WORKING PAPER 24-08

Does Lock-up Lead to Stability? Implications for Runs in the Proof-of-Stake Protocol

Samuel Hempel
Federal Reserve Board of Governors
sam.hempel@frb.gov

Gregory Phelan
Department of Economics, Williams College
gp4@williams.edu

Thomas Ruchti
Office of Financial Research
thomas.ruchti@ofr.treasury.gov

OFR
OFFICE OF FINANCIAL RESEARCH

October 31, 2024

Why These Findings Are Important

Stakers invest capital to facilitate transactions under the Proof-of-Stake (PoS) protocol, which is replacing Proof-of-Work (PoW) for most crypto currencies. PoS is less energy intensive but requires more capital. Because stakers put capital at risk with PoS, there is a higher risk that they may coordinate an exit—a form of run. A run on a major PoS blockchain, like Ethereum, could undermine security for the crypto asset and disrupt broader crypto markets dependent on that blockchain. In this paper, the authors examine the effect of margin, lock-up periods, and low rewards to stakers on the risk of runs.

Key Findings

1. **Low rewards for staking increase run risk.**
2. **Using leverage, or margin, when proposing blocks can also increase run risk**
3. **Longer lock-up periods reduce but do not eliminate run risk.**

How the Authors Reached These Findings

The authors model different scenarios that incorporate run risks to demonstrate how runs can occur in any protocol that relies on voluntary lock-up periods to validate transactions. Those protocols with more liquid funds are shown to have less run risk. In theory, low rewards improve security of PoS by reducing the incentive for neutral parties to participate in co-opting the blockchain by nefarious actors. However, low rewards exacerbate run risk.

Does lock-up lead to stability?
Implications for runs in the Proof-of-Stake protocol

Samuel Hempel, Gregory Phelan, and Thomas Ruchti*

October 31, 2024

Abstract

Blockchains increasingly rely on the capital-intensive Proof-of-Stake protocol over the energy-intensive Proof-of-Work protocol to propose blocks, putting those blockchains at risk of capital withdrawal that could undermine consensus and security. We model a population of investors who decide to stake, reaping staking rewards, or exit, liquidating their crypto-asset holdings. Runs on staking are more common for weak protocols, when price impacts of protocol failure are high, or when rewards to staking are low. We extend the model to leveraged staking, where margin calls exacerbate run dynamics. In examining staking lock-up periods, we find that a longer lock-up period can reduce runs but does not eliminate them. Previous work demonstrates that consensus and security of a crypto-asset depend on low staking rewards, but our results highlight that low rewards induce runs. If a run were to occur on a major Proof-of-Stake backed blockchain, such as Ethereum, this would undermine security, potentially disrupting crypto markets dependent on that blockchain and the Decentralized Finance networks that depend on it.

*Hempel: Federal Reserve Board of Governors, email: sam.hempel@frb.gov. Phelan: Department of Economics, Williams College, email: gp4@williams.edu. Ruchti: Office of Financial Research, Department of the Treasury, email: thomas.ruchti@ofr.treasury.gov. We are grateful for feedback from Dagmar Chiella, Jean Flemming, Ben Gillen, Xuesong Huang, Jessica Hurst Francisco Ilabaca, Charles Kahn, Mark Paddrik, Nagpurnanand Prabhala, Sriram Rajan, Danylo Rakowsky, Romina Ruprecht, Fahad Saleh, and Kevin Zhao. We also appreciate questions and feedback from seminar participants at the OFR and from conference participants at the 2024 Federal Reserve Summer Workshop on Money, Banking, Payments, and Finance. Any errors are our own.

1 Introduction

On September 15, 2022, Ethereum, the second most valuable crypto-asset by market capitalization, executed The Merge, instituting a form of the Proof-of-Stake (PoS) protocol. Under PoS protocols, new blocks in a decentralized ledger are proposed by computers directly or indirectly operated by owners of the associated coin.[1] Validators, or *stakers*, contribute tokens to a staking contract. They are then tasked with maintaining and operating a computer node, which holds a record of the public ledger, or blockchain, of previous transactions.[2] Practitioners have touted improved security and reduced energy requirements of PoS, allowing for greater scalability while imposing no explicit cost on validators.[3,4] Concomitant with these benefits, PoS protocols require substantial capital from validators to ensure consensus in transactions and security from attacks.[5] Given the preponderance of the PoS protocol, a sudden asset devaluation—as in a run on the staking contract—could precipitate accompanying vulnerability of consensus and security. Such a failure would disrupt crypto markets, undermining the many Decentralized Finance (DeFi) networks that depend on it.

[1] In practice, the owner of a coin can engage in solo staking, pay for a service to propose blocks on their behalf, or invest their coin in a staking pool.

[2] While algorithms vary, validators are, proportionate to their stake, offered the opportunity to propose new blocks, wherein they bundle transactions and propose them en masse to the public ledger. Validator nodes are also recruited to perform checks on the work of other validator nodes.

[3] https://www.cnn.com/2022/09/15/tech/ethereum-merge-crytocurrency-energy-consumption-hnk-intl; https://www.reuters.com/technology/ethereums-energy-saving-merge-upgrade-2022-09-15/; https://time.com/nextadvisor/investing/cryptocurrency/proof-of-work-vs-proof-of-stake/; https://medium.com/coinmonks/the-merge-ethereums-vision-to-more-scalability-security-and-sustainability-9176cd3b6a78

[4] A common claim at the time of The Merge was that PoS would result in faster transaction speeds, while in fact reduced latency involves separate changes to the protocol, such as those available with blockchain sharding (https://fortune.com/crypto/2022/10/04/vitalik-buterin-lays-out-ethereum-post-merge-roadmap/).

[5] Full staking rewards accruing only to those locking up significant amounts. The level for Ethereum is 32 ETH—roughly $50,000 at the time of The Merge.

In this paper, we identify and examine run risk as an important trade off present in the PoS protocol, even when such protocols include mandatory lock-up periods, restricting stakers from withdrawing staked capital. Our approach demonstrates that run risk is a concern in any protocol that depends on the voluntary locking up of otherwise liquid funds to aid in incentive compatible transaction validation. Thus, while PoS protocols have the potential for improved security, we show that the inherent design of these protocols creates incentives for run risk, thus diminishing the potential security gains for PoS.

We first establish run risks in a simple, static model of staking. In the model, investors can choose to stake or exit their positions. Investors are therefore rewarded for staking and trade this off with the risk that the crypto-asset fails due to an attack on the protocol. The security of the protocol induces a coordination problem in the choice to stake or exit. We find that runs on staking are more common when the crypto-asset's protocol is weak and susceptible to scale-based attacks and forking of the chain, when the price impacts of protocol failure are high, or when rewards to staking are low.

Our static model is motivated by the well-understood possibility of a 51% attack on a public ledger protocol (Sayeed and Marco-Gisbert, 2019). In this attack, a malevolent actor or group of actors acquires 51% of validation power, allowing them to alter the chain in a way that diverts currency away from other owners. Such attacks are a greater risk if stakers exit.[6] Security is therefore dependent both on the number of validators and their consistent participation in the validation of transactions. Given the PoS protocol's depen-

[6]Most attacks on crypto-assets depend on distorting the transactions recorded on the blockchain, something that is easier to accomplish if total staking is low.

dence on capital, a run could materialize in which investors pull their tokens, depleting the pool of validators, which in turn would inhibit the protocol's security and the participation of validator nodes. Security issues of transactions or discrepancies over the finality of transactions could lead to a decrease in the value of the crypto asset, which could, in turn, lead to further depletion of the pool of validators.

This analysis supposes that stakers can instantaneously exit the pool, but in practice, PoS protocols typically place implicit or explicit restrictions on unstaking or removing staked coins from the protocol's mechanism for validating transactions. These restrictions may be implicit, such as having to enter a queue to remove the stake, or explicit, wherein restrictions are meant to reduce the risk that too many of the protocol's stakers exit at once. The concern is that too many exits at once leave the protocol without validators for transactions, collapsing the protocol. We model these restrictions in a continuous time framework in which investors trade off validating transactions by entering a queue to exit the protocol, similar to the process for the Ethereum chain.[7] Applying insights from Guimaraes (2006), we show that lowering the liquidity of the protocol, or speed with which stakers can exit, can increase the time it takes for a run to lead to failure of the protocol. This increased run time is accompanied by reduced run risk as well. In other words, there is a greater chance that a longer run is abandoned, reducing the incentive to run in the first place. However, the value of a PoS currency is lower when withdrawal restrictions are greater.

One of the central problems for financial institutions is how to accurately and effi-

[7]When Ethereum executed the Shanghai-Capella upgrade on April 12, 2023, it allowed validators to fully exit, but only after completing necessary steps, such as providing a withdrawal address and broadcasting a "voluntary exit" message. Once these steps are completed, exiting stakers are placed in a queue. Exit is contingent on congestion at the time.

ciently execute, clear, and settle transactions. Valid transactions conform to a set of pre-specified rules and reflect the actions and intent of the respective parties. Validating transactions requires an entity or protocol to ensure that a transfer of funds conforms to these rules. The integrity of this process requires that something dear to the validator is at stake. In traditional payments systems, transactions pass through centralized bodies that act as guarantors of transactions. Centralized validators put their reputation on the line when accepting or rejecting proposed transactions.

In contrast, in blockchains, transactions are proposed to a public ledger, or blockchain, and they are deemed valid through the general acceptance of participating parties in the updated public ledger. Validation under PoW, employed for cryptocurrencies such as Bitcoin, rewards fastest computed solutions to complex hash functions. In this case, computer hardware and electricity consumption is at stake for validators. PoW is arguably secure (Nakamoto, 2008) and theory posits that there exists a longest chain equilibrium that achieves consensus (Biais et al., 2019a,b). However, the fact that the PoW protocol puts electricity consumption at stake means that it has the potential to consume a significant amount of electricity.

PoS protocols such as Ethereum, in contrast, are energy efficient. Instead of hardware use and electricity consumption, PoS requires that validators put significant capital at stake. On the one hand, the literature has shown that sufficiently small staking rewards allow PoS protocols to achieve security (John et al., 2021) and consensus (Saleh, 2021). On the other hand, our results demonstrate that small staking rewards increase the risk of runs. In other words, using the PoS protocol produces a tradeoff between having a

secure, consensus-inducing blockchain and having low run risk.[8] [9]

We contribute to the literature on the economics of staking protocols. Saleh (2021) shows that for PoS to achieve consensus, where a single dominant branch exists for a coin, requires sufficiently small staking rewards. John et al. (2021) argue that maximizing the participation in a coin, a pre-condition for security, also requires staking rewards that are not so high as to preclude coin adoption by investors with greater trading needs. In this paper, we find that reducing the risk of runs on a funding-based protocol like PoS depend on high staking rewards. Considering our results in conjunction with the findings in these papers, a negative result emerges in the design of a PoS crypto-asset. A coin can have increased consensus and security from attacks via low rewards, but it then will exhibit higher run risk. Conversely, a coin can reduce its run risk with high rewards, but this results in reduced consensus and security of the coin.[10] We investigate whether there are financial risks to PoS mechanisms. PoS is dependent on validators staking their coins, with full staking rewards accruing only to those locking up significant amounts.

[8] Proof-of-History (PoH) protocols depend on a shared and verifiable history among nodes via a verifiable delay function that hashes incoming events, noting when events occur. This protocol is energy efficient and is not susceptible to runs but is not as secure as other chains, as evidenced by the history of bot attacks flooding protocols like Solana with transactions. For details on the Saturday, April 30, 2022 attack on Solana, see https://finance.yahoo.com/news/solana-loses-consensus-bots-flood-023947602.html.

[9] If a protocol failure or a run were to occur on the Ethereum network, it would be potentially disruptive to the significant share of decentralized finance (DeFi) that depends on the Ethereum blockchain.

[10] We do not embed either Saleh (2021) or John et al. (2021) in our analysis. However, the presence of run risks could exacerbate short-term reward-driven incentives in Saleh (2021) and could harm the participation in a PoS coin, as studied in John et al. (2021). Uncertainty over the threats to security studied in both papers serve as the impetus to runs in our setting.

2 Background

2.1 Proof-of-Stake adoption

Proof-of-Stake (PoS) was first proposed by King and Nadal (2012), who viewed it as part of a hybrid alternative to Proof-of-Work (PoW) that would be more energy-efficient. As Saleh (2022) notes, the first pure PoS platform was the Nxt blockchain launched in 2013. In 2018, The New York Times published an article examining the energy consumption of Bitcoin (a PoW blockchain), with discussion of Ethereum potentially transitioning from PoW to PoS in the future.[11]

In October 2020, the staking contract for Ethereum was launched, and in November 2020, the staking contract began accepting deposits from users in increments of 32 ETH.[12] Following a series of upgrades over the next 22 months, Ethereum's Mainnet chain, validated via PoW, was merged with its Beacon chain, validated via PoS, with a shared history. PoS now validates transactions on Ethereum, the second largest cryptocurrency by marketcap. This fact alone makes PoS systemically important to the crypto and digital assets sector, given it supports most decentralized finance (DeFi), decentralized apps (dApps) activity, and non-fungible tokens (NFTs), comprising a vast array of Layer 2 protocols all built on the Ethereum chain.

Beyond Ethereum and the Merge, PoS is commonplace. There are 11 blockchains with a market capitalization over $1 billion USD that run on PoS but only nine that run on

[11] Nathaniel Popper. There Is Nothing Virtual About Bitcoin's Energy Appetite. The New York Times. January 21, 2018.

[12] See Etherscan for the precise details of the Ethereum deposit contract, including the source code and all known transactions. https://etherscan.io/address/0x00000000219ab540356cBB839Cbe05303d7705Fa

PoW.[13] The total market cap of PoS blockchains is $639 billion, whereas the total market cap of PoW blockchains is $1.38 trillion. However, if you remove the Bitcoin blockchain, the aggregate size of PoW blockchains is only $190 billion.

2.2 How staking works

There are several ways to participate in staking. Fundamental staking by investors is often referred to as *solo* staking, or home staking. This is the only method by which investors can directly stake and that does not require a third party. It therefore allows for an investor to participate in all the benefits and costs of staking. It entails both the contribution of staked coin, such as 32 ETH (approximately $125,000 as of this writing) on the Ethereum platform, and validation services via the contribution of hardware or a *node*, which we discuss in greater detail below.

Staking as a service (Saas) is another way to stake, which typically employs third party provided hardware and bandwidth while the owner controls keys and the assets staked. Finally, pooled staking involves the contribution of a non-standard amount to a broader pool that is managed and staked on the investors' behalf. If a significant number of individuals withdraw their deposits from pooled staking, this could result in a liquidity-based run for the fundamental staker. We examine such a scenario theoretically in Appendix B.4.

Using Ethereum as an example, fundamental or solo staking requires that an investor runs a node or computer hardware validator of the coin's relevant blockchain. To fully

[13] We use the term blockchain as a shorthand for the crypto-assets native to a particular blockchain. For example, BNB (Binance Coin) is the native coin of the Binance Chain, which means its market capitalization is a lower bound on the sum of all activity hosted on the Binance Chain. See Cryptoslate.com for data by consensus mechanism (Irresberger et al., 2023). Data as of March 7, 2024.

participate without being dropped from staking, stakers must run their node at all times. The node runs an execution layer client, which collects transactions, executes them, and records the up-to-date state of the blockchain. The node also runs a consensus layer client, which allows the network of nodes to achieve agreement based on validated transactions data. Finally, stakers generate keys that allow them to deposit coins in the staking contract, essentially locking them up for future use.

Coins remain staked until they are withdrawn. Withdrawing coins from the PoS protocol is not immediate. In the case of Ethereum, stakers must first provide a withdrawal address, broadcasting that they are voluntarily exiting the protocol via aforementioned validator keys via the node's validator client. Block-proposing validators determine if there are withdrawals by sweeping through other validators by their indexed number, starting at 0. A queue is determined, producing congestion in validator exit.

While there are no explicit conditions locking up staked coins, such a procedure slows down if there is congestion. Only so many validators can exit at a time, meaning that a staker could have to wait multiple days to exit the protocol. In the case of Ethereum, queues have reached upwards of a week to two weeks at a time, depending on outflows.[14]

2.3 Block rewards under Proof-of-Stake

Across most consensus mechanisms employed by blockchains, there is a fundamental goal to incentivize honest participation in the validation process while thwarting nefarious activity. For PoS, this consensus mechanism is built on the premise that validators

[14]https://markets.businessinsider.com/news/currencies/ethereum-validators-forced-to-wait-days-to-unstake-amid-celsius-withdraws-1032946687

have a *stake* in the viability of the token. To ensure that potential validators have a stake, PoS protocols typically require validators to hold some amount of the token in an illiquid deposit contract. Then transactions can be validated only by those participants who have a stake in the protocol.

In exchange for locking up some portion of their token holdings in the deposit contract, stakers in PoS protocols are allowed to participate in a lottery-like process where opportunities to propose new blocks are distributed randomly to stakers, a "Follow the Satoshi" mechanism wherein a coin among all staked coins is chosen at random with the probability of being selected equal to the staker's market share of total deposits in the contract .[15] An immediate consequence of this design is that the largest stakers are likely to dominate the consensus process, raising potential concerns about concentration risk (Irresberger and Yang, 2023).

There are several ways in which validators receive rewards in a PoS mechanism. Using Ethereum as an example, once a staker's proposed block receives attestations and signatures, coins are given as a reward. Rewards are also given for checking new blocks and attesting to them if they are valid, wherein rewards are given if the staker votes with the majority. Furthermore, stakers may earn greater rewards for proposing blocks that include decentralized exchange arbitrage, providing execution to collateralized borrowing, and the front-running of large trades, for which there is a considerable execution premium.

[15] A full description of the PoS mechanism is more complex, but this lottery-like process based on deposited token holdings is the core feature that we examine in our paper. See Xiao et al. (2020) for a much more detailed description of PoS protocols.

2.4 Proof-of-Stake security: the role of scale

A well-understood phenomenon in many transaction validation protocols is the 51% attack (Sayeed and Marco-Gisbert, 2019).[16] In this attack, a malevolent actor or group of actors acquires 51% of validation power, allowing them to alter the chain in a way that diverts currency away from other owners. In a PoW protocol, if a large subset of the mining infrastructure were to coordinate, it could expropriate the remaining share of transaction validators and eventually the rest of the chain. In the PoS protocol, the same could occur if a large subset of stakers were to coordinate, expropriating validators and the entire chain accordingly.[17]

While a greater than majoritarian share of the validation technology is not required for such an attack, attaining that share would make an attack's viability rise to near certainty. In the case of a PoS protocol attack, as stakers exit, a scale-based attack is mechanically more viable. That is, the size of the staking contract is directly associated with the security of the protocol. As a staking contract rises in size, this makes it increasingly financially costly for a malevolent actor or group of actors to execute an attack. As a staking contract falls in size or sees stakers exit, this makes it increasingly financially viable for a malevolent actor or group of actors to execute an attack.

[16] Other threats to security include wallet attacks, wherein malevolent actors steal wallet information, and hard forks, wherein there are disagreements among users and developers in a coin's broader community.

[17] While a 51% attack is more commonly known, it is by no means the only scale-relevant attack on a protocol. For example, the attack on Terra Luna was more straightforwardly prompted by a lump-sum removal of stake and sale of crypto-assets on the open market Liu et al. (2023).

3 Staking Runs in a Static Model

We first model a static game to highlight how the PoS protocol produces run incentives. Investors make a choice between exiting the pool, selling their stake, or remaining and receiving staking rewards. Security strength of the protocol declines as stakers leave the staking contract. In this simple setup, stakers can instantly leave the pool. However, an important element of most staking protocols is the rate at which stakers can exit (i.e., lock-up periods). We consider this issue in the dynamic model in Section 4.

3.1 Players, Actions, and Prices

There is a single period with two subperiods. There is a coin, or token, that has an exogenous initial price p and an endogenous end-of-period price p_1. We will use "coin" and "token" interchangeably in our analysis.

There is a unit measure of players called stakers endowed with a single unit of the coin. In the first subperiod, players have the option of staking their coins to earn rewards or of selling their coins at the current market price. In the second subperiod, the protocol succeeds or fails, and payoffs are realized.

Stakers can take one of two actions: exit or stake. If a player chooses to exit, they sell their coin and receive an expected price p_1^e, which we describe below. If a player decides to stake through the end of the period, a player receives the opportunity to earn rewards by appending new blocks to the chain. Staking generates expected rewards given by a fraction $\rho > 0$ of a coin for all stakers, similar to block rewards in PoS protocol coins. PoS generally implements a "Follow the Satoshi" randomization wherein a coin among

all staked coins is chosen at random. We assume, without loss of generality, that stakers always opt to append a new block to the chain.[18]

The final price p_1 depends on the endogenous size of the staking contract. If a fraction $\ell \in [0, 1]$ of stakers exit the pool, the price of the coin is assumed to be

$$p_1 = p(1 - c\ell), \tag{1}$$

where $c \in (0, 1]$ is a parameter governing the relationship between staking supply and the price (we require $c \leq 1$ so that the price cannot fall below zero). This relationship could capture a number of real-life mechanisms. In the extreme, reduced staking may reduce liquidity, decreasing the utility of the coin and its market value in turn. A greater risk may be to the security of transactions and the stability of the PoS protocol. If fewer stakers are present, this increases scale-based attacks, which could lower a coin's implicit value due to, for example, multiple branches in the blockchain leading to a lack of consensus or even expropriation of one set of coin holders by another set of coin holders.

Theoretically, the size of the staking contract is directly related to the security of the protocol (John et al., 2020, 2021). Additionally, in practice, PoS protocols are susceptible to a variety of attacks. The simplest example is a 51% attack (Sayeed and Marco-Gisbert, 2019), in which a staker or group of stakers achieves a 51% share of staked coins and is assured, over time, to dominate the blockchain and therefore determine which digital

[18]Because not appending a block results in lower rewards to staking, in equilibrium, runs will be more likely if stakers are assumed to decide whether or not to append. In practice, protocols may allow participants to receive rewards for proposing blocks, attesting blocks, crossing limit orders to produce maximum extractable value (MEV, i.e., market making fees), tips, or gas fees. To the extent that stake is correlated with rewards, ρ should approximate any marginal remuneration of remaining in the staking protocol.

wallets hold what. Other attacks on such a protocol exist, such as short-range reorganizations, adversarial delay, or potentially ideologically-driven, long-range reorganizations of the blockchain (Schwarz-Schilling et al., 2021). In all cases, the susceptibility to an attack becomes greater whenever stakers exit.

Given this, we also introduce protocol risk or the fundamental strength of the protocol. We let θ denote the protocol strength. Following the intuition of scale-based attacks, if too many stakers exit the protocol, transaction security falls. Formally, we suppose that the protocol automatically fails when $\ell > \theta$. This results in a decline in the price by the additional fraction η. Failure represents an undermining of consensus and finality of the blockchain. This condition is analogous to the classic "attack" equilibria in standard global games (e.g., currency crises). Note that a low θ corresponds to a weak protocol. If $\theta = 0$, then the protocol fails if any stakers exit the pool, and if $\theta < 0$ the protocol fails exogenously. If $\theta = 1$, then the protocol survives even if all stakers exit (if $c > 0$ then there would still be a price consequence).

In practice, exit from a staking protocol may be rationed, delaying execution. This is particularly relevant when the market microstructure involves execution based on fees or depends on the speed with which remaining stakers validate trades. To capture this dynamic in the first subperiod, we suppose that trades are executed sequentially and that a player's position in line is stochastic. For a fraction ℓ of stakers that choose to exit, each seller's position in the queue is uniformly distributed on $[0, \ell]$. Stakers rationally anticipate potentially front-running sales made by other stakers, and so the expected price

from selling is

$$p_1^e = p\left(1 - \frac{c}{2}\ell\right). \tag{2}$$

If the protocol fails, then failure occurs in the second subperiod, after trades are executed. That means that the expected second subperiod price upon protocol failure, $\ell > \theta$, is

$$p_1^{\text{failure}} = p\left(1 - c\ell - \eta\right). \tag{3}$$

Because all prices are multiplicative in the initial price p, without loss of generality we can normalize $p = 1$ to simplify exposition.

We suppose that $1 - c - \eta \geq 0$ to ensure that prices are bounded by zero.

3.2 Strategic interaction and equilibrium

As described above, stakers make a simultaneous decision to stake their coins, earning rewards, or to exit (withdraw), selling their coins.

We characterize the net payoff to staking relative to exiting and selling, denoted by $\pi(\ell, \theta)$, a function of the fraction of stakers ℓ that exit as well as a function of the fundamental, θ. The payoff to selling is $p_1^e = 1 - \frac{c}{2}\ell$. The payoff to staking is $(1+\rho)(1-c\ell)$ when $\ell \leq \theta$ and $(1+\rho)(1-c\ell-\eta)$ when $\ell > \theta$, reflecting the final value of the initial

coin and the value of coins received from staking rewards. Therefore, when $\ell \leq \theta$,

$$\pi(\ell, \theta) = \underbrace{(1+\rho)(1-c\ell)}_{\text{Payoff to staking}} - \underbrace{\left(1 - \frac{c}{2}\ell\right)}_{\text{Payoff to selling}}$$

and when $\ell > \theta$,

$$\pi(\ell, \theta) = \underbrace{(1+\rho)(1-c\ell-\eta)}_{\text{Payoff to staking}} - \underbrace{\left(1 - \frac{c}{2}\ell\right)}_{\text{Payoff to selling}}$$

That is,

$$\pi(\ell, \theta) = \begin{cases} \rho - c\left(\frac{1}{2} + \rho\right)\ell & \ell \leq \theta \\ \rho - c\left(\frac{1}{2} + \rho\right)\ell - (1+\rho)\eta & \ell > \theta. \end{cases} \tag{4}$$

3.3 Equilibria without protocol risk

We first suppose that there is no protocol risk, $\theta = 1$, to focus on how incentives create the possibility of runs even in the absence of protocol risk. We also suppose that

$$c\left(\frac{1}{2} + \rho\right) - \rho \geq 0, \tag{5}$$

so that, with common knowledge, if all stakers exit the pool then it is optimal to exit.

The equilibria of the game among stakers are governed by the payoff gain $\pi(\ell, \theta)$. Importantly, the players' actions feature *strategic complementarities*. That is, $\frac{\partial \pi(\ell, 1)}{\partial \ell} < 0$, which means that the incentive to stake is decreasing in the fraction of players that exit, or equivalently, the incentive to stake and sell is increasing in the fraction of other players

that stake and sell. Under complete information, there are three candidates for Bayesian Nash equilibria (BNE): a stable, staking (hold) equilibrium in which the market continues to function as normal, an unstable run equilibrium in which all stakers pull their tokens and sell into an illiquid market, and a mixed-strategy equilibrium. With the above parameter condition, the game admits multiple equilibria, as stated in the following proposition.

Proposition 1 (Multiple equilibria with common knowledge in a riskless protocol)**.** *Suppose* $\frac{c}{2} \geq \rho(1-c)$*. With perfection information, the game among stakers features multiple BNE: a hold equilibrium, a run equilibrium, and a mixed-strategy equilibrium.*

Proof. The proof is straightforward. First, consider the hold equilibrium. If the incentive to exit is negative when no stakers leave the pool, that is if $\pi(0,1) \geq 0$, then it is a pure-strategy equilibrium for no stakers to sell (equilibrium $\ell^* = 0$). Note that $\pi(0,1) = \rho > 0$ and therefore the hold equilibrium always exists. The intuition is straightforward: if all stakers remain in the pool, there is no price-risk to the protocol and staking earns positive rewards.

Second, consider the run equilibrium. If the incentive to stake is negative when all stakers leave the pool, that is if $\pi(1,1) \leq 0$, then it is a pure-strategy equilibrium for all stakers to sell (equilibrium $\ell^* = 1$). The run equilibrium exists if $\pi(1,1) \leq 0$:

$$\rho - c\left(\frac{1}{2} + \rho\right) \leq 0 \implies \frac{c}{2} \geq \rho(1-c).$$

If $c = 1$ then a run equilibrium always exists for any level of ρ because the price collapses

to zero if all stakers leave the pool. For $c \in (0,1)$, a run equilibrium exists whenever

$$\rho \leq \bar{\rho} \equiv \frac{c}{2(1-c)}.$$

A run equilibrium can exist whenever the staking rewards are sufficiently low that they do not compensate for the protocol risk from stakers exiting the pool. In other words, a run equilibrium can exist if the price impact of low staking is sufficiently high or the benefit to staking is sufficiently low.

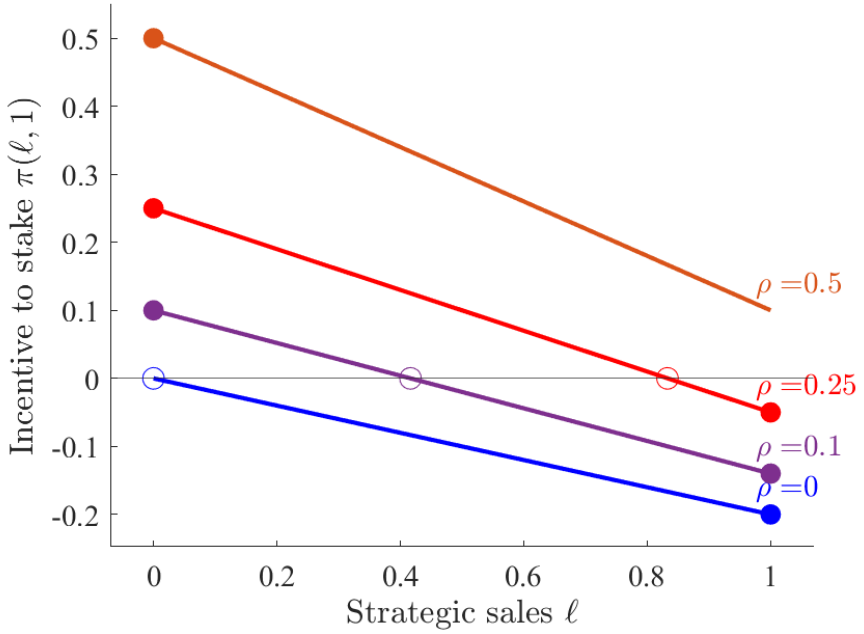
Finally, consider the mixed-strategy equilibrium. If the incentive to exit (stake) is zero when a fraction of stakers leave the pool, that is if $\pi(\ell^*, 1) = 0$ for some equilibrium $\ell^* \in (0,1)$, then it is a mixed-strategy equilibrium for all stakers to exit with probability ℓ^*. This immediately follows given the monotonicity of π. □

In Figure 1, we demonstrate how stake, run, and mixed-strategy equilbria vary with rewards rates. The incentive to stake, $\pi(\ell, 1)$, declines in the level of strategic sales, ℓ. Supposing a declining value of $c = 0.4$ with withdrawals, we see that only a staking equilibrium exists when rewards are sufficiently high, as shown in the graph by a rewards rate of $\rho = 0.5$. Stake, run, and mixed-strategy equilibria exist for lower rewards rates and only run and mixed-strategy equilibria exist when rewards are taken to nothing, $\rho = 0$.

The analysis so far depends on two simple ingredients to illustrate how the staking incentives can create a run equilibrium without protocol risk. First, when stakers exit the pool new stakers do not immediately enter. This is an empirically realistic assumption because staking requires familiarity with the protocol and the ability to append blocks to the chain via computer nodes. Second, the coin price depends on the size of the staking

Figure 1: **Multiple equilibria with no protocol risk**

This figure shows the equilibria, stake, run, and mixed-strategy, for various levels of rewards, ρ. As a baseline, the declining value of the price as a function of withdrawals is $c = 0.4$.



(Source: Authors' analysis)

contract, which is likely empirically realistic.

However, showing that runs exist is not enough. We endeavor to demonstrate when runs are likely to occur. To do so, we explore equilibrium selection in a global game setting to characterize how runs depend on the fundamentals of protocol risk.

3.4 Equilibria with protocol risk

We now explore the setting when $\theta < 1$, meaning that the strength of the protocol depends on the fraction of stakers leaving the pool. More specifically, the fundamental θ is drawn at the beginning of the period from a distribution F on $(0, 1)$. There are potentially

several forums to access information about opinions over the strength of the protocol at any given time and users may access these forums at different times. We therefore suppose that there is imperfect information about θ and each player i observes an idiosyncratic signal $\hat{\theta}_i = \theta + \sigma_\varepsilon \varepsilon_i$, where the mean-zero signal noise ε_i is i.i.d. across all i with distribution G_ε and $\sigma_\varepsilon > 0$. As a result of the noise in signals, a staker faces fundamental uncertainty about the strength of the protocol θ, as well as strategic uncertainty about the fraction ℓ of other stakers who sell their coins.

We focus on the limit of vanishing signal noise, $\sigma_\varepsilon \to 0$, and therefore treat θ as non-random in the exposition except when deriving the global game equilibrium. Under complete information, there can be multiple equilibria in the strategic interaction among stakers. Introducing noise into stakers' payoffs breaks the common knowledge underpinning any multiplicity of equilibria, as we describe below.

As before, we assume the parameter restriction in (5). We make use of the standard global games framework (e.g., Morris and Shin, 2003) to derive a unique BNE for the game among stakers.

Proposition 2 (Unique global game equilibrium)**.** *For signal noise $\sigma_\varepsilon \to 0$, the unique Bayesian Nash equilibrium among stakers is in switching strategies around a threshold θ^* defined by $\int_0^1 \pi(\ell, \theta^*)\, \mathrm{d}\ell = 0$:*

$$\theta^* = \frac{\frac{c}{2}\left(\frac{1}{2} + \rho\right) - \rho}{(1+\rho)\eta} + 1. \tag{6}$$

For protocol strength below the threshold, $\theta < \theta^*$, all players sell their coins (exit) and the protocol suffers a run. For protocol strength above the threshold, $\theta \geq \theta^*$, all players stake their coins and the protocol is stable.

The proof is in Appendix A. Note that a protocol is more robust, or stable, whenever θ^* is low. In this case, stakers will not run unless the fundamental θ is very low—runs require fundamental weakness. In contrast, when θ^* is high, stakers will run even if the protocol has (relatively) strong fundamentals. In addition, the protocol becomes more stable with a higher rewards rate.

Corollary 1. *A protocol is more stable when rewards rates are higher (i.e., θ^* is decreasing in ρ).*

We now consider a benchmark case to derive intuition for how exactly rewards, ρ, affect stability. First, suppose that as long as the protocol survives, then the price is unaffected by ℓ, but that the price falls to zero if the protocol fails (i.e., $c = 0$ and $\eta = 1$). The pricing function is therefore $p_1 = p$ if $\ell \leq \theta$ but $p_1 = 0$ if $\ell > \theta$. With this assumption, the payoff to selling is p (fixed) and the payoff to staking is $(1+\rho)p_1$, where p_1 depends on whether or not the protocol fails. If $\ell \leq \theta$ and therefore the protocol succeeds, then selling leads to an opportunity cost of ρ units of the token ($\pi(\ell, \theta) = \rho$). If the protocol fails because $\ell > \theta$, then selling protects the investment ($\pi(\ell, \theta) = -1$). In this case,

$$\int_0^1 \pi(\ell, \theta) \mathrm{d}\ell = \theta\rho - (1-\theta) = \theta(1+\rho) - 1,$$

which has a unique solution

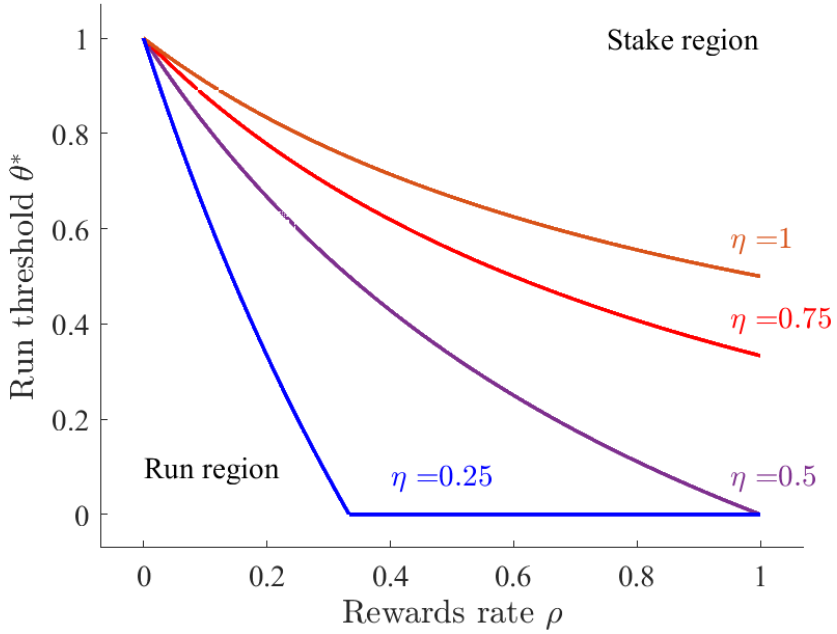
$$\theta^* = \frac{1}{1+\rho}.$$

When $\theta < \theta^*$, then $\int_0^1 \pi(\ell, p) \mathrm{d}\ell < 0$, and so it is dominant to exit. Thus, the protocol will fail (all stakers exit) whenever $\theta < \theta^* = \frac{1}{1+\rho}$.

In this and other potential cases, increasing the benefit to staking, ρ, makes the proto-

Figure 2: **Run threshold θ^* as a function of rewards ρ**

This figure shows the run threshold, θ^*, as a function of rewards, ρ, for various declines in the asset's value upon protocol failure, η. For protocol strength θ above this threshold, the equilibrium is to stake. For protocol strength θ below this threshold, the equilibrium is to run.



(Source: Authors' analysis)

col more robust. These features can be seen in Figure 2. As a quantitative exercise, one might want to know how high ρ might have to be to prevent runs. A natural benchmark to consider is a 51% attack, wherein $\theta^* = 0.51$.[19] While a 51% attack does not depend on exit of the other 49% of stakers, their exit would make a 51% attack successful almost surely. For the protocol to be stable up until the 51% threshold, the protocol requires a rewards rate $\rho = 100\%$. If rewards are lower, then players will endogenously choose to exit, thus facilitating a 51% attack, for example.

[19]We describe a 51% attack as being equal to 51% of staking share for expositional purposes. The nature of a 51% attack is that there is coordination among $50\% + \varepsilon$ for some $\varepsilon > 0$. More surreptitious attacks are left as examples that can be explored by the reader.

Second, continue to suppose that the price effects from weakened security are small ($c \approx 0$), but now suppose that the protocol loses only a fraction of its value in failure. With $c = 0$, the threshold becomes

$$\theta^* = \frac{1 - \rho(1 - \eta)}{(1 + \rho)\eta} < \frac{1}{1 + \rho},$$

which intuitively implies a more stable protocol since the price drop in failure is less severe. Now consider an empirically plausible (though still high) case with $\rho = 20\%$, and suppose that the protocol loses half its value in failure, $\eta = 0.5$. Then runs occur whenever $\theta < 2/3$. If instead $\rho = 10\%$, and failure means losing 90% of value, then runs occur whenever $\theta < 89/99 = 0.8989$. Note that the 51% protocol clearly falls into these ranges: if 49% of the stakers leave, then a 51% attack can occur. In addition, the threshold increases if there are price effects from weakened security even when the protocol survives.

Summary The global game analysis highlights the important tension in the PoS protocol. Decreasing the probability of runs (i.e., increasing θ^*) requires increasing the staking rewards ρ. In other words, a robust protocol requires sufficiently *high* staking rewards. However, Saleh (2021) shows that for PoS to achieve consensus, where a single dominant branch exists for a coin, requires sufficiently *small* staking rewards. John et al. (2021) argue that maximizing the participation in a coin, a pre-condition for security, also requires staking rewards that are not so high as to preclude coin adoption by investors with greater trading needs.

PoW produces consensus at the cost of electricity expenditure. PoS can provide consensus without this energy cost but it is subject to runs. Ensuring consensus requires that rewards are sufficiently low, but minimizing the likelihood of runs requires rewards that are sufficiently high. Thus, there exist protocol strengths such that a protocol can have a consensus-preserving chain or be run proof but not both. A protocol with high fundamentals could maintain stability and consensus with low staking rewards. But a change in fundamentals could trigger protocol failure. We consider this possibility in the dynamic model.

The analysis so far was intentionally simple and did not include realistic and important features. In the Appendix, we consider extensions to our model to accommodate relevant institutional details important to evaluating the stability of PoS coins. First, in Appendix B.1, staking rewards for each individual staker are proportional to the aggregate share of staked coins so that there are potentially higher rewards when the staking contract shrinks. In our primary analysis, we abstract away from this feature. We find that run risks still materialize even in this case. Second, Appendix B.2 considers how leveraged staking could contribute to run risk. Many cryptocurrency investors use leverage as a part of their overall strategies (Pelster et al., 2019). Our model's primary analysis shows that run risks do not depend on price, but in practice leveraged staking could lead to greater runs if prices decline, producing margin calls and the need to sell positions. Our analysis confirms this possibility.

4 Staking runs in a dynamic model

The static model supposes that the staking contract is perfectly liquid, meaning that stakers can immediately sell their coins to exit the staking contract. In reality, these contracts can be illiquid, meaning stakers cannot immediately sell their coins to exit the staking contract. We now consider how runs can occur with illiquid staking contracts. Our model follows Guimaraes (2006), who uses an analogous setting for a currency attack. In our setting, the PoS protocol can suffer from failure risk if enough stakers leave the pool.

4.1 Players, actions, and prices

Time is continuous and infinite. As before, there is a unit measure of stakers. Stakers discount the future at rate r. Staking earns benefit at a rate ρ.

The staking contract is *illiquid*. Motivated by some of the DeFi protocol designs, we suppose that stakers who wish to exit the staking contract receive stochastic opportunities to withdraw their tokens, and otherwise they stay staked. Opportunities arrive at exponential rate δ, at which point tokens can be withdrawn or rolled-over. The contract is more liquid the higher is δ. Hence, the fastest stakers can exit the pool is at rate δ, which occurs when all stakers withdraw as soon as an opportunity arises.

As before, we suppose that the price of the token p falls with the pool of stakers. We now cast the model in terms of the size of the staking contract rather than the fraction who have exited as we did in the static model. Let A denote the size of the staking contract. For simplicity,

$$p(A) = A^{\gamma}, \tag{7}$$

where $\gamma \in [0, 1]$ denotes the elasticity of the price to the staking contract.

As before, the fundamental strength of the protocol is given by θ, which corresponds to the fraction of stakers that can exit before the protocol collapses. It is convenient to define the fundamental strength of the protocol relative to the pool of stakers. Let $\phi \equiv 1 - \theta$ denote the "inverse" of strength. If the measure of stakers falls below ϕ, then the price collapses to zero. Weak fundamentals therefore correspond to low θ and a high ϕ.

Let ϕ^* denote the value of ϕ below which a run begins, and let $\hat{\phi}$ denote the value of ϕ below which the protocol fails. The model will yield two insights. First, the model will determine when a run occurs, given by the value ϕ^* when stakers begin exiting the pool. Second, and related, the model will determine the strength of the protocol when it fails, given by the value $\hat{\phi}$ when $A \leq \hat{\phi}$. The characteristics of the staking contract, notably the rewards rate and the illiquidity, will determine these thresholds.

In reality, the fundamental strength of a crypto protocol waxes and wanes with market conditions, competing protocols, and malicious attacks. For tractability, we model the drift of protocol strength via regime switching. We suppose that $\mathrm{d}\theta_t = -\mu_\theta \mathrm{d}t$, where μ_θ can take two possible values:

$$\mu_1 < 0, \text{ and } \mu_0 > 0.$$

We suppose that the state switches at a Poisson rate λ.[20]

In state 1 the protocol strengthens over time; recall that failure occurs when $A < \phi$, so a low ϕ (high θ) is better. To simplify our analysis, we suppose that $|\mu_1| > \delta$, which means that a run is not possible because the fundamental improves at a faster rate than a

[20] Appendix B.3 consider the model with a Brownian process for μ_t.

run can occur. Thus, a run can only occur in state 0.

We suppose that there is a long-run value of staking a dollar in the stable state 1 which is $v > 1$ per dollar. To maintain tractability, we let v be exogenous. We also suppose there is no discounting ($r = 0$), which is without loss of generality.

4.2 Strategy and Equilibrium

A strategy for an agent yields a decision, either to stake or to run, for every pair (A, θ). A threshold θ^* (equivalently, ϕ^*) is a function of A that defines the regions to stake or to exit. Given the assumptions so far, the model permits a unique threshold equilibrium, ϕ^*.

For notational purposes, we will "start the clock" when the run begins (i.e., $\phi_t = \phi^*$). Thus, t denotes the time since the run began. Because stakers accumulate tokens at rate ρ and the price falls at rate $\delta\gamma$ as stakers exit, the net benefit to staking accumulates at a rate $\nu \equiv \rho - \delta\gamma$, where we define ν as the net benefit. Therefore, $e^{\nu t}$ is the total value accumulated from staking so long as the protocol has not failed (at which point the value of the token is zero).

For agents to have an incentive to stake, the net rewards must accumulate faster than the discount rate r, because otherwise the present value of the accumulated tokens would decrease over time even in the absence of a run. We, therefore, assume

$$\nu > 0 \implies \rho > \delta\gamma, \tag{8}$$

in order to make staking viable. Furthermore, players must receive sufficient opportunity

to exit the pool once a run begins. This requires δ to be sufficiently high:

$$\delta > \nu. \tag{9}$$

Note for example that if prices did not depend on the size of the staking contract ($\gamma = 0$), then we require $\delta + r > \rho > r$.

Suppose we are in state 0 and $\phi = \phi^*$, triggering a run. A run could now potentially end for exogenous reasons because the regime switches to state 1. Let T denote the length of the run if failure occurs *before* a regime switch. The payoff to continuing to stake the token is complicated by what may happen next. A staker receives a payoff (exiting the run) in two potential ways. First, at rate δ an opportunity will arise in the future to withdraw the token. Second, at rate λ the regime will switch and the run will end.

The first event occurs if a withdrawal opportunity occurs before a regime switch. The probability of a regime switch occurring before time t is

$$\int_0^t \lambda e^{-\lambda s} \mathrm{d}s = 1 - e^{-\lambda t}.$$

Thus, the probability of a regime switch not occurring before time t is $e^{-\lambda t}$. Hence, the probability of a withdrawal opportunity occurring at time t when a regime switch has not yet occurred is $\delta e^{-\delta t} e^{-\lambda t}$. Similarly, the probability of a withdrawal opportunity occurring before time t is

$$\int_0^t \delta e^{-\delta s} \mathrm{d}s = 1 - e^{-\delta t},$$

And, thus, the probability of a withdrawal opportunity not occurring before t is $e^{-\delta t}$.

Hence, the probability of a regime switch occurring at time t when a withdrawal opportunity has not yet occurred is $\lambda e^{-\delta t} e^{-\lambda t}$.

We characterize equilibrium in terms of how long a run lasts (i.e., how long until the protocol fails once a run occurs). We denote this time to failure by T. At failure, the staking contract has measure

$$\hat{A} \equiv e^{-\delta T},$$

and the fundamental equals $\hat{\phi} = \hat{A}$, allowing us to determine the threshold ϕ^* when the attack begins using

$$T\mu_\theta = \hat{\phi} - \phi^*.$$

We now characterize the expected payoff to staking or withdrawing. Consider the marginal agent when $\phi = \phi^*$ and the run will last T. Withdrawing the token earns 1. Continuing to stake is slightly more complicated. Because $\phi = \phi^*$, if the player stakes, she knows she will withdraw the next chance she gets. If she withdraws at $t < T$ before the protocol fails, she will get $e^{\nu t}$. That is, her tokens accumulate at rate ρ, but the price falls at rate $\delta\gamma$, for a total accumulation of ν, defined above. Let π denote the expected payoff of staking relative to withdrawing. Then we have the following result:

Lemma 1. *Let T be the time until the protocol fails. Then the net value to staking relative to withdrawing is*

$$\pi = \frac{\delta + \lambda v}{\lambda + \delta - \nu} \left(1 - e^{-(\lambda+\delta-\nu)T}\right) - 1. \tag{10}$$

At the threshold ϕ^*, players are indifferent between staking and exiting ($\pi = 0$), which allows us to solve for T and characterize the fundamental strength at failure.

Proposition 3. *In equilibrium, the run lasts*

$$T = \frac{1}{\delta + \lambda - \nu} \log \frac{\delta + \lambda v}{\nu + \lambda(v-1)}, \tag{11}$$

and at failure the protocol fundamental equals

$$\hat{\phi} = \left(\frac{\delta + \lambda v}{\nu + \lambda(v-1)} \right)^{-\frac{\delta}{\delta + \lambda - \nu}}.$$

The run length T is positive so long as

$$\delta + \lambda v > \nu + \lambda(v-1) \implies \delta + \lambda > \nu,$$

which holds given our assumption that $\delta > \nu$. We calculate the equilibrium threshold

$$\phi^* = \hat{A} - \mu_0 T. \tag{12}$$

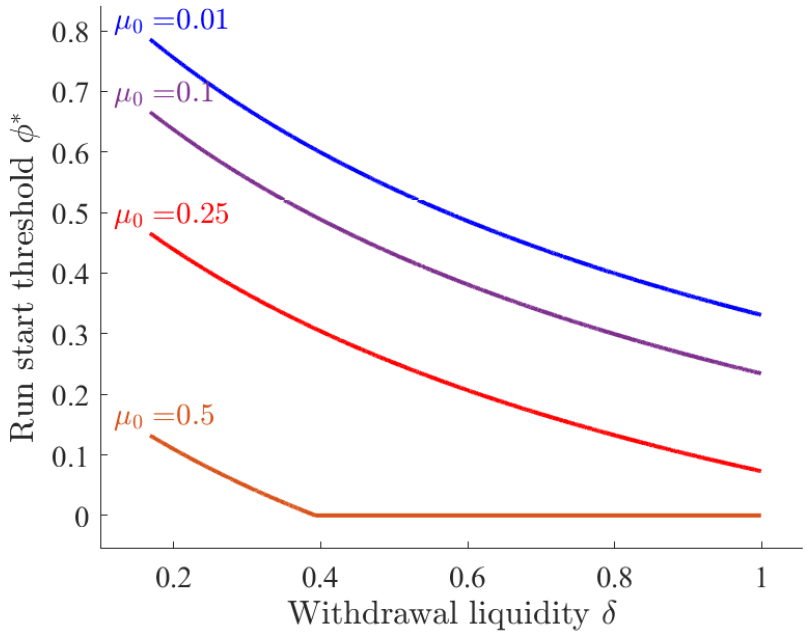
We plot the dynamic threshold in Figure 3 as a function of liquidity, δ, for different protocol decline rates, μ_0. Once protocol decline is high enough, $\phi > \phi^*$, the run begins. As can be seen, a more illiquid contract leads to a higher ϕ^*, meaning lower likelihood of runs.

4.3 Comparative dynamics

We can now characterize how the staking contract affects the run thresholds. For tractability, we let $\lambda \to 0$ in order to analytically characterize the comparative dynamics.

Figure 3: **Dynamic run threshold ϕ^* as a function of liquidity δ**

This figure shows the dynamic run threshold, ϕ^*, as a function of liquidity, δ, for various rates of protocol decline in the bad state, μ_0. Once protocol decline reaches the plotted lines, the run begins. Parameters are rewards ($\rho = 0.2$), price concavity ($\gamma = 0.2$), regime switching Poisson arrival rate ($\lambda = 0.4$), and long-run staking value in the good state ($v = 2$).



(Source: Authors' analysis)

First, we define $\psi \equiv \frac{\delta}{v}$. With $\lambda \to 0$, we have

$$\hat{\phi} = \hat{A} = \psi^{-\frac{\psi}{\psi-1}}. \tag{13}$$

Differentiating, and using that $\log \psi < \psi - 1$, we have $\frac{d\hat{A}}{d\psi} < 0$. We can therefore define the equilibrium threshold, analogously to Equation (12)

$$\phi^* = \psi^{-\frac{\psi}{\psi-1}} - \mu_\theta T = \psi^{-\frac{\psi}{\psi-1}} - \frac{\mu_\theta}{r+\delta-v} \log \frac{\delta}{v-r},$$

and it is immediate that

$$\frac{dT}{d\psi} < 0.$$

This yields the following comparative dynamics. First, as staking illiquidity goes to zero (i.e., as $\delta \to \infty$), attacks are instantaneous:

$$\lim_{\psi \to \infty} T = 0, \quad \lim_{\psi \to \infty} \hat{A} = 0.$$

Recall that the protocol fails when $A_t < \phi_t$ and that ϕ_t is increasing in the bad state, 0. If stakers do not withdraw, the protocol will not fail until $\phi_t = 1$, or when the fundamental falls to 0. But as $\psi \to \infty$, the protocol fails as soon as $\phi_t = 0$, which is early. This means that a liquid staking contract (high δ) causes the protocol to fail early. Since $\lim_{\delta \to \infty} T = 0$ and $\lim_{\delta \to \infty} \hat{A} = 0$, we must have that $\phi^* \to 0$, so runs begin early as well.

A staking contract in which agents can withdraw easily (high δ) decreases the size of the pool when the protocol fails and decreases the length of the run. Agents want to stay in the staking contract to earn the net benefits $\nu = \rho - \delta\gamma$. A high δ means that they can quickly withdraw—as can others—and so agents wait to withdraw until the last minute (a low T) but then withdraw quickly, drawing down the staking contract to a low $\hat{A}$.

In sum, we find that lock-ups both ward off runs and delay them. In equilibrium, more opportunities to leave the bad state occur in a slower run, so runs are less likely to occur. However, lock-ups could have implications for liquidity, which we discuss in Appendix B.4.

5 Conclusion

In this paper, we identify and examine run risk as an important tradeoff present in the PoS protocol. We first establish run risks in a simple, static model of staking. Investors choose to stake or exit their positions and are rewarded for staking, trading this off with the risk that the crypto-asset fails due to an attack on the protocol. The protocol's security serves as an impetus to coordination in the choice to stake or exit. We find that runs on staking are more common when the crypto-asset's protocol is weaker, when the price impacts of protocol failure are high, or when rewards to staking are low.

Our approach demonstrates that run risk is a concern in any protocol that depends on the voluntary locking up of otherwise liquid funds to aid in incentive compatible transaction validation. In a dynamic model, we show that the inherent design of these protocols creates incentives for run risk, even when a protocol includes a lock-up period for staked coins. Greater lock-up periods do mitigate run risk but do not eliminate it entirely.

Our findings contribute to our understanding of the economic viability of PoS mechanisms. In conjunction with results in the extant literature, we examine tradeoffs in transaction validation. Traditional transaction validation puts a central intermediary's reputation at stake. In PoW, validators consume electricity in the effort to solve a secure hash problem, putting this energy expenditure at stake. PoS validators do not consume nearly as much electricity but must put significant capital at stake. While PoS achieves energy efficiency, it introduces other tradeoffs. Because PoS's security (John et al., 2021) and consensus (Saleh, 2021) depend on low rewards and low rewards induce more run risk, PoS poses a risk to financial stability due to its prevalence and the fact that it either lacks a

secure, consensus-inducing blockchain or is susceptible to staker runs.

References

BIAIS, B., C. BISIERE, M. BOUVARD, AND C. CASAMATTA (2019a): "The blockchain folk theorem," *Review of Financial Studies*, 32, 1662–1715.

BIAIS, B., C. BISIÈRE, M. BOUVARD, AND C. CASAMATTA (2019b): "Blockchains, coordination, and forks," in *AEA Papers and Proceedings*, vol. 109, 88–92.

FRANKEL, D. AND A. PAUZNER (2000): "Resolving indeterminacy in dynamic settings: the role of shocks," *The Quarterly Journal of Economics*, 115, 285–304.

GUIMARAES, B. (2006): "Dynamics of currency crises with asset market frictions," *Journal of International Economics*, 68, 141–158.

IRRESBERGER, F., K. JOHN, P. MUELLER, AND F. SALEH (2023): "The public blockchain ecosystem: An empirical analysis," *Working Paper*.

IRRESBERGER, F. AND R. YANG (2023): "Coin concentration of proof-of-stake blockchains," *Economics Letters*, 111219.

JOHN, K., T. J. RIVERA, AND F. SALEH (2020): "Economic implications of scaling blockchains: Why the consensus protocol matters," *Working paper*.

——— (2021): "Equilibrium staking levels in a proof-of-stake blockchain," *Working paper*.

LIU, J., I. MAKAROV, AND A. SCHOAR (2023): "Anatomy of a run: The terra luna crash," Tech. rep., Working paper.

MORRIS, S. AND H. S. SHIN (2003): "Global games: Theory and applications," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 1*, Cambridge University Press, 56–114.

——— (2004): "Liquidity black holes," *Review of Finance*, 8, 1–18.

NAKAMOTO, S. (2008): "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Business Review*, 21260.

PELSTER, M., B. BREITMAYER, AND T. HASSO (2019): "Are cryptocurrency traders pioneers or just risk-seekers? Evidence from brokerage accounts," *Economics Letters*, 182, 98–100.

SALEH, F. (2021): "Blockchain without waste: Proof-of-stake," *Review of Financial Studies*, 34, 1156–1190.

SAYEED, S. AND H. MARCO-GISBERT (2019): "Assessing blockchain consensus and security mechanisms against the 51% attack," *Applied sciences*, 9, 1788.

SCHWARZ-SCHILLING, C., J. NEU, B. MONNOT, A. ASGAONKAR, E. N. TAS, AND D. TSE (2021): "Three Attacks on Proof-of-Stake Ethereum," *arXiv preprint arXiv:2110.10086*.

XIAO, Y., N. ZHANG, W. LOU, AND Y. T. HOU (2020): "A survey of distributed consensus protocols for blockchain networks," *IEEE Communications Surveys & Tutorials*, 22, 1432–1465.

A Proofs

Proof of Proposition 2. In order to apply the standard global game result that there is a unique equilibrium and that it is in switching strategies, we have to show that the payoff gain $\pi(\ell, \theta)$ satisfies certain properties (Morris and Shin, 2003). Because we have defined π as the net payoff to staking as a function of the players exiting, we need the following 5 conditions to apply global games techniques:

1. Action monotonicity: $\pi(\ell, \theta)$ is non-increasing in ℓ;
2. State monotonicity: $\pi(\ell, \theta)$ is non-decreasing in θ;
3. Laplacian State monotonicity: there exists a unique θ^* solving $\int_0^1 \pi(\ell, \theta^*) \mathrm{d}\ell = 0$;
4. Limit dominance: there exist $\underline{\theta}$ and $\overline{\theta}$ such that $\pi(\ell, \theta) > 0 \, \forall \ell$ if $\theta \leq \underline{\theta}$ and $\pi(\ell, \theta) < 0$ $\forall \ell$ if $\theta \geq \overline{\theta}$ (i.e., above/below the bounds there is strict dominance in one strategy); and
5. Continuity.

Our assumption on the timing of trades ensures action monotonicity and state monotonicity. We get local dominance with $\underline{\theta} = 0$ and $\overline{\theta} = 1$; in the first case, the protocol always fails and in the second it never fails. Continuity is immediate. Thus, we can apply the global games methodology. Recall that

$$\pi(\ell, \theta) = \begin{cases} \rho - c\left(\frac{1}{2} + \rho\right)\ell & \ell \leq \theta \\ \rho - c\left(\frac{1}{2} + \rho\right)\ell - (1+\rho)\eta & \ell > \theta. \end{cases}$$

We therefore have

$$\begin{aligned} \int_0^1 \pi(\ell, \theta) \mathrm{d}\ell &= \theta\left(\rho - c\left(\frac{1}{2} + \rho\right)\ell\right) + (1-\theta)\left(\rho - c\left(\frac{1}{2} + \rho\right)\ell - (1+\rho)\eta\right), \\ &= \rho - \frac{c}{2}\left(\frac{1}{2} + \rho\right) - (1-\theta)(1+\rho)\eta. \end{aligned}$$

This equals zero when

$$\theta^*(1+\rho)\eta = \frac{c}{2}\left(\frac{1}{2}+\rho\right) - \rho + (1+\rho)\eta,$$
$$\theta^* = \frac{\frac{c}{2}\left(\frac{1}{2}+\rho\right) - \rho}{(1+\rho)\eta} + 1.$$

The run equilibrium occurs when $\theta < \theta^*$. □

Proof of Corollary 1. Differentiating equation (6),

$$\begin{aligned}\frac{\partial \theta^*}{\partial \rho} =& \frac{(1+\rho)\left(\frac{c}{2}-1\right) - \left(\frac{c}{2}\left(\frac{1}{2}+\rho\right) - \rho\right)}{(1+\rho)^2\eta},\\ =& \frac{-(1+\rho)\left(1-\frac{c}{2}\right) - \frac{c}{4} + \rho\left(1-\frac{c}{2}\right)}{(1+\rho)^2\eta},\\ =& \frac{-\left(1-\frac{c}{2}\right) - \frac{c}{4}}{(1+\rho)^2\eta} < 0,\end{aligned}$$

since $c \in (0,1]$. Thus, a higher ρ decreases the exit threshold, improving stability.

□

Proof of Lemma 1. When a withdrawal opportunity arises at t, the payoff is $e^{\nu t}$. When a regime switch occurs, the payoff is $e^{\nu t}v$. Altogether, the expected payoff to staking a token if a run will last T periods barring a regime switch is

$$\begin{aligned}V &= \int_0^T \left(\delta e^{-\delta t}e^{-\lambda t}e^{\nu t} + \lambda e^{-\delta t}e^{-\lambda t}e^{\nu t}v\right)\mathrm{d}t,\\ &= \int_0^T (\delta + \lambda v)\, e^{-(\delta+\lambda-\nu)t}\mathrm{d}t,\\ &= \frac{\delta+\lambda v}{\lambda+\delta-\nu}\left(1 - e^{-(\lambda+\delta-\nu)T}\right),\end{aligned}$$

and, therefore, the net payoff to staking instead of withdrawing is

$$\pi = \frac{\delta+\lambda v}{\lambda+\delta-\nu}\left(1 - e^{-(\lambda+\delta-\nu)T}\right) - 1.$$

□

Proof of Proposition 3. At the threshold, the expected net value of staking is zero.

$$\begin{aligned}
\frac{\delta+\lambda v}{\lambda+\delta-\nu}\left(1-e^{-(\lambda+\delta-\nu)T}\right) &= 1, \\
1-e^{-(\lambda+\delta-\nu)T} &= \frac{\lambda+\delta-\nu}{\delta+\lambda v}, \\
e^{-(\lambda+\delta-\nu)T} &= 1-\frac{\lambda+\delta-\nu}{\delta+\lambda v}, \\
e^{-(\lambda+\delta-\nu)T} &= \frac{\nu+\lambda(v-1)}{\delta+\lambda v}, \\
-(\lambda+\delta-\nu)T &= \log\frac{\nu+\lambda(v-1)}{\delta+\lambda v},
\end{aligned}$$

Hence, we have

$$T = \frac{1}{\lambda+\delta-\nu}\log\frac{\delta+\lambda v}{\nu+\lambda(v-1)},$$

We, therefore, have that

$$\hat{A} = \left(\frac{\delta+\lambda v}{\nu+\lambda(v-1)}\right)^{-\frac{\delta}{\delta+\lambda-\nu}}.$$

□

B Model extensions

This section considers extensions to the main models in the body of the paper.

B.1 Proportional Rewards in Static Model

We now consider the static model when staking rewards depend on the pool size. In most cryptographic protocols, including Proof-of-Work (PoW) and Proof-of-Stake (PoS) protocols, if transactions stay the same, then the benefits of staking increase when the staking contract size decreases. That is, a set number of transactions is spread among a smaller and smaller number of validators. When stakers leave the pool, the rewards to remaining stakers proportionally increase. This effect works in reverse of strategic complementarities: if the pool is sufficiently small, then the gains increase and provide

incentives to continue staking. We now investigate runs with this relevant institutional feature included, deriving necessary conditions for runs with proportional rewards.

Suppose that the staking benefit is given by the following function

$$\rho(\ell) = \frac{\tau}{1-\ell} \ ,$$

where τ is the total transactions and $1-\ell$ is the remaining size of the staking contract. In this case, the rewards to staking are explicitly spread among the remaining validators. Consider the benefits to selling for $\ell < \theta$. Then we have

$$\pi(\ell, \theta) = \frac{\tau}{1-\ell} - c\ell\left(\frac{1}{2} + \frac{\tau}{1-\ell}\right).$$

Strategic complementarities require that $\dfrac{\partial \pi(\ell,\theta)}{\partial \ell} < 0$. We note that

$$\frac{\partial \pi(\ell,\theta)}{\partial \ell} = \frac{\rho(\ell)}{1-\ell}(1-c\ell) - c\left(\frac{1}{2} + \rho(\ell)\right),$$

is negative whenever

$$\frac{c(1-\ell)^2}{2(1-c)} > \tau.$$

Note that as $\ell \to 1$, the potential gains from validating become infinite *so long as the protocol survives*. But for an ℓ that is sufficiently large, the protocol fails, and therefore $1-\ell$ is bounded. The marginal gains from validating are bounded as well.

Consider the benchmark considered earlier in which the price is unaffected by ℓ, but the price falls to zero if the protocol fails (i.e., $c = 0$ and $\eta = 1$). The equilibrium cutoff satisfies

$$\begin{aligned}\int_0^1 \pi(\ell,\theta)\mathrm{d}\ell &= \int_0^\theta \rho(\ell)\mathrm{d}\ell - \int_\theta^1 1\mathrm{d}\ell, \\ &= -\left(1 - \theta + \tau\log(1-\theta)\right).\end{aligned}$$

For all τ, this has a zero $\theta^* \in (0,1)$, and thus, equilibrium runs can occur with propor-

tional rewards. Furthermore, θ^* is decreasing in τ. Thus, when τ is large and, therefore, θ^* is low, the protocol is relatively strong because runs only occur when $\theta < \theta^*$. Thus, just as higher ρ strengthens the protocol, a higher τ (more transactions to validate) strengthens the protocol.

B.2 Leveraged Stakers in Static Model

We now consider the addition of leverage in the static model. The staking decision is not independent of other financial contracts. For example, stakers may borrow coins to fund their stake. In many borrowing arrangements, the decline in the value of a purchased asset may precipitate payment to a broker so that the investor's equity rises to a maintenance threshold (margin calls). If this margin maintenance necessitates that some stakers exit the pool to sell their staked coin, then the fraction of exiting stakers would be a function of current price.

Thus, leverage decisions aside from the decision to stake could induce greater run risk. We explore this in two separate extensions to the model. In the first, the fraction of forced exiters, as a function of current price of the staked coin, is unknown. In the second, the stop-loss price is unknown, which alters the coordination problem that stakers must solve.

B.2.1 Fraction unknown

Denote the fraction of forced exiters by $\theta(p)$ or simply θ. We will ignore protocol risk for now. Thus, a fraction θ is forced to sell and of those not forced to sell, a fraction ℓ sell strategically.

In terms of interpretation θ captures any exogenous liquidity needs in general. Agents could have liquidity needs because they have borrowed on margin, or they could have maturity mismatch if the staked token backs short-term liabilities. In either case, θ captures pressure to liquidate for non-strategic reasons.

Then we have

$$\pi(\ell, \theta) = c\left(\frac{1}{2} + r\right)(\theta + (1-\theta)\ell) - r.$$

and hence,

$$\begin{aligned}\int_0^1 \pi(\ell, \theta) d\ell &= c\left(\frac{1}{2} + r\right)\left(\theta + (1-\theta)\frac{1}{2}\right) - r, \\ &= c\left(\frac{1}{2} + r\right)\left(\frac{1+\theta}{2}\right) - r, \\ \theta^* &= \frac{4r}{c(1+2r)} - 1,\end{aligned}$$

and runs occur when $\theta > \theta^*$. Since $\theta(p)$ is a decreasing function of p, this means runs occur when $p < p^* \equiv p^{-1}(\theta^*)$.

The broader point here is that liquidation risk is driving instability in this setting. Stabilizing the protocol would require ensuring that stakers are not subject to liquidity risks, meaning that tokens are not staked as part of a liquidity transformation institution, nor are staked tokens funded with collateralized debt.

B.2.2 Stop-loss Price unknown

Following Morris and Shin (2004), we explore a setting in which leveraged stakers are forced to liquidate at a "strike" price or loss limit q_i. In other words, they must sell if $p < q_i$ and they get zero payoff (lose their holdings). Suppose that strikes are correlated

$$q_i = \theta + \eta_i,$$

i.e., θ is the common threshold, and individual agents have a strike distributed idiosyncratically around that threshold, given by η_i. Let the variance of η_i go to zero and we approximate common knowledge, per usual. For simplicity we use additive pricing.

Upper and lower dominance also hold in this setting. First, if $p < q_i$ then traders must sell regardless of what others do; thus, there exists a price such that agents will sell even if nobody else does. But note that if all agents sell, then the price cannot fall below $p - c$,

and thus, if $p - c > q_i \implies p > q_i + c$, then the price will never cause a forced sale. From our earlier analysis, if everybody else sells but the agent's loss limit is not breached (i.e., the same analysis as before!), then an agent would hold so long as

$$c\left(\frac{1}{2}+r\right) < rp \implies p > \overline{p} \equiv \frac{c}{r}\left(\frac{1}{2}+r\right).$$

Thus, we can have an always-hold equilibrium so long as the price is sufficiently high.

The loss limit complicates the payoffs because now payoffs to buying and holding depend on whether the loss limit is breached. Let $\hat{\ell}_i$ denote the maximum quantity of sales before a loss limit is breached for agent i, which satisifies

$$q_i = p - c\hat{\ell}_i.$$

If $\ell < \hat{\ell}_i$ then aggregate sales are insufficient to trigger a stop loss and the payoff is as above. But if $\ell > \hat{\ell}_i$, then the agent will be forced out of the position and have zero payoff. The payoff to holding is

$$u(\ell, p) = \begin{cases} (1+r)(p - c\ell) & \ell < \hat{\ell}_i \\ 0 & \ell > \hat{\ell}_i, \end{cases} \tag{14}$$

and the payoff to selling is

$$w(\ell, p) = \begin{cases} p - \frac{c}{2}\ell & \ell < \hat{\ell}_i \\ \frac{\hat{\ell}_i}{\ell}\left(p - \frac{c}{2}\hat{\ell}_i\right) & \ell > \hat{\ell}_i, \end{cases} \tag{15}$$

since with probability $\frac{\hat{\ell}_i}{\ell}$ the sale gets executed with a price above the loss limit q_i and, otherwise, the price is below the loss limit, yielding a payoff of zero. Therefore, we have

$$\pi(\ell) = \begin{cases} c\left(\frac{1}{2}+r\right)\ell - rp & \ell < \hat{\ell}_i \\ \frac{\hat{\ell}_i}{\ell}\left(p - \frac{c}{2}\hat{\ell}_i\right) & \ell > \hat{\ell}_i. \end{cases} \tag{16}$$

Note that

$$\hat{\ell}_i = \frac{p - q_i}{c}.$$

By uniformity, the probability that $\ell < \hat{\ell}_i$ is $\hat{\ell}_i$. We can write

$$\begin{aligned}
\int_0^1 \pi(\ell) d\ell &= \int_0^{\hat{\ell}_i} \left(c\left(\frac{1}{2} + r\right)\ell - rp \right) d\ell + \int_{\hat{\ell}_i}^1 \frac{\hat{\ell}_i}{\ell}\left(p - \frac{c}{2}\hat{\ell}_i\right) d\ell, \\
&= c\left(\frac{1}{2} + r\right)\frac{\hat{\ell}_i}{2} - \hat{\ell}_i rp + \hat{\ell}_i\left(p - \frac{c}{2}\hat{\ell}_i\right)\int_{\hat{\ell}_i}^1 \frac{1}{\ell} d\ell, \\
&= c\left(\frac{1}{2} + r\right)\frac{\hat{\ell}_i}{2} - \hat{\ell}_i rp - \hat{\ell}_i\left(p - \frac{c}{2}\hat{\ell}_i\right)\log\left(\hat{\ell}_i\right).
\end{aligned}$$

Plugging in for $\hat{\ell}$ we have

$$\begin{aligned}
\int_0^1 \pi(\ell) d\ell &= \frac{c}{2}\left(\frac{1}{2} + r\right)\frac{p-q}{c} - \frac{p-q}{c} rp - \frac{p-q}{c}\left(p - \frac{p-q}{2}\right)\log\left(\frac{p-q}{c}\right), \\
&= \frac{c}{2}\left(\frac{1}{2} + r\right)\frac{p-q}{c} - \frac{p-q}{c} rp - \frac{p-q}{c}\left(\frac{p+q}{2}\right)\log\left(\frac{p-q}{c}\right), \\
&= \frac{p-q}{c}\left(\frac{c}{2}\left(\frac{1}{2} + r\right) - rp - \left(\frac{p+q}{2}\right)\log\left(\frac{p-q}{c}\right)\right).
\end{aligned}$$

We know that $p > q$ (you are forced to sell if $p < q$ so we must be considering strategic decisions only if not forced), and thus we can drop the first term. Thus, $\int_0^1 \pi(\ell) d\ell = 0$ if

$$\begin{aligned}
0 &= \frac{c}{2}\left(\frac{1}{2} + r\right) - rp - \left(\frac{p+q}{2}\right)\log\left(\frac{p-q}{c}\right), \\
\left(\frac{p+q}{2}\right)\log\left(\frac{p-q}{c}\right) &= \frac{c}{2}\left(\frac{1}{2} + r\right) - rp, \\
\log\left(\frac{p-q}{c}\right) &= \frac{\frac{c}{2}\left(\frac{1}{2} + r\right) - rp}{\frac{p+q}{2}}, \\
p &= q + c * exp\left(\frac{c\left(\frac{1}{2} + r\right) - 2rp}{p+q}\right).
\end{aligned}$$

Since the exponential function is strictly positive, this means that agents sell at some price $p > q$, i.e., they sell preemptively before they are forced.

B.3 Vanishing Noise in Dynamic Model

We now extend the dynamic model to include a potentially richer stochastic process. Suppose that the protocol strength evolves stochastically according to

$$\mathrm{d}\theta_t = \mu_\theta \mathrm{d}t + \sigma_\theta \mathrm{d}Z_t,$$

where Z_t is a Brownian motion. In contrast to the main analysis in the body of the paper, the drift μ_θ is fixed, but Brownian shocks can increase or decrease the protocol strength.

Following the argument in Frankel and Pauzner (2000), there is a unique equilibrium $\phi^*(A)$ whenever $\sigma_\theta > 0$. We can derive the cutoff $\phi^*(A)$ explicitly if we let $\mu_\theta, \sigma_\theta \to 0$. Fundamental risk goes to zero, but we maintain the friction in staking. By the same arguments in Frankel and Pauzner (2000) and Guimaraes (2006), there is a unique threshold $\phi^*(A_0)$ that generates a run. With vanishing shocks, we can then conclude that once the system moves into the region of the state space with a run, it will continue until the protocol fails. Specifically, the system starts at (A_0, ϕ^*) with $A_0 > \phi^*$. Then a run is triggered and A_t could decline until $A_t = \phi^*$, at which point the protocol fails. Indeed, equilibrium is given precisely by Proposition 3 in this model.

B.4 Connecting value and liquidity

The Proof-of-Stake (PoS) protocol could be strengthened by changing the value v that agents receive in the absence of a run. Note that

$$\begin{aligned}
\frac{dT}{dv} &\propto \frac{\lambda}{\delta + \lambda v} - \frac{\lambda}{\nu + \lambda(v-1)}, \\
&\propto \lambda \left(\nu + \lambda(v-1) - (\delta + \lambda v)\right), \\
&\propto \nu - \delta - \lambda = -(\lambda + \delta - \nu) < 0,
\end{aligned}$$

which is therefore negative given our assumption on parameters. Thus, increasing v decreasing the length of the run. This also increases $\hat{A}$. Hence, ϕ^* is also increasing in v (because $\hat{A}$ is higher, and we subtract off a lower T). Thus, a higher v strengthens the

protocol by increasing the run threshold.

Value and liquidity shocks In practice, stakers are investors that face an opportunity cost when using funds for staking. In practice, investors may have a liquidity need, or in the case of pooled staking, may see a substantial customer withdrawal of funds otherwise supplied for staking purposes. We examine such liquidity needs here.

Suppose that investors might be subject to liquidity risks at a rate β. Specifically, there is a normal state in which there are no liquidity risks and agents earn a flow utility ρ from staking. However, with rate β, they may switch to a liquidity risk state, due to exogenous liquidity needs. This state will persist until another shock occurs, also at rate β. For tractability of these arguments, we suppose that if stakers can withdraw their token before the next Poisson shock, then they consume and get the value ρ. If they do not get an opportunity to withdraw, then they receive zero.

We let v_l denote the value stakers derive when staking in a protocol in the liquidity risk state. We let v_0 denote the value stakers derive when staking in a protocol in the normal state. Because an opportunity to exit arrives at a rate δ (i.e., the probability of receiving a withdrawal opportunity before the next Poisson shock), the expected value of staking in the liquidity risk state is

$$v_l = \frac{\delta}{\delta + \beta}\rho.$$

The value function in the normal state satisfies

$$0 = \rho + \beta(v_l - v_0) \implies v = \frac{\rho}{\beta} + v_l.$$

Hence, the value is

$$v_0 = \rho\left(\frac{1}{\beta} + \frac{\delta}{\delta + \beta}\right).$$

Note that

$$\frac{dv_0}{d\delta} = \frac{\beta\rho}{(\delta + \beta)^2} > 0,$$

which means that decreasing the withdrawal rate will decrease utility. This is intuitive: if

it is harder to withdraw, then liquidity opportunities will be missed more frequently. This also means there is a tradeoff in setting the contract rate δ. If liquidity risks are sufficiently important (i.e., if β is sufficiently high), then increasing δ may not stabilize the protocol.