed from position-level data of the Aave V2 Main Market (Aave), the leading lending protocol on the Ethereum network that relies on CPFs. It comprises 146,620 capitalization choices across 15,107 positions and 31 CPFs over a 30-month period (January 2021 to June 2023). This dataset provides a well-defined setting to analyze how oracle design parameters influence user decisions regarding the composition and risk profile of borrowing positions. Together, these datasets not only underpin our empirical analyses but also constitute a resource that may be useful for future research in this area. In what follows, we describe the sample construction and data collection in more detail.

4.1. Chainlink Price Feed data

The construction of the dataset begins with all 88 CPFs for Crypto Asset/USD pairs on the Ethereum blockchain listed in the Chainlink registry as of August 28, 2023. To include a broad range of crypto assets while avoiding selection biases, we restricted the sample to 40 CPFs that (i) have verified status according to the Chainlink registry, (ii) have been continuously available throughout our observation period, and (iii) quote assets that are non-stablecoins, and (iv) are available as tokens on Ethereum, either natively or in a widely accepted bridged form such as BTC or DOGE. Table 1 lists the 40 CPFs included in this study along with their trigger criteria.

For each CPF in scope, we obtained the Ethereum address of the static proxy contract and the trigger criteria (heartbeat and target corridor) from the Chainlink registry, as these parameters represent key accuracy signals of a CPF. To collect on-chain CPF data, we analyzed the source code of deployed smart contract instances and developed a tailored data extraction approach. Using a self-hosted Erigon archive node complemented by the Infura API, we reconstructed the history of aggregator contracts and their configurations for each CPF over the observation period. From the transaction history of these contracts, we extracted all successful price updates and the corresponding transmission reports, which include the participating reporters and their individual price inputs. In the rare cases where multiple updates occurred within the same block, we retained only the final update. This dataset enables us to trace oracle price developments for each CPF on a block-by-block basis and analyze the evolution of reporter behavior over time.

As a proxy for ground truth asset prices, we use historical price data from *Binance*, obtained via its public API. Binance is the largest centralized crypto asset exchange by trading volume during our observation period, making it a suitable reference source. We collected high-frequency price and volume data at 1-s intervals for each asset tracked by the CPFs in scope, with prices denominated in TetherUSD (USDT), Binance's primary quote asset. These data allow us to assess the price deviations of each CPF relative to the ground truth, identify target corridor violations, and compute asset-specific measures of market volatility at the block level. For conversions between USD and USDT, we rely on the USDT/USD CPF.

We complement our dataset with block-level statistics collected from our Erigon archive node. In particular, we record the median gas price per block as an indicator of block space demand and, by extension, as a proxy for network congestion. Additionally, we collect block timestamps, which are necessary for re-sampling and aligning the ground truth price and market volatility data from Binance with the corresponding block intervals.

While the compiled dataset represents one of the most comprehensive block-level oracle price panels to date, we acknowledge certain limitations. First, using Binance data as the basis for ground truth prices and market volatility measures captures only a fraction of overall market activity and may be affected by wash trading (see, for example, Cong et al., 2023) or other forms of market manipulation. Nevertheless, Binance was the largest centralized crypto asset exchange during our sample period in terms of trading volume, liquidity, and asset coverage, making it the most viable and consistent data source. Second, regarding reporter behavior, our analysis focuses exclusively on observable on-chain oracle activity. We cannot capture details about reporters' underlying data sources or the coordination processes within the CPF off-chain reporting network, which may influence both update timing and content. To the best of our knowledge, such information is not publicly available. Third, for each block, we retain only the final valid CPF transmission. While this approach ensures consistency and avoids double-counting multiple updates within the same block, it may obscure more granular intra-block dynamics. Finally, we assume that the target corridor parameters for each CPF remain constant throughout the sample period. An analysis of the update behavior of each CPF, analogous to Fig. 1, did not reveal any regime shifts that would indicate changes to the trigger criteria.

Table 2 presents summary statistics for the dependent and independent variables used in the empirical analyses addressing **RQ1** and **RQ2**. Additional details on the construction of these variables from the collected data are provided in Appendix A.1.

[2] Transaction fees for position updates, i.e., adjustment of collateral or repayment of debt, are volume independent.

Table 1
Overview of the 40 Chainlink Price Feeds (CPFs) included in the study. The table lists each asset/USD pair, heartbeat parameter, target corridor width, and corresponding Ethereum proxy contract address. All CPFs are verified according to the Chainlink registry and were continuously active throughout the observation period (March 2022–August 2023).

Asset	Pair	Target corridor (bp)	Heartbeat (s)	Price feed proxy address (Ethereum)
0x	ZRX/USD	100	3600	0x2885d15b8Af22648b98B122b22FDF4D2a56c6023
1inch	1INCH/USD	100	86 400	0xc929ad75B72593967DE83E7F7Cda0493458261D9
Aave	AAVE/USD	100	3600	0x547a514d5e3769680Ce22B2361c10Ea13619e8a9
Adex	ADX/USD	200	86 400	0x231e764B44b2C1b7Ca171fa8021A24ed520Cde10
Ankr	ANKR/USD	200	86 400	0x7eed379bf00005CfeD29feD4009669dE9Bcc21ce
Avalanche	AVAX/USD	200	86 400	0xFF3EEb22B5E3dE6e705b44749C2559d704923FD7
Badger DAO	BADGER/USD	200	86 400	0x66a47b7206130e6FF64854EF0E1EDfa237E65339
Balancer	BAL/USD	200	86 400	0xdF2917806E30300537aEB49A7663062F4d1F2b5F
Bancor	BNT/USD	500	86 400	0x1E6cF0D433de4FE882A437ABC654F58E1e78548c
Bitcoin	BTC/USD	50	3600	0xF4030086522a5bEEa4988F8cA5B36dbC97BeE88c
BNB	BNB/USD	100	86 400	0x14e613AC84a31f709eadbdF89C6CC390fDc9540A
Chainlink	LINK/USD	100	3600	0x2c1d072e956AFFC0D435Cb7AC38EF18d24d9127c
Compound	COMP/USD	100	3600	0xdbd020CAeF83eFd542f4De03e3cF0C28A4428bd5
Convex finance	CVX/USD	200	86 400	0xd962fC30A72A84cE50161031391756Bf2876Af5D
Curve DAO	CRV/USD	100	86 400	0xCd627aA160A6fA45Eb793D19Ef54f5062F20f33f
Decentraland	MANA/USD	200	86 400	0x56a4857acbcfe3a66965c251628B1c9f1c408C19
Dogecoin	DOGE/USD	50	86 400	0x2465CefD3b488BE410b941b1d4b2767088e2A028
Enjin coin	ENJ/USD	200	86 400	0x23905C55dC11D609D5d11Dc604905779545De9a7
Ethereum	ETH/USD	50	3600	0x5f4eC3Df9cbd43714FE2740f5E3616155c5b8419
Ethereum name service	ENS/USD	200	86 400	0x5C00128d4d1c2F4f652C267d7bcdD7aC99C16E16
ForTube	FOR/USD	200	86 400	0x456834f736094Fb0AAD40a9BBc9D4a0f37818A54
Frax share	FXS/USD	200	86 400	0x6Ebc52C8C1089be9eB3945C4350B68B8E4C2233f
Kyber network	KNC/USD	100	86 400	0xf8fF43E991A81e6eC886a3D281A2C6cC19aE70Fc
Maker	MKR/USD	100	3600	0xec1D1B3b0443256cc3860e24a46F108e699484Aa
Omisego	OMG/USD	200	86 400	0x7D476f061F8212A8C9317D5784e72B4212436E93
Orchid protocol	OXT/USD	500	86 400	0xd75AAaE4AF0c398ca13e2667Be57AF2ccA8B5de6
Perpetual protocol	PERP/USD	200	86 400	0x01cE1210Fe8153500F60f7131d63239373D7E26C
Polygon (MATIC)	MATIC/USD	100	3600	0x7bAC85A8a13A4BcD8abb3eB7d6b4d632c5a57676
Reserve rights	RSR/USD	200	86 400	0x759bBC1be8F90eE6457C44abc7d443842a976d02
Solana	SOL/USD	200	86 400	0x4ffC43a60e009B551865A93d232E33Fce9f01507
Spell token	SPELL/USD	100	86 400	0x8c110B94C5f1d347fAcF5E1E938AB2db60E3c9a8
Sushi	SUSHI/USD	100	3600	0xCc70F09A6CC17553b2E31954cD36E4A2d89501f7
Swipe	SXP/USD	100	86 400	0xFb0CfD6c19e25DB4a08D8a204a387cEa48Cc138f
Synthetix network	SNX/USD	100	86 400	0xDC3EA94CD0AC27d9A86C180091e7f78C683d3699
The graph	GRT/USD	200	86 400	0x86cF33a451dE9dc61a2862FD94FF4ad4Bd65A5d2
The sandbox	SAND/USD	200	86 400	0x35E3f7E558C04cE7eEE1629258EcbbA03B36Ec56
Tomochain	TOMO/USD	200	86 400	0x3d44925a8E9F9DFd90390E58e92Ec16c996A331b
Truefi	TRU/USD	200	86 400	0x26929b85fE284EeAB939831002e1928183a10fb1
Uniswap	UNI/USD	100	3600	0x553303d460EE0afB37EdFf9bE42922D8FF63220e
Yearn finance	YFI/USD	100	3600	0xA027702dbb89fbd58938e4324ac03B58d812b0E1

Table 2
Summary statistics of dependent and independent variables used in the empirical analyses addressing RQ1 (determinants of deviations) and RQ2 (violation recovery); constructed from block-level oracle data and Binance market data. The dependent Variable is Deviation, which measures the absolute proportional deviation of the Chainlink oracle price from the Binance ground truth price in basis points. Independent variables include CPF deviations, target corridor width, heartbeat frequency, reporter participation and uncertainty, gas prices, market volatility measures, and trading volume. The sample consists of 40 CPFs observed over 3,740,230 blocks between March 2022 and August 2023.

Variable	Mean	SD	Min	Median	Max
Deviation (bp)	56.5156	55.6970	0	40.3347	6596.4308
Fast Heartbeat	0.2750	0.4465	0	0	1
Target Corridor (bp)	163.7500	95.3990	50	200	500
Binance USDT Volume (m)	0.0194	0.1841	0	<0.0001	99.1221
Reporter Participation (%)	0.9959	0.0178	0.6842	1	1
Reporter Uncertainty	0.0002	0.0078	0	<0.0001	3.1060
Median Price per Gas	30.7813	92.5395	0	21.7940	10 027.8910
Target Corridor Violation (t-1)	0.0248	0.1556	0	0	1.0000
Binance Move (t-2)	2.7615	5.8546	0	0.4585	4328.2721
Binance Move (t-24)	19.1677	29.2143	0	11.9685	21 519.0031
Over Deviation (bp, t-1)	0.7547	9.9179	0	0	6396.4308
Violation Sequence Length	0.2515	3.2431	0	0	700
Post-Merge	0.6686	0.4707	0	1	1

Note: All variables have the same number of observations ($40 \cdot 3{,}740{,}230 = 149{,}609{,}200$).

4.2. Capitalization choice data

The construction of our second dataset begins with the 37 assets that were valid reserve assets in Aave during the observation period. To maintain a broad yet reliable sample, we exclude assets that (i) were not directly integrated into the Aave oracle system through an Asset/ETH CPF but instead relied on meta-layers combining multiple CPFs or other third-party sources, or (ii) experienced major adverse events during the observation window. The remaining 31 assets included in our dataset are listed in Table 3, together with information on the corresponding CPF, as obtained from the Chainlink registry.

Table 3
List of Aave V2 Main Market reserve assets included in the capitalization-choice analysis (RQ3). The table reports each asset, its category (crypto asset or stablecoin), the underlying Chainlink Price Feed (CPF) used by Aave for pricing, and the corresponding trigger criteria (target corridor in basis points and heartbeat interval in seconds). Assets were selected based on continuous availability and direct integration into Aave's oracle system during the observation period (January 2021–June 2023).

Asset	Category	Underlying CPF	Target corridor (bp)	Heartbeat (s)
0x	Crypto	ZRX/ETH	200	86 400
1inch	Crypto	1INCH/ETH	200	86 400
Aave	Crypto	AAVE/ETH	200	86 400
Ampleforth	Stable	AMPL/ETH	200	86 400
Balancer	Crypto	BAL/ETH	200	86 400
Basic attention token	Crypto	BAT/ETH	200	86 400
Binance USD	Stable	BUSD/ETH	100	86 400
Wrapped bitcoin	Crypto	BTC/ETH	200	86 400
Curve DAO	Crypto	CRV/ETH	200	86 400
Convex finance	Crypto	CVX/ETH	200	86 400
Dai	Stable	DAI/ETH	100	86 400
DeFi pulse index	Crypto	DPI/ETH	200	86 400
Enjin coin	Crypto	ENJ/ETH	200	86 400
Wrapped ether	Crypto	n/a	–	–
Fei USD	Stable	FEI/ETH	200	86 400
Filecoin	Crypto	FIL/ETH	200	86 400
Frax dollar	Stable	FRAX/ETH	200	86 400
Kyber network crystal	Crypto	KNC/ETH	200	86 400
ChainLink token	Crypto	LINK/ETH	100	21 600
Decentraland	Crypto	MANA/ETH	200	86 400
Maker	Crypto	MKR/ETH	100	86 400
RAI reflex index	Stable	RAI/ETH	200	86 400
Republic token	Crypto	REN/ETH	200	86 400
Synthetix network	Crypto	SNX/ETH	200	86 400
Synthetix USD	Stable	SUSD/ETH	100	86 400
TrueUSD	Stable	TUSD/ETH	100	86 400
Uniswap	Crypto	UNI/ETH	200	86 400
USD coin	Stable	USDC/ETH	100	86 400
Pax dollar	Stable	USDP/ETH	200	86 400
Tether USD	Stable	USDT/ETH	100	86 400
Yearn finance	Crypto	YFI/ETH	100	86 400

Based on an analysis of Aave smart contracts, we collect blockchain data to reconstruct the complete state of each borrowing position active in Aave. Each time collateral or debt is adjusted, we record a new observation for that position. This event-based structure allows us to track changes in capitalization and risk preferences over time, rather than relying on static snapshots. We restrict the sample to economically meaningful and analytically valid observations. First, we retain only active positions, i.e., those with non-zero collateral and debt balances, and exclude positions where the dominant collateral or debt asset accounts for less than 95% of the total. This ensures that asset-specific metadata, such as volatility measures, oracle parameters, and asset category can be consistently attributed to a single asset on each side of the position. We also exclude collateralization updates that are triggered by liquidations, involve identical assets on both sides (e.g., ETH collateral and ETH debt), or result in debt of less than 0.001 ETH. These steps ensure that our analysis reflects user-driven preferences and focuses on genuine borrowing activity.

We restrict the observation period to exclude the final weeks of the protocol's phase-out in favor of Aave v3. To capture changes in risk preferences over time, we retain only positions that were updated at least once after being opened during this period. Consequently, each position included in the sample has at least two observations.

At the core of the analysis is the user's capitalization choice, defined as the decision to supply a specific amount of collateral relative to the debt value of a position. This decision is observed through the *health factor* of a position, which is derived from the overcollateralization ratio for a given asset combination, as introduced conceptually in the theoretical framework in Section 2.3. A higher health factor indicates greater collateralization and lower liquidation risk, while a lower value implies a higher risk of liquidation.

We focus our analysis on economically relevant behavior near the margin of liquidation risk, i.e., close to a health factor of 1.0. To this end, we discard the positions in the top 10% of the health factor distribution. Extremely high health factors provide little meaningful variation, as heavily overcollateralized positions are unlikely to be actively managed. Beyond a certain level, differences

in health factor become economically insignificant and would distort our results. After applying this filter, the maximum health factor in the dataset is 4.55, corresponding to an excess collateralization of 355%.

Next, we construct a continuous measure of user sophistication using the first principal component (PC1) from a principal component analysis (PCA) of features including collateral size, debt size, and whether the position is controlled by a smart contract or an externally owned account (EOA). This continuous score serves as a proxy for differences in user capability, information access, or risk appetite. Positions are further categorized by the stablecoin/crypto composition of collateral and debt, distinguishing between Crypto-Stable, Stable-Crypto, Crypto-Crypto, and Stable-Stable configurations.

For relative market price volatility, we again rely on data from the Binance API. For each asset, we compute short-term and long-term volatility measures using hourly and daily price data, respectively. For positions involving one stablecoin and one volatile asset, volatility is measured using only the volatile component. For crypto-to-crypto positions, volatility is calculated based on relative price changes, while positions consisting exclusively of stablecoins on both sides are assigned a volatility of zero.

Table 4 reports summary statistics for the variables used in our analysis of **RQ3**. A detailed description of the variables and their construction is provided in Appendix A.2.

Table 4
Summary statistics of dependent and independent variables used in the empirical analysis of user capitalization choices on Aave V2 (RQ3). The dependent variable is the Health Factor, which measures the degree of overcollateralization relative to protocol-imposed minimums (values above 1 indicate lower liquidation risk). Independent variables include oracle design parameters (Heartbeat and Target Corridor), a CPF version indicator (Agg. V4), a user sophistication score derived from principal component analysis, position type dummies capturing the collateral-debt composition (crypto-stable, stable-crypto, crypto-crypto), and short-term (7-day) and long-term (1-year) volatility measures computed from Binance price data. The sample consists of 146,620 capitalization decisions from 15,107 unique addresses between January 2021 and June 2023.

Variable	Mean	SD	Min	Med	Max
Health Factor	1.8204	0.7564	0.0082	1.5846	4.5542
Heartbeat (h)	15.4314	5.5692	0.0000	12.0000	24.0000
Target Corridor (bp)	80.9753	41.3257	0.0000	50.0000	200.0000
Agg. V4	0.3498	0.4769	0.0000	0.0000	1.0000
Soph. Score	0.0081	1.4035	−0.2630	−0.2193	84.4294
Crypto - Stable	0.7351	0.4413	0.0000	1.0000	1.0000
Stable - Crypto	0.1315	0.3380	0.0000	0.0000	1.0000
Crypto - Crypto	0.0530	0.2240	0.0000	0.0000	1.0000
Volatility (7d)	0.0098	0.0063	0.0000	0.0087	0.0499
Volatility (1y)	0.0390	0.0158	0.0000	0.0395	0.1208
Post Merge	0.2466	0.4311	0.0000	0.0000	1.0000

Note: All variables have the same number of observations (146,620).

5. Empirical analysis and results

5.1. Deviation analysis

We can now use our data to address the first research question and test hypotheses **H1.1–H1.6**. Initially, we conduct a pooled OLS regression incorporating observable time-invariant differences among CPFs. This step provides an initial understanding of the relationship between the dependent and independent variables. We estimate the following model:

$$\begin{aligned}\texttt{Deviation (bp)}_{it} = {} & \beta_0 + \beta_1\,\texttt{Fast Heartbeat}_{it} + \beta_2\,\texttt{Target Corridor (bp)}_{it} + \beta_3\,\texttt{Reporter Participation (\%)}_{it} + \\ & \beta_4\,\texttt{Reporter Uncertainty}_{it} + \beta_5\,\texttt{Median Price per Gas}_{it} + \beta_6\,\texttt{Target Corridor Violation (t}-1)_{it} \\ & + \beta_7\,\texttt{Binance Move (t}-2)_{it} + \beta_8\,\texttt{Binance Move (t}-24)_{it} + \sum_{i=9}^{10}\beta_i\,\texttt{Control Variables}_{it} + \varepsilon_{it},\end{aligned} \tag{5}$$

where the `Control Variables` are `Post Merge` and `Binance USDT Volume (m)`. With `Post Merge`'s natural breakpoint, we control for structural differences before and after the Ethereum Merge, which was a fundamental change to the blockchain's consensus mechanism. The use of trading volume as a control variable is explained in Section 3.

We acknowledge the potential for unobserved heterogeneity among CPFs. For instance, there might be minor variations in off-chain processes or differences associated with the reporters of a specific CPF. If this unobservable heterogeneity remains constant over time, we can account for it by including individual-fixed effects for each CPF. This fixed effects regression is modeled as follows:

$$\begin{aligned}\texttt{Deviation (bp)}_{it} = {} & \beta_1\,\texttt{Reporter Participation (\%)}_{it} + \beta_2\,\texttt{Reporter Uncertainty}_{it} + \beta_3\,\texttt{Median Price per Gas}_{it} + \\ & \beta_4\,\texttt{Target Corridor Violation (t}-1)_{it} + \beta_5\,\texttt{Binance Move (t}-2)_{it} + \beta_6\,\texttt{Binance Move (t}-24)_{it} + \\ & \sum_{i=7}^{8}\beta_i\,\texttt{Control Variables}_{it} + \alpha_i + \varepsilon_{it}\end{aligned} \tag{6}$$

where α_i represents the individual (cross-sectional) fixed effects. This term accounts for any unobserved variables that are constant over time but vary across CPFs. Since this includes the CPF trigger criteria `Fast Heartbeat` and `Target Corridor (bp)`, these variables are no longer included. The remaining structure of the equation remains the same as in the pooled OLS regression equation.

The regression results are shown in Table 5. Model specifications (1) to (4) show the pooled OLS regression results and (5) to (8) show the results for the fixed effects regression. The table contains the regression coefficients and standard errors in brackets. Note that we use a heteroskedasticity and autocorrelation (HAC) consistent covariance matrix in line with Newey and West (1987). The reported standard errors have been corrected accordingly.

Table 5
Regression results analyzing the determinants of Chainlink Price Feed (CPF) deviations from benchmark market prices (RQ1). Columns (1)–(4) report pooled OLS estimates with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors following Newey and West (1987). Columns (5)–(8) report entity fixed effects regressions that account for unobserved time-invariant CPF heterogeneity. The dependent variable is deviation in basis points relative to Binance reference prices. Independent variables include heartbeat frequency, target corridor width, reporter participation and uncertainty, gas price (as a proxy for congestion), lagged target corridor violations, and short- and medium-term price changes. Control variables account for trading volume and Ethereum's Merge. Coefficients are reported with HAC-robust standard errors in parentheses.

	Dependent variable: `Deviation (BP)`							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
`Intercept`	26.121*** (0.038)	28.135*** (0.593)	11.487*** (0.583)	8.518*** (0.557)				
`Individual FE`					yes	yes	yes	yes
`Fast Heartbeat`	−12.768*** (0.020)	−12.774*** (0.020)	−11.583*** (0.020)	−11.376*** (0.019)				
`Target Corridor (bp)`	0.244*** (0.0002)	0.244*** (0.0002)	0.244*** (0.0002)	0.254*** (0.0002)				
`Reporter Participation (%)`		−2.037*** (0.595)	2.203*** (0.578)	3.366*** (0.554)		−7.361*** (0.601)	−6.421*** (0.565)	−3.441*** (0.561)
`Reporter Uncertainty`		104.775*** (4.829)	58.957*** (3.145)	65.903*** (2.880)		109.610*** (4.875)	100.461*** (4.711)	69.713*** (2.940)
`Median Price per Gas`	0.012*** (0.0002)	0.012*** (0.0002)	0.005*** (0.0002)	0.003*** (0.0001)	0.012*** (0.0002)	0.012*** (0.0002)	0.007*** (0.0001)	0.003*** (0.0001)
`Target Corridor Violation (t-1)`				87.437*** (0.127)			106.016*** (0.080)	89.065*** (0.131)
`Binance Move (t-2)`			0.470*** (0.004)	0.361*** (0.005)				0.362*** (0.005)
`Binance Move (t-24)`			0.409*** (0.003)	0.279*** (0.002)				0.272*** (0.002)
`Binance USDT Volume (m)`	3.003*** (0.031)	2.977*** (0.031)	0.908*** (0.024)	−0.853*** (0.023)	6.987*** (0.050)	7.002*** (0.050)	2.560*** (0.031)	0.082*** (0.029)
`Post Merge`	−9.619*** (0.020)	−9.589*** (0.020)	−4.994*** (0.032)	−3.657*** (0.026)	−9.600*** (0.020)	−9.553*** (0.020)	−6.229*** (0.018)	−3.646*** (0.027)
Observations	149,609,200	149,609,200	149,608,240	149,608,240	149,609,200	149,609,200	149,609,160	149,608,240
$\bar{R}^2$ (*pooled*)	0.227	0.227	0.282	0.335				
$\bar{R}^2$ (*within*)					0.010	0.010	0.121	0.150

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

We find that the *p*-values associated with all variables in both regressions are significantly lower than the 0.01 threshold. This leads us to reject the null hypothesis, which posits no relationship between the respective variables and `Deviation (bp)` exists. However, given our use of an unusually large dataset, caution is warranted when exclusively looking at *p*-values. In large datasets, even very small effects can achieve statistical significance. Consequently, we will place particular emphasis on the magnitude and robustness of the coefficients across our various model specifications. When discussing effect sizes, our focus is on evaluating the relative magnitude and consistency of the coefficients across various model specifications, rather than interpreting them as exact causal estimates.

Let us first look at the CPF update trigger criteria `Fast Heartbeat` and `Target Corridor (bp)`. Both variables have robust coefficients across all pooled OLS model specifications. The findings indicate that a heartbeat interval of 1 h, compared to a 24-h period, correlates with a decrease in deviation by roughly 12 bp. Furthermore, we estimate that a 4 bp increase of the target corridor leads to approximately 1 bp increase in deviation. For instance, CPFs with a 20 bp target corridor are estimated to have on average a 25 bp higher deviation than CPFs with a 100 bp target corridor. These are both significant effects and in line with our initial expectations. Considering the very low *p*-values, we are confident to reject H_0 for both **H1.1** and **H1.2**.

Next, we observe `Reporter Participation (%)`. We elaborated in Section 3 that there could be competing effects for **H1.3a**. An increased participation might provide better resilience to outliers, but in turn affect the timeliness of the update. While the coefficient is significant at the 0.01 level, the p-value is inconsistent across model specifications and comparatively low. The robustness of the result is further called into question by the high inconsistency of the coefficients and their tendency to change signs. All things considered, we are not confident to draw any conclusion from this association.

The coefficient of `Reporter Uncertainty` is positive and seems to decrease in model specifications where we include variables that capture market volatility. This is in line with our expectations, as `Reporter Uncertainty` is highly correlated with different metrics of market volatility. The low p-value and the persistence of this effect across model specifications that include various variables designed to capture volatility lead us to confidently reject H_0 for **H1.3b**.

The `Median Price per Gas` has a consistently positive, yet very small coefficient. Moreover, the effect further decreases with the inclusion of variables that capture market volatility. Note that a coefficient of 0.005 suggests that a gas price increase of 200 GWEI is associated with an average increase of just 1 basis point. While the p-value is very small and would technically allow us to reject H_0 for **H1.4**, the size of the coefficient suggests that the association is negligible. This is not in line with our initial expectations and we will further discuss **H1.4** in Section 7.

The coefficients for the volatility measures `Binance Move (t-2/24)` are both positive and relatively robust across various model specifications. This suggests a clear association between short-/mid-term price changes and CPF deviation from the ground truth. We are confident to reject H_0 for **H1.5**. The observed association may stem from delayed CPF updates or could be driven by general market uncertainty. This question will be further explored when we look at CPF response time to violations of the target corridor in the next section.

Lastly, the large and relatively robust coefficient for the lagged variable `Target Corridor Violation (bp, t-1)` indicates that the autoregressive component plays on average a more dominant role than the increased update pressure following a violation. The low p-value and large coefficient leads us to confidently reject H_0 for **H1.6**.

5.2. Transition probability analysis

In this section, we examine our second research question alongside hypotheses **H2.1** and **H2.2.** Initially, we model our transition probabilities as a Markov chain and conduct a numerical analysis of the state transitions. Subsequently, we extend the model to accommodate multiple, distinct transition paths for recovering from a target corridor violation. Lastly, we employ a Multinomial Logistic Regression to empirically test the stated hypotheses.

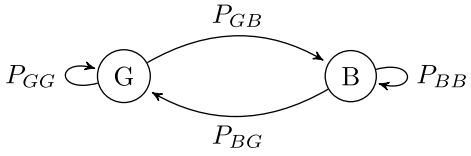


Fig. 3. Markov chain model for Chainlink Price Feed (CPF) deviations relative to the target corridor. The diagram distinguishes between the "Good" state (G), where oracle-reported prices remain within the corridor, and the "Bad" state (B), where the deviation exceeds the threshold. Transition probabilities between states (P_{GG}, P_{GB}, P_{BG}, P_{BB}) represent the likelihood of remaining in or moving between states in consecutive blocks.

We can model a CPF's performance as a Markov chain with two distinct states: the "Bad" state (B), defined as the state where the oracle is in violation, i.e., outside its designated target corridor, and the "Good" state (G), where the CPF price is within its target corridor. This is illustrated in Fig. 3. The transition probability from G to B (G → B) is denoted by P_{GB}, indicating the probability of a new target corridor violation. In contrast, P_{BG} quantifies the probability of transitioning from B to G. Finally, the self-transition probabilities, P_{GG} and P_{BB}, correspond to the probabilities of the oracle maintaining its current state, either within or outside the target corridor. The probabilities P_{GG} and P_{GB} as well as P_{BG} and P_{BB} each sum up to 1 respectively.

The transition matrix P for the Markov chain is given by:

$$P = \begin{bmatrix} P_{GG} & P_{GB} \\ P_{BG} & P_{BB} \end{bmatrix}.$$

We can observe these probabilities across our data for 40 CPFs over 18 months, resulting in the following transition matrix:

$$P = \begin{bmatrix} P_{GG} & P_{GB} \\ P_{BG} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0.9965 & 0.0035 \\ 0.1360 & 0.8640 \end{bmatrix}.$$

When a CPF is in state B, there are two distinct ways to transition back to state G: either via a CPF update (c) or via random ground truth price movement (r). The transition probabilities for these paths are denoted as P^c_{BG} and P^r_{BG}, respectively. The overall transition probability from B to G is the sum of these probabilities, expressed as $P_{BG} = P^c_{BG} + P^r_{BG}$. This extended model is illustrated in Fig. 4.

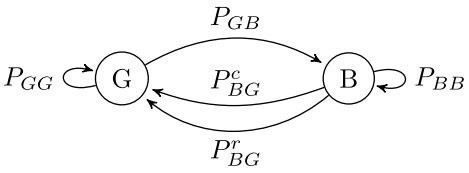


Fig. 4. Extended Markov chain representation of Chainlink Price Feed (CPF) target corridor violations and recovery paths (RQ2). The "Good" state (G) indicates that the reported price lies within the target corridor, while the "Bad" state (B) indicates a deviation beyond the threshold. From B, we distinguish between two transition paths back to G: recovery via a CPF update (P^c_{BG}) or recovery through random ground truth price movements that realign reported and benchmark prices (P^r_{BG}).

We can again infer the probabilities by observing our data to arrive at the following, extended transition matrix:

$$P' = \begin{bmatrix} P_{GG} & P_{GB} \\ P^c_{BG} + P^r_{BG} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0.9965 & 0.0035 \\ 0.0497 + 0.0863 & 0.8640 \end{bmatrix}.$$

At first glance, we see that more B → G transitions seem to be caused by random price movements instead of CPF updates. This is also visualized in Fig. 5(a). However, by looking at Fig. 5(b), which only contains violations with an over deviation of 2 basis points or more, we see that significantly less of these transitions are caused by random price movements. Further examination of even higher levels of over deviation in Fig. 5(c) and Fig. 5(d) reveals that an increase in over deviation seems to correlate with a greater likelihood of transitioning through a CPF update.

To more accurately estimate the effects of over deviation and violation sequence length on the transition probabilities P^c_{BG} and P^r_{BG}, we estimate a Multinomial Logistic Regression model as follows, where $j \in \{c, r\}$ represents the two transition paths.

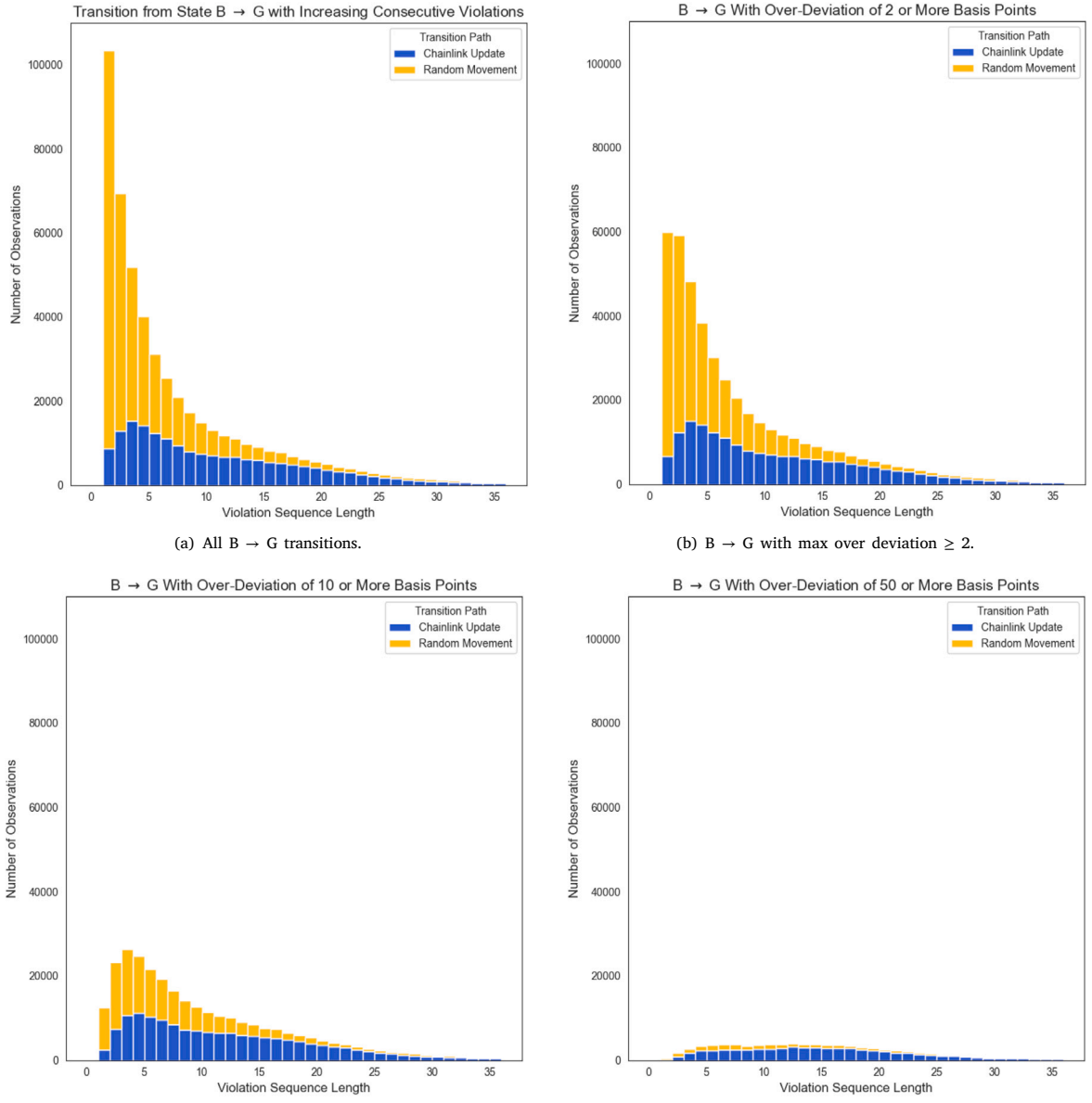
$$\ln\left(\frac{P^j_{BG}}{P_{BB}}\right) = \beta_{0j} + \beta_{1j}\ \texttt{Over Deviation}\ (\texttt{bp}, \texttt{t}-1) + \beta_{2j}\ \texttt{Violation Sequence Length} + \sum_{i=3}^{8} \beta_{ij}\ \texttt{Control Variables} \tag{7}$$

Note that each transition path has its distinct coefficients, representing the marginal effects on the conditional state transition probability P^c_{BG} or P^r_{BG}, compared to the reference group. The analysis results are summarized in Table 6, showing regression coefficients with standard errors in brackets.

Table 6
Multinomial logistic regression results for state transitions from a target corridor violation. The table reports coefficients for the likelihood of recovering via a CPF update or through random ground truth price movement (relative to staying in violation). Independent variables include lagged over-deviation, violation sequence length, heartbeat frequency, target corridor width, gas prices, and market volatility measures. Robust standard errors are reported in parentheses.

	Multinomial logistic regression for state transitions from violation (Reference = Stay)							
	Model Spec. 1:		Model Spec. 2:		Model Spec. 3:		Model Spec. 4:	
	CPF	Random	CPF	Random	CPF	Random	CPF	Random
Over Deviation (bp, t-1)	0.004***	−0.059***	0.004***	−0.059***	0.005***	−0.059***	0.008***	−0.066***
	(0.00003)	(0.0002)	(0.00003)	(0.0002)	(0.00003)	(0.0002)	(0.0001)	(0.0002)
Violation Sequence Length	0.004***	−0.007***	0.004***	−0.007***	0.004***	−0.007***	0.003***	−0.002***
	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
Fast Heartbeat			0.168***	0.028***	0.157***	0.038***	0.161***	0.030***
			(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
Target Corridor (bp)			−0.003***	0.0001***	−0.003***	0.00001	−0.003***	−0.001***
			(0.00005)	(0.00003)	(0.00005)	(0.00003)	(0.00005)	(0.00004)
Median Price per Gas					−0.0003***	0.0002***	−0.0003***	0.0001***
					(0.00002)	(0.00001)	(0.00002)	(0.00001)
Binance Move (t-2)							−0.007***	0.007***
							(0.0002)	(0.0002)
Binance Move (t-24)							−0.002***	0.004***
							(0.00004)	(0.00003)
Binance USDT Volume (m)					0.095***	−0.092***	0.104***	−0.114***
					(0.002)	(0.005)	(0.002)	(0.005)
Constant	−3.089***	−1.648***	−2.774***	−1.669***	−2.790***	−1.661***	−2.729***	−1.818***
	(0.003)	(0.003)	(0.007)	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)
Akaike Inf. Crit.	3,627,535	3,627,535	3,618,053	3,618,053	3,615,182	3,615,182	3,589,921	3,589,921

Note: *p < 0.1; **p < 0.05; ***p < 0.01.



(a) All B → G transitions.

(b) B → G with max over deviation ≥ 2.

(c) B → G with max over deviation ≥ 10.

(d) B → G with max over deviation ≥ 50.

Fig. 5. Transition paths from the "Bad" state (B) to the "Good" state (G) following a target corridor violation. Panels (a)–(d) show subsamples conditioned on the maximum magnitude of the deviation past the target corridor ("over deviation"): all B → G transitions, over deviations ≥2 basis points, ≥10 basis points, and ≥50 basis points. Each panel differentiates between two distinct recovery mechanisms: resolution through a CPF update or resolution through random ground truth price movements that realign reported and benchmark prices.

Similar to the previous regressions, we find that the p-values associated with all variables are significantly lower than the 0.01 threshold. This leads us to reject the null hypothesis, which posits no relationship between the respective variables and the transition probabilities exists.

In addition to our two main independent variables, we include various other independent variables in several model specifications as a way to control for the effects of CPF configurations and market conditions and to assess the robustness of the main coefficients.

Let us begin by examining the `Violation Sequence Length` coefficients with respect to the different transition paths. We observe a robust positive coefficient for P^c_{BG} and a robust negative coefficient for P^r_{BG} across the first three model specifications. In

the fourth model specification, where we also control for ground truth price volatility, the coefficients are reduced in magnitude, but the sign remains robust and the p-value is very low for all model specifications. When translating the coefficients to relative risk ratios, the result suggests that the longer we stay in the violation state B, the more likely it is to transition via a CPF update and the less likely it is to transition via random price movements. We are therefore confident to reject H_0 for **H2.1**.

The `Over Deviation (bp, t-1)` coefficients in regards to the different transition paths are robust and show a positive coefficient for P^c_{BG} as well as a negative coefficient for P^r_{BG} across all model specifications. When translating the coefficients to relative risk ratios, the result suggests that the more severe the violation, the more likely it is to transition via a CPF update and the less likely it is to transition via random price movements. The low p-value and large relative risk ratios lead us to confidently reject H_0 for **H2.2**.

5.3. Capitalization choice analysis

To examine how oracle characteristics influence user behavior in decentralized lending, we test hypotheses **H3.1** to **H3.4** using position-level panel data from Aave. We estimate position fixed effects regressions to control for time-invariant heterogeneity across positions, such as individual risk preferences or portfolio management styles. This framework allows us to isolate how differences in oracle configuration, particularly heartbeat and target corridor parameters, affect capitalization choices:

$$\begin{aligned}\texttt{Position Health Factor}_{it} = {} & \beta_1\ \texttt{Heartbeat (h)}_{it} + \beta_2\ \texttt{Target Corridor (bp)}_{it} + \beta_3\ \texttt{Soph. Score}_{it} + \beta_4\ \texttt{Crypto–Stable}_{it} + \\ & \beta_5\ \texttt{Stable–Crypto}_{it} + \beta_6\ \texttt{Crypto–Crypto}_{it} + \beta_7\ \texttt{Volatility 7d}_{it} + \beta_8\ \texttt{Volatility 1y}_{it} + \\ & \sum_{i=9}^{10} \beta_i\ \texttt{Control Variables}_{it} + \alpha_i + \varepsilon_{it}\end{aligned} \tag{8}$$

We estimate three model specifications with a progressively richer set of variables. The first specification includes only the two main variables of interest: `Heartbeat (h)` and `Target Corridor (bp)`. The second specification accounts for user sophistication and adds categorical indicators for the asset category composition of each position, where `Stable` denotes a stablecoin asset and `Crypto` refers to a non-stablecoin asset. The third specification incorporates realized price volatility between each position's collateral and debt assets, using both short-term (7-day) and long-term (1-year) historical volatility. All specifications include controls for the CPF update logic (via a dummy for the aggregator version) and for structural changes associated with the Ethereum Merge.

The results of the capitalization choice analysis are summarized in Table 7, which reports regression coefficients with heteroskedasticity-robust standard errors in brackets. We use robust standard errors rather than Newey–West-type HAC corrections because the panel structure exhibits irregular time intervals between observations and a relatively small number of observations per position, which violates the assumptions underlying HAC estimators.

The `Heartbeat (h)` coefficient is negative across all specifications, suggesting that faster oracle update frequencies are associated with higher health factors. In other words, users relying on more frequently updated price feeds maintain more conservative capitalization levels. This finding is contrary to the expectation stated in **H3.1**, which hypothesized that faster updates would reduce perceived exposure to stale prices and enable lower collateral buffers. As highlighted in our CPF update behavior analysis (Fig. 1), heartbeats account for only a small share of actual updates, with most updates triggered by target corridor violations. If users are aware of this mechanism, they may place less emphasis on the nominal heartbeat. Moreover, the significance of the coefficient declines once user sophistication and volatility variables are included, indicating that these factors mediate the relationship. The subsample analyses in the next section further confirm that the effect of `Heartbeat (h)` is not robust across specifications. Overall, we find no evidence supporting **H3.1** and cannot reject the null hypothesis.

The `Target Corridor (bp)` coefficient is positive, stable, and highly significant across all three model specifications, providing strong support for hypothesis **H3.2**. Positions relying on oracles with wider target corridors tend to exhibit higher health factors, suggesting that users respond to increased tolerance for price deviation by increasing their capitalization levels. This is consistent with the theoretical expectation that wider corridors increase exposure to oracle mispricing between updates, thereby exposing users to a greater risk of oracle deviations. Although the magnitude of the coefficient slightly decreases with additional variables, the effect remains robust. We therefore reject the null hypothesis for **H3.2**.

The `Sophistication Score` coefficient is negative and highly significant in both specifications where it is included, providing strong support for hypothesis **H3.3**. This result indicates that users classified as more sophisticated tend to maintain lower health factors, and in turn more aggressive and capital-efficient positions. This behavior is consistent with the interpretation that sophisticated users are better equipped to monitor risk and respond dynamically to market conditions. They can therefore operate closer to liquidation thresholds. The effect remains stable when accounting for asset volatility. Overall, the evidence strongly supports **H3.3**, and we reject the null hypothesis.

Evidence for **H3.4** is observed in the asset category dummy variables in the second specification. Positions spanning different asset categories exhibit significantly higher health factors compared to `Stable–Stable` positions, consistent with the idea that cross-category pairs face higher relative price risk and therefore require greater capitalization. When direct volatility measures are added in the third specification, the magnitude and significance of these dummies declines, indicating that they capture perceived volatility effects in the absence of explicit volatility variables. This interpretation is reinforced by the realized volatility variables: the coefficient on 1-year volatility is positive and highly significant, showing that users increase capitalization in response to long-term volatility, whereas the coefficient on 7-day volatility is small and insignificant. This suggests that users base decisions on structural risk rather than short-term price noise. Overall, we find strong support for **H3.4** and reject the null hypothesis.

Table 7
Panel regression results with entity fixed effects for borrowing positions on Aave V2. The dependent variable is the position health factor, a measure of excess capitalization beyond protocol-imposed minimums. Independent variables include oracle heartbeat, target corridor, user sophistication score, volatility measures, and dummy variables capturing collateral-debt asset class combinations. Results are reported across three specifications, with robust standard errors clustered at the entity level.

	Dependent variable: Position health factor		
	(1)	(2)	(3)
Heartbeat (h)	−0.0144***	−0.0044	−0.0071**
	(0.0019)	(0.0028)	(0.0028)
Target Corridor (bp)	0.0028***	0.0016***	0.0017***
	(0.0002)	(0.0003)	(0.0003)
Agg. V4	−0.0859***	−0.0588***	−0.0549***
	(0.0174)	(0.0177)	(0.0177)
Soph. Score		−0.0341***	−0.0335***
		(0.0038)	(0.0037)
Crypto-Stable		0.2555***	0.0767***
		(0.0213)	(0.0287)
Stable-Crypto		0.2374***	0.0564*
		(0.0219)	(0.0293)
Crypto-Crypto		0.1974***	0.0237
		(0.0367)	(0.0408)
Volatility (7d)			0.0832
			(0.4323)
Volatility (1y)			3.7792***
			(0.4097)
Post Merge	−0.0496***	−0.0475***	−0.0330***
	(0.0076)	(0.0076)	(0.0077)
Observations	146 620	146 620	146 620
N. of groups	15 107	15 107	15 107
R^2	0.0026	0.0077	0.0087

Note: Robust standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

6. Robustness

In this section, we evaluate the robustness and limitations of our baseline findings. Specifically, we address potential sources of endogeneity, examine the sensitivity of our results to outliers, and assess the stability of our models across different subsamples of the data.

6.1. Endogeneity

Recognizing the possibility of endogeneity, we employ various empirical strategies designed to mitigate bias arising from reverse causality, omitted variables, and measurement error.

In our first analysis (**RQ1**), the main regressors enter the model in lagged form, capturing the effects of static CPF design parameters, past reporter participation, and lagged price dynamics on current deviations. This temporal structure reduces concerns about reverse causality, as the regressors are predetermined with respect to the dependent variable. Statistical inference is conducted using heteroskedasticity- and autocorrelation-consistent standard errors, which account for serial correlation inherent in the high-frequency panel structure of the data. In addition, CPF fixed effects are included to control for time-invariant unobserved heterogeneity across CPFs. This specification captures structural differences in oracle configuration, unobserved liquidity conditions, and other asset-specific characteristics that could otherwise bias the estimated coefficients. To assess the robustness of our findings, we explore alternative lag-levels and verify that the main results remain stable across model variants.

Endogeneity concerns are less pronounced in **RQ2** and **RQ3** due to differences in empirical design. In the former, we model discrete state transitions following target corridor violation events. The explanatory variables are predetermined by construction, as they are measured at the time of entry into the violation state, while the outcome reflects the subsequent resolution mechanism. This ordering rules out reverse causality, and the multinomial logistic framework further limits sensitivity to extreme values in the predictors. In the latter, the dependent variable is the user's capitalization choice, while the explanatory variables include advertised oracle design characteristics and past market conditions. Here too, the use of lagged regressors and position fixed effects addresses

potential concerns related to simultaneity and omitted variable bias. We also estimate specifications that include daily fixed effects in addition to the position entity fixed effects, which absorb more granular forms of unobserved heterogeneity, including time-varying differences in oracle behavior or market conditions. As in the first analysis, we assess robustness across a range of model specifications and variable definitions.

Measurement error is unlikely to pose a material concern. All on-chain variables, including oracle updates and individual lending market positions, are retrieved directly via RPC calls to Ethereum archive nodes, ensuring both accuracy and completeness. The only exception is the reference price data from centralized exchanges, which are obtained through the respective exchanges' official APIs. In these cases, we rely on the integrity of the external data provider. Finally, to reduce the risk of measurement error and variable misspecification, we incorporate multiple variables that capture similar economic constructs, apply alternative data treatments such as outlier adjustment (see next subsection), and assess consistency across subsamples. The main results remain robust across these variations, limiting reliance on any single measurement choice or modeling assumption.

6.2. Outlier treatment

Outliers are retained in the baseline analysis for **RQ1**, as we consider them potentially informative and reflective of economically meaningful phenomena. Nonetheless, to evaluate the sensitivity of our results to extreme observations, we conduct a robustness check using winsorized regressors. Specifically, all relevant variables are winsorized at the 95th percentile. For the Binance USDT Volume variable, we additionally standardize the data for each asset such that the within-asset mean equals one prior to winsorization. This approach allows us to capture relative volume fluctuations at the asset level, while differences in baseline trading activity across assets are absorbed by CPF fixed effects.

We re-estimate all eight baseline model specifications using the winsorized regressors. The results are shown in the Appendix in Table C.8 and remain broadly consistent with the baseline findings, suggesting that the primary relationships are not driven by extreme observations. Two exceptions merit attention. First, the coefficient on `Binance Move (t-2)` becomes negative, which is consistent with the notion that extreme short-term price movements disproportionately drive large deviations. The volatility effect appears to be captured instead by the mid-term volatility measure, `Binance Move (t-24)`, which increases in magnitude. Second, the coefficient on `Reporter Participation (%)` undergoes a changes sign in specification (2). As discussed in the main analysis, this variable is subject to opposing theoretical effects and should be interpreted with caution. All other coefficients remain stable in both sign and statistical significance, supporting the robustness of the main empirical results.

Outliers are less of a concern in the context of **RQ2** for both theoretical and econometric reasons. The analysis conditions on the presence of a target corridor violation and models transition probabilities across discrete outcomes. Large values in key predictors, such as over deviation magnitude or sequence length, are inherent to the violation state and therefore reflect the dynamics of interest rather than noise. Moreover, the multinomial logistic specification maps predictor values onto bounded probabilities, which reduces sensitivity to extreme observations. As a result, excluding or trimming these values would be both conceptually inappropriate and statistically unnecessary.

Outliers are addressed more explicitly in the sample selection for the baseline analysis addressing **RQ3**, where the dependent variable (`Health Factor`) can take on extreme values due to highly overcollateralized positions. Such values offer limited marginal insight into users' risk behavior and can disproportionately influence regression estimates. To ensure the baseline analysis captures economically relevant capitalization choices, we discard the top 10% of the health factor distribution in all specifications related to this analysis. This leaves the maximum value for the health factor at 4.55, or 355% excess collateralization. No additional outlier treatment is applied to the regressors, as most are either bounded, discrete, or derived from meaningful behavioral or market metrics. The one exception is the sophistication score, which exhibits right-skewness due to a small number of high-value observations. To address this, we re-estimate the baseline model using a log-transformed sophistication score. The results are in the Appendix in Table E.21 and show that all coefficients retain their sign, approximate magnitude and significance across specifications.

6.3. Subsamples

To further examine the validity of our main results, we create subsamples for both datasets and re-estimate the models in Section 5, maintaining the original specifications as closely as possible.

To evaluate the stability of the baseline findings for **RQ1** and **RQ2**, we first test whether the results are consistent across different heartbeat clusters. Specifically, we re-estimate each model separately for two subsamples: one consisting of price feeds with a fast heartbeat frequency (1 h) and the other restricted to price feeds with a slow heartbeat frequency (24 h). The corresponding regression outputs are reported in Appendices C.2 and D.1.

For the deviation analysis, the estimated coefficients and significance levels remain largely stable across specifications and subsamples. A notable exception is the `Reporter Participation (%)` variable, for which both the sign and statistical significance differ between the fast and slow heartbeat subsamples. This variation is not unexpected given the variable's low cross-sectional variance (mean participation rate of 99.6%), small coefficient magnitudes, and limited statistical relevance in the baseline model. These characteristics suggest that `Reporter Participation (%)` possesses limited prediction power in this context. With respect to `Reporter Uncertainty`, we find robust and statistically significant coefficients in the fast heartbeat subsample. However, the relationship weakens in the slow heartbeat group, with coefficients generally losing statistical significance. With similar reasoning as above, this indicates a relatively limited role of `Reporter Uncertainty` in explaining expected deviation.

For the transition probability analysis, the variable Target Corridor (bp) does not display robustness for the random recovery path across heartbeat subsamples. This outcome is consistent with the baseline findings, where the variable exhibited low coefficient magnitudes and lacked statistical significance. Given that the random transition path is not systematically related to oracle behavior, the absence of a robust association is conceptually coherent. By contrast, for the CPF transition path, the variable remains statistically significant and directionally consistent across both subsamples, suggesting a more stable relationship in non-random recovery mechanisms.

Next, we examine whether the structural changes induced by the Ethereum Merge coincide with systematic shifts in model performance. The Merge altered Ethereum's consensus mechanism from proof of work to proof of stake and replaced stochastic block intervals with a deterministic 12-s block time. These changes have the potential to influence network conditions, such as update predictability and congestion, which may in turn affect oracle behavior. Approximately 39% of our sample period corresponds to the pre-Merge regime, with the remainder covering the post-Merge period. To evaluate the stability of our findings, we re-estimate all model specifications for each consensus period separately. The results are reported in Appendices C.3 and D.2.

The deviation analysis results reveal consistent coefficient estimates and significance levels for most explanatory variables across both the pre- and post-Merge samples. In line with the heartbeat-based subsample analysis, the explanatory power of Reporter Participation (%) remains limited, and Reporter Uncertainty also exhibits weak predictive strength, particularly in the post-Merge subsample.

By contrast, the recovery analysis reveals more pronounced differences across subsamples. In the post-Merge period, we observe faster adjustment following violations through the oracle update (non-random) path. This shift is visible both in the estimated state transition probabilities and in the regression-based results. To illustrate these dynamics, we replicate the figures and transition probabilities from Section 5.2 by subsample with the disaggregated outputs presented in Appendix D.3.

These findings should be interpreted with caution. Although the Merge introduced more predictable block intervals and implemented measures aimed at reducing congestion, we cannot fully isolate its effects from other contemporaneous changes. The observed improvements in violation recovery may partly reflect ongoing upgrades to the oracle infrastructure or other time-dependent trends. A comprehensive causal assessment of these effects lies beyond the scope of this paper and represents a promising avenue for future research.

Finally, our proxy for network congestion, Median Price Per Gas, remains statistically significant and directionally consistent across all specifications and subsamples.

Overall, the subsample analyses provide strong evidence that the main findings for the first two research questions remain valid under varying heartbeat frequencies and structural shifts in Ethereum's protocol. These results reinforce both the credibility and external validity of the baseline estimations.

To assess the robustness of our findings regarding **RQ3**, we re-estimate the model specifications with the addition of daily time fixed effects. This adjustment controls for unobserved, time-varying factors such as market-wide sentiment shifts or protocol events, that could influence user capitalization behavior. As a result, the Post Merge dummy is excluded due to collinearity with the time dummies. The results, presented in Table E.22 in the Appendix, are largely consistent with those from the baseline analysis. Key differences include increased statistical significance of the Heartbeat (h) coefficient in models (2) and (3)and a further decline in the significance of the asset category combination dummies in model (3). Notably, the Agg. V4 dummy remains highly significant across all models, despite limited temporal variation. This finding reinforces the interpretation that the introduction of the new aggregation logic may have had a lasting impact on capitalization choices.

Next, we investigate heterogeneity in capitalization behavior by splitting the sample by collateral category, distinguishing positions backed by stablecoins from those backed by volatile crypto assets. The results are reported in Table E.23. Since certain asset combinations do not occur within each subsample, the categorical dummy structure is adjusted accordingly. In the stable collateral group, only Stable-Stable and Stable-Crypto positions appear, so the Crypto-Stable and Crypto-Crypto dummies are omitted. In the crypto collateral group, only Crypto-Stable and Crypto-Crypto combinations appear, with Crypto-Stable serving as the reference category. Due to the shift in reference category across subsamples, the dummy coefficients capture different comparisons and are not directly comparable to the baseline model or to each other. Nevertheless, we include these dummies to capture within-group differences in asset-pair risk, to ensure consistency with the baseline model, and to facilitate a clearer interpretation of capitalization choices within each collateral category.

We find substantial variation in the magnitude and significance of the Heartbeat (h) coefficient across subsamples, raising concerns about the generalizability of its effect. While the coefficient is significant in the baseline model, it becomes insignificant in the stable collateral group once more variables are added, and regains significance only in the crypto collateral sample. This inconsistency reinforces the conclusion from Section 5.3 that nominal heartbeat frequency plays a limited role in user decision-making, likely because most price updates are triggered by target corridor violations rather than elapsed time. Consequently, the Heartbeat (h) coefficient should be interpreted with caution for the third research question.

Particularly noteworthy is the stable collateral group, where the Post Merge coefficient is consistently significant and reverses sign compared to the baseline analysis. This suggests that the aggregate negative effect of the Ethereum Merge on capitalization levels is largely driven by positions with crypto-denominated collateral. Furthermore, the diminished significance of the Target Corridor (bp) variable in models (2) and (3) for the Stable Collateral group may reflect differences in capitalization strategies across markets.

To further explore heterogeneity in user behavior, we divide the sample based on the Sophistication Score, splitting the sample at the median. The results, presented in Table E.24, highlight additional divergences. For the high sophistication group, the Target Corridor (bp) variable loses significance in the extended specifications, while the Post Merge effect weakens.

However, the asset category dummies remain highly significant across all models, including those directly accounting for realized volatility. In contrast, the low sophistication group exhibits weaker significance for the asset category dummies in model (2), with both significance and direction of coefficients changing in model (3) once realized volatility variables are introduced.

Taken together, these subsample analyses for **RQ3** indicate meaningful heterogeneity in user behavior and preferences. The observed differences along collateral type and user sophistication suggest that agents respond differently to oracle characteristics and market conditions when determining capitalization levels. While the main patterns remain visible, the variability in effect sizes and statistical significance highlights the importance of context when interpreting the determinants of capitalization choices. Furthermore, the inconsistent results for the `Heartbeat (h)` variable across both the baseline and subsample analyses reinforce the conclusion that it has no consistent or economically meaningful effect.

7. Conclusion

Many financial blockchain applications depend on external data sources for asset prices. These data providers are commonly referred to as price oracles. The accuracy of the reported data is crucial, as it directly affects the performance and reliability of smart contract-based financial infrastructure. Deviations from the ground truth price can lead to unexpected results, create frictions, increase uncertainty, and misalign incentives. In cases of significant deviations, inaccurate oracles can pose an existential threat to protocols that rely on them. We outline this relationship and examine how expected oracle price deviations influence the risk parameter selection of both protocols and users. While more conservative risk parameters increase resilience to oracle deviations, they reduce capital efficiency. Consequently, the economic viability of many blockchain-based financial applications depends on the availability of sufficiently accurate price oracles.

From a financial risk perspective, price oracles can be regarded as critical external dependencies, and it is essential to assess and understand the associated risks. When a particular oracle becomes widely adopted, the resulting dependency can become a source of systemic risk. In such cases, improving our understanding of oracle accuracy and reliability is relevant not only for individual protocols but also for the stability of the DeFi ecosystem as a whole.

Our analysis focuses on *Chainlink*, the dominant oracle provider, and addresses three research questions concerning its price feeds: (1) their accuracy over time and the factors associated with it, (2) their ability to regain expected accuracy levels following target corridor violations, and (3) the extent to which oracle design parameters affect user capitalization choices in DeFi lending markets.

In the first part, we observe the accuracy of 40 CPFs over an 18-month period. We find an average price deviation of approximately 57 basis points. Our pooled OLS and fixed effects regression models suggest that CPFs with lower heartbeat values and narrower target corridors tend to perform better. Short- and medium-term market volatility are strongly associated with price deviations, while gas prices appear to have little impact on oracle performance. This result is noteworthy, as one might expect higher gas prices to impair oracle responsiveness during periods of network congestion. Our findings, however, indicate that CPFs maintain stable performance with respect to gas price fluctuations.

In the second part of our analysis, we study how CPFs handle target corridor violations and the associated recovery dynamics. We find that 97.5% of all observations remain within the target corridor defined by CPF price deviation trigger criteria. When violations occur, the average time to recovery is around 10 blocks (~2 min and 10 s). Smaller violations are both more frequent and recovered more quickly. In many cases, the recovery is driven by a move in the ground truth price rather than a CPF update. Larger violations persist longer and are more likely to require a direct update from the CPF. As the time spent outside the target corridor increases, the probability of random recovery declines, while the probability of a CPF-triggered correction rises.

Our third analysis underscores the role of oracle accuracy configurations in shaping user behavior and capital efficiency. We find that oracle target corridor width has a strong positive effect on health factors, suggesting that users increase collateral buffers when exposed to lower advertised oracle accuracy. In contrast, the nominal heartbeat frequency shows no consistent effect, likely because most updates are triggered by target corridor violations rather than elapsed time. Long-term volatility is an important driver of capitalization decisions, indicating that users adjust their collateralization in response to structural risk rather than short-term fluctuations. Furthermore, we find pronounced differences between sophisticated and less sophisticated users, with the former operating closer to liquidation thresholds while maintaining effective risk control. These findings highlight that oracle design parameters and market volatility not only affect protocol-level outcomes but also influence individual user risk management, with implications for systemic stability.

Our research sheds light on an important but empirically underexplored area. CPFs are of great relevance for blockchain-based financial markets, and understanding their performance and risk implications is essential for both protocol designers and users. We believe that our findings provide useful insights for researchers, policy makers and practitioners, while emphasizing that this work constitutes an initial step in exploring these issues.

Future research could investigate heartbeat-triggered updates in more detail, analyze individual reporter behavior, or leverage mempool data to gain deeper insights into price update delays. In addition, similar analyses could be conducted for other DONs once sufficient data becomes available. A cross-chain comparison, including layer-2 networks, could further clarify how differences in block times or consensus mechanisms affect oracle performance. Lastly, we encourage applied research that explores the practical implications of our findings, particularly regarding how oracle performance and design parameters influence the selection of protocol risk parameters.

CRediT authorship contribution statement

Matthias Nadler: Writing – review & editing, Writing – original draft, Visualization, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Katrin Schuler:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Fabian Schär:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the anonymous reviewer as well as Mitchell Goldberg and Dario Thürkauf for their valuable inputs and suggestions.
Funding: This work was supported by the WWZ Förderverein [grant number FV-104].

Appendix A. Variable description

A.1. Variables for Chainlink Price Feed analyses

This section provides a comprehensive summary and explanation of the variables employed in our data analysis with the Chainlink price feed data: Sections 5.1 and 5.2. We use a selection of observable variables, derived from our hypotheses, in conjunction with variables commonly employed in traditional models for evaluating market conditions. Each variable is either clearly exogenous, or used with a lag of one (or more periods). Summary statistics for all variables can be found in Table 2.

`Deviation (bp)` This variable measures the absolute proportional deviation of the Chainlink oracle price (p_o) from the Binance ground truth price (p_g) and is calculated as $|\frac{p_o}{p_g} - 1|$. We use this variable to quantify the accuracy of the oracle price in basis points (bp).

`Fast Heartbeat` In our data, the CPFs have heartbeats set to either 1 h (3600 s) or 24 h (86,400 s). CPF reporters are required to transmit a new price report if no update has been recorded within the specified heartbeat interval. For the purpose of our study, we employ a dummy variable, assigning a value of 1 for CPFs associated with a 1-h heartbeat interval, indicative of a faster update frequency, and a value of 0 for those with a 24-h interval, representing a slower update frequency.

`Target Corridor (bp)` The target corridor sets the maximum allowed deviation in basis points between the oracle price and the ground truth. When this deviation is exceeded, the CPF should update. For instance, a target corridor of 100 indicates CPF reporters are required to transmit a new price update if they detect a deviation of 1% or more between the most recent oracle price and the currently observed ground truth. Each asset in our analysis falls within one of these specified target corridor ranges: 50, 100, 200, or 500 bp.

`Binance USDT Volume (in Millions)` The *single-counted* trading volume in the Binance spot markets. For each token the <TOKEN>USDT k-line spot data with a 1-s frequency was resampled to include all volumes from the timestamp of the last block to the timestamp of the current block. This variable is expressed in million USDT.

`Reporter Participation (%)` At any given time, each CPF has a pre-defined set of reporters. For an update report to be considered valid, it must contain price observations from at least two-thirds of these reporters. This variable indicates the percentage of reporters who submitted a price observation during the last oracle update.

`Reporter Uncertainty` This variable measures the level of uncertainty or disagreement among the reporters regarding the most recent oracle price update. It does this by determining the variance across all reported and normalized prices that are part of the most recent update.

`Median Price per Gas` This variable represents the median gas price of the current block, specifically computed as the median USDT price for one billion units of gas, commonly referred to as the GWEI gas price. When calculating the median, transactions with a Gas Price of 0 are ignored, as these do not reflect the market price of gas, but instead are included by the validator; for example to perform internal accounting. For data before the Merge, we look at the total Gas Price for a transaction. For data after the Merge, we calculate the Gas Price as `Base Fee` plus `Priority Fee`. Gas prices are set through a mix of blockchain protocol rules and the demand for block space, making this measure a reliable indicator of both short-term and long-term fluctuations in demand for block space, or in other words, network congestion.

Target Corridor Violation (t-1) This represents a target corridor violation with a one-block lag. It is a dummy variable where the value is set to 1 if the oracle price was outside the target corridor in the previous block, and 0 if it was within the target corridor.

Binance Move (t-n) These variables serve as indicators of past changes in the Binance ground truth price. More formally, they are defined as $|1 - \frac{p_{g,t-n}}{p_{g,t-1}}|$, where p_g represents the ground truth price and $t - n$ denotes a lag of n blocks. Essentially, this measures the proportional change in the ground truth price from the nth previous to the previous block and is an indicator for price volatility. A value of $n = 24$ corresponds to an approximate lag of 5 min.

Over Deviation (bp, t-1) Lagged by one block, this variable quantifies the extent of the current target corridor violation in basis points, calculated as the total deviation minus the target corridor. It is zero when no violation occurs.

Violation Sequence Length This variable tracks the length of time spent outside the target corridor. It is set to 0 if the current block falls within the target corridor. However, when a target corridor violation occurs (a move out of the target corridor), the variable's value starts at 1 and then increments by 1 for each successive block that continues to stay outside the corridor.

Post Merge A dummy variable that is 0 for any block before the Ethereum Merge and 1 for any block thereafter. The Merge occurred in block 15537394 on September 15, 2022.

A.2. Variables for capitalization choice analysis

This section provides a comprehensive summary and explanation of the variables employed in our analysis with AAVE lending market data: Section 5.3. We use a selection of observable variables derived from our hypotheses in conjunction with variables commonly employed in traditional models for evaluating market conditions. Each variable is clearly exogenous. Summary statistics for all variables can be found in Table 7.

Health Factor This variable represents the capitalization of a borrowing position relative to the liquidation threshold. It indicates how safe the position is from being liquidated, based on the value of the collateral relative to the borrowed assets. Formally, it is calculated as: (Coll. Value×Liq. Threshold)/Debt Value, where the Liquidation Threshold is a protocol-defined percentage that determines the maximum loan-to-value for a specific collateral asset type. A health factor of 1 or greater means the position is healthy (safe), while a position with a health factor of below 1 is undercollateralized and eligible for liquidation.

Heartbeat (h) All of the CPFs used in the Aave protocol have heartbeats set to either 1 h, 6 h or 24 h. CPF reporters are required to transmit a new price report if no update has been recorded within the specified heartbeat interval. Since Ether is the internal unit of account in the Aave protocol, it relies on asset oracle prices denominated in ETH. In our regression model, the entities are individual Aave positions, each defined by a specific asset pair consisting of one collateral and one debt asset. The heartbeat for a position is computed as the average of the heartbeats of its collateral and debt asset, with ETH (and by extension WETH) modeled with a heartbeat of 0, as it does not rely on a CPF (1 ETH is always 1 ETH).

Target Corridor (bp) The target corridor sets the maximum allowed deviation in basis points between the oracle price and the ground truth. When this deviation is exceeded, the CPF should update. For instance, a target corridor of 100 indicates CPF reporters are required to transmit a new price update if they detect a deviation of 1% or more between the most recent oracle price and the currently observed ground truth. By similar reasoning as above, the corridor is calculated as the average of the target corridors of the collateral and debt assets with the target corridor for both ETH and WETH defined as 0.

Agg. V4 CPF oracles are periodically upgraded to improve efficiency and security. A potentially relevant paradigm shift was the introduction of off-chain aggregation with the deployment of version 4 Chainlink price aggregator contracts. This upgrade did not occur simultaneously across all CPFs. Agg. V4 is a dummy variable equal to 1 if both the collateral and debt oracles for a given position use version 4 aggregators, and 0 otherwise.

Soph. Score The sophistication score is derived from a principal component analysis (PCA) of three variables: total collateral (in ETH), total debt (in ETH), and an indicator for whether the position is controlled by an externally owned account (EOA). The first principal component (PC1) is extracted after standardizing these variables and is interpreted as a proxy for user sophistication. Higher scores typically reflect more involved capital and greater likelihood of smart contract control. This value is calculated at each point of observation, and may vary for a position over time; for example, in case of significant changes to the amount locked as collateral or drawn as debt.

Crypto – Stable, Stable – Crypto, Crypto – Crypto These three dummy variables capture the asset category combination of a position's asset collateral and debt. Specifically: Crypto – Stable indicates non-stablecoin collateral and stablecoin debt; Stable – Crypto indicates stablecoin collateral and non-stablecoin debt; and Crypto – Crypto indicates that both assets are non-stablecoins. The omitted category and baseline value in this setup, Stable – Stable, corresponds to positions where both collateral and debt are stablecoins. These variables are defined at the position level based on boolean flags for asset type.

`Volatility (7d)` This variable measures short-term relative price volatility for the asset pair in a given Aave position. It is calculated using 7 days of hourly Binance price data prior to the block timestamp. If either the collateral or debt asset is a stablecoin, the volatility is based solely on the non-stable asset's log return standard deviation. If both assets are non-stablecoins, the relative return between the two assets is computed, and the volatility is based on the standard deviation of its log returns. For fully stablecoin positions, the value is set to zero.

`Volatility (1y)` This variable measures long-term relative price volatility using daily Binance data over the 365 days preceding the block timestamp. The computation follows the same logic as the 7-day version: using the standard deviation of log returns for the non-stable asset, or the relative return between two non-stablecoins. If both assets are stablecoins, the volatility is defined as zero.

`Post Merge` A dummy variable that is 0 for any block before the Ethereum Merge and 1 for any block thereafter. The Merge occurred in block 15537394 on September 15, 2022.

Appendix B. Chainlink Price Feed architecture

This appendix provides a technical introduction to the operational mechanics CPFs. Each CPF consists of a set of smart contracts deployed on-chain and supported by an off-chain peer-to-peer communication network among designated reporters, as illustrated in Fig. B.6. The n participating reporters are approved by Chainlink and publicly disclosed on its website, alongside the proxy contract address and the trigger criteria that determine when data are transmitted.

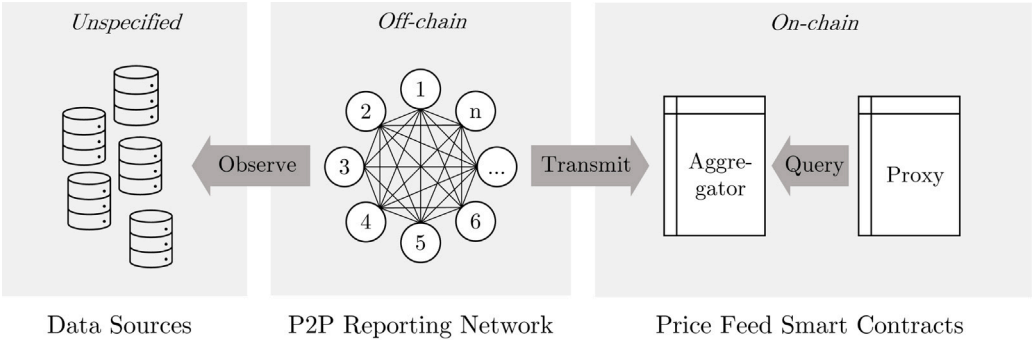


Fig. B.6. Operational flow of a Chainlink Price Feed (CPF) on Ethereum. The diagram shows how a decentralized oracle network (DON) of authorized reporters observes the off-chain reference price, submits cryptographically signed values, and aggregates them through the CPF's smart contract logic.

Each CPF provides a price quote for a specific asset pair. Reporters independently monitor their proprietary, undisclosed price sources and periodically share their observations through the off-chain reporting network, which is coordinated by a dedicated protocol. When the participating reporters agree that at least one CPF trigger criteria has been met, all observations are compiled into a report, cryptographically signed by the reporters, and submitted to the CPF's aggregator contract as the calldata of a `transmit` transaction (Breidenbach et al., 2021).

Upon receipt of a `transmit` transaction, the aggregator verifies the report's integrity, timeliness, and proper authorization. The necessary information, such as authorized reporter addresses or the required quorum, are part of the aggregator configuration variables. In case all checks pass, the latest price variable in the aggregator contract is updated with the approximated median of the observations transmitted in the report. From this point on, the new price point is returned, whenever queried via the CPF's proxy contract, which serves as static gateway for DeFi applications relying on the CPF.

Appendix C. Robustness deviation analysis

C.1. Winsorized data

Table C.8
Pooled OLS (columns 1–4) and individual fixed effects (columns 5–8) estimates of Chainlink Price Feed deviations from Binance reference prices, using winsorized regressors. The dependent variable is Deviation (basis points). Regressors include Fast Heartbeat (1 h vs. 24 h), Target Corridor (basis points), Reporter Participation (%), Reporter Uncertainty, Median Price per Gas, lagged Target Corridor Violation, short- and mid-horizon returns (Binance Move at t-2 and t-24), Binance USDT Volume, and a Post-Merge indicator; fixed-effects specifications omit time-invariant trigger parameters. All relevant regressors are winsorized at the 95th percentile, and Binance USDT Volume is standardized within asset before winsorization, as described in Section 6.2. Coefficients are reported with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors.

	Dependent variable: Deviation (BP)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	19.719***	19.027***	16.513***	11.572***				
	(0.041)	(0.587)	(0.584)	(0.554)				
Individual FE					yes	yes	yes	yes
Fast Heartbeat	−12.851***	−10.593***	−10.528***	−10.479***				
	(0.020)	(0.021)	(0.020)	(0.019)				
Target Corridor (bp)	0.247***	0.245***	0.246***	0.256***				
	(0.0002)	(0.0002)	(0.0002)	(0.0002)				
Reporter Participation (%)		−2.819***	−1.553***	0.617		−6.757***	−6.199***	−5.514***
		(0.588)	(0.585)	(0.555)		(0.594)	(0.562)	(0.562)
Reporter Uncertainty (wins.)		39,282.020***	34,157.350***	27,064.340***		36,919.680***	27,444.000***	24,606.080***
		(86.199)	(83.132)	(76.464)		(88.113)	(79.407)	(78.513)
Median Price per Gas (wins.)	0.087***	0.073***	0.055***	0.034***	0.085***	0.073***	0.044***	0.034***
	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.0005)	(0.0005)
Target Corridor Violation (t-1)				97.582***			101.265***	98.907***
				(0.071)			(0.077)	(0.069)
Binance Move (t-2, wins.)			−0.683***	−0.400***				−0.377***
			(0.009)	(0.008)				(0.008)
Binance Move (t-24, wins.)			1.331***	0.820***				0.789***
			(0.008)	(0.007)				(0.007)
Binance USDT Vol. (std., wins.)	5.249***	4.551***	3.639***	2.252***	5.510***	4.811***	2.963***	2.436***
	(0.008)	(0.008)	(0.007)	(0.006)	(0.008)	(0.008)	(0.006)	(0.006)
Post Merge	−7.781***	−6.930***	−6.199***	−4.245***	−7.718***	−6.916***	−4.622***	−4.225***
	(0.020)	(0.020)	(0.020)	(0.018)	(0.020)	(0.020)	(0.018)	(0.018)
Observations	149,609,200	149,609,200	149,609,120	149,609,120	149,609,200	149,609,200	149,609,160	149,609,120
$\bar{R}^2$ *(pooled)*	0.236	0.245	0.252	0.323				
$\bar{R}^2$ *(within)*					0.023	0.032	0.131	0.134

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

C.2. Heartbeat subsample

Table C.9
Sample restricted to CPFs with a slow (= low-frequency) heartbeat. Pooled OLS with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Target Corridor (bp)	0.243***	0.243***	0.244***	0.255***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Reporter Participation (%)		−0.287	2.872***	4.073***
		(0.843)	(0.827)	(0.799)
Reporter Uncertainty		119.376***	71.399***	77.300***
		(5.640)	(3.621)	(3.414)
Median Price per Gas	0.012***	0.012***	0.006***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Target Corridor Violation (t-1)				97.137***
				(0.163)
Binance Move (t-2)			0.441***	0.340***
			(0.005)	(0.006)
Binance Move (t-24)			0.391***	0.270***
			(0.003)	(0.003)
Binance USDT Volume (m)	95.969***	95.888***	34.072***	9.748***
	(1.100)	(1.098)	(0.606)	(0.502)
Post-Merge	−8.878***	−8.855***	−4.610***	−3.373***
	(0.026)	(0.026)	(0.037)	(0.034)
Intercept	25.394***	25.669***	10.925***	7.421***
	(0.041)	(0.840)	(0.829)	(0.800)
Observations	108,466,670	108,466,670	108,465,974	108,465,974
R^2	0.159	0.159	0.208	0.261
Adjusted R^2	0.159	0.159	0.208	0.261

Note: HAC standard errors. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table C.10
Sample restricted to CPFs with a fast (= high-frequency) heartbeat. Pooled OLS with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Target Corridor (bp)	0.390***	0.390***	0.277***	0.287***
	(0.0004)	(0.0004)	(0.0004)	(0.001)
Reporter Participation (%)		−3.600***	1.498***	2.236***
		(0.544)	(0.473)	(0.404)
Reporter Uncertainty		28.164***	−4.706***	0.564
		(4.922)	(1.340)	(1.814)
Median Price per Gas	0.010***	0.010***	0.003***	0.001***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Target Corridor Violation (t-1)				66.035***
				(0.122)
Binance Move (t-2)			0.559***	0.429***
			(0.003)	(0.004)
Binance Move (t-24)			0.478***	0.311***
			(0.002)	(0.003)
Binance USDT Volume (m)	4.499***	4.500***	0.896***	0.020
	(0.034)	(0.034)	(0.024)	(0.023)
Post-Merge	−10.840***	−10.822***	−5.574***	−4.479***
	(0.022)	(0.022)	(0.025)	(0.025)
Intercept	0.828***	4.379***	−3.298***	−4.292***
	(0.033)	(0.543)	(0.473)	(0.404)
Observations	41,142,530	41,142,530	41,142,266	41,142,266
R^2	0.095	0.095	0.275	0.403
Adjusted R^2	0.095	0.095	0.275	0.403

Note: *p < 0.1; **p < 0.05; ***p < 0.01.

Table C.11
Sample restricted to CPFs with a slow (= low-frequency) heartbeat. Fixed effects at the CPF-level with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Reporter Participation (%)		−9.699***	−9.671***	−6.709***
		(0.849)	(0.810)	(0.806)
Reporter Uncertainty		120.616***	111.935***	81.110***
		(5.658)	(5.480)	(3.489)
Median Price per Gas	0.012***	0.012***	0.008***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Target Corridor Violation (t-1)			115.962***	99.544***
			(0.106)	(0.169)
Binance Move (t-2)				0.340***
				(0.006)
Binance Move (t-24)				0.261***
				(0.003)
Binance USDT Volume (m)	117.621***	117.485***	62.568***	22.455***
	(1.306)	(1.303)	(0.874)	(0.554)
Post-Merge	−8.814***	−8.760***	−5.775***	−3.319***
	(0.026)	(0.026)	(0.024)	(0.035)
Observations	108,466,670	108,466,670	108,466,641	108,465,974
R^2	0.010	0.010	0.108	0.132
Adjusted R^2	0.010	0.010	0.108	0.132

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Table C.12
Sample restricted to CPFs with a fast (= high-frequency) heartbeat. Fixed effects at the CPF-level with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Reporter Participation (%)		−2.238***	0.231	2.907***
		(0.543)	(0.429)	(0.406)
Reporter Uncertainty		32.800***	23.231***	0.652
		(4.946)	(4.701)	(1.813)
Median Price per Gas	0.010***	0.010***	0.005***	0.001***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Target Corridor Violation (t-1)			83.743***	65.984***
			(0.082)	(0.122)
Binance Move (t-2)				0.431***
				(0.004)
Binance Move (t-24)				0.309***
				(0.003)
Binance USDT Volume (m)	5.015***	5.021***	2.168***	0.068***
	(0.037)	(0.037)	(0.024)	(0.024)
Post-Merge	−10.836***	−10.819***	−7.227***	−4.493***
	(0.022)	(0.022)	(0.017)	(0.026)
Observations	41,142,530	41,142,530	41,142,519	41,142,266
R^2	0.039	0.039	0.293	0.365
Adjusted R^2	0.039	0.039	0.293	0.365

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

C.3. Pre- & post-Merge subsample

Table C.13
Sample restricted to observations from the pre-Merge consensus regime. Pooled OLS with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Fast Heartbeat	−8.714***	−8.782***	−7.113***	−7.314***
	(0.040)	(0.040)	(0.037)	(0.035)
Target Corridor (bp)	0.273***	0.272***	0.275***	0.292***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Reporter Participation (%)		−17.806***	−10.189***	−5.004***
		(0.882)	(0.854)	(0.809)
Reporter Uncertainty		63.676***	50.725***	51.539***
		(2.767)	(2.601)	(2.569)
Median Price per Gas	0.011***	0.011***	0.005***	0.003***
	(0.0002)	(0.0002)	(0.0002)	(0.0001)
Target Corridor Violation (t-1)				89.573***
				(0.130)
Binance Move (t-2)			0.400***	0.300***
			(0.003)	(0.003)
Binance Move (t-24)			0.459***	0.282***
			(0.001)	(0.002)
Binance USDT Volume (m)	2.646***	2.628***	0.755***	−1.007***
	(0.051)	(0.051)	(0.035)	(0.037)
Intercept	20.320***	38.123***	16.490***	9.491***
	(0.068)	(0.881)	(0.855)	(0.813)
Observations	49,565,520	49,565,520	49,564,560	49,564,560
R^2	0.211	0.211	0.290	0.375
Adjusted R^2	0.211	0.211	0.290	0.375

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Table C.14
Sample restricted to observations from the post-Merge consensus regime. Pooled OLS with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Fast Heartbeat	−14.781***	−14.682***	−13.772***	−13.363***
	(0.023)	(0.024)	(0.024)	(0.023)
Target Corridor (bp)	0.229***	0.229***	0.229***	0.235***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Reporter Participation (%)		26.587***	27.578***	25.896***
		(0.740)	(0.721)	(0.704)
Reporter Uncertainty		413.738***	109.603***	142.017***
		(44.567)	(18.001)	(16.997)
Median Price per Gas	0.039***	0.039***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Target Corridor Violation (t-1)				85.312***
				(0.205)
Binance Move (t-2)			0.566***	0.448***
			(0.007)	(0.009)
Binance Move (t-24)			0.363***	0.272***
			(0.004)	(0.004)
Binance USDT Volume (m)	3.281***	3.250***	1.100***	−0.662***
	(0.037)	(0.037)	(0.030)	(0.029)
Intercept	18.669***	−7.929***	−15.148***	−14.025***
	(0.045)	(0.739)	(0.723)	(0.706)
Observations	100,043,680	100,043,680	100,043,680	100,043,680
R^2	0.230	0.231	0.271	0.305
Adjusted R^2	0.230	0.231	0.271	0.305

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Table C.15
Sample restricted to observations from the pre-Merge consensus regime. Fixed effects at the CPF-level with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Reporter Participation (%)		−17.839***	−12.822***	−9.159***
		(0.921)	(0.856)	(0.850)
Reporter Uncertainty		72.816***	67.680***	57.642***
		(2.862)	(2.786)	(2.661)
Median Price per Gas	0.011***	0.011***	0.006***	0.003***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Target Corridor Violation (t-1)			107.521***	90.805***
			(0.096)	(0.131)
Binance Move (t-2)				0.303***
				(0.003)
Binance Move (t-24)				0.276***
				(0.002)
Binance USDT Volume (m)	6.614***	6.630***	1.918***	−0.098***
	(0.084)	(0.084)	(0.044)	(0.037)
Observations	49,565,520	49,565,520	49,565,480	49,564,560
R^2	0.002	0.002	0.176	0.209
Adjusted R^2	0.002	0.002	0.176	0.209

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Table C.16
Sample restricted to observations from the post-Merge consensus regime. Fixed effects at the CPF-level with heteroskedasticity- and autocorrelation-consistent (HAC) standard errors; columns (1)–(4) mirror the baseline specifications.

	Dependent variable: Deviation (bp)			
	(1)	(2)	(3)	(4)
Reporter Participation (%)		19.855***	17.203***	18.819***
		(0.743)	(0.715)	(0.707)
Reporter Uncertainty		390.691***	339.169***	129.347***
		(43.538)	(40.749)	(16.296)
Median Price per Gas	0.039***	0.038***	0.028***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Target Corridor Violation (t-1)			104.283***	87.205***
			(0.136)	(0.213)
Binance Move (t-2)				0.443***
				(0.009)
Binance Move (t-24)				0.264***
				(0.004)
Binance USDT Volume (m)	7.427***	7.400***	3.252***	0.375***
	(0.055)	(0.055)	(0.042)	(0.046)
Observations	100,043,680	100,043,680	100,043,680	100,043,680
R^2	0.001	0.002	0.072	0.099
Adjusted R^2	0.001	0.002	0.072	0.099

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Appendix D. Robustness violation recovery analysis

D.1. Heartbeat subsample

Table D.17
Sample restricted to CPFs with a slow (= low-frequency) heartbeat. Multinomial logistic regression for transitions out of a violation state with two alternatives: CPF-update ("CPF") vs. random ground truth price movement ("Random"); "Stay" is the reference. Columns (1)–(4) mirror the baseline specifications.

	Multinomial logistic regression for state transitions from violation (Reference = Stay)							
	Model Spec. 1:		Model Spec. 2:		Model Spec. 3:		Model Spec. 4:	
	CPF	Random	CPF	Random	CPF	Random	CPF	Random
Over Deviation (bp, t-1)	0.004***	−0.051***	0.004***	−0.051***	0.004***	−0.051***	0.007***	−0.058***
	(0.00003)	(0.00023)	(0.00003)	(0.00023)	(0.00004)	(0.00023)	(0.00007)	(0.00023)
Violation Sequence Length	0.003***	−0.012***	0.003***	−0.012***	0.003***	−0.012***	0.002***	−0.005***
	(0.00011)	(0.00026)	(0.00011)	(0.00026)	(0.00011)	(0.00026)	(0.00012)	(0.00024)
Target Corridor (bp)			−0.00224***	0.00003	−0.00220***	0.00000	−0.00199***	−0.00069***
			(0.00005)	(0.00004)	(0.00005)	(0.00004)	(0.00005)	(0.00004)
Median Price per Gas					−0.00026***	0.00023***	−0.00028***	0.00015***
					(0.00002)	(0.00001)	(0.00002)	(0.00001)
Binance Move (t-2)							−0.00587***	0.00702***
							(0.00018)	(0.00017)
Binance Move (t-24)							−0.00144***	0.00372***
							(0.00004)	(0.00004)
Binance USDT Volume (m)					0.1920***	−0.1758***	0.2661***	−0.6800***
					(0.0216)	(0.0302)	(0.0209)	(0.0365)
Constant	−3.1809***	−1.6711***	−2.8807***	−1.6741***	−2.8807***	−1.6764***	−2.8248***	−1.8273***
	(0.0035)	(0.0033)	(0.0070)	(0.0057)	(0.0071)	(0.0058)	(0.0071)	(0.0060)
Akaike Inf. Crit.	2,344,512		2,341,964		2,341,342		2,323,485	

Note: *p < 0.1; **p < 0.05; ***p < 0.01.

Table D.18
Sample restricted to CPFs with a fast (= high-frequency) heartbeat. Multinomial logistic regression for transitions out of a violation state with two alternatives: CPF-update ("CPF") vs. random ground truth price movement ("Random"); "Stay" is the reference. Columns (1)–(4) mirror the baseline specifications.

	Multinomial logistic regression for state transitions from violation (Reference = Stay)							
	Model Spec. 1:		Model Spec. 2:		Model Spec. 3:		Model Spec. 4:	
	CPF	Random	CPF	Random	CPF	Random	CPF	Random
Over Deviation (bp, t-1)	0.006***	−0.081***	0.007***	−0.082***	0.007***	−0.082***	0.0110***	−0.0882***
	(0.00007)	(0.0004)	(0.00007)	(0.0004)	(0.00007)	(0.0004)	(0.00013)	(0.0004)
Violation Sequence Length	0.007***	0.010***	0.007***	0.010***	0.008***	0.009***	0.005***	0.013***
	(0.00021)	(0.00035)	(0.00023)	(0.00037)	(0.00024)	(0.00038)	(0.00023)	(0.00033)
Target Corridor (bp)			−0.0192***	0.0016***	−0.0192***	−0.0002	−0.0188***	−0.0019***
			(0.00018)	(0.00017)	(0.00020)	(0.00020)	(0.00020)	(0.00020)
Median Price per Gas					−0.0007***	0.0001***	−0.0005***	0.0001***
					(0.00005)	(0.00001)	(0.00004)	(0.00002)
Binance Move (t-2)							−0.0084***	0.0066***
							(0.00036)	(0.00036)
Binance Move (t-24)							−0.0025***	0.0046***
							(0.00009)	(0.00007)
Binance USDT Volume (m)					0.0098***	−0.0877***	0.0208***	−0.1142***
					(0.0030)	(0.0050)	(0.0029)	(0.0053)
Constant	−2.981***	−1.590***	−1.259***	−1.735***	−1.245***	−1.559***	−1.153***	−1.657***
	(0.005)	(0.004)	(0.016)	(0.016)	(0.018)	(0.019)	(0.018)	(0.019)
Akaike Inf. Crit.	1,272,991		1,262,437		1,261,428		1,253,954	

Note: HAC standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

D.2. Pre- & post-Merge subsample

Table D.19
Sample restricted to the pre-Merge consensus regime. Multinomial logistic regression for transitions out of a violation state with two alternatives: CPF-update ("CPF") vs. random ground truth price movement ("Random"); "Stay" is the reference. Columns (1)–(4) mirror the baseline specifications.

	Multinomial logistic regression for state transitions from violation (Reference = Stay)							
	Model Spec. 1:		Model Spec. 2:		Model Spec. 3:		Model Spec. 4:	
	CPF	Random	CPF	Random	CPF	Random	CPF	Random
Over Deviation (bp, t-1)	0.005***	−0.058***	0.005***	−0.058***	0.005***	−0.058***	0.010***	−0.065***
	(0.00004)	(0.0002)	(0.00004)	(0.0002)	(0.00004)	(0.0002)	(0.00009)	(0.0003)
Violation Sequence Length	0.014***	−0.003***	0.016***	−0.003***	0.018***	−0.003***	0.014***	0.004***
	(0.0002)	(0.0004)	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)	(0.0003)
Fast Heartbeat	–	–	0.198***	−0.018***	0.203***	−0.006	0.203***	−0.006**
			(0.007)	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)
Target Corridor (bp)			−0.004***	−0.0003***	−0.004***	−0.0003	−0.003***	−0.001***
			(0.00007)	(0.00005)	(0.00007)	(0.00005)	(0.00006)	(0.00005)
Median Price per Gas					−0.0005***	0.0002***	−0.0005***	0.0001***
					(0.00002)	(0.000009)	(0.00002)	(0.00001)
Binance Move (t-2)							−0.007***	0.008***
							(0.0002)	(0.0002)
Binance Move (t-24)							−0.003***	0.004***
							(0.00006)	(0.00005)
Binance USDT Volume (m)					0.100***	−0.067***	0.109***	−0.084***
					(0.003)	(0.006)	(0.003)	(0.006)
Constant	−3.465***	−1.636***	−3.151***	−1.599***	−3.158***	−1.596***	−3.050***	−1.796***
	(0.004)	(0.004)	(0.01)	(0.007)	(0.010)	(0.007)	(0.01)	(0.008)
Akaike Inf. Crit.	2,046,958		2,040,818		2,038,571		2,021,476	

Note: *p < 0.1; **p < 0.05; ***p < 0.01.

Table D.20
Sample restricted to the post-Merge consensus regime. Multinomial logistic regression for transitions out of a violation state with two alternatives: CPF-update ("CPF") vs. random ground truth price movement ("Random"); "Stay" is the reference. Columns (1)–(4) mirror the baseline specifications.

	Multinomial logistic regression for state transitions from violation (Reference = Stay)							
	Model Spec. 1:		Model Spec. 2:		Model Spec. 3:		Model Spec. 4:	
	CPF	Random	CPF	Random	CPF	Random	CPF	Random
Over Deviation (bp, t-1)	0.003***	−0.063***	0.004***	−0.063***	0.003***	−0.064***	0.005***	−0.071***
	(0.00004)	(0.00037)	(0.00004)	(0.00037)	(0.00004)	(0.00037)	(0.00008)	(0.00038)
Violation Sequence Length	−0.002***	−0.011***	−0.001***	−0.011***	−0.001***	−0.011***	−0.002***	−0.005***
	(0.00018)	(0.00033)	(0.00017)	(0.00034)	(0.00017)	(0.00034)	(0.00018)	(0.00029)
Fast Heartbeat			0.233***	0.076***	0.210***	0.091***	0.212***	0.077***
			(0.00729)	(0.00681)	(0.00738)	(0.00685)	(0.00739)	(0.00687)
Target Corridor (bp)			−0.00319***	0.00052***	−0.00307***	0.00035***	−0.00297***	−0.00025***
			(0.00007)	(0.00005)	(0.00007)	(0.00005)	(0.00007)	(0.00005)
Median Price per Gas					0.00188***	0.00227***	0.00194***	0.00069***
					(0.00007)	(0.00008)	(0.00007)	(0.00009)
Binance Move (t-2)							−0.00625***	0.00494***
							(0.00023)	(0.00028)
Binance Move (t-24)							−0.00024***	0.00378***
							(0.00005)	(0.00005)
Binance USDT Volume (m)					0.06960***	−0.12363***	0.07520***	−0.14517***
					(0.00344)	(0.00693)	(0.00343)	(0.00723)
Constant	−2.728***	−1.642***	−2.435***	−1.728***	−2.506***	−1.768***	−2.476***	−1.856***
	(0.00399)	(0.00404)	(0.00957)	(0.00836)	(0.00977)	(0.00871)	(0.00984)	(0.00886)
Akaike Inf. Crit.	1,561,905		1,556,459		1,554,363		1,546,087	

Note: *p < 0.1; **p < 0.05; ***p < 0.01.

D.3. Pre- & post-Merge supplementary data

Total transition probabilities (from Section 5.2):

$$P' = \begin{bmatrix} P_{GG} & P_{GB} \\ P^c_{BG} + P^r_{BG} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0.9965 & 0.0035 \\ 0.0497 + 0.0863 & 0.8640 \end{bmatrix}.$$

Pre-Merge transition probabilities:

$$P'_{pre\_merge} = \begin{bmatrix} P_{GG} & P_{GB} \\ P^c_{BG} + P^r_{BG} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0.9948 & 0.0052 \\ 0.0398 + 0.0829 & 0.8773 \end{bmatrix}.$$

Post-Merge transition probabilities:

$$P'_{post\_merge} = \begin{bmatrix} P_{GG} & P_{GB} \\ P^c_{BG} + P^r_{BG} & P_{BB} \end{bmatrix} = \begin{bmatrix} 0.9980 & 0.0020 \\ 0.0658 + 0.0918 & 0.8424 \end{bmatrix}.$$

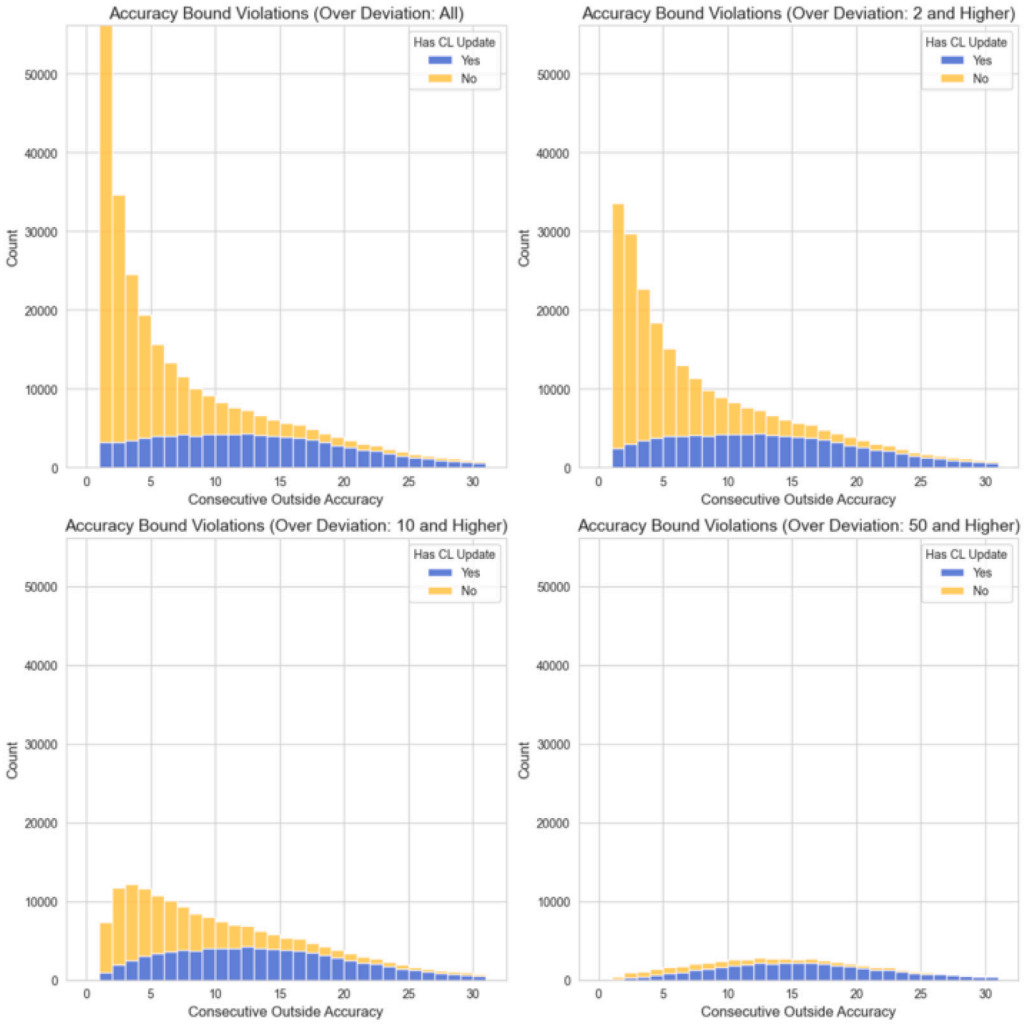


Fig. D.7. State transitions from "Bad" (target corridor violation) to "Good" (within corridor) for Chainlink Price Feeds (CPFs) observed before the Ethereum Merge. Bars split recoveries into CPF-update resolutions versus random market price realignments. Panels condition on violation sequence length and maximum over-deviation.

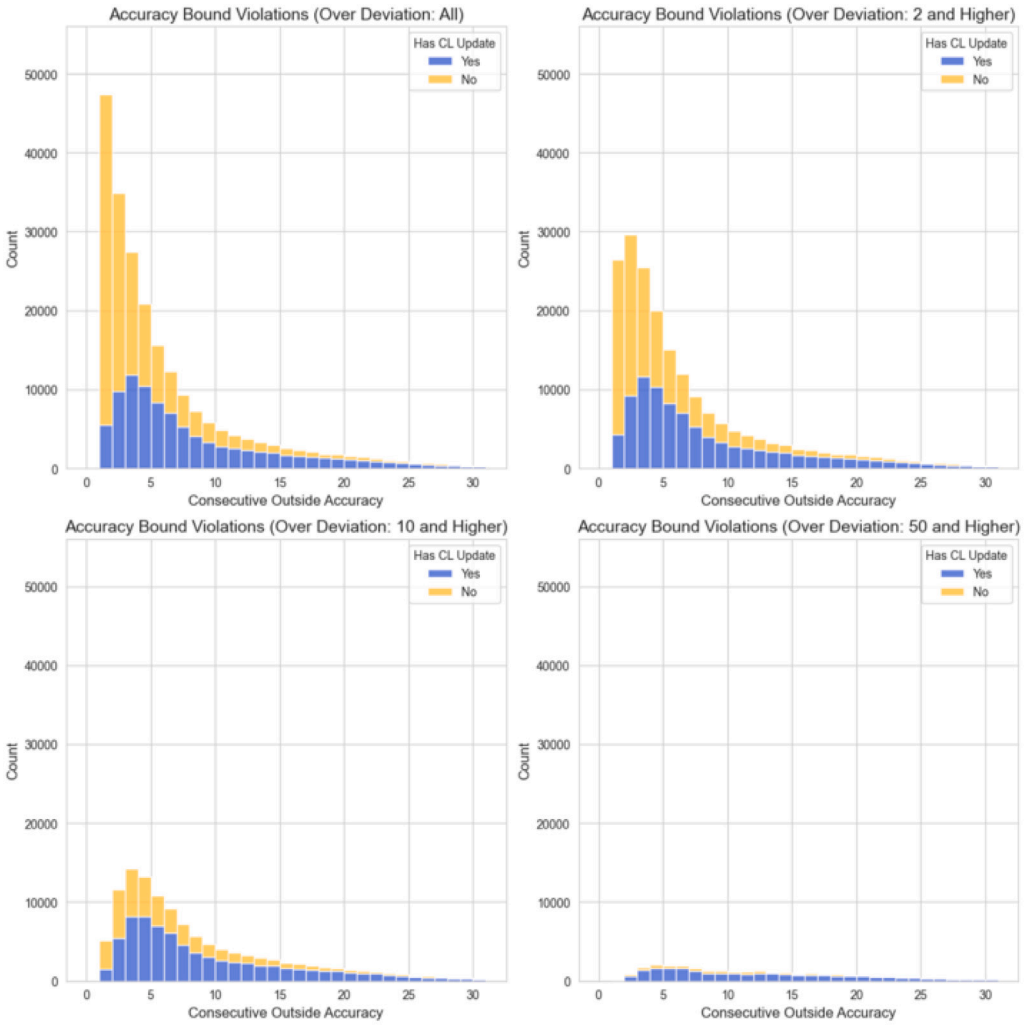


Fig. D.8. State transitions from "Bad" (target corridor violation) to "Good" (within corridor) for Chainlink Price Feeds (CPFs) observed after the Ethereum Merge. Bars split recoveries into CPF-update resolutions versus random market price realignments. Panels condition on violation sequence length and maximum over-deviation.

Appendix E. Robustness capitalization choice analysis

E.1. Outlier treatment

Table E.21
This specification is identical to the baseline, except that the sophistication score (first principal component of collateral size, debt size, and account type) is log-transformed to reduce the influence of extreme outliers. Independent variables include CPF heartbeat interval, target corridor, position-type dummies (crypto-stable, stable-crypto, crypto-crypto), and volatility measures (7-day and 1-year), with Post-Merge and aggregator version indicators as controls. Robust standard errors are reported.

	Dependent variable: hf	
	(2)	(3)
Heartbeat (h)	−0.0036	−0.0063**
	(0.0027)	(0.0028)
Target Corridor (bp)	0.0012***	0.0013***
	(0.0003)	(0.0003)
Agg. V4	−0.0451**	−0.0407**
	(0.0178)	(0.0178)
Log Soph. Score	−0.4808***	−0.4766***
	(0.0152)	(0.0152)
Crypto-Stable	0.2388***	0.0602**
	(0.0212)	(0.0284)
Stable-Crypto	0.2109***	0.0306
	(0.0218)	(0.0289)
Crypto-Crypto	0.1727***	0.0004
	(0.0365)	(0.0405)
Volatility (7d)		0.1949
		(0.4276)
Volatility (1y)		3.7441***
		(0.4042)
Observations	146 620	146 620
N. of groups	15 107	15 107
R^2	0.0188	0.0198

Note: Robust standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

E.2. Subsample analyses

Table E.22
This specification is identical to the baseline, except that it includes time fixed effects (Daily bins, 7200 blocks). Independent variables include CPF heartbeat interval, target corridor, sophistication score, position-type dummies (crypto-stable, stable-crypto, crypto-crypto), and volatility measures (7-day and 1-year), with Post-Merge and aggregator version indicators as controls. Robust standard errors are reported.

	Dependent variable: hf		
	(1)	(2)	(3)
Heartbeat (h)	−0.0146***	−0.0046*	−0.0075***
	(0.0019)	(0.0028)	(0.0028)
Target Corridor (bp)	0.0028***	0.0016***	0.0017***
	(0.0002)	(0.0003)	(0.0003)
Agg. V4	−0.0878***	−0.0609***	−0.0560***
	(0.0175)	(0.0177)	(0.0177)
Soph. Score		−0.0340***	−0.0335***
		(0.0038)	(0.0037)
Crypto-Stable		0.2567***	0.0634**
		(0.0213)	(0.0285)
Stable-Crypto		0.2370***	0.0418
		(0.0219)	(0.0291)
Crypto-Crypto		0.1954***	0.0089
		(0.0367)	(0.0407)
Volatility (7d)			0.2796
			(0.4312)
Volatility (1y)			4.0344***
			(0.4041)
Observations	146 620	146 620	146 620
N. of groups	15 107	15 107	15 107
R^2	0.0022	0.0073	0.0085

Note: Robust standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Table E.23
This specification is identical to the baseline, except that it splits the sample into stable and nonstable collateral subsets. This reduces the number of position-type dummies from four to two (stable-crypto and stable-stable, or crypto-stable and crypto-crypto respectively). The reference category for this dummy variable is different for the two models and for the nonstable collateral model (columns 4-6), the reference category (crypto-stable) also differs from the baseline specification. Other Independent variables include CPF heartbeat interval, target corridor, sophistication score, and volatility measures (7-day and 1-year), with Post-Merge and aggregator version indicators as controls. Robust standard errors are reported.

	Dependent variable: hf					
	Stable collateral			Nonstable collateral		
	(1)	(2)	(3)	(4)	(5)	(6)
Heartbeat (h)	−0.0266***	0.0061	0.0030	−0.0028	−0.0077**	−0.0111***
	(0.0061)	(0.0068)	(0.0068)	(0.0026)	(0.0034)	(0.0034)
Target Corridor (bp)	0.0050***	0.0011*	0.0013*	0.0013***	0.0019***	0.0021***
	(0.0006)	(0.0007)	(0.0007)	(0.0003)	(0.0004)	(0.0004)
Agg. V4	−0.1693***	−0.1553***	−0.1481***	−0.0423**	−0.0506**	−0.0432**
	(0.0551)	(0.0542)	(0.0545)	(0.0202)	(0.0205)	(0.0205)
Soph. Score		−0.1501***	−0.1507***		−0.0254***	−0.0248***
		(0.0095)	(0.0096)		(0.0031)	(0.0030)
Stable-Crypto		0.2159***	0.0765	–	–	–
		(0.0337)	(0.0567)			
Crypto-Crypto	–	–	–		−0.0642**	−0.0620**
					(0.0267)	(0.0267)
Volatility (7d)			1.5526			0.1248
			(1.4812)			(0.4590)
Volatility (1y)			2.5436**			4.5480***
			(1.0298)			(0.4863)
Post Merge	0.1620***	0.1657***	0.1740***	−0.1034***	−0.1030***	−0.0833***
	(0.0173)	(0.0171)	(0.0174)	(0.0086)	(0.0086)	(0.0088)
Observations	31 071	31 071	31 071	115 549	115 549	115 549
N. of groups	3522	3522	3522	12 640	12 640	12 640
R^2	0.0095	0.0252	0.0258	0.0022	0.0039	0.0053

Note: Robust standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Table E.24
This specification is identical to the baseline, except that it splits the sample into above and below median sophistication score subsets. Independent variables include CPF heartbeat interval, target corridor, sophistication score, position-type dummies (crypto-stable, stable-crypto, crypto-crypto), and volatility measures (7-day and 1-year), with Post-Merge and aggregator version indicators as controls. Robust standard errors are reported.

	Dependent variable: hf					
	High sophistication			Low sophistication		
	(1)	(2)	(3)	(4)	(5)	(6)
Heartbeat (h)	−0.0111***	0.0050	0.0025	−0.0143***	−0.0110***	−0.0130***
	(0.0029)	(0.0042)	(0.0043)	(0.0027)	(0.0039)	(0.0039)
Target Corridor (bp)	0.0025***	0.0006	0.0008	0.0025***	0.0022***	0.0022***
	(0.0003)	(0.0005)	(0.0005)	(0.0003)	(0.0005)	(0.0005)
Agg. V4	−0.1153***	−0.0757***	−0.0709**	−0.0536**	−0.0608**	−0.0593**
	(0.0280)	(0.0284)	(0.0284)	(0.0237)	(0.0242)	(0.0242)
Soph. Score		−0.0329***	−0.0326***		−19.4036***	−19.2729***
		(0.0036)	(0.0035)		(0.7614)	(0.7595)
Crypto-Stable		0.3522***	0.2431***		0.0887***	−0.1199***
		(0.0305)	(0.0411)		(0.0320)	(0.0435)
Stable-Crypto		0.3297***	0.2190***		0.0822**	−0.1298***
		(0.0309)	(0.0416)		(0.0349)	(0.0463)
Crypto-Crypto		0.3092***	0.1980***		0.0128	−0.1818***
		(0.0535)	(0.0599)		(0.0546)	(0.0612)
Volatility (7d)			0.3776			−0.2582
			(0.6058)			(0.6551)
Volatility (1y)			2.2320***			4.5095***
			(0.5735)			(0.6362)
Post Merge	−0.0228**	−0.0180*	−0.0085	−0.0840***	−0.0823***	−0.0682***
	(0.0103)	(0.0102)	(0.0105)	(0.0124)	(0.0123)	(0.0124)
Observations	73 310	73 310	73 310	73 310	73 310	73 310
N. of groups	5393	5393	5393	12 136	12 136	12 136
R^2	0.0023	0.0126	0.0130	0.0024	0.0156	0.0168

Note: Robust standard errors. *p < 0.1; **p < 0.05; ***p < 0.01.

Data availability

Data will be made available on request.

References

Al-Breiki, H., Rehman, M.H.U., Salah, K., Svetinovic, D., 2020. Trustworthy blockchain oracles: Review, comparison, and open research challenges. IEEE Access 8, 85675–85685.

Albizri, A., Appelbaum, D., 2021. Trust but verify: The oracle paradox of blockchain smart contracts. J. Inf. Syst. 35 (2), 1–16.

Angeris, G., Chitra, T., 2020. Improved price oracles: Constant function market makers. In: Proceedings of the 2nd ACM Conference on Advances in Financial Technologies. pp. 80–91.

Aspembitova, A.T., Bentley, M.A., 2023. Oracles in decentralized finance: Attack costs, profits and mitigation measures. Entropy 25 (1).

Aspris, A., Foley, S., Svec, J., Wang, L., 2021. Decentralized exchanges: The "wild west" of cryptocurrency trading. Int. Rev. Financ. Anal. 77, 101845.

Bartoletti, M., Chiang, J.H.-y., Lafuente, A.L., 2021. SoK: lending pools in decentralized finance. In: Financial Cryptography and Data Security. FC 2021 International Workshops: CoDecFin, DeFi, VOTING, and WTSC. pp. 553–578.

Bessembinder, H., Maxwell, W., Venkataraman, K., 2006. Market transparency, liquidity externalities, and institutional trading costs in corporate bonds. J. Financ. Econ. 82 (2), 251–288.

Breidenbach, L., Cachin, C., Coventry, A., Juels, A., Miller, A., 2021. Chainlink off-chain reporting protocol. Version 1.2. (Accessed 31 July 2023).

Brennan, M.J., Jegadeesh, N., Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. Rev. Financ. Stud. 6 (4), 799–824.

Buterin, V., 2017. On path independence. (Accessed 28 August 2023).

Cong, L.W., Li, X., Tang, K., Yang, Y., 2023. Crypto wash trading. Manag. Sci. 69 (11), 6427–6454.

Cong, L.W., Prasad, E.S., Rabetti, D., 2025. Financial and Informational Integration Through Oracle Networks. NBER Working Paper 33639, National Bureau of Economic Research.

D'Souza, J.M., Ramesh, K., Shen, M., 2010. The interdependence between institutional ownership and information dissemination by data aggregators. Account. Rev. 85 (1), 159–193.

Duffie, D., Stein, J.C., 2015. Reforming LIBOR and other financial market benchmarks. J. Econ. Perspect. 29 (2), 191–212.

Duley, C., Gambacorta, L., Garratt, R., Koo Wilkens, P., 2023. The Oracle Problem and the Future of DeFi. BIS Bulletin 76, Bank for International Settlement.

Epps, T.W., Epps, M.L., 1976. The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. Econometrica 44 (2), 305–321.

Eskandari, S., Salehi, M., Gu, W.C., Clark, J., 2021. Sok: Oracles from the ground truth to market manipulation. In: Proceedings of the 3rd ACM Conference on Advances in Financial Technologies. pp. 127–141.
Gu, W.C., Raghuvanshi, A., Boneh, D., 2021. Empirical measurements on pricing oracles and decentralized governance for stablecoins. Cryptoeconomic Syst. 1 (2).
Heimbach, L., Wattenhofer, R., 2022. Sok: Preventing transaction reordering manipulations in decentralized finance. In: AFT '22: Proceedings of the 4th ACM Conference on Advances in Financial Technologies. pp. 47–60.
Kaleem, M., Shi, W., 2021. Demystifying pythia: A survey of ChainLink oracles usage on ethereum. In: FC 2021: Financial Cryptography and Data Security. pp. 115–123.
Lehar, A., Parlour, C.A., 2022. Systemic Fragility in Decentralized Markets. BIS Working Paper 1062, Bank for International Settlement - Monetary and Economic Department.
Lehar, A., Parlour, C.A., 2025. Decentralized exchange: The uniswap automated market maker. J. Financ. 80 (1), 321–374.
Lin, L.X., Budish, E., Cong, L.W., He, Z., Bergquist, J.H., Panesir, M.S., Kelly, J., Lauer, M., Prinster, R., Zhang, S., et al., 2019. Deconstructing decentralized exchanges. Stanf. J. Blockchain Law Policy 2 (1), 58–77.
Liu, B., Szalachowski, P., Zhou, J., 2021. A first look into DeFi oracles. In: 2021 IEEE International Conference on Decentralized Applications and Infrastructures. DAPPS, pp. 39–48.
Mackinga, T., Nadahalli, T., Wattenhofer, R., 2022. TWAP oracle attacks: Easier done than said? In: 2022 IEEE International Conference on Blockchain and Cryptocurrency. ICBC.
Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–708.
Pasdar, A., Lee, Y.C., Dong, Z., 2023. Connect API with blockchain: A survey on blockchain oracle implementation. ACM Comput. Surv. 55 (10), 1–39.
Perez, D., Werner, S.M., Xu, J., Livshits, B., 2021. Liquidations: DeFi on a knife-edge. In: Borisov, N., Diaz, C. (Eds.), FC 2021: Financial Cryptography and Data Security. pp. 457–476.
Qin, K., Zhou, L., Livshits, B., Gervais, A., 2021. Attacking the DeFi ecosystem with flash loans for fun and profit. In: International Conference on Financial Cryptography and Data Security. pp. 3–32.
Schär, F., 2021. Decentralized finance: On blockchain- and smart contract-based financial markets. Rev. Fed. Reserv. Bank St Louis 103 (2), 153–174.
Schuler, K., Nadler, M., Schär, F., 2023. Contagion and loss redistribution in crypto asset markets. Econom. Lett. 231, 111310.
Wang, S.-H., Wu, C.-C., Liang, Y.-C., Hsieh, L.-H., Hsiao, H.-C., 2021. ProMutator: Detecting vulnerable price oracles in DeFi by mutated transactions. In: 2021 IEEE European Symposium on Security and Privacy Workshops. pp. 380–385.
Werner, S., Perez, D., Gudgeon, L., Klages-Mundt, A., Harz, D., Knottenbelt, W., 2022. Sok: Decentralized finance (DeFi). In: Proceedings of the 4th ACM Conference on Advances in Financial Technologies. pp. 30–46.
Woo, S., Song, J., Park, S., 2020. A distributed oracle using intel SGX for blockchain-based IoT applications. Sensors 20 (9).
Zhang, F., Cecchetti, E., Croman, K., Juels, A., Shi, E., 2016. Town crier: An authenticated data feed for smart contracts. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. pp. 270–282.
Zhao, Y., Kang, X., Li, T., Chu, C.-K., Wang, H., 2022. Toward trustworthy DeFi oracles: Past, present, and future. IEEE Access 10, 60914–60928.
Zhou, L., Xiong, X., Ernstberger, J., Chaliasos, S., Wang, Z., Wang, Y., Qin, K., Wattenhofer, R., Song, D., Gervais, A., 2023. Sok: Decentralized finance (DeFi) attacks. In: 2023 IEEE Symposium on Security and Privacy. SP, pp. 2444–2461.