12-2023

The economics of financial scams: Evidence from Initial Coin Offerings

Kenny PHUA

Bo SANG
Singapore Management University, bo.sang.2017@pbs.smu.edu.sg

Chi Shen WEI

Yang YU
Singapore Management University, gloriayu@smu.edu.sg

The Economics of Financial Scams: Evidence from Initial Coin Offerings*

Kenny Phua Bo Sang Chishen Wei Gloria Yang Yu

Abstract

We examine the economics of financial scams by analyzing the market for initial coin offerings (ICOs). Using data snapshots of 5,873 ICOs, we find that irregularities in ICO characteristics across listing websites predict higher scam risk and are likely intentional. These patterns are consistent with a model where malicious issuers maximize profits by using irregularities to screen for naïve investors. Almost half of the ICOs in our sample may be scams, amounting to more than U.S. $6 billion in losses. Our results draw attention to the frequent use of screening mechanisms in financial scams.

Keywords: Financial scams, Forensic finance, Cryptocurrency
JEL classification: D40, D84, G12, G14

*Yu (corresponding author): gloriayu@smu.edu.sg, Singapore Management University. Phua: kenny.phua@uts.edu.au, University of Technology Sydney. Sang: bo.sang@bristol.ac.uk, University of Bristol. Wei: chishen.wei@polyu.edu.hk, Hong Kong Polytechnic University. We are grateful for insightful comments from Mykola Babiak, Thomas Bourveau, Shaen Corbett, Stephen Dimmock, John Griffin, Hogyu Jhang, Leo Liu, Thomas Matthys, Marco Navone, Tālis Putniņš, Daniel Rabetti, Kanis Saengchote, Wanyi Wang, Jing Xu, and conference/seminar participants at American Finance Association (AFA) Annual Meeting 2023, Asian Bureau of Finance and Economic Research (ABFER) Annual Conference 2022, Asian Finance Association Annual Conference 2022, Boca Corporate Finance and Governance Conference 2022, Financial Markets and Corporate Governance (FMCG) Conference 2022, Global AI Finance Research Conference 2021, International Cardiff Fintech Conference 2023, UWA Blockchain and Cryptocurrency Conference 2021, Vietnam Symposium in Banking and Finance 2021, Australian National University, Hong Kong Polytechnic University, Massey University, Monash University, Nanyang Technological University, Queensland University of Technology, Singapore Management University, University of Adelaide, University of Melbourne, University of Queensland, University of Sydney, and University of Technology Sydney. We thank Windy Chung Mai, Yao Yumi Tong, and Ruochen Yin for excellent research assistance. This study is funded by the Singapore Ministry of Education Grant 18-C207-SMU-007. This paper was previously circulated under the title "Don't Trust, Verify: The Economics of Scams in Initial Coin Offerings".

The Economics of Financial Scams: Evidence from Initial Coin Offerings

Abstract

We examine the economics of financial scams by analyzing the market for initial coin offerings (ICOs). Using data snapshots of 5,873 ICOs, we find that irregularities in ICO characteristics across listing websites predict higher scam risk and are likely intentional. These patterns are consistent with a model where malicious issuers maximize profits by using irregularities to screen for naïve investors. Almost half of the ICOs in our sample may be scams, amounting to more than U.S. $6 billion in losses. Our results draw attention to the frequent use of screening mechanisms in financial scams.

`Keywords:` Financial scams, Forensic finance, Cryptocurrency
`JEL classification:` D40, D84, G12, G14

1 Introduction

According to the U.S. Federal Trade Commission, consumer losses to fraud and scams in 2022 alone totaled U.S. $8.8 billion. Overall welfare losses are likely more severe because victims often suffer depression, shame, and unemployment.[1] To reduce and hopefully prevent such harm, researchers are investigating the economics of financial crime and the conditions that lead to the prevalence of fraud. For example, recent studies uncover the scale and determinants of financial advisor misconduct (Egan, Matvos, and Seru, 2019; Dimmock, Gerken, and Van Alfen, 2021) and shed light on the anatomy of organized cybercrimes (Cong, Harvey, Rabetti, and Wu, 2023). However, evidence on financial scams is scarce because we rarely observe the techniques used to commit such crimes, and victims are often reluctant to come forward.

We examine the economics of financial scams by exploiting a unique setting in the market for initial coin offerings (ICOs). ICOs are a form of crowdfunding for blockchain projects and have raised an estimated U.S. $50 billion dollars through 2020 (PriceWaterhouseCoopers, 2020). ICOs can help reduce frictions relating to asymmetric information and agency costs in the early stages of cutting-edge entrepreneurial ventures (Howell, Niessner, and Yermack, 2020; Benedetti and Kostovetsky, 2021; Davydiuk, Gupta, and Rosen, 2022; Lee, Li, and Shin, 2022). But there is also widespread belief that scams, fraud, and abuse are commonplace in the ICO market. This market offers an ideal laboratory to study financial scams because we can systematically (i) observe how issuers market their offerings to prospective investors and (ii) analyze blockchain data that are publicly available in immutable ledgers.

To investigate how ICOs were marketed to investors, we collect point-in-time snapshots of self-reported ICO data from five leading ICO listing websites. ICO data do not reside in a centralized repository and are instead scattered across various listing websites.[2] Consistent with Lyandres, Palazzo, and Rabetti (2022), we find rampant cross-site irregularities in ICO data. For example, the AdHive ICO was marketed on three websites with conflicting material disclosures (see, Figure 1). Strikingly, over a

[1] Button, Lewis, and Tapley (2009) examine how victims fare in the aftermath of scams.

[2] Listing websites host the ICO listing information and are distinct from cryptocurrency exchanges or brokerages.

third of 5,873 ICOs have such irregularities at their first appearances in our sample.

- Figure 1 here -

To understand the prevalence of irregularities, we model the behavior of malicious and honest ICO issuers who aim to maximize profits. Issuers face a pool of naïve and astute investors, whose types are ex ante unobservable. Naïve investors do not conduct due diligence and are vulnerable to funding ICO scams. In contrast, astute investors will carefully evaluate the ICO offering and demand more information on public forums. These demands increase the risk that a malicious issuer is caught or exposed as a scam (Becker, 1968) and impose an opportunity cost on their time (Ehrlich, 1973). Astute investors are sophisticated and eventually refrain from funding scams. For a malicious issuer, astute investors are hence undesirable because they impose costs but provide no funds to the offering.

In this setting, irregularities act as an effective screening device for malicious issuers to filter out astute investors as early as possible. Most astute investors notice the irregularities, deduce that the ICO is fraudulent, and dismiss the offering without imposing any costs on the malicious issuer.[3] The remaining pool of investors would be mostly naïve investors—the ideal targets of the malicious issuer. Honest issuers are marketing legitimate investments and have nothing to hide. Therefore, they have no incentives to screen with irregularities.

We test the model prediction that ICOs with irregularities have higher scam risk. To identify ICO scams, we collect crowdsourced scams from `DeadCoin.com`. We corroborate these records with multiple sources and characterize them using zero-shot learning (ZSL), a machine learning method based on large language models. Estimates from Cox regressions reveal that the odds of an ICO being a scam more than double in the presence of irregularities. At the intensive margin, an additional irregularity increases the odds of a scam by 11.6%. We also affirm the consistency of our conclusions in supplementary OLS estimations.

The open-access feature of blockchain technology offers a unique opportunity to

[3]The screening process may be imperfect, so some astute investors could still ask questions and impose costs on the malicious issuer.

evaluate the screening mechanism by performing on-chain analysis of wallets on the Ethereum network. We collect data on token holdings to characterize the sophistication of the typical token holder in every ICO. We find that wallets holding tokens of ICOs with irregularities (i) have lower portfolio values, (ii) are less diversified, and (iii) are less active. Consistent with our screening mechanism, ICOs with irregularities also seem to filter out more investors who might have demanded more information. We find that `Reddit` message boards of ICOs with irregularities have fewer questions, comments, and unique users. These patterns suggest that ICOs with irregularities are more likely to attract naïve investors.

Although the ICO market was fairly short-lived, the economic insights from our study could hold external validity for the design of other types of financial scams. The authorities and financial institutions in Table 1 warn that perpetrators often use misspellings and grammatical errors in scams. While these errors are obvious to many, our findings suggest that these irregularities are designed to target inattentive victims and quickly screen out savvy people who are unlikely to fall prey. The design of the irregularity may vary with the context, but we provide an economic explanation for the widespread use of this tactic in scams.

- Table 1 here -

We evaluate alternative explanations for our results. First, we verify that irregularities are produced by ICO issuers, not by the listing websites. To minimize the risk of data errors or oversight, we only collect ICO data from high-quality and prominent listing websites. We also check that these websites rely on self-reported and open-access submission of data by the issuers. Our investigation suggests that a substantial portion of ICO data is self-reported by issuers. In any case, irregularities due to data collection errors of listing websites likely reflect noise and should not predict ICO scam risk.

Second, we examine whether irregularities are simply unintentional mistakes. For this analysis, we track advisors who are hired to provide support in marketing, fundraising, and technical execution. We create a network by linking ICOs based on their common advisors. If irregularities were idiosyncratic, unintentional mistakes, they should

be randomly distributed throughout this network. Instead, we find that the distribution of irregularities is too systematic to be explained by chance. Using network analysis, our evidence suggests that the use of irregularities is learned from or passed along through common advisors.

Third, we consider the view that irregularities are a symptom of low issuer quality, not malice. Controlling for various proxies of ICO quality from Bourveau, De George, Ellahie, and Macciocchi (2021) and Davydiuk, Gupta, and Rosen (2022), we show that irregularities continue to predict ICO scam risk. Furthermore, if the motives behind irregularities are nefarious, regulatory scrutiny should deter malicious issuers from entering the ICO market. Indeed, we find that ICOs launched shortly after news of regulatory action in cryptocurrency markets are less likely to have irregularities.

While our findings indicate that malicious issuers use irregularities as a screening device, other tactics to target naïve investors may also be in play. We examine two potential tactics. First, we find that malicious issuers promote their ICOs on listing websites that generate more traffic from paid advertisements, referral links, and search engines. Such website traffic tends to be composed of more naïve investors. Second, consistent with investor warnings issued by the Securities Exchange Commission (SEC), we find that celebrity endorsements are strongly associated with ICO scam risk. In both tests, we still find that irregularities incrementally predict ICO scam risk with comparable economic magnitudes.

Finally, we perform a welfare analysis on the potential financial losses from ICO scams. A key challenge in this analysis is that many scams may go undetected because victims are often reluctant to report losses. To address the underdetection of ICO scams, we use detection-controlled estimation (DCE) methods (Feinstein, 1990). Our DCE results indicate that up to 2,893 ICOs (49.3% of ICOs) in our sample could be scams with associated financial losses exceeding U.S. $6 billion. These estimates represent a ten-fold increase over the scam rate implied from the `Deadcoins` data, highlighting the potential underdetection problem typical of scams.

Our study contributes to a growing literature on the economics of financial crime. Egan, Matvos, and Seru (2019) find that financial advisors who "specialize" in miscon-

duct tend to target unsophisticated investors and work at firms that tolerate misconduct. Dimmock, Gerken, and Graham (2018) find that financial advisors learn how to commit misconduct from their colleagues. Likewise, our analysis also shows that ICO scams aim to target less sophisticated investors, and that common ICO advisors seem to facilitate the use of irregularities. We add to this literature by demonstrating how perpetrators could engineer a screening strategy to target naïve victims for profit. By providing an economic explanation for the widespread use of irregularities in financial scams, our paper can help authorities understand how fraud is carried out.

Our paper also adds to the evidence on the controversies surrounding cryptocurrencies (Yermack, 2015). For example, Griffin and Shams (2020) find that Tether, a digital currency pegged to the U.S. dollar, is used to manipulate Bitcoin prices. Li, Shin, and Wang (2021) and Dhawan and Putniņš (2022) document choreographed pump-and-dump trading schemes in cryptocurrencies. Aloosh and Li (2019), Amiram, Lyandres, and Rabetti (2020), and Cong, Li, Tang, and Yang (2020) find evidence of wash trading that artificially boosts trading volumes on crypto-exchanges. Foley, Karlsen, and Putniņš (2019) and Makarov and Schoar (2021) find that Bitcoin is used for trading and speculative purposes, but it also facilitates illicit activities. Although our results may shine an unflattering light on cryptocurrencies, our primary aim is to advance our understanding of the economics behind financial crime and misconduct.

2 Why are irregularities so prevalent?

We develop a simple model to rationalize the use of irregularities as a screening device in the ICO market. Our model illustrates the profit-maximizing motive of ICO issuers, incorporating elements of rational choice (Becker, 1968) and the opportunity cost of time (Ehrlich, 1973) from the literature on the economics of crime. It also embeds the targeting strategy of scammers (Herley, 2012) in a market with honest and malicious issuers. We show that there is a separating equilibrium where malicious issuers opt to screen with irregularities but honest issuers do not.

2.1 Setup

There are two types of ICO issuers $\theta \in \{H, S\}$ (honest, scam). Issuers face a mass of p investors, of which there are q naïve investors and $p-q$ astute investors.[4] Investor types $\tau \in \{N, A\}$ (naïve, astute) are ex ante unobservable. We define x as the number of irregularities that an investor tolerates, above which she immediately dismisses the ICO as a scam. The x parameter captures the degree to which an investor is careful or performs due diligence. The conditional PDF of x for an investor-type is $\phi(x \mid \tau)$, which has support over the set of nonnegative, real numbers $x \in [0, \infty)$.

Our model has three periods. In period one, an issuer chooses the number of irregularities $\mathcal{I}(\theta)$, which acts as a cutoff to target a pool of investors. Given a $\mathcal{I}(\theta)$, the fraction of astute investors who immediately dismiss the ICO is $\int_0^{\mathcal{I}(\theta)} \phi(x \mid \tau = A)\, dx$. Some astute investors fail to dismiss the ICO and remain potential targets. The conditional complementary CDF (CCDF) $\int_{\mathcal{I}(\theta)}^{\infty} \phi(x \mid \tau = A)\, dx$ gives the fraction of astute investors who remain. Likewise, the fraction of naïve investors who remain potential targets is $\int_{\mathcal{I}(\theta)}^{\infty} \phi(x \mid \tau = N)\, dx$. Essentially, the choice of $\mathcal{I}(\theta)$ characterizes the issuer's targeting strategy. A higher (lower) $\mathcal{I}(\theta)$ targets lower (higher) fractions of both investor-types because CCDFs are decreasing in x by definition.

In period two, the remaining astute investors—who have not dismissed the ICO—raise queries about the offering on public forums. Without loss of generality, we assume that naïve investors are unsophisticated and fail to raise queries. All issuers incur a marginal operational cost $\kappa^O > 0$ to address every query raised by the astute investors. This cost relates to the notion of an opportunity cost of time faced by perpetrators (Ehrlich, 1973). For malicious issuers, every query imposes an additional exposé cost $\kappa^E > 0$ because queries risk prematurely "blowing the cover" of scam ICOs. Honest issuers are

[4] To focus on the screening mechanism, our model abstracts away from investors' incentives to participate in the ICO market. There are at least two reasons why investors may be willing to fund ICOs despite the prevalence of scams. First, investors may be attracted to the high skewness in the distribution of ICO returns. Conditional on successful listings on cryptocurrency exchanges, Lyandres, Palazzo, and Rabetti (2022) find that the average (maximum) ICO return on the first trading day is 384% (3,870%). These patterns imply that investors may also be willing to make many losing bets in hopes of capturing an investment that yields outsized returns. Second, overconfident investors (Daniel, Hirshleifer, and Subrahmanyam, 1998; Odean, 1998) may be willing to participate in the ICO market because they overestimate their ability to evaluate ICOs and avoid scams.

not fraudulent, so they face no exposé costs ($\kappa^E = 0$).

In the final period, every investor decides whether to fund a positive finite $f \in \mathbb{R}_{>0}$, which represents the expected benefit to the issuer. For a malicious issuer, the marginal cost to target an astute investor is $\kappa^O + \kappa^E$. We assume per-investor funds exceed these costs (i.e., $f > \kappa^O + \kappa^E$). Astute investors only fund honest ICOs because they ultimately become suspicious of scams and refrain from funding them. In other words, when malicious issuers target an astute investor, they receive zero funds and suffer a loss of $\kappa^O + \kappa^E$. Naïve investors fund both ICO types. The matrix below summarizes the net profits received by each issuer-type θ from each investor-type τ.[5]

	Issuer type	
Investor type	Honest (H)	Scam (S)
Naïve (N)	f	f
Astute (A)	$f - \kappa^O$	$0 - \kappa^O - \kappa^E$

2.2 Optimal targeting strategies

Given the above payoff matrix, each issuer decides on an optimal targeting strategy to maximize profits given the gain-versus-loss tradeoff (Becker, 1968). We use a receiver operating characteristics (ROC) curve to analyze the issuers' binary classification problem. Figure 2 presents a hypothetical ROC curve. A given point on the curve represents a targeting strategy $\mathcal{I}(\theta)$ that trades off profitable naïve investors against costly astute investors. The slopes at points on the ROC curve characterize these tradeoffs.

To offer some intuition, we consider two extreme targeting strategies on the ROC curve. Point (i) has a flat slope and represents an indiscriminate targeting strategy

[5]Our assumption of perfect discernment (gullibility) in astute (naïve) investors is without loss of generality and only acts to ease mathematical exposition. We could have allowed targeted astute (naïve) investors to stochastically fall for (detect) ICO scams and fund them. Our key insights will hold as long as malicious issuers experience a positive (negative) payoff from targeting a(n) naïve (astute) investor.

($\mathcal{I}(\theta) = 0$) that targets all naïve investors and retains all costly astute investors. Point (ii) has a steep slope and represents a conservative targeting strategy ($\mathcal{I}(\theta) \to \infty$) that avoids all costly astute investors but forgoes all profitable naïve investors.

- Figure 2 here -

Definition 1. *For a given $\mathcal{I}(\theta)$, the corresponding point on the ROC curve has a nonnegative slope that is the first derivative of the fraction of naïve investors targeted (i.e. true positive rate) with respect to the fraction of astute investors targeted (i.e., false positive rate).*

$$\begin{aligned}\frac{\partial(\int_{\mathcal{I}(\theta)}^{\infty}\phi(x \mid \tau = N)\,dx)}{\partial(\int_{\mathcal{I}(\theta)}^{\infty}\phi(x \mid \tau = A)\,dx)} &= \frac{(1-\Phi(\mathcal{I}(\theta) \mid \tau = N))'_{\mathcal{I}(\theta)}}{(1-\Phi(\mathcal{I}(\theta) \mid \tau = A))'_{\mathcal{I}(\theta)}} \\ &= \frac{\phi(\mathcal{I}(\theta) \mid \tau = N)}{\phi(\mathcal{I}(\theta) \mid \tau = A)} \geq 0 \\ \textit{where} \quad \Phi(\mathcal{I}(\theta) \mid \tau) &\coloneqq \int_0^{\mathcal{I}(\theta)} \phi(\mathcal{I}(\theta) \mid \tau)\,dx\end{aligned}$$

Definition 2. *By the general properties of the ROC curves, the slopes increase monotonically in $\mathcal{I}(\theta)$.*

$$\frac{\partial}{\partial \mathcal{I}(\theta)}\left(\frac{\phi(\mathcal{I}(\theta) \mid \tau = N)}{\phi(\mathcal{I}(\theta) \mid \tau = A)}\right) \geq 0$$

We now derive the optimal targeting strategy of the malicious issuer. Defining the density of naïve investors in the population as $z = q/p \in [0,1]$, the malicious issuer's expected net profits $\pi(S, \mathcal{I}(S))$ are:

$$\frac{\pi(S,\mathcal{I}(S))}{p} = zf\overbrace{\int_{\mathcal{I}(S)}^{\infty}\phi(x \mid \tau = N)\,dx}^{\text{frac. naïve targeted}} - (1-z)\,(\kappa^O + \kappa^E)\overbrace{\int_{\mathcal{I}(S)}^{\infty}\phi(x \mid \tau = A)\,dx}^{\text{frac. astute targeted}} \quad (1)$$

Proposition 1. *The malicious issuer's optimal targeting strategy resides at the point on the ROC curve where the slope is:*

$$\frac{1-z}{z} \cdot \frac{\kappa^O + \kappa^E}{f} \geq 0$$

The detailed proof is in Appendix A.

We discuss comparative statics of the malicious issuer's optimal targeting strategy. When the population density of naïve investors is higher ($z \uparrow$), the malicious issuer's optimal targeting strategy moves rightwards on the ROC curve where the slope is flatter. Similarly, when funds per investor ($f \uparrow$) are higher, the issuer prefers a more aggressive targeting strategy (rightward) because the trade-off between naïve and astute investors—both in absolute numbers and payoffs—is more favorable. On the contrary, higher costs ($\kappa^O \uparrow, \kappa^E \uparrow$) prescribe a more conservative targeting strategy with a steeper slope as targeting astute investors becomes more expensive relative to the gains f from naïve investors.

Remark 1. *A malicious issuer who is more concerned about exposé costs ($\kappa^E \uparrow$) optimally chooses to operate at a steeper slope on the ROC curve by choosing a more positive $\mathcal{I}(S)$.*

Proposition 2. *For $(z < 1) \wedge (\kappa^O + \kappa^E > 0) \wedge (f \in \mathbb{R}_{>0})$, the malicious issuer optimally uses irregularities as a screening device: $\mathcal{I}^*(S) > 0$. The detailed proof is in Appendix A.*

The malicious issuer maximizes profits by using some irregularities ($\mathcal{I}^*(S) > 0$) to screen out some costly astute investors while forgoing some profitable naïve investors. There are more naïve investors to the right of the optimal targeting strategy. However, Definition 2 implies that the incremental costs imposed by the additional astute investors outweigh the gains from capturing these naïve investors.

To examine the honest issuer's optimal targeting strategy, we first express her expected net profits $\pi(H, \mathcal{I}(H))$ as:

$$\frac{\pi(H,\mathcal{I}(H))}{p} = zf \overbrace{\int_{\mathcal{I}(H)}^{\infty} \phi(x \mid \tau = N)\, dx}^{\text{frac. naïve targeted}} + (1-z)\,(f-\kappa^O) \overbrace{\int_{\mathcal{I}(H)}^{\infty} \phi(x \mid \tau = A)\, dx}^{\text{frac. astute targeted}} \quad (2)$$

Proposition 3. *For $f > \kappa^O$, the honest issuer optimally refrains from a screening strategy by choosing no irregularities: $\mathcal{I}(H) = 0$. The detailed proof is in Appendix A.*

Every additional investor, regardless of type τ, always provides the honest issuer a positive profit: $\min(f, f - \kappa^O) > 0$. Thus, she is always better off by targeting an incremental investor. This intuition implies that she optimally adopts an indiscriminate targeting strategy (i.e., $\mathcal{I}^*(H) = 0$) by operating at the point on the ROC curve where the slope is zero.

2.3 Equilibrium

Our model predicts that malicious issuers use irregularities to screen out some astute investors while honest issuers abstain from irregularities. This separating equilibrium is stable if issuers have no incentives to deviate from their optimal targeting strategies.

Proposition 4. *Suppose (i) some investors are astute ($z < 1$); (ii) investor queries are costly ($\kappa^O + \kappa^E > 0$); and (iii) funds raised per investor are positive and finite ($f \in \mathbb{R}_{>0}$). Then there exists a separating equilibrium in which* $\Pr(\mathcal{I}^*(H) = 0 \mid \theta = H) = 1$ *and* $\Pr(\mathcal{I}^*(S) > 0 \mid \theta = S) = 1$. *The detailed proof is in Appendix A.*

We first show by contradiction that the malicious issuer can do better by deviation if and only if it were profitable to adopt an indiscriminate targeting strategy in the first place. That is, the conditions laid out in Proposition 4 must fail to hold. Next, we show that the honest issuer must forgo some investors if she screens with irregularities. However, every targeted investor regardless of type τ yields a positive profit $\min(f, f - \kappa^O) > 0$ for the honest issuer. Thus, any screening strategy must be suboptimal to her.

2.4 Discussion

Our model formalizes the intuition behind the commonly observed pattern that scams—including non-ICO ones—often contain irregularities. From the perpetrator's point of view, it is generally optimal to screen out some unprofitable segments of the victim pool when it is marginally costly to target certain victims.

In the ICO setting, malicious issuers devise a strategy that induces naïve investors to self-identify. Investors who overlook irregularities and remain viable victims are

likely naïve—the ideal targets of malicious issuers. Thus, the primary prediction of our model is that ICOs with irregularities are more likely to be scams. Moreover, REMARK 1 predicts that the relation between irregularities and scam risk should also hold on the intensive margin. A higher number of irregularities corresponds to higher (perceived) exposé risk, which likely reflects greater scam risk.

Our model can help rationalize the operational designs of other scams. For example, typographical and grammatical errors are common in email hoaxes and SMS phishing attacks. At first impression, these irregularities are perplexing because most people are astute enough to spot them and simply dismiss the scam. However, the scammers want to only retain naïve people who overlook these red flags and are therefore more likely to be viable victims. Notably, vigilantes disrupt the operations of scammers by asking them many questions and monopolizing their time. Indeed, the advice from regulators to ask more questions acts not only to inform investors, but also to increase the costs to scammers.

3 Data, variables, and descriptive statistics

This section summarizes the institutional features of ICO listing websites, our data collection process, the main variables, and the descriptive statistics of our sample.

3.1 Institutional features of ICO listing websites

ICO listing websites provide a platform for issuers to market their offerings to the general public.[6] To list an ICO, the issuer often submits ICO information directly to the website for approval (Lyandres, Palazzo, and Rabetti, 2022). Submissions typically require minimal technical sophistication. For example, Figure 3 contains a screenshot of the sign-up page on a representative listing website. Listings are typically free, but listing websites may promote and rate an ICO for a fee (Cohney, Hoffman, Sklaroff, and Wishnick, 2019).

- Figure 3 here -

[6]The Internet Appendix contains a detailed overview of ICOs.

3.2 ICO data from listing websites

We systematically collect point-in-time ICO data snapshots from five major websites that aggregate ICO listings—(i) `ICOBench`, (ii) `ICOCheck`, (iii) `ICOData`, (iv) `ICODrops`, and (v) `ICORating`. We select these five listing websites based on (i) their popularity reported by Alexa Traffic Rank on August 15th 2018, (ii) the number of ICOs covered, and (iii) the technical feasibility of scraping the websites.[7] On the 15th of every month from August 2018 to August 2019, we scrape ICO data from these five websites. In total, we have 13 data collection events and a time series of ICO characteristics for every ICO-website pair. Because ICO identifying information may vary across websites, we manually cross-check all ICOs and designate a set of unique identifiers to every ICO in our sample. To resolve residual conflicts in our collected data, we hand-check our data against other Internet sources. Thus, we alleviate concerns of variation in ICO names, misspellings, and name changes. Overall, our sample contains 5,873 matched ICOs.[8]

We investigate whether ICO data were primarily self-reported by issuers by retrieving the legal disclaimers of listing websites from the `Wayback Machine`. We find that `ICORating` (perma.cc/65XV-376Q) and `ICODrops` (perma.cc/2SLU-ZUHG) explicitly relied on issuers to report accurate ICO data.

> By sending information, the Client confirms its accuracy and validity. This information is used as a basis for writing an Audit report [...]
> - `ICORating` (5th February 2019)

> All information, data, white papers and other materials concerning a particular token sale is prepared solely by its organizer, and such person is solely responsible for the accuracy of all statements it has made. [...]
> - `ICODrops` (24th January 2019)

The `ICOBench` website had no such disclaimers but boasted of an ICO submission process that was "as easy as ABC" (perma.cc/4DHB-LFHZ).

> (1) Create a free personal account using your corporate email [...] (4) Fill

[7] Based on the Alexa Traffic Rank on November 30th 2018, Lyandres, Palazzo, and Rabetti (2022) obtain ICO data from `ICOBench`, `ICODrops`, `ICORating`, `ICOMarks`, and `ICOData`. We replace `ICOMarks` with `ICOCheck` for the latter two considerations.

[8] The numbers of unique ICOs covered by the listing websites are: `ICORating` (4,166), `ICOBench` (4,021), `ICOData` (1,896), `ICODrops` (625), and `ICOCheck` (580).

in with accurate data and submit the form. [. . .]
- ICOBench (31st March 2019)

The ICOCheck and ICOData websites provide open-access submission of ICO data but do not provide any legal disclaimers or guarantees on the accuracy of the information.

Overall, it appears that a substantial portion of the ICO data was self-reported by issuers. Although we cannot categorically rule out the scenario that listing websites collected ICO data independently and made their own mistakes, it is unlikely then that irregularities would predict ICO scam risk.

3.3 Identifying ICO scams

We collect ICO scam allegations from a prominent, crowdsourced anti-fraud project hosted on the now-defunct DeadCoins.com. This data has been used in other ICO studies (e.g., Davydiuk, Gupta, and Rosen, 2022). The DeadCoins website (perma.cc/AEV2-KW27) curated a list of alleged ICO scams, accompanied by reasons behind the allegations. Using a machine-learning technique, known as zero-shot learning (ZSL), we find that these reasons include signs of dishonesty, product concerns, reputation issues, communication stoppage, and regulatory action.[9] For example, the Shopin ICO token was marked as "dead" (i.e., inactive) on Deadcoins following a SEC complaint. Subsequently, the founders behind the Shopin ICO were charged with securities fraud and violations of registration processes.

To reduce Type-I errors, we corroborate every Deadcoin scam allegation with several media sources. Notably, the Deadcoin website also prominently displays a form to contest scam allegations. First, we check whether the ICO is reported by regulatory authorities (e.g., SEC, DoJ). Second, we search on Factiva for press coverage (e.g., news articles, website articles, journal articles) of the ICO scam. Third, we search popular online forums and social media (e.g., Reddit, Cryptocompare) for mentions of the ICO scam. We admit an alleged ICO scam into our sample only if it is found on at least one of the above three media channels. In total, we match 243 ICO scams

[9]The Internet Appendix contains more details on the ZSL analysis of these ICO scam allegations.

to our sample.

3.4 Variables

Our key independent variable is *irregularities*, which is defined as the total number of cross-website discrepancies in the characteristics of an ICO at its first appearance in our sample. Specifically, we collect the common set of 13 characteristics that are reported across the five websites. The characteristics are *banned*, *whitelist*, *presale*, *hardcap*, *softcap*, *accept BTC*, *accept ETH*, *accept USD*, *ticker*, *start date*, *end date*, *duration*, and *country*. We define these characteristics below. Figure 4 visualizes the proportion of ICOs with at least one cross-website discrepancy by these characteristics at first appearances in our sample. Irregularities commonly occur in *whitelist* (30.3%), *start date* (24.0%), *end date* (24.1%), *presale* (19.6%), and *banned* (13.7%). Irregularities in *softcap*, *ticker*, and *country* are uncommon.

- Figure 4 here -

In our empirical tests, we control for variables that describe the fundraising structure of an ICO. The following control variables are coded as indicators that switch on if the ICO has the corresponding fundraising features. An ICO is *banned* if it is banned by at least one regulatory authority. A *whitelist* allows an ICO issuer to limit the sale of tokens to a selected group of registered investors. An ICO can hold a *presale* round to sell tokens before the public fundraising campaign is launched. The *hardcap* is the maximum amount of funds that can be sold in an ICO. The *softcap* is the minimum amount of funds that must be raised in an ICO, or else funds are returned to investors and the project is discontinued. We control for payment options in the ICO with *accept BTC* (*ETH*, *USD*). The last indicator is *SEC filing*, which switches on if the ICO has regulatory filings with the SEC. The *duration* of an ICO is the length of its fundraising period in days. Finally, we control for the regulatory environment in the ICO's country of registration with the *enforcement* and *disclosure* indices from La Porta, Lopez-De-Silanes, Shleifer, and Vishny (2000).

We search for regulatory filings (Form D, Form 1-A, and Form C) of ICOs that are available on the SEC EDGAR database. We search the database using the keywords

"token", "ICO", "initial coin offering", "coin", and "crypto". Then we manually determine whether every filing is ICO-related. We first read the filing document and check whether it pertains to an initial coin offering or other types of offering. If this information is not stated, we then use the firm name written in the document combined with the keywords "ICO", "offering", "token" to perform a search on SEC EDGAR. All else failing, we use the names of persons (i.e., founders, CEOs, and directors) in the filing combined with the above keywords to perform another search on SEC EDGAR. In our sample, 77, two, and eight ICOs have filed for a Form D, Form 1-A, and Form C, respectively.

3.5 Descriptive statistics

Table 2 reports summary statistics of our sample. Panel A reports that the average ICO has 1.30 irregularities, and 35% of ICOs have at least one irregularity. 95% of ICOs are banned in at least one country, which is unsurprising as ICOs are illegal in several countries (e.g., China, Egypt, Morocco). About half of ICOs impose selectivity in their investor clientele or fundraising structures as 55% of ICOs have an investor *whitelist*, and 48% of them have *presale* rounds. While most ICOs (70%) have a *hardcap* in their fundraising structures, only a minority (31%) have a *softcap*. ETH (USD) is the most (least) popular payment currency among ICO issuers. Fewer than 1% of ICOs in our sample have regulatory filings with the SEC. The fundraising period for the average (median) ICO is 54 (36) days. Panel B reports the Pearson pairwise correlations among our variables. Our key variable *irregularities* is weakly correlated with most variables, except for *presale* (31%), *hardcap* (28%), and *accept ETH* (30%).

- Table 2 here -

Table 3 produces a comparison between ICOs with and without irregularities. This simple exercise provides initial evidence in support of our main hypothesis. ICOs with at least one irregularity are more likely to incur a scam allegation (7% vs. 2%). ICOs with irregularities also have weaker governance—they are less likely to have a investor *whitelist* (46% vs. 60%) and are more likely to hold a *presale* funding round (69% vs. 37%). Such ICOs tend to signal limited supply of tokens with shorter fundraising periods (*duration* of 47 days vs. 57 days) and greater likelihood of a *hardcap* (89% vs. 60%). ICO irregularities also correlate with a wider range of payment options.

- Table 3 here -

4 Irregularities and ICO scams

This section presents our main findings on the hypothesis that issuers use irregularities to screen for naïve investors.

4.1 Survival analysis: ICO scam risk

Using survival analysis, we test whether ICOs with more irregularities are more likely to be scams. We track the survival time of an ICO as the time elapsed between its entry into our sample and the occurrence of a scam allegation. Our empirical setting has two features that call for survival analysis. First, scam allegations have a time dimension in that it takes time to discover the scam. Second, our dependent variable is right-censored in that an ICO survives until the end of our observation window if it does not have a scam allegation. However, right-censored ICOs are not necessarily scam-free—they could be scams that are yet to be detected.

We first plot the proportion of surviving ICOs—the survival function $S(t)$—with respect to survival time t for four groups of ICOs, sorted by their *irregularities*. We compute the survival function within every group as

$$S(t) = \begin{cases} \frac{r_t - f_t}{r_t} \times S(t-1), & \text{for } t > 0 \\ 1, & \text{for } t = 0 \end{cases} \tag{3}$$

where r_t represents the number of surviving and uncensored ICOs instantaneously before time t, and f_t is the number of ICOs that incur scam allegations. Figure 5 shows that all four groups begin with $S(0) = 1$ because our sample excludes ICOs that are immediately known to be scams. As time passes, the survival functions of all four groups decline as ICO scams are reported on the `DeadCoin` website. We observe that the survival rate in the high-*irregularities* group decreases rapidly, while the survival rate in the low-*irregularities* group decreases much more slowly. This pattern suggests that irregularities are positively associated with the incidence of ICO scams.

- Figure 5 here -

Next, we use Cox regressions to estimate the effect of irregularities on the incidence of ICO scams. We define $h(t) = -\frac{\delta}{\delta t} \log S(t)$ as the expected hazard that denotes the rate of ICO scams conditional on survival up to time t, and $h_0(t)$ as the baseline hazard when all covariates equal zero. We estimate specification (4) as follows.

$$h_i(t) = h_0(t) \exp\left(\beta_1 \mathbb{1}(\textit{irregularities}_i > 0) + \mathbf{X}_i^\top \boldsymbol{\beta}\right) + \epsilon_i \tag{4}$$

The vectors $\mathbf{X}$ and $\boldsymbol{\beta}$ represent vectors of control variables and their corresponding estimated coefficients, respectively. For ease of interpretation, we express estimated coefficients as hazard ratios. A hazard ratio of one implies that an increase in the covariate has no effect on the hazard of ICO scams. If the hazard ratio is above (below) one, then the covariate is associated with an increase (decrease) in the hazard of ICO scams.

- Table 4 here -

Our estimates in Table 4 show that ICOs with more *irregularities* are more likely to be scams. Column 1 shows that the presence of *irregularities* more than doubles ($t = 6.03$) the hazard of ICO scams. In the intensive margin, column 2 shows that an additional irregularity is associated with a 22.7% ($t = 8.69$) rise in the hazard of ICO scams. We further include coverage quartile fixed effects and stratify our ICOs by their calendar-quarter cohorts in column 3 to address two concerns.[10] First, the coverage fixed effects takes into account the possibility that *irregularities* are mechanically related to the number of websites that an ICO is listed on. Second, the stratification allows ICOs to have cohort-specific baseline hazards $h_0(t)$ to absorb heterogeneity in the hazard of ICO scams across cohorts. In this augmented specification, we find that an additional irregularity increases the hazard of ICO scams by 11.6%. ($t = 2.78$).

In the Internet Appendix, we show that our results are similar using OLS regressions. Overall, the evidence suggests that irregularities of ICO attributes across listing websites are a powerful ex-ante predictor of scams. One implication of our results is

[10]Coverage is the number of listing websites that an ICO is listed on. Two ICOs are in the same cohort if their ICO start dates are in the same calendar quarter.

that simple cross-website verification of ICO attributes can act as an effective form of due diligence for prospective investors.

4.2 Assessing the screening mechanism

Next, we extract Ethereum blockchain data .[11] The data contain token holdings and transaction activities of cryptocurrency wallets (henceforth, wallets). Using wallet-level data, we examine the relation between the sophistication of the typical token holder and ICO irregularities. The Internet Appendix contains details of data collection in this test.

We characterize the (lack of) sophistication of a typical token holder by computing three wallet-level measures. First, we define the *value* of a wallet by computing the total portfolio value in U.S. dollars of all tokens held. To the extent that wealth positively correlates with sophistication, we expect unsophisticated investors to have lower wallet values. Second, we define *diversity* as the number of distinct tokens held. Unsophisticated investors may possess less diversified wallets with fewer distinct ICO tokens. Third, we define *activity* as the number of wallet transactions. Unsophisticated investors with less technical or trading expertise may make fewer transactions. We aggregate these measures at the ICO level by taking the medians of every measure.

To test whether malicious issuers successfully use irregularities to screen for naïve investors, we estimate Poisson regressions in specification (5) because our outcome variables are non-negative (Cohn, Liu, and Wardlaw, 2022). The dependent variable is *value*, *diversity*, or *activity*. The key independent variable is $\mathbb{1}(\textit{irregularities} > 0)$—an indicator that switches on if the ICO has at least one irregularity at its first appearance in our sample. Our models include ICO calendar-quarter cohort fixed effects, and standard errors are clustered by these cohorts. For ease of interpretation, we express

[11] The Etherium blockchain is a digitally distributed, decentralized, public ledger of all transactions that occurred on the network. Most ICO tokens adopt the ERC-20 (Ethereum Request for Comments 20) standard, which facilitates interoperability with other tokens on the Ethereum network.

the estimated coefficients as incidence rate ratios.

$$\{\log(value_i), \log(diversity_i), \log(activity_i)\} = \alpha + \beta_1 \mathbb{1}(irregularities_i > 0) + \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i \tag{5}$$

Panel A of Table 5 shows that less sophisticated investors are more likely to hold tokens of ICOs with irregularities. Column 1 indicates that the typical investor in such ICOs has a 60.1% ($t = 2.61$) lower wallet *value*. In column 2, switching on $\mathbb{1}(irregularities > 0)$ is associated with a 19.7% ($t = 2.88$) decline in *diversity*. Column 3 shows that transaction *activity* of investors in ICOs with *irregularities* is lower by 9.0% ($t = 2.62$). Overall, our results suggest that ICOs with *irregularities* attract investors with less sophistication, measured by wallet value, diversity, and transaction frequency.

- Table 5 here -

We further test our screening mechanism by hand-matching ICOs to their `Reddit` message boards and tracking user activity with the `Pushshift` API. Our model predicts that an ICO with irregularities should receive fewer queries if the screening strategy is successful. Consistent with this prediction, Panel B of Table 5 shows that `Reddit` message boards of such ICOs have fewer (i) comments ($-48.8\%, t = 4.67$), (ii) questions ($-46.6\%, t = 2.95$), and (iii) unique users ($-20.0\%, t = 1.96$). Overall, our findings are consistent with a screening motive behind irregularities.

5 Are irregularities unintentional mistakes?

Our findings thus far are consistent with the view that malicious issuers use irregularities as a tactic to screen for naïve investors. But we recognize that the true motives of ICO issuers are ultimately unobservable. This section provides three tests to evaluate the alternative explanation that irregularities are unintentional.

5.1 Systematic patterns of irregularities

If ICO irregularities are idiosyncratic, unintentional mistakes, they should be randomly distributed among ICOs. However, if issuers strategically use irregularities for the malicious purposes we claim, there should be systematic footprints in the ICO ecosystem.

To test this assertion, we examine whether ICO advisers (henceforth, advisors) play a role in promoting irregularities behavior. ICO issuers hire advisors to provide technical, marketing, and economic expertise. They often work on multiple ICOs but are controversial. Some advisors have been convicted of illegal touting and tax evasion, while others have allegedly failed to perform basic due diligence on client ICOs. About 60% of ICOs hire advisors. We hypothesize that advisors may intentionally propagate irregularities behavior in direct and indirect ways through social transmission.[12] Advisors may directly convey know-how about the use of irregularities. Or, they may learn and adopt such behavior if it is an acceptable norm among the advisor community.

To test this hypothesis empirically, we construct an ICO network by linking two ICOs if they share at least one common advisor. We managed to match 2,110 advisors with 2,271 ICOs using data extracted from the `ICOBench` listing website. This test has a smaller sample because we must exclude ICOs that either have no advisors or are unlinked to any ICOs. Next we apply the Ballester, Calvó-Armengol, and Zenou (2006) network model of behavior propagation. If irregularities were unintentional and randomly distributed, we should observe no correlation between *irregularities* and network structure. Otherwise, the model predicts that ICOs with higher Katz centrality in the network should exhibit more irregularities. Katz centrality is a popular network measure used to study the diffusion of microfinance, medical knowledge, microeconomic shocks, and education outcomes (Banerjee, Chandrasekhar, Duflo, and Jackson, 2013, 2019; Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi, 2012; Calvó-Armengol, Patacchini, and Zenou, 2009).

- Figure 6 here -

[12] Illegal behaviors propagate through social transmission in a variety of contexts (Case and Katz, 1991; Damm and Dustmann, 2014; Dimmock, Gerken, and Graham, 2018).

To visualize the results, Figure 6 presents a circular layout of this network. ICOs are arranged according to their *irregularities* on the circumference of the circle. Starting at the 12 o'clock point, ICOs have more *irregularities* as we move along the circumference in the clockwise direction. The lines within the circle represent links between ICOs. By quick visual inspection, ICOs with more *irregularities* tend to locate in regions with higher densities of links. Generally, such ICOs are also more central in the network.

- Table 6 here -

For a more rigorous examination of the relation between Katz centrality and *irregularities*, we estimate Poisson regressions in Table 6. Estimated coefficients are presented as incidence rate ratios. Consistent with our model predictions, column 1 shows that a 10% increase in Katz centrality is associated with a 4.6% ($t = 2.27$) rise in *irregularities*.[13] Next, we conjecture that the transmission of *irregularities* behavior is stronger between two ICOs if they share more common advisors. Thus, we also construct a weighted ICO network, in which links are weighted by the number of common advisors. In column 2, we find a quantitatively similar effect using weighted links—a 10% increase in Katz centrality is associated with a 5.4% increase ($t = 2.17$) in *irregularities*. As a robustness check, we also construct an indicator $\mathbb{1}(\textit{high centrality})$ that switches on if an ICO has an above-median Katz centrality. Using this measure, columns 3 and 4 report that central ICOs have 6.1% ($t = 1.96$) and 6.7% ($t = 2.25$) higher *irregularities* than peripheral ICOs, respectively.

Consistent with predictions from the Ballester, Calvó-Armengol, and Zenou (2006) network model of behavior propagation, central ICOs use more irregularities. Our findings suggest that ICO irregularities are unlikely to be idiosyncratic or unintentional. Overall, while advisors could be valuable information and service intermediaries in the ICO market, some may facilitate the spread of malignant behaviors.

5.2 Irregularities and ICO quality

Are irregularities merely careless mistakes? Suppose low quality issuers fail to exert effort to accurately market their offerings on listing websites. If such issuers also

[13] We calculate this economic magnitude as follows: $\log(1.1) \times (1.485 - 1) = 0.046$.

produce poorer blockchain projects, irregularities may present as a mere symptom of low ICO quality—not necessarily malicious intent. To address this alternative interpretation, we aim to disentangle ICO quality effects from our screening mechanism.

To measure ICO quality, we use a comprehensive list of variables that may correlate with ICO quality. High quality ICOs tend to have a whitepaper, a `Github` code repository, experienced team members, and venture funding (Davydiuk, Gupta, and Rosen, 2022). Such ICOs may also differentiate themselves through costly, voluntary disclosure (Bourveau et al., 2021) by publicizing the source code and code audits of their projects. Given the information asymmetry in the then-nascent ICO market, issuers could signal their quality by retaining a substantial stake at fundraising and by setting a long vesting duration (Davydiuk, Gupta, and Rosen, 2022). Finally, we use the amount of funds raised as an ex post, market-based measure of ICO quality.

We perform factor analysis on the above variables to recover a latent, common component of ICO quality. Panel A of Table 7 shows how these variables load on the first three factors.[14] All 10 quality-related variables load positively on the first factor. With the exception of venture funding, we find likewise in the second factor. Some loadings in the third factor are negatively signed. So, the third factor is unlikely to capture ICO quality. The first factor alone explains 56.4% of variance across all the variables, dwarfing the incremental variance explained by the other factors. For parsimony and ease of interpretation, we thus retain the first factor as the *quality* factor.

- Table 7 here -

To disentangle the ICO quality effect from our screening mechanism, we horserace *quality* against *irregularities* in Cox regressions of scam risk. In column 1 of Panel B, we find that higher ICO *quality* is associated with lower scam risk. This pattern helps validate our *quality* measure. However, the link between *quality* and scam risk is noisy ($t = 0.15$). Notably, the association between *irregularities* and scam risk remains positive and statistically significant (+20.3%, $t = 5.88$). In column 2, our findings hold even as we saturate the model with control variables used in Table 4. As a robustness

[14] We extract 10 factors but do not present loadings on factors 4 through 10 for brevity.

check, column 3 deploys the 10 quality-related variables in lieu of the *quality* factor. Here, we continue to find that ICOs with *irregularities* are more likely to be scams. Taken together, these findings suggest that our screening mechanism has a distinct effect from ICO quality.

5.3 Regulatory scrutiny and irregularities

If irregularities are nefarious, then the threat of regulatory action should deter malicious issuers from entering the ICO market. We hypothesize that ICOs launched after periods of higher regulatory scrutiny will have fewer irregularities, on average. To test the deterrence effect, we collect news of regulatory actions taken by the U.S. authorities. As Appendix B shows, these regulatory actions primarily involve ICO fraud and conflicts of interest. None of these actions specifically mention inaccurate disclosures on listing websites.

We first define *regulatory scrutiny* as an indicator that switches on if there are regulatory news articles released in the month prior to the first appearance of an ICO in our sample. Next, we test the effect of *regulatory scrutiny* on the use of irregularities. We estimate logistic and Poisson regressions according to specification (6).

$$\left\{\log\left(\frac{p_i}{1-p_i}\right), \log\left(\textit{irregularities}_i\right)\right\} = \alpha + \beta_1 \textit{regulatory scrutiny}_i + \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i \quad (6)$$

The first outcome variable in this specification is the logit of p, which is the probability that the ICO has at least one irregularity at its first appearance in our sample. To estimate the intensive margin, we use *irregularities* as the second outcome variable. The vectors $\mathbf{X}$ and $\boldsymbol{\beta}$ represent vectors of control variables and their corresponding estimated coefficients, respectively.

- Table 8 here -

Our results in Table 8 show that ICOs that launch immediately after regulatory scrutiny have fewer irregularities. Column 1 shows that *regulatory scrutiny* decreases the odds of an ICO having irregularities in the next month by 44.4% ($t = 2.23$). On the intensive margin, column 2 shows that ICOs have 35.6% ($t = 3.31$) fewer

irregularities in the presence of regulatory scrutiny. These patterns suggest that the use of irregularities likely reflects strategic, malicious behavior. Notably, we find a link between *regulatory scrutiny* and irregularities although our sample of news articles does not mention the latter.

Nevertheless, these patterns could simply reflect more careful behavior by ICO issuers in the face of regulatory scrutiny. If regulatory scrutiny simply spurs greater conscientiousness among issuers, corrections of prior irregularities should be more likely after the news events. Using our point-in-time data snapshots, we track whether issuers correct their ICO irregularities from month to month. We test this alternative story with $\mathbb{1}(\Delta irregularities < 0)$—an indicator that switches on when an ICO has a decrease in *irregularities* from the previous month. Column 3 shows no statistically significant link between *regulatory scrutiny* and $\mathbb{1}(\Delta irregularities < 0)$, suggesting that regulatory scrutiny does not just spur issuers to take greater care.

6 Other screening tactics

While malicious ICO issuers use irregularities to target naïve investors, they may also use other tactics to screen for investor sophistication. We collect data on two examples of such actions—celebrity endorsements and web traffic data of listing websites—and test whether these tactics predict ICO scams.

Celebrity endorsements are more likely to attract naïve investors who learn of investment opportunities through celebrities on social media.[15] The U.S. SEC created an investor education website to warn investors that celebrity endorsements of ICOs are a prominent red flag of investment scams.[16] Another tactic to better target naïve investors is to choose listing websites based on the sophistication of its users. Malicious issuers may opt to list on websites that have a higher amount of passive web traffic. This type of traffic is usually generated by less sophisticated investors, who access

[15] We collect data on celebrity endorsements by performing web searches using combinations of these keywords: "celebrity"/"promoter"/"influencer" and "ICO"/"initial coin offering"/"token". Next, we read all relevant search results and identify ICOs that are promoted by celebrities. To ensure completeness of our search efforts, we also search for the same combinations of keywords in the Factiva database. Our sample includes celebrities who span the entertainment, sports, business and media sectors.

[16] Source: `https://www.investor.gov/ico-howeycoins`

the site through paid advertisements, third-party referral links, or search engines. We define the *web traffic ratio* of an ICO as the ratio of passive traffic to active traffic, aggregated across the listing websites that list it in the month prior to its start date.[17]

- Table 9 here -

To test whether celebrity endorsements and strategic choices of listing websites predict ICO scams, we estimate Cox regressions in Table 9. We express estimated coefficients as hazard ratios. The key independent variable in column 1 is $\mathbb{1}(\textit{celebrity})$—an indicator that switches on if an ICO is endorsed by a celebrity. Here, we find that the scam risk of an ICO with a celebrity endorsement is more than 12 times ($t = 9.34$) that of an ICO without one. This finding echoes the warning issued by the SEC on celebrity endorsements in investment scams. In column 2, we examine whether celebrity endorsements subsume the predictive effect of *irregularities* on ICO scam risk. They do not. While $\mathbb{1}(\textit{celebrity})$ remains a strong predictor of ICO scam risk, we find that an additional irregularity raises the odds of a scam by 10.9% ($t = 2.93$). This result suggests that irregularities and celebrity endorsements are separate tactics in the malicious issuer's repertoire, and that investors should be on the lookout for both.

Column 3 shows that a unit increase in *web traffic ratio* is associated with a 27.3% ($t = 2.94$) higher odds of an ICO scam. This pattern suggests that malicious issuers strategically choose listing websites that receive a relatively larger share of passive web traffic.[18] In column 4, we find that *irregularities* remains a positive and statistically significant predictor of ICO scam risk. Thus, irregularities have an incremental screening effect to that of the choice of listing websites.

Overall, malicious issuers appear to use other tactics in complement with irregularities to target naïve investors. Irregularities continue to have a distinct predictive effect on ICO scam risk. To identify potential ICO scams, investors could perform simple cross-site due diligence and look for red flags such as celebrity endorsements.

[17] Using data from a web traffic analytics vendor `SEMrush`, we measure the quantities of passive and active web traffic in each of the five listing websites over time. Active web traffic counts visitors who access a listing website by directly typing its Uniform Resource Locator (URL) or through the use of saved browser bookmarks.

[18] In our model (Section 2), this strategic choice is akin to choosing an investor mass with a higher density z of naïve investors. In turn, a higher z increases the issuer's expected profits, ceteris paribus.

7 Estimating the true prevalence of ICO scams

This section provides estimates of the true prevalence of ICO scams. ICO scams are likely to be underreported because victims of financial fraud are often reluctant to step forward (Gee and Button, 2019). During this period, regulatory oversight was lax, and legal jurisdiction of cryptocurrencies was unclear. With limited legal recourse, victims may have been unwilling to report ICO scams to the authorities. For example, the ICO advisory firm Satis Group estimates in an industry report that 78% of ICOs were scams (Dowlat, 2018), but many go unreported.[19]

As a starting benchmark, our `Deadcoins` sample identifies 243 verified ICO scams. From this sample, the implied lower bound of the scam rate is 4.1%. Next, we search for functioning websites in our sample of 5,873 ICOs. An active web presence suggests that the ICO issuer is still paying for web hosting services and is not trying to evade investors or regulators. As of September 2023, only 37.7% (2213/5,873) of ICOs have functioning websites, so 62.3% of ICOs may have failed. This estimate is likely an upper bound on the true scam rate because an ICO may cease operations for legitimate reasons. Overall, our tallied statistics imply a scam rate that ranges between 4.1% and 62.3%. In the next section, we use a structural approach to estimate the true scam rate.

7.1 Detection controlled estimation

To address the underdetection of frauds and scams, researchers use a structural econometric technique called detection controlled estimation (DCE), which simultaneously estimates a system of two equations.[20] In our setting, the first equation models ICO scam probabilities, while the second equation models detection on `DeadCoins` conditional on the occurrence of ICO scams. We estimate the DCE model using the maximum likelihood method.

[19] The Satis Group report uses a smaller and earlier sample, a different definition of ICO scams, and a different estimation methodology.

[20] For example, see Wang, Winton, and Yu (2010), Comerton-Forde and Putniņš (2014), and Foley, Karlsen, and Putniņš (2019).

For identification, the DCE model requires instrumental variables that are uniquely associated with the probability of either ICO scams or detection. We hypothesize that malicious issuers opportunistically launch ICOs during periods of strong sentiment in crypto-markets. To instrument for scam probabilities, we construct three measures of crypto-market sentiment in the month prior to the ICO start date. The instrument *app downloads* is the log-transformed number of downloads of mobile applications for cryptocurrency exchange (Auer, Cornelli, Doerr, Frost, and Gambacorta, 2022). A high level of *app downloads* reflects a large increase in retail cryptocurrency investors, many of whom may be naïve. We also use as instruments *altcoin returns* and *BTC returns*—the cumulative returns of non-Bitcoin cryptocurrencies and Bitcoin, respectively.

These instruments for ICO scam probabilities are arguably unassociated with detection probabilities for three reasons. First, to the extent that detection is ICO-specific, our measures of marketwide sentiment should be orthogonal to detection probabilities. Second, if ICO scams are easier to detect when sentiment is high, then we should expect detection to be quick. However, we find that typically several months elapse between the end date of an ICO scam and its subsequent detection on the `DeadCoins` website. Third, the reasons provided for scam allegations on the `DeadCoins` website do not refer to sentiment timing.

- Table 10 here -

Table 10 reports estimates from our DCE models. The first two columns labeled Model A use *app downloads* as the instrument in the scam stage. Column 1 shows that *app downloads* is positively and significantly associated with ICO scams. This pattern supports the view that malicious issuers time their ICO launches during periods of strong sentiment in cryptocurrency markets. Using *altcoin returns* and *BTC returns* as instruments in Models B and C, we continue to find support for the hypothesis that malicious issuers opportunistically launch ICOs during periods of strong sentiment.

Our DCE models also address concerns that the underdetection of ICO scams is nonrandom. For example, an unobserved characteristic may jointly lead ICO scams to (i) have more irregularities and (ii) be more prone to detection on the `DeadCoins`

website. So, the predictive effect of *irregularities* on ICO scams in our main results could be spurious. Our DCE results suggest otherwise because the estimated loadings on *irregularities* in the scam equations of Models A to C are positive and statistically significant. Thus, irregularities remain a powerful predictor of ICO scams even when we account for nonrandom underdetection.

7.2 Welfare analysis of ICO scams

We fit the models in columns 1, 3, and 5 of Table 10 to structurally estimate the prevalence of ICO scams. Models A to C estimate that 44.7%, 49.3%, and 49.2% of ICOs in our sample are scams, respectively. Thus, our DCE models estimate true scam rates that are within the lower and upper bound ranges (4.1% to 62.3%) from our earlier tallied statistics.

Next, we perform a back-of-envelope welfare analysis. Suppose that 49% of the 5,873 ICOs are scams. The median ICO in our sample raises U.S. $2.2 million. Therefore, potential losses based on a 49% scam rate are estimated to be U.S. $2.2 million $\times$ $0.49 \times 5{,}873$ = U.S. $6.3 billion. However, it is more challenging to estimate the *social* welfare effects of ICO scams because some individuals may view ICO investments as gambles. If skewness-loving individuals substitute traditional gambling devices with ICO investments, the net welfare losses from ICO scams could be smaller. For example, the U.S. Census Bureau reports that state-administered lottery funds alone generated U.S. $76.4 billion in sales in 2018. That said, regulated gambling revenues are often channeled towards productive uses in society, but not so for ICO scams. Overall, it is difficult to estimate the net social welfare loss due to ICO scams. Nevertheless, we hope that our sobering estimates offer some insights into the potential risks within the rapidly evolving cryptocurrency landscape.

8 Concluding remarks and implications beyond ICOs

We exploit the data-rich setting of the ICO market to systematically study the economics of financial scams. We discover that malicious issuers use cross-site irregularities as a tactic to target naïve investors and screen out astute investors. ICO scam

risk more than doubles in the presence of irregularities. We develop a screening model to formalize the economic mechanism. Our model shows that the optimal strategy is a function of the population density of viable victims and the potential net profit per victim. When there are more viable victims and higher potential profits, perpetrators cast a wider net to capture more victims. This prediction is consistent with the finding in Egan, Matvos, and Seru (2019) of more financial advisor misconduct in counties with more elderly, less educated, and wealthier populations.

Our findings provide an economic explanation for the common use of irregularities in scams. Authorities and financial institutions often warn that grammatical errors, misspellings, and odd phrasings are red flags of scams (see Table 1). Our study rationalizes the puzzling prevalence of such irregularities in scams: Perpetrators employ various types of irregularities (i.e., obvious mistakes) to quickly screen out savvy individuals who make unviable victims. Thus, while the ICO market was short-lived, it provides a unique setting to systematically analyze the economics of financial scams.

Although we estimate that almost half of ICOs could be scams with associated losses of more than U.S. $6 billion, we caution that our exercise does not encompass all facets of investor welfare. For example, investors may obtain nonpecuniary utility by investing in ICOs. Or they may invest in a broad range of ICOs, some of which are fraudulent, as part of an investment strategy. Nevertheless, our study speaks to the idea that less sophisticated consumers in novel financial systems may be disadvantaged or, knowing this risk, abstain from participation (Makarov and Schoar, 2022). Thus, more robust consumer protection efforts could faciliate the broad adoption of open-access, decentralized finance.

Appendix A Proofs

Proof of Proposition 1. To solve for the malicious issuer's optimal targeting strategy, we first expand equation (1) as:

$$\frac{\pi(S,\mathcal{I}(S))}{p} = zf\left[1-\Phi(\mathcal{I}(S)\mid\tau=N)\right]-(1-z)(\kappa^O+\kappa^E)\left[1-\Phi(\mathcal{I}(S)\mid\tau=A)\right]$$
$$\text{where}\quad \Phi(\mathcal{I}(\theta)\mid\tau) \coloneqq \int_0^{\mathcal{I}(\theta)} \phi(\mathcal{I}(\theta)\mid\tau)\,dx \tag{A.1}$$

The malicious issuer maximizes profits by choosing optimal $\mathcal{I}^*(S)$, which satisfies this first-order condition:

$$-zf\phi(\mathcal{I}^*(S)\mid\tau=N)+(1-z)(\kappa^O+\kappa^E)\phi(\mathcal{I}^*(S)\mid\tau=A)=0 \tag{A.2}$$

Rearranging the first-order condition yields:

$$\frac{1-z}{z}\cdot\frac{\kappa^O+\kappa^E}{f}=\frac{\phi(\mathcal{I}^*(S)\mid\tau=N)}{\phi(\mathcal{I}^*(S)\mid\tau=A)}\geq 0 \tag{A.3}$$

which is the slope on the ROC curve per Definition 1. ■

Proof of Proposition 2. From Proposition 1, it is straightforward to see that the malicious issuer's optimal targeting strategy has a nonzero slope if and only if all of the following three conditions hold: (i) Some investors are astute ($z<1$); (ii) investor queries are costly ($\kappa^O+\kappa^E>0$); and (iii) funds raised per investor are positive and finite ($f\in\mathbb{R}_{>0}$).

$$\frac{1-z}{z}\cdot\frac{\kappa^O+\kappa^E}{f}\begin{cases}>0 & (z<1)\wedge(\kappa^O+\kappa^E>0)\wedge(f\in\mathbb{R}_{>0})\\ =0 & \text{otherwise}\end{cases} \tag{A.4}$$

Otherwise, an indiscriminate targeting strategy, which targets all investors, corresponds to Point (i) on Figure 2 where the slope is zero and $\mathcal{I}(S)=0$. Thus, the malicious issuer's optimal targeting strategy has a nonzero slope that resides leftwards of Point

(i) and entails $\mathcal{I}^*(S) > 0$. ■

Proof of Proposition 3. To solve for the honest issuer's optimal targeting strategy, we first expand equation (2) as:

$$\frac{\pi(H, \mathcal{I}(H))}{p} = zf\left[1 - \Phi(\mathcal{I}(H) \mid \tau = N)\right] + (1-z)(f - \kappa^O)\left[1 - \Phi(\mathcal{I}(H) \mid \tau = A)\right]$$
$$\text{where} \quad \Phi(\mathcal{I}(\theta) \mid \tau) \coloneqq \int_0^{\mathcal{I}(\theta)} \phi(\mathcal{I}(\theta) \mid \tau)\, dx \tag{A.5}$$

The honest issuer maximizes profits by choosing optimal $\mathcal{I}^*(H)$, which satisfies this first-order condition:

$$-zf\phi(\mathcal{I}^*(H) \mid \tau = N) - (1-z)(f - \kappa^O\phi(\mathcal{I}^*(H) \mid \tau = A) = 0 \tag{A.6}$$

Rearranging the first-order condition yields her optimal targeting strategy on the ROC curve with slope:

$$-\frac{1-z}{z} = \frac{\phi(\mathcal{I}^*(H) \mid \tau = N)}{\phi(\mathcal{I}^*(H) \mid \tau = A)} \geq 0 \tag{A.7}$$

By Definition 1, any point on a ROC curve must yield a nonnegative slope. Thus, the solution to the honest issuer's optimal targeting strategy is $z = 1$. Equation A.4 implies that she hence optimally operates at the zero-slope point on the ROC curve.

$$\left(-\frac{1-z}{z} \geq 0, \forall z \in [0,1] \iff z = 1\right) \Rightarrow \frac{\phi(\mathcal{I}^*(H) \mid \tau = N)}{\phi(\mathcal{I}^*(H) \mid \tau = A)} = 0 \tag{A.8}$$

From equation (A.4), this point corresponds to an indiscriminate targeting strategy. Thus, she optimally refrains from a screening strategy by choosing $\mathcal{I}^*(H) = 0$. ■

Proof of Proposition 4. For the separating equilibrium to be stable, we show that the malicious and honest issuers have no incentives to deviate from their respective optimal targeting strategies. For the malicious issuer, we aim to show the following by contradiction.

$$\frac{\pi(S, \mathcal{I}(S))}{p} > \frac{\pi(S, 0)}{p}, \quad \forall \mathcal{I}(S) > 0 \tag{A.9}$$

Suppose the malicious issuer is better off by mimicking the honest issuer's optimal choice of zero irregularities.

$$\frac{\pi(S,0)}{p} - \frac{\pi(S,\mathcal{I}(S))}{p} > 0, \quad \forall \mathcal{I}(S) > 0 \tag{A.10}$$

Expanding the above expression, we get:

$$\begin{aligned} & zf - (1-z)(\kappa^O + \kappa^E) \\ & \quad - zf\left[1 - \Phi(\mathcal{I}(S), \tau = N)\right] + (1-z)(\kappa^O + \kappa^E)\left[1 - \Phi(\mathcal{I}(S), \tau = A)\right] \\ & \quad = zf\Phi(\mathcal{I}(S), \tau = N) - (1-z)\kappa(S)\Phi(\mathcal{I}(S), \tau = A) > 0, \quad \forall \mathcal{I}(S) > 0 \\ & \quad \text{where} \quad \Phi(\mathcal{I}(\theta) \mid \tau) \coloneqq \int_0^{\mathcal{I}(\theta)} \phi(\mathcal{I}(\theta) \mid \tau)\, dx \end{aligned} \tag{A.11}$$

Rearranging the terms, we obtain:

$$\begin{aligned} & z > \frac{(\kappa^O + \kappa^E)\Phi(\mathcal{I}(S), \tau = A)}{f\Phi(\mathcal{I}(S), \tau = N) + (\kappa^O + \kappa^E)\Phi(\mathcal{I}(S), \tau = A)}, \quad \forall \mathcal{I}(S) > 0 \\ & \iff (z = 1) \vee (\kappa^O + \kappa^E) = 0 \vee (f \to \infty) \end{aligned} \tag{A.12}$$

Notice that the above expression is true for all $\mathcal{I}(S)$ if and only if at least one of these three conditions holds: (i) All investors are naïve ($z = 1$); (ii) the malicious issuers face no costs to address investor queries ($\kappa^O + \kappa^E = 0$); or (iii) funds raised per investor tend to infinity ($f \to \infty$). These conditions are inconsistent with those laid out in Proposition 4.

$$(z = 1 \vee (\kappa^O + \kappa^E) = 0 \vee f \to \infty) \nvdash (z < 1 \wedge (\kappa^O + \kappa^E) > 0 \wedge f \in \mathbb{R}_{>0}) \tag{A.13}$$

Thus, the contradiction cannot hold, implying that the malicious issuer can do no better by deviating to $\mathcal{I}(S) = 0$. For the honest issuer, we show the following.

$$\frac{\pi(H,0)}{p} \geq \frac{\pi(H,\varepsilon)}{p}, \quad \forall \varepsilon > 0 \tag{A.14}$$

Suppose $\mathcal{I}^*(H) = 0$ is not the optimal targeting strategy, and the honest issuer does better by choosing $\varepsilon > 0$. By definition, the CDF $\Phi(\cdot)$ is monotonically increasing in

m regardless of investor-type τ. So, equation (1) shows that the honest issuer must forgo some astute and naïve investors by choosing ε.

$$\frac{\pi(H,0)}{p} - \frac{\pi(H,\varepsilon)}{p} = zf\Phi(\varepsilon \mid \tau = N) + (1-z)(f-\kappa^O)\Phi(\varepsilon \mid \tau = A) > 0, \quad \forall \varepsilon > 0 \tag{A.15}$$

For $f > \kappa^O$, the honest issuer receives no greater profits by forgoing investors, regardless of their types. Thus, she has no incentive to deviate from $\mathcal{I}^*(H) = 0$. ■

Appendix B News of regulatory actions taken by U.S. authorities

Date	Title	News summary
16th Jun 2018	SEC: Fraud surrounds initial coin offerings, blockchain security notwithstanding.	SEC has a unit that monitors ICO scams.
21st Jun 2018	Members of the House will now be required to disclose bitcoin, other cryptocurrency holdings; Ethics Committee strongly encourage House members who are considering investing in an ICO to seek guidance.	Ethics Committee have taken actions to regulate House members in ICO investments.
27th Jun 2018	Facebook to accept cryptocurrency ads again; January's blanket ban is reversed, though crypto firms will have to get case-by-case approval.	Tech companies such as Facebook banned cryptocurrencies ads. Promotional efforts for cryptocurrencies have come under fire from federal and state regulators.
15th Aug 2018	Even free tokens face regulatory heat as coin offerings scrutinized; SEC punishes company that didn't sell any tokens, saying potential investors were misled about details of oil-drilling project.	The SEC punished a firm that did not sell any tokens to crack down on fraud in the market for initial coin offerings.
12th Sep 2018	SEC takes first action against hedge fund over cryptocurrency investments; In a separate case that's another first, agency penalizes brokers who ran an "ICO superstore".	The SEC fined a hedge fund manager who falsely advertised his cryptocurrency fund as the first regulated crypto-fund in the United States. Separately, the SEC also fined two men who ran a website that connects investors with initial coin offerings.

(To be continued)

Date	Title	News summary
12th Sep 2018	Judge lets cryptocurrency fraud case go forward, in win for SEC; For first time a federal court weighs in on the government's jurisdiction over ICOs in a criminal case.	The SEC scored a victory in their crackdown on cryptocurrency fraud as a judge ruled that initial coin offerings are subject to U.S. securities laws.
11th Oct 2018	SEC says stop ICOs that falsely claimed SEC approval.	SEC's complaint charges Blockvest and Ringgold with violating federal securities laws.
22nd Oct 2018	SEC suspends trading in company for making false cryptocurrency-related claims about SEC regulation and registration.	SEC suspended trading in the securities of a company for making false cryptocurrency-related claims.
16th Nov 2018	SEC settles enforcement actions over two initial coin offerings	Two startups agreed to comply with investor protection rules and offer money back to thousands of people who bought their digital tokens.
30th Nov 2018	Boxer Mayweather Jr., producer DJ Khaled agree to settle SEC crypto charges.	Celebrity endorsements of coin offerings may be illegal if the promoters fail to disclose the source and amount of their compensation.
21st May 2019	SEC obtains emergency order halting alleged diamond-related ICO Scheme targeting hundreds of investors.	SEC halted a Ponzi scheme, which was purportedly a cryptocurrency business.
5th Jun 2019	SEC challenges Canada firm's coin offering	SEC sued Kik for not providing investors with full and fair disclosure about its token and its business.

Table B.1. News of regulatory actions taken by U.S. authorities (Aug '18–Aug '19)

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). "The network origins of aggregate fluctuations". Econometrica 80, 1977–2016.

Aloosh, A. and Li, J. (2019). "Direct evidence of Bitcoin wash trading". Available at SSRN 3362153.

Amiram, D., Lyandres, E., and Rabetti, D. (2020). "Cooking the order books: Information manipulation and competition among crypto exchanges". Working paper.

Auer, R., Cornelli, G., Doerr, S., Frost, J., and Gambacorta, L. (2022). "Crypto trading and Bitcoin prices: Evidence from a new database of retail adoption". Bank for International Settlements Working Paper.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). "Who's who in networks. Wanted: The key player". Econometrica 74, 1403–1417.

Banerjee, A., Chandrasekhar, A., Duflo, E., and Jackson, M. O. (2013). "The diffusion of microfinance". Science 341, 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2019). "Using gossips to spread information: Theory and evidence from two randomized controlled trials". Review of Economic Studies 86, 2453–2490.

Becker, G. (1968). "Crime and punishment: An economic approach". Journal of Political Economy 76, 169–217.

Benedetti, H. and Kostovetsky, L. (2021). "Digital tulips? Returns to investors in initial coin offerings". Journal of Corporate Finance 66, 101786.

Bourveau, T., De George, E., Ellahie, A., and Macciocchi, D. (2021). "The role of disclosure and information intermediaries in an unregulated capital market: Evidence from initial coin offerings". Journal of Accounting Research 60, 129–167.

Button, M., Lewis, C., and Tapley, J. (2009). "A better deal for fraud victims". National Fraud Authority, United Kingdom.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). "Peer Effects and Social Networks in Education". Review of Economic Studies 76, 1239–1267.

Case, A. and Katz, L. (1991). "The company you keep: The effects of family and neighborhood on disadvantaged youths". NBER Working Paper.

Cohn, J. B., Liu, Z., and Wardlaw, M. I. (2022). "Count (and count-like) data in finance". Journal of Financial Economics 146, 529–551.

Cohney, S., Hoffman, D., Sklaroff, J., and Wishnick, D. (2019). "Coin-operated capitalism". Columbia Law Review 119, 591–676.

Comerton-Forde, C. and Putniņš, T. (2014). "Stock price manipulation: Prevalence and determinants". Review of Finance 18, 23–66.

Cong, L. W., Harvey, C., Rabetti, D., and Wu, Z.-Y. (2023). "An anatomy of crypto-enabled cybercrimes". NBER Working Paper.

Cong, L. W., Li, X., Tang, K., and Yang, Y. (2020). "Crypto wash trading". Available at SSRN 3530220.

Damm, A. and Dustmann, C. (2014). "Does growing up in a high crime neighborhood affect youth criminal behavior?" American Economic Review 104, 1806–32.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). "Investor psychology and security market under- and overreactions". Journal of Finance 53, 1839–1885.

Davydiuk, T., Gupta, D., and Rosen, S. (2022). "De-crypto-ing signals in initial coin offerings: Evidence of rational token retention". Management Science, forthcoming.

Dhawan, A. and Putniņš, T. (2022). "A new wolf in town? Pump-and-dump manipulation in cryptocurrency markets". Review of Finance, forthcoming.

Dimmock, S., Gerken, W., and Graham, N. (2018). "Is fraud contagious? Coworker influence on misconduct by financial advisors". Journal of Finance 73, 1417–1450.

Dimmock, S., Gerken, W., and Van Alfen, T. (2021). "Real estate shocks and financial advisor misconduct". Journal of Finance 76, 3309–3346.

Dowlat, S. (2018). "Cryptoasset market coverage initiation: Network creation".

Egan, M., Matvos, G., and Seru, A. (2019). "The market for financial adviser misconduct". Journal of Political Economy 127, 233–295.

Ehrlich, I. (1973). "Participation in illegitimate activities: A theoretical and empirical investigation". Journal of Political Economy 81, 521–565.

Feinstein, J. (1990). "Detection controlled estimation". Journal of Law and Economics 33, 233–276.

Foley, S., Karlsen, J., and Putniņš, T. (2019). "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?" Review of Financial Studies 32, 1798–1853.

Gee, J. and Button, M. (2019). "The financial cost of fraud 2019: The latest data from around the world". Crowe UK and University of Portsmouth Research Report.

Griffin, J. and Shams, A. (2020). "Is Bitcoin really untethered?" Journal of Finance 75, 1913–1964.

Herley, C. (2012). "Why do Nigerian scammers say they are from Nigeria?" *WEIS*.

Howell, S., Niessner, M., and Yermack, D. (2020). "Initial coin offerings: Financing growth with cryptocurrency token sales". Review of Financial Studies 33, 3925–3974.

La Porta, R., Lopez-De-Silanes, F., Shleifer, A., and Vishny, R. (2000). "Agency problems and dividend policies around the world". Journal of Finance 55, 1–33.

Lee, J., Li, T., and Shin, D. (2022). "The wisdom of crowds in FinTech: Evidence from initial coin offerings". Review of Corporate Finance Studies 11, 1–46.

Li, T., Shin, D., and Wang, B. (2021). "Cryptocurrency pump-and-dump schemes". Available at SSRN 3267041.

Lyandres, E., Palazzo, B., and Rabetti, D. (2022). "Initial coin offering (ICO) success and post-ICO performance". Management Science 68, 8658–8679.

Makarov, I. and Schoar, A. (2021). "Blockchain analysis of the bitcoin market". NBER Working Paper.

— (2022). "Cryptocurrencies and decentralized finance (DeFi)". NBER working paper.

Odean, T. (1998). "Volume, volatility, price, and profit when all traders are above average". Journal of Finance 53, 1887–1934.

PriceWaterhouseCoopers (2020). "6th ICO/STO Report: A strategic perspective".

Wang, T. Y., Winton, A., and Yu, X. (2010). "Corporate fraud and business conditions: Evidence from IPOs". Journal of Finance 65, 2255–2292.

Yermack, D. (2015). "Is Bitcoin a real currency? An economic appraisal". *Handbook of Digital Currency*. Elsevier, 31–43.

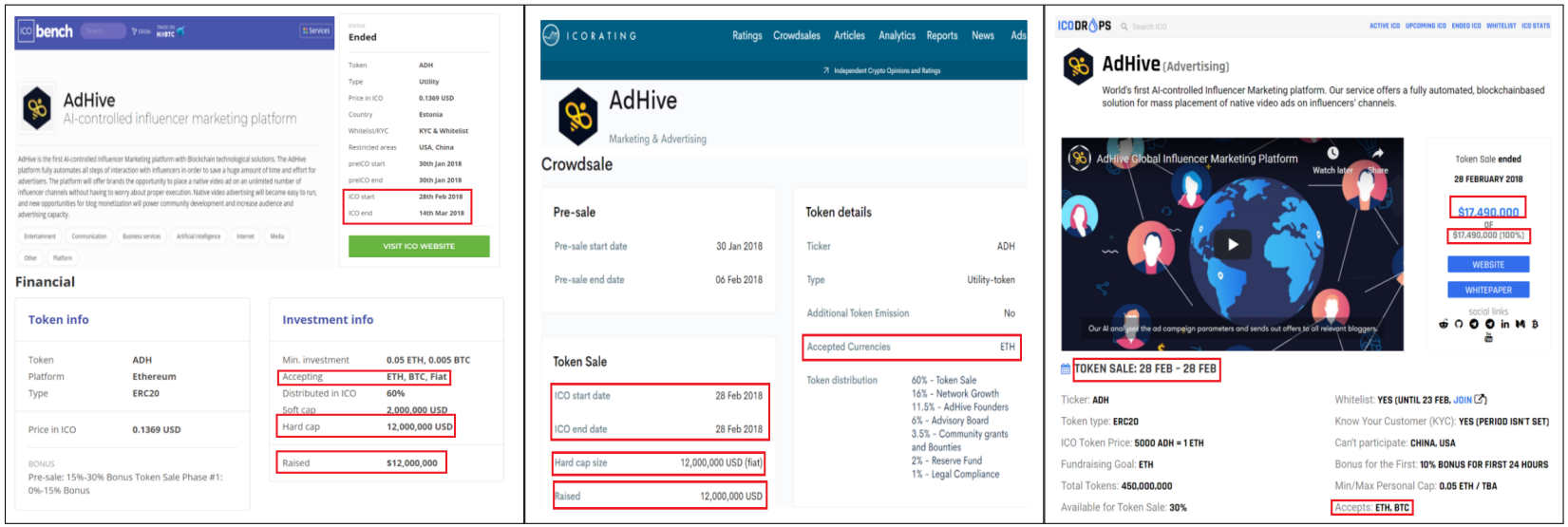

Figure 1. This figure presents screenshots of the AdHive ICO information pages on three ICO listing websites—`ICOBench.com`, `ICORating.com`, and `ICODrops.com`. There are discrepancies in the end date, hardcap, raised funds, and accepted payment modes across the three listing websites.

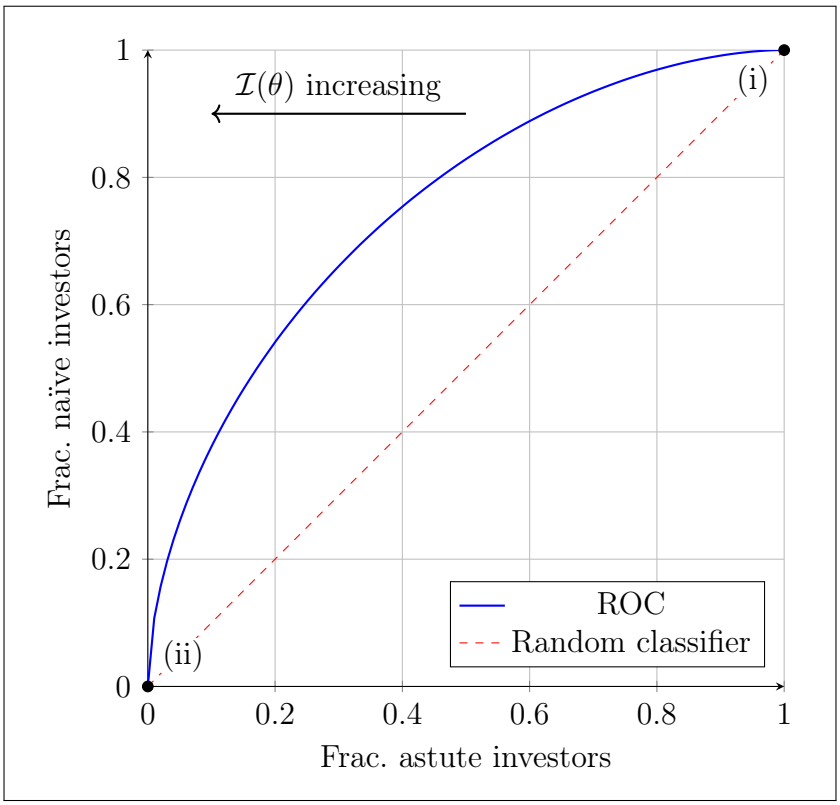

Figure 2. This figure plots a hypothetical receiver operating characteristics (ROC) curve and a random classifier benchmark. Every point on the ROC curve corresponds to a targeting strategy characterized by $\mathcal{I}(\theta) \in [0, \infty)$. For a given $\mathcal{I}(\theta)$, the issuer targets some fraction of naïve investors (y-axis) and some fraction of astute investors (x-axis). These fractions are the conditional complementary CDFs of investors' tolerance to irregularities. Point (i) represents an indiscriminate targeting strategy ($\mathcal{I}(\theta) = 0$) such that all naïve and astute investors are targeted. Point (ii) represents a conservative targeting strategy ($\mathcal{I}(\theta) \to \infty$) such that the issuer avoids all astute investors but also forgoes all naïve investors.

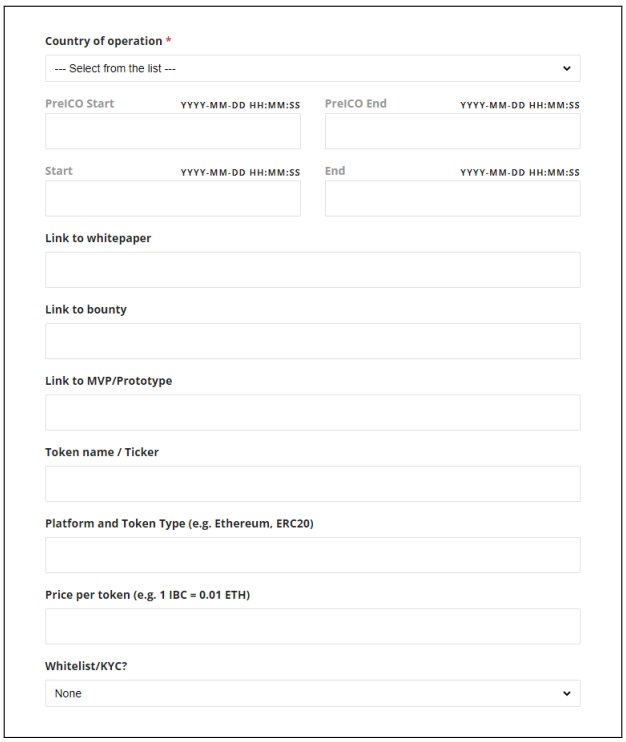

Figure 3. This figure presents a partial screenshot of the `ICOBench.com` webpage on which issuers self-report ICO data.

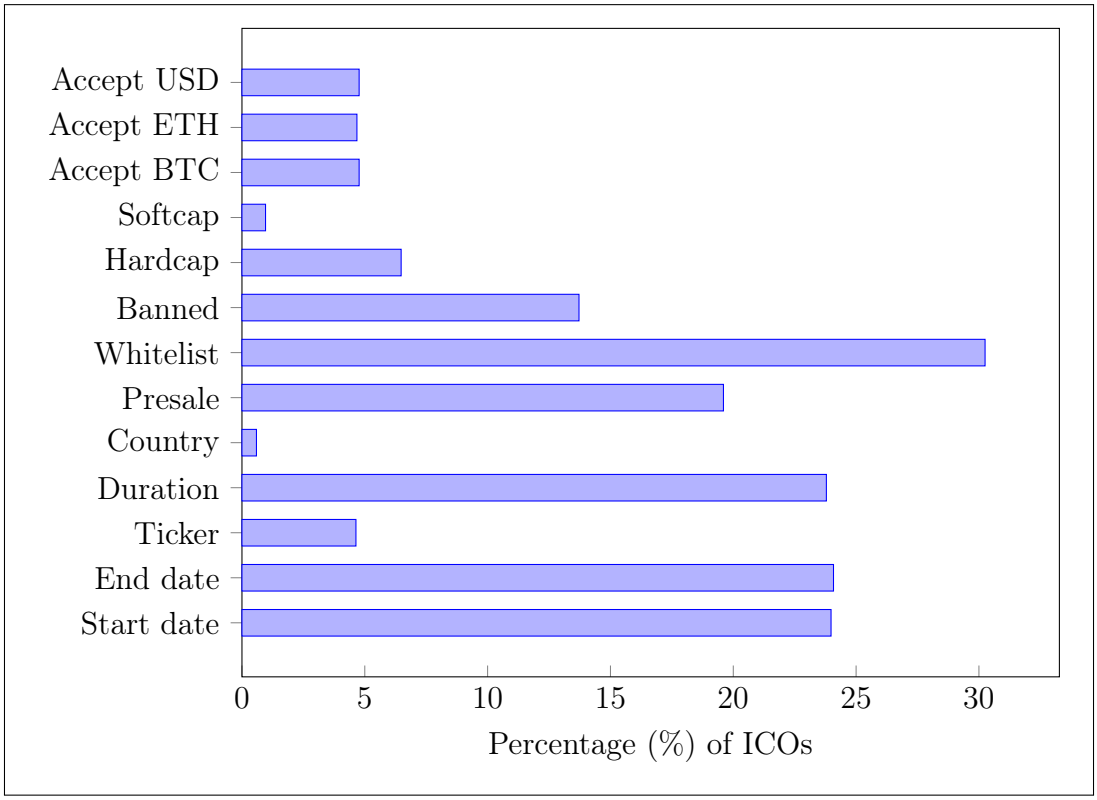

Figure 4. This figure presents the proportion of ICOs with at least one cross-website discrepancy (i.e., irregularity) in a particular characteristic at first appearance in our sample.

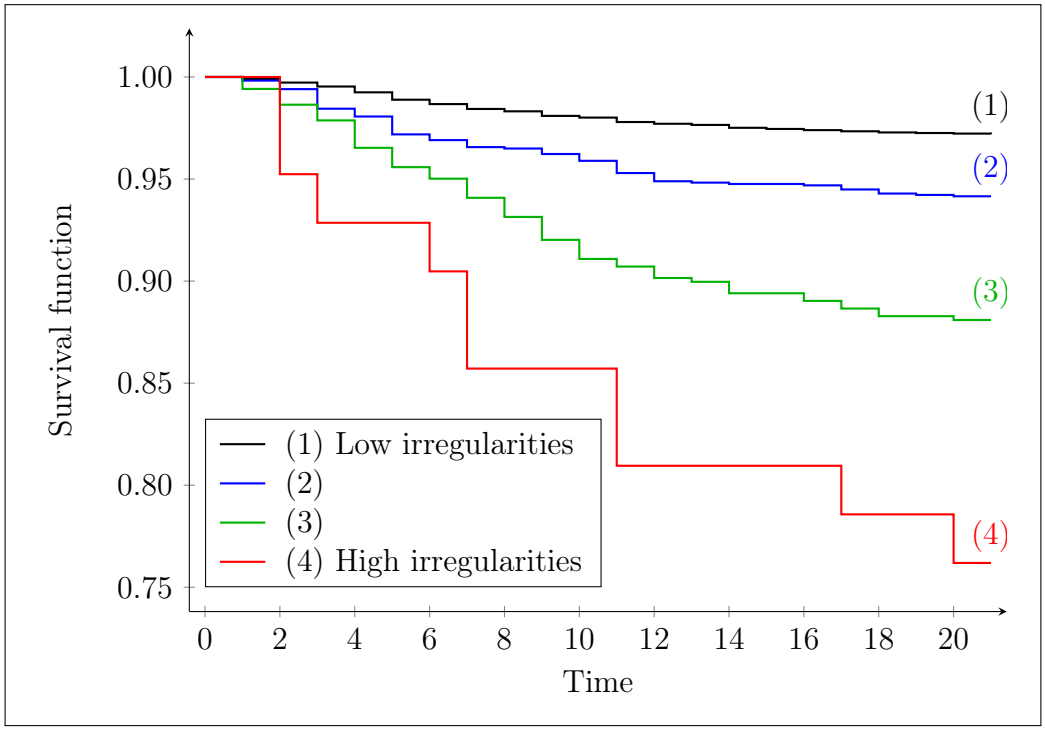

Figure 5. This figure presents the survival functions of ICOs in our sample. We assign every ICO into one of four groups based on its number of cross-website discrepancies in its characteristics at its first appearance in our sample (i.e., *irregularities*). The x-axis is the time-to-event—months elapsed from the time of entry into our sample. The y-axis is the groupwise proportion of ICOs that are not identified as scams on `DeadCoin.com` (i.e., survive) at a given time.

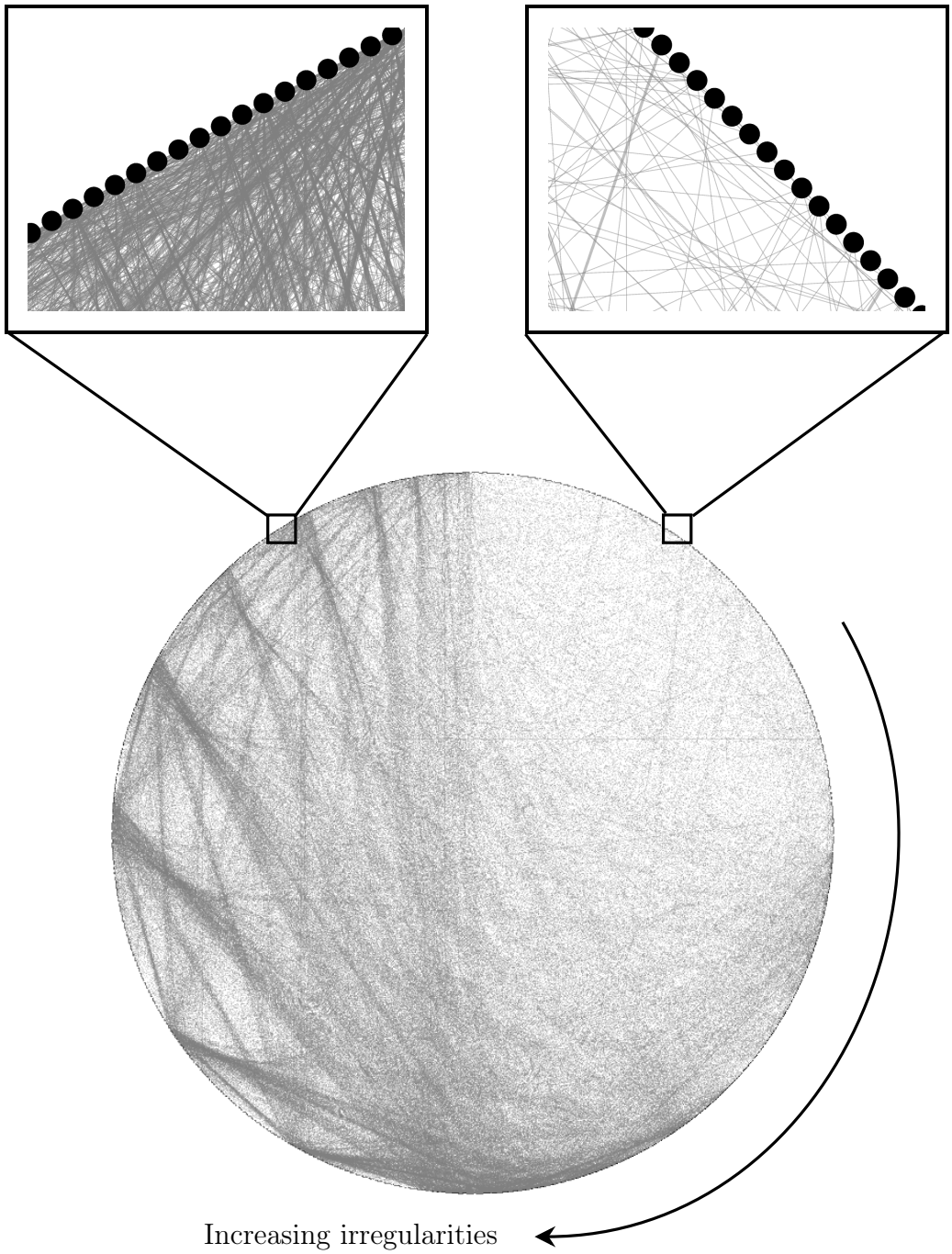

Figure 6. This figure presents a circular layout of the advisor-linked ICO network described in Section 5.1. The ICOs are arranged according to their *irregularities* on the circumference of the circle. The ICO at the 12 o'clock position has the fewest *irregularities*. As we move along the circumference in the clockwise direction, the ICOs have more *irregularities*. Lines inside the circle represent network links between ICOs.

Table 1. Examples of irregularities in various financial scams

(1) Entity	(2) Scam type	(3) Advisory
American Bankers Association https://perma.cc/A5AV-R89X	Phishing attacks	Look for Scam Tip-Offs [...] Grammatical errors or something just seems fishy or not right.
Federal Trade Commission (FTC) https://perma.cc/9APT-NVQR	Employment scams	[...] indicators that should raise your suspicions – for example, email from personal accounts not affiliated with a company, poor spelling and grammar [...]
Financial Crimes Enforcement Network (FinCen) https://perma.cc/25PM-CCHU	Investment scams	Technical Red Flags [...] poor spelling or grammatical structure, dubious customer testimonials, or a generally amateurish site design.
Financial Industry Regulatory Authority (FINRA) https://perma.cc/DT33-Y9QE	Imposter scams	Investors should look for the typical mistakes, such as poor grammar, misspellings, odd or awkward phrasings, or misuse of investor terminology.
Internal Revenue Service (IRS) https://perma.cc/Q5CN-FNE7	Identity theft	How to Spot a Scam [...] Uses incorrect grammar or odd phrasing.
J.P. Morgan Chase Bank https://perma.cc/VWE2-AWL3	Wire fraud	Commercial real estate wire fraud warning signs [...] Grammatical errors or spelling mistakes within communications.
National Anti-Scam Centre https://perma.cc/GV58-PZ2Y	Mail delivery scams	It has spelling and grammatical errors. These types of errors are a sign that it could be a scam.
Norton (cybersecurity firm) https://perma.cc/E92W-MRSF	Phishing attacks	Some phishing emails or texts might look unprofessional to you, using poor grammar or asking you to click on links with odd-looking URLs.

This table presents examples of advisories on the tell-tale signs of scams and fraud. Column 1 lists the names of the authorities and businesses that issued these advisories, along with links to the full documents. Column 2 categorizes the type of scam. Column 3 provides selected excerpts from the advisory texts.

Table 2. Descriptive statistics

Panel A. Summary statistics ($N = 5{,}873$)

	μ	σ	p10	p50	p90
Irregularities	1.30	2.19	0	0	4
$\mathbb{1}$(Irregularities > 0)	0.35	0.48	0	0	1
Banned	0.95	0.21	1	1	1
Whitelist	0.55	0.50	0	1	1
Duration (days)	54	50	15	36	107
Presale	0.48	0.50	0	0	1
Hardcap	0.70	0.46	0	1	1
Softcap	0.31	0.46	0	0	1
Accept BTC	0.28	0.45	0	0	1
Accept ETH	0.59	0.49	0	1	1
Accept USD	0.10	0.30	0	0	0
SEC filing	0.01	0.10	0	0	0
Enforcement	0.26	0.42	0	0	1
Disclosure	1.20	1.23	0	0.73	2.92

Panel B. Pairwise correlations (pct.pt.)

		A	B	C	D	E	F	G	H	I	J	K	L
A	Irregularities												
B	Banned	(2)											
C	Whitelist	(7)	9										
D	Presale	31	(1)	18	(6)								
F	Hardcap	28	2	(20)	(6)	12							
G	Softcap	3	(4)	6	10	16	35						
H	Accept BTC	16	0	8	6	23	9	18					
I	Accept ETH	30	0	14	(1)	42	16	16	44				
J	Accept USD	5	0	7	6	14	7	13	38	22			
K	SEC filing	4	1	2	1	4	2	1	5	3	5		
L	Enforcement	11	(2)	(4)	(1)	6	5	7	(1)	3	2	(3)	
M	Disclosure	13	(11)	(3)	3	5	2	8	(3)	(1)	1	6	31

Panel A reports the summary statistics of the irregularities measures and ICO characteristics. Panel B presents Pearson pairwise correlations between variables at the ICO level. Correlations are rounded to their nearest integers and expressed in percentage points. Variables are extracted from the first appearances of ICOs in our 13-month observation window.

Table 3. Differences in means

	(1) Irregularities > 0	(2) Irregularities $= 0$	Δ (1) $-$ (2)	$\|t\text{-stat}\|$
ICO scam	0.07	0.02	0.05	8.09
Banned	0.95	0.95	-0.01	1.36
Whitelist	0.46	0.60	-0.15	10.90
Duration (days)	47	57	-10	8.36
Presale	0.69	0.37	0.32	24.69
Hardcap	0.89	0.60	0.28	26.93
Softcap	0.33	0.30	0.03	2.68
Accept BTC	0.39	0.23	0.16	12.55
Accept ETH	0.80	0.47	0.33	27.75
Accept USD	0.12	0.09	0.03	3.91
SEC filing	0.01	0.01	0.01	1.82
Enforcement	0.33	0.22	0.11	9.64
Disclosure	1.44	1.07	0.37	11.16

This table presents differences in ICO scam rates and characteristics between ICOs with and without irregularities. Column (1) contains ICOs with at least one irregularity. Column (2) contains ICOs with no irregularities. We report differences in means (Δ) and their associated t-statistics.

Table 4. Irregularities and ICO scams

Event: ICO scam

	(1)	(2)	(3)
$\mathbb{1}$(Irregularities > 0)	2.544		
	(6.03)		
Irregularities		1.227	1.116
		(8.69)	(2.78)
Banned	1.107	1.117	1.094
	(0.31)	(0.34)	(0.24)
Whitelist	1.403	1.251	1.537
	(2.38)	(1.60)	(2.14)
Duration	0.997	0.998	0.999
	(1.39)	(0.91)	(0.42)
Presale	0.869	0.777	0.948
	(0.95)	(1.64)	(0.47)
Hardcap	1.990	1.826	1.795
	(3.40)	(2.94)	(2.70)
Softcap	0.792	0.817	0.958
	(1.53)	(1.33)	(0.31)
Accept BTC	0.984	0.954	0.918
	(0.10)	(0.30)	(0.38)
Accept ETH	1.220	1.207	1.305
	(1.20)	(1.13)	(2.14)
Accept USD	1.109	1.145	1.247
	(0.47)	(0.62)	(0.55)
Enforcement	0.822	0.827	0.816
	(1.20)	(1.17)	(1.20)
Disclosure	0.983	0.974	0.964
	(0.31)	(0.48)	(0.82)
SEC filing	0.674	0.609	0.591
	(0.55)	(0.70)	(2.17)
# ICOs	5,873	5,873	5,873
Cohort strata	N	N	Y
Coverage-quartile FE	N	N	Y
Clustered SE	N	N	Y

This table presents estimates from Cox regressions. Estimated coefficients are expressed as hazard ratios. The failure event in these regressions is *ICO scam*. An ICO triggers the event if the `DeadCoin` site identifies it as a scam. Otherwise, it is right-censored. The key independent variables in our regressions are *irregularities* and $\mathbb{1}$(*irregularities* > 0). The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The indicator $\mathbb{1}$(*irregularities* > 0) equals one if the ICO has at least one *irregularities*, and equals zero otherwise. t-statistics are reported in parentheses.

Table 5. Assessing the screening mechanism

Panel A. Irregularities and wallet characteristics

	(1)	(2)	(3)
Dependent variable:	Value	Diversity	Activity
$\mathbb{1}$(Irregularities > 0)	0.399	0.803	0.910
	(2.61)	(2.88)	(2.62)
Controls	Y	Y	Y
# ICOs	1,996	1,996	1,996
Cohort FE	Y	Y	Y
Clustered SE	Y	Y	Y

Panel B. Irregularities and investor queries

	(1)	(2)	(3)
Dependent variable:	Average # per post		
	Comments	Questions	Users
$\mathbb{1}$(Irregularities > 0)	0.512	0.534	0.800
	(4.67)	(2.95)	(1.96)
log (# Posts)	0.920	0.811	1.042
	(0.83)	(4.30)	(0.33)
log (Community size)	1.347	1.251	1.273
	(4.38)	(3.97)	(3.99)
Controls	Y	Y	Y
# ICOs	541	541	541
Cohort FE	Y	Y	Y
Clustered SE	Y	Y	Y

Panels A and B present estimates from Poisson regressions. Estimated coefficients are expressed as incidence rate ratios. The dependent variables in Panel A are *value*, *diversity*, and *activity*. The *value* of an ICO is the median portfolio value (in U.S. dollars) of wallets that hold its tokens. The *diversity* of an ICO is the median number of distinct tokens held in wallets that hold its tokens. The *activity* of an ICO is the median number of blockchain transactions performed by wallets that hold its tokens. The dependent variables in Panel B relate to investors' activity on `Reddit` subforums (i.e., subreddits) of ICOs up until the ICO end date. The *avg. # comments per post* (*avg. # questions per post*, *avg. # users per post*) is the number of user comments (questions, unique users), divided by the number of posts on the subreddit. The key independent variable is $\mathbb{1}$(*irregularities* > 0). The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The indicator $\mathbb{1}$(*irregularities* > 0) equals one if the ICO has at least one *irregularities*, and equals zero otherwise. t-statistics are reported in parentheses. We include in our models other control variables used in Table 4 but suppress their estimated coefficients for brevity.

Table 6. Irregularities and central ICOs

Dependent variable: Irregularities

	(1)	(2)	(3)	(4)
Weighted links	N	Y	N	Y
log (Centrality)	1.485	1.567		
	(2.27)	(2.17)		
𝟙(High centrality)			1.061	1.067
			(1.96)	(2.25)
Banned	0.974	0.974	0.974	0.974
	(0.48)	(0.47)	(0.45)	(0.46)
Whitelist	1.134	1.134	1.133	1.133
	(1.85)	(1.85)	(1.82)	(1.82)
Duration	0.999	0.999	0.999	0.999
	(1.56)	(1.56)	(1.56)	(1.57)
Presale	1.590	1.591	1.588	1.587
	(7.47)	(7.49)	(7.60)	(7.62)
Hardcap	1.598	1.599	1.596	1.597
	(6.75)	(6.77)	(6.98)	(6.90)
Softcap	0.996	0.996	0.996	0.997
	(0.29)	(0.26)	(0.30)	(0.22)
Accept BTC	1.065	1.065	1.067	1.067
	(1.31)	(1.31)	(1.32)	(1.35)
Accept ETH	1.249	1.249	1.245	1.243
	(2.31)	(2.31)	(2.25)	(2.26)
Accept USD	1.033	1.034	1.036	1.035
	(0.76)	(0.77)	(0.81)	(0.79)
Enforcement	1.023	1.022	1.023	1.025
	(0.73)	(0.72)	(0.74)	(0.76)
Disclosure	1.001	1.001	1.000	1.000
	(0.08)	(0.08)	(0.02)	(0.02)
SEC filing	0.947	0.946	0.942	0.944
	(0.62)	(0.62)	(0.66)	(0.63)
# ICOs	2,271	2,271	2,271	2,271
Cohort FE	Y	Y	Y	Y
Clustered SE	Y	Y	Y	Y

This table presents estimates from Poisson regressions. Estimated coefficients are expressed as incidence rate ratios. The dependent variable is *irregularities*. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The key independent variables are log (*centrality*) and 𝟙(*high centrality*). The variable log (*centrality*) is the log-transformed Katz centrality of the ICO. The variable 𝟙(*high centrality*) is an indicator that equals one if the ICO has a higher Katz centrality than the median Katz centrality in the sample, and equals zero otherwise. t-statistics are reported in parentheses.

Table 7. Irregularities and ICO quality

Panel A. Factor analysis

	Factor 1	Factor 2	Factor 3
𝟙(Whitepaper)	0.081	0.086	0.194
𝟙(Github)	0.075	0.022	0.135
𝟙(Experienced)	0.061	0.032	0.106
𝟙(Advisors)	0.145	0.128	0.507
𝟙(Venture funding)	1.704	−0.086	−0.248
𝟙(Code posted)	0.071	0.645	−0.028
𝟙(Code audited)	0.065	0.920	−0.143
Retention	0.050	0.011	0.028
Vesting	0.071	0.117	0.261
Funds raised	0.098	0.026	0.052

	Variance explained (%)	Cumulative (%)
Factor 1	56.4	56.4
Factor 2	24.9	81.3
Factors 3–10	18.7	100

Panel A presents results from factor analysis of variables that are related to ICO quality. We present factor loadings of these variables on the first three factors. We also present the (cumulative) proportion of variance explained by factors 1 through 10. The indicator 𝟙(*whitepaper*) switches on if the ICO has a whitepaper. The indicator 𝟙(*Github*) switches on if the ICO issuer posts code to a Github repository. The indicator 𝟙(*experienced*) switches on if an ICO team member has previously worked on another ICO. The indicator 𝟙(*advisors*) switches on if the ICO lists advisors in its whitepaper or website. The indicator 𝟙(*venture funding*) switches on if the ICO receives venture funding. The indicators 𝟙(*code posted*) and 𝟙(*code posted*) switch on if the ICO posts, respectively, the source code of its smart contract and a security audit of the source code to `Etherscan.io`. *Retention* is the proportion of issued tokens that cannot be sold to outside investors during an ICO. *Vesting duration* is the number of months an ICO issuer declares in a vesting schedule for tokens. *Funds raised* is the amount of capital raised in U.S. dollars.

Table 7. (cont'd) Irregularities and ICO quality

Panel B. Cox regressions (Event: ICO scam)

	(1)	(2)	(3)
Irregularities	1.203	1.085	1.073
	(5.88)	(1.82)	(1.79)
Quality	0.994	0.990	
	(0.15)	(0.34)	
𝟙(Whitepaper)			1.150
			(0.31)
𝟙(Github)			1.007
			(0.03)
𝟙(Experienced)			0.960
			(0.23)
𝟙(Advisors)			1.207
			(2.52)
𝟙(Venture funding)			0.845
			(0.70)
𝟙(Code posted)			0.680
			(0.60)
𝟙(Code audited)			2.850
			(1.27)
Retention			1.001
			(0.23)
Vesting			1.071
			(0.87)
Funds raised			1.000
			(2.92)
Controls	Y	Y	Y
# ICOs	3,281	3,281	3,281
Cohort strata	N	Y	Y
Coverage-quartile FE	N	Y	Y
Clustered SE	N	Y	Y

Panel B presents estimates from Cox regressions. Estimated coefficients are expressed as hazard ratios. The failure event in these regressions is *ICO scam*. An ICO triggers the event if the `DeadCoin` site identifies it as a scam. Otherwise, it is right-censored. The key independent variables are *irregularities* and *quality*. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The *quality* of an ICO is the first extracted factor from the factor analysis in Panel A. t-statistics are reported in parentheses.

Table 8. Regulatory scrutiny and irregularities

	(1)	(2)	(3)
Dependent variable:	$\mathbb{1}$(Irregularities >0)	Irregularities	$\mathbb{1}(\Delta\text{Irregularities}<0)$
Regulatory scrutiny	0.556	0.644	1.177
	(2.23)	(3.31)	(1.53)
Banned	0.727	0.921	1.352
	(2.04)	(1.68)	(2.98)
Whitelist	0.507	0.949	0.609
	(4.03)	(1.08)	(5.38)
Duration	0.998	0.998	1.000
	(1.97)	(2.89)	(0.87)
Presale	4.409	2.469	4.694
	(7.83)	(9.39)	(5.44)
Hardcap	4.370	3.220	2.211
	(9.28)	(21.23)	(4.69)
Softcap	0.846	1.010	0.862
	(1.83)	(0.57)	(3.88)
Accept BTC	1.276	1.142	1.037
	(2.07)	(2.81)	(0.50)
Accept ETH	4.951	2.453	0.434
	(5.76)	(8.10)	(4.39)
Accept USD	0.878	0.998	1.038
	(1.13)	(0.06)	(0.23)
Enforcement	1.549	1.162	1.212
	(3.59)	(5.33)	(1.42)
Disclosure	1.367	1.142	1.141
	(4.34)	(6.61)	(3.24)
SEC filing	0.942	0.950	1.597
	(0.18)	(0.44)	(1.28)
Unit of observation	ICO	ICO	ICO-month
# observations	5,873	5,873	57,617
Cohort FE	Y	Y	Y
Clustered SE	Y	Y	Y

This table present estimates from logistic (columns 1 and 3) and Poisson (column 2) regressions. Estimated coefficients in columns 1 and 3 (2) are expressed as odds (incidence rate) ratios. The dependent variables are $\mathbb{1}(\textit{irregularities}>0)$, *irregularities*, and $\mathbb{1}(\Delta\textit{irregularities}<0)$. The indicator $\mathbb{1}(\textit{irregularities}>0)$ equals one if the ICO has at least one cross-site discrepancies of its characteristics at its first appearance in our sample, and equals zero otherwise. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The indicator $\mathbb{1}(\Delta\textit{irregularities}<0)$ equals one if an ICO has a reduction in cross-site discrepancies from the previous month, and equals zero otherwise. The key independent variable is *regulatory scrutiny*—an indicator that switches on if there are regulatory news articles released within the prior calendar month. t-statistics are reported in parentheses.

Table 9. Other screening tactics

Event: ICO scam

	(1)	(2)	(3)	(4)
𝟙(Celebrity)	12.390	12.027		
	(9.34)	(8.20)		
Web traffic ratio			1.273	1.263
			(2.94)	(2.84)
Irregularities		1.109		1.111
		(2.93)		(2.75)
Controls	Y	Y	Y	Y
# ICOs	5,873	5,873	5,873	5,873
Cohort strata	Y	Y	Y	Y
Coverage-quartile FE	Y	Y	Y	Y
Clustered SE	Y	Y	Y	Y

This table presents estimates from Cox regressions. Estimated coefficients are expressed as hazard ratios. The failure event in these regressions is *ICO scam*. An ICO triggers the event if the `DeadCoin` site identifies it as a scam. Otherwise, it is right-censored. The key independent variables in our regressions are 𝟙(*celebrity*), *web traffic ratio*, and *irregularities*. The indicator 𝟙(*celebrity*) equals one if an ICO is endorsed by a celebrity, and equals zero otherwise. To compute *web traffic ratio* of an ICO, we first classify web traffic to listing websites into two categories—passive and active. Passive web traffic counts visitors referred to a listing website via third-party referral links, paid advertisements, and search engines. Active web traffic counts visitors who access a listing website by directly typing its Uniform Resource Locator (URL) or through the use of saved browser bookmarks. Next, we define the *web traffic ratio* of an ICO as the ratio of passive traffic to active traffic, aggregated across the listing websites that list it in the month prior to its start date. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

Table 10. Estimating the true prevalence of ICO scams

Detection controlled estimation (DCE)

	(1)	(2)	(3)	(4)	(5)	(6)
	Model A		Model B		Model C	
	Scam	Detect	Scam	Detect	Scam	Detect
Irregularities	0.105	0.119	0.084	0.122	0.076	0.124
	(6.59)	(5.32)	(3.46)	(8.37)	(3.29)	(8.54)
Instruments						
App downloads	0.119					
	(2.18)					
Altcoin returns			0.402			
			(4.51)			
BTC returns					0.581	
					(4.25)	
Controls	Y		Y		Y	
# ICOs	5,873		5,873		5,873	
Est. # (%) Scams	2,625 (44.7%)		2,893 (49.3%)		2,888 (49.2%)	

This table presents estimates from detection controlled estimation (DCE) models, which are implemented as bivariate probit models. We simultaneously model the scam and detection processes of ICO scams. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The instruments for the scam processes are *app downloads*, *altcoin returns*, and *BTC returns*. The variable *app downloads* is the log-transformed number of downloads of cryptocurrency exchange mobile applications in the month prior to the ICO start date. The variable *altcoin returns* (*BTC returns*) is the cumulative returns of non-Bitcoin cryptocurrencies (Bitcoin) in the month prior to the ICO start date. To probabilistically identify ICO scams, we fit the scam stage (columns 1, 3, and 5) of every model. Thereafter, we compute the number (proportion) of ICOs identified as scams. t-statistics are reported in parentheses under the estimated coefficients.

For Online Publication

Internet Appendix to: The Economics of Financial Scams: Evidence from Initial Coin Offerings

Kenny Phua Bo Sang Chishen Wei Gloria Yang Yu

The Internet Appendix contains supplementary information and additional tests for the paper "The Economics of Financial Scams: Evidence from Initial Coin Offerings". The contents of the Internet Appendix are organized as follows.

For Online Publication

Internet Appendix to: The Economics of Financial Scams: Evidence from Initial Coin Offerings

The Internet Appendix contains supplementary information and additional tests for the paper "The Economics of Financial Scams: Evidence from Initial Coin Offerings". The contents of the Internet Appendix are organized as follows.

Section I Overview of ICOs.

Section II Details on ICO scam allegations extracted from the `Deadcoins` website.

Section III Further details to Subsection 4.2 of the main text.

Section IV Robustness checks of the link between irregularities and ICO scam risk.

I Overview of ICOs

An ICO allows entrepreneurs to raise capital via cryptographically secured tokens. Typically, an issuer resorts to an ICO when other sources of capital (e.g., venture capital and private equity) are prohibitively expensive or inaccessible. Thus, an ICO is a risky crowdfunding operation, in which the issuer sells tokens that will serve as the payment medium for the products or services of the start-up. There are several stages in the ICO process. First, the issuer creates fundraising campaign materials. Next, the issuer sets the pricing terms and markets the offering on listing websites. Finally, if the ICO financing goals are met, the issuer creates and distributes tokens to investors.

I.I Fundraising campaign: Listing websites

The fundraising campaign entails (i) producing a whitepaper, (ii) hosting a website to provide additional information, (iii) maintaining an active social media presence, and (iv) listing the token on ICO listing websites. A whitepaper describes the project goals, objectives, and development milestones. But whitepapers often lack details of business operations and rarely contain financial disclosures.

To list an ICO on a listing website, the issuer directly submits token information on the website and awaits approval. Such submissions require little technical sophistication. Listings are typically free. But, for an additional fee, the website can prominently feature and promote the ICO. The issuer may also hire advisors to promote and market the ICO. These advisors usually have technical or marketing expertise and may alleviate information asymmetry between the issuer and potential investors. However, celebrities with little or no blockchain expertise are also employed as advisors to promote the ICO. The SEC has warned that celebrity endorsements are often associated with ICO scams.

I.II ICO pricing and listing on secondary markets

The pricing structures of ICOs are often opaque. On listing websites, issuers advertise a subscription price to the general public. But many ICOs invite privileged

investors to an earlier presale offering. While details on the presale pricing structure are not publicly available, Fahlenbrach and Frattaroli (2020) find that presales offer a significant discount to the subsequent public offering price. Presale funding rounds are controversial. They may signal strong demand from informed investors but are also used to manipulate the sentiments of the general public. The SEC has also warned that presales are often associated with ICO scams.

The issuer may set funding goals in the ICO. The softcap is the minimum amount of funds raised to continue the project. An issuer may also specify a hardcap, which is the maximum number of tokens that can be sold in the ICO. The hardcap limits the amount of funds that can be raised in the ICO. If the softcap is met and the project is successful, the issuer will create and distribute the tokens to investors. Subsequently, investors may trade the tokens in the secondary market or use the tokens for its utility (e.g., access products or services funded by the ICO). Investors tend to have short holding periods and flip the tokens on cryptocurrency exchanges (Fahlenbrach and Frattaroli, 2020).

I.III Regulatory environment

The ICO regulatory environment differs across countries. Some countries impose outright bans on ICOs (e.g., China and South Korea), while other countries adopt regulatory guidelines (e.g., Australia and the United States). The SEC of the United States uses the Howey Test framework to determine whether a digital asset qualifies as a security.[1] Specifically, a digital asset is a security if (i) there is an investment of money and (ii) expectation of profits; (iii) the investment of money is in a common enterprise; and (iv) any profit comes from the efforts of a promoter or third party. The SEC Chairman Gary Gensler and his predecessor Jay Clayton believe most ICOs pass the Howey Test and are hence subject to U.S. securities laws.

Issuers of security tokens can register with SEC via Form S-1 or apply for registration exemptions. Although most ICOs should arguably be classified as security offerings, fewer than 100 tokens in our sample are registered with the SEC potentially

[1] For details on the Howey Test framework, refer to:
`https://www.sec.gov/corpfin/framework-investment-contract-analysis-digital-assets`

due to the high compliance costs. For exemptions, Regulation D applies if funds are raised from only accredited investors; Regulation A and A+ apply if funds are raised from a broader set of investors but the offering is less than $50 million; and issuers can also make token sales under Regulation Crowdfunding.

II Details on ICO scam allegations

This section provides details on the ICO scam allegations extracted from `Deadcoins`. These allegations are crowdsourced from the investor community and are reported as unstructured texts. Crowdsourced scam allegations provide valuable insights but may also contain biases or unverified claims. To reduce false positive errors, we further verify the allegations according to the process in Section 3.3 of the main text. The final sample consists of 243 allegations.

We construct a word cloud from the 100 most frequent words, excluding common stopwords and very short words. Figure I shows that the word "scam" is the most frequent. Allegations often contain references to attributes of the "project" and "team", and also mention "Telegram", which is an online messaging application with enhanced privacy and encryption features.

We manually read the 243 scam allegations and find mentions of these five topics in decreasing frequencies: (i) signs of dishonest behavior, (ii) product concerns, (iii) reputation of ICO team/issuers, (iv) stoppage of communications, and (v) regulatory issues. To analyze the data in a systematic way, we employ a machine learning technique called zero-shot learning (ZSL) developed in computer science. We utilize ZSL because our sample of scam allegations is too small to employ machine learning text classifiers (e.g., Latent Dirichlet Allocation topic models), and there are no off-the-shelf classifiers specifically trained to detect textual mentions of these five topics.

A ZSL model is designed to recognize and categorize items or topics that it has "never seen before" in its training dataset. Specifically, we use the `bart-large-mnli` model (Yin, Hay, and Roth, 2019), which is a large language model with 407 million parameters, trained on 433,000 annotated sentence pairs from the Multi-Genre Natural Language Inference (MultiNLI) dataset (Williams, Nangia, and Bowman, 2017).

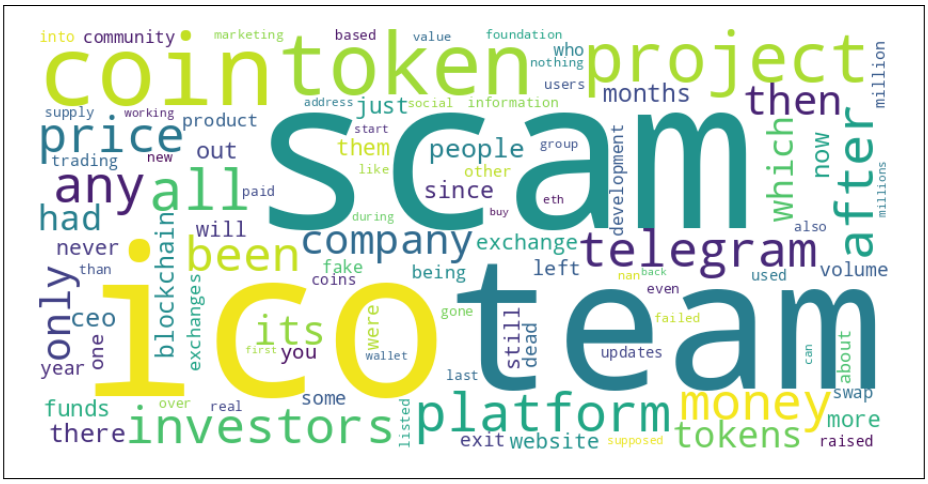

Figure I. This figure presents the word cloud of ICO scam allegations extracted from the now-defunct `Deadcoins` website (perma.cc/AEV2-KW27). The word cloud is generated from the 100 most frequent words across the scam allegations, excluding common stopwords and words with fewer than three characters.

We prompt the `bart-large-mnli` model to detect mentions of the five topics by providing these keywords: (i) dishonest (ii) product, (iii) reputation, (iv) communications, and (v) regulations. The model scans the allegations for these specific keywords but, more importantly, links contextual references in the allegations to the keywords. For example, it could infer that the sentences "*The ICO team stopped posting updates*" and "*The team ran another scam previously*" relate to communications issues and reputation concerns, respectively. For every scam allegation, the model assigns probability scores corresponding to each of the five keywords. A score for a keyword represents the probability that the related topic is mentioned in the allegation. Because an allegation may contain mentions of multiple topics, we allow the sum of probability scores to exceed one. Thereafter, we average the probability scores for each keyword in all 243 allegations and present their distributions in Figure II.

Dishonesty is the most common justification behind allegations (78.3%). Product concerns feature heavily in the allegations (47.0%) because ICOs often raise funds in the ideation stage of their entrepreneurial ventures, Allegations also frequently cite concerns about the issuers' reputations (39.8%) and stoppage of communications (38.8%). Investors are infrequently concerned with regulatory issues around the ICO (9.0%), per-

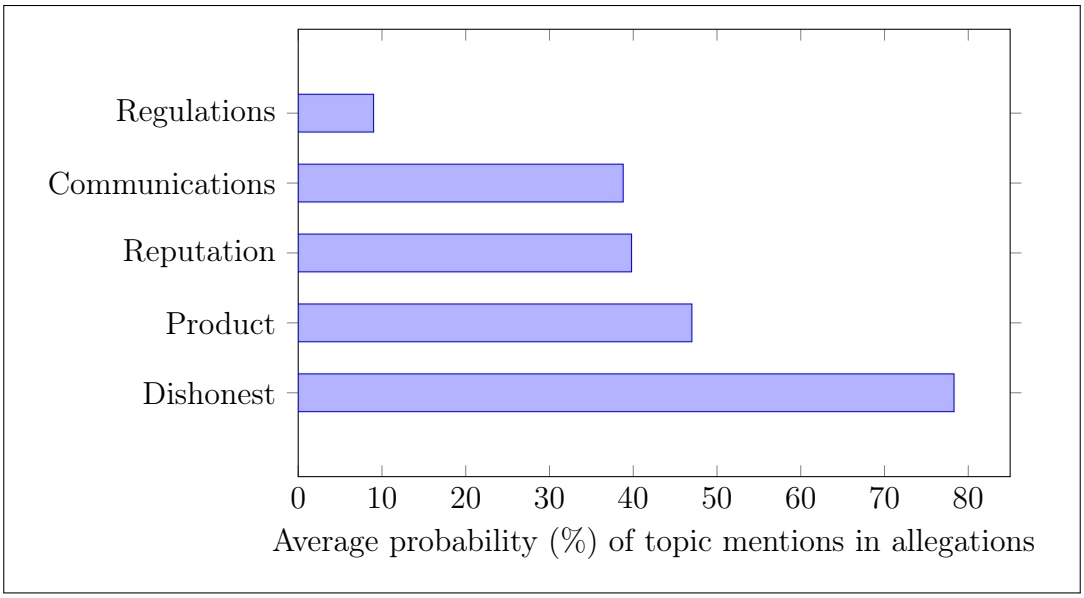

Figure II. This figure presents the average probabilities of topic mentions in the 243 ICO scam allegations extracted from the now-defunct `Deadcoins` website (perma.cc/AEV2-KW27).

haps due to the lack of clear, robust regulatory oversight. Finally, we verify that the ZSL classification yields highly similar topic mentions with our initial manual reading.

III More details: Assessing the screening mechanism

We describe how we match an ICO token to its contract address on the Ethereum network. To begin our matching process, we search for either the name or ticker of the ICO on the `Etherscan.io/tokens` website. In the ideal case, we would find only one match from this search. In that case, we collect the contract address stated on the `Etherscan` page. To increase the likelihood that we capture the token contract, we check that there is a number of "Holders" stated on the page. If we find multiple matches, we pick the contract address with the highest number of "Holders". Sometimes, the name/ticker of an ICO is exceedingly common and will match with many ICO projects. Due to time and resource constraints, we abandon a token match in cases where our search turns up 100 or more potential matches. For example, Figure III contains screenshots of the `Etherscan` page for the BNB ICO token.

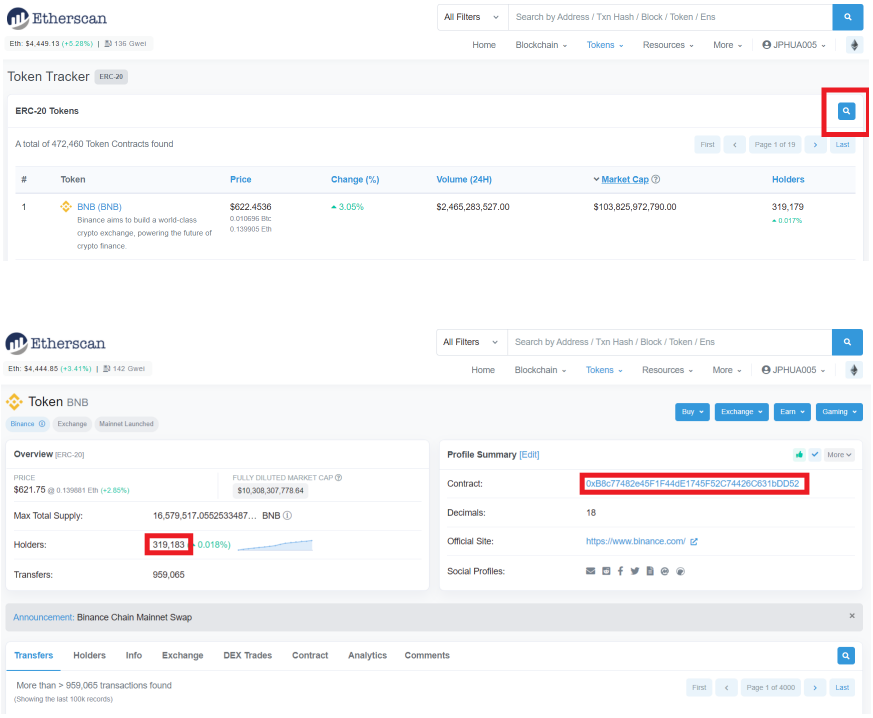

Figure III. This figure presents screenshots of the BNB ICO token page on `Etherscan`.

We use the following procedure to characterize the sophistication of wallet-users. First, we find ICO contract addresses by manually searching the ICO name and ticker on the website `Etherscan.io`. Every ICO token has a unique contract address on the Ethereum blockchain. Next, we find the wallet addresses that hold each ICO tokens at the Ethereum block height corresponding to 10 days after the ICO end date using the `Covalent Unified` API. We focus on the top 100 wallet addresses of every ICO by token holdings and exclude an ICO if it has fewer than 30 wallets holding its tokens. Our final sample in this test comprises 110,607 wallets holding tokens of 1,996 ICOs.

IV Robustness checks

In this section, we show that the relation between irregularities and ICO scam risk is robust to various econometric specifications.

IV.I Irregularities and ICO scam risk

We perform various robustness checks on the results presented in Table 3 of the main text. Overall, our robustness checks produce estimates that are quantitatively and qualitatively similar.

ICO COHORT DEFINITION. In column 1 of Table I, we reconstruct ICO cohorts at the calendar month level and stratify our Cox regression as such. Compared to column 3 of Table 3 in the main text, we find a quantitatively similar effect of *irregularities* on the hazard of ICO scams ($+11.8\%, t = 2.74$).

MULTIPLE LISTINGS ONLY. By construction, irregularities could be mechanically driven by the number of websites that an ICO is listed on. Pre-empting this concern, we have saturated our Cox regression models in Table 3 of the main text with coverage quartile fixed effects. In column 2 of Table I, we further address this concern by removing from our sample ICOs that are listed on a single website. In this restricted sample, we continue to find a positive and statistically significant loading on *irregularities* ($+12.1\%, t = 2.46$).

OLS ESTIMATOR. In Section 4.1 of the main text, we use Cox regressions to model ICO scam risk to handle the time-to-event nature of our data. Specifically, the time it takes to discover a scam could differ across ICOs, and many ICOs remain without a scam allegation by the end of our observation window. Moreover, the residuals of time-to-event data often depart from normality assumptions. Because of these features in our data, an OLS estimator may yield biased estimates and invalid standard errors. However, for robustness, we test whether our findings hold with an OLS estimation in column 3 of Table I. We find that an additional irregularity increases the probability of an ICO scam by +1.1 ($t = 3.18$) percentage points, on average. Given that only 4.1% (243/5,873) of our sample ICOs are labeled as scams, this estimate is economically significant. Nevertheless, we obtain an OLS R^2 of only 4.2% because ICO scams are rare relative to the occurrence of irregularities in our sample. In Section 7 of the main text, we explain that—typical of data on scams—ICO scams are likely underdetected in our sample and take steps to econometrically adjust for this issue.

Table I. Robustness checks: Irregularities and ICO scams

	(1)	(2)	(3)
Estimator	Cox	Cox	OLS
Specification	(A)	(B)	(C)
Irregularities	1.118	1.121	0.011
	(2.74)	(2.46)	(3.18)
# ICOs	5,873	2,433	5,873
Cohort strata	Y	Y	N
Cohort FE	N	N	Y
Coverage-quartile FE	Y	Y	Y
Cluster SE	Y	Y	Y
R^2	-	-	4.2%

Columns 1 and 2 (Column 3) of this table present estimates from Cox (OLS) regressions. Estimated coefficients in columns 1 and 2 are expressed as hazard ratios. The key independent variables are *irregularities* and *retention*. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. The variable *retention* is the fraction of all issued tokens that cannot be sold to outside investors during an ICO (Davydiuk, Gupta, and Rosen, 2022). Section 3 of the main text contains variable definitions. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

IV.II Other screening tactics

Using OLS estimation, we examine how ICO scam risk is related to celebrity endorsements, choice of listing websites, and irregularities. Table II presents our results, which are quantitatively and qualitatively similar to those in Table 8 in the main text. Columns 1 and 2 show that celebrity endorsements increase the probability of an ICO scam by about +40 percentage points. In columns 3 and 4, we find that malicious issuers tend to list on websites that derive more web traffic from likely unsophisticated individuals. We continue to find a statistically and economically significant link between *irregularities* and ICO scam risk.

Table II. Robustness checks: Other screening tactics

Dependent variable: 𝟙(ICO scam)				
	(1)	(2)	(3)	(4)
𝟙(Celebrity)	0.409	0.403		
	(7.03)	(6.81)		
Web traffic ratio			0.006	0.005
			(2.29)	(2.20)
Irregularities		0.010		0.011
		(3.28)		(3.13)
Controls	Y	Y	Y	Y
# ICOs	5,873	5,873	5,873	5,873
Cohort FE	Y	Y	Y	Y
Coverage-quartile FE	Y	Y	Y	Y
Clustered SE	Y	Y	Y	Y
R^2	5.7%	6.2%	3.8%	4.3%

This table presents estimates from OLS regressions. The dependent variable 𝟙(*ICO scam*) is an indicator that equals one if the ICO is ever identified as a scam on `Deadcoins`, and equals zero otherwise. The key independent variables in our regressions are 𝟙(*celebrity*), *web traffic ratio*, and *misrep*. The indicator 𝟙(*celebrity*) equals one if an ICO is endorsed by a celebrity, and equals zero otherwise. To compute *web traffic ratio* of an ICO, we first classify web traffic to listing websites into two categories—passive and active. Passive web traffic counts visitors referred to a listing website via third-party referral links, paid advertisements, and search engines. Active web traffic counts visitors who access a listing website by directly typing its Uniform Resource Locator (URL) or through the use of saved browser bookmarks. Next, we define the *web traffic ratio* of an ICO as the ratio of passive traffic to active traffic, aggregated across the listing websites that list it in the month prior to its start date. The *irregularities* of an ICO is the total number of cross-site discrepancies of its characteristics at its first appearance in our sample. Standard errors are clustered by ICO cohorts. t-statistics are reported in parentheses.

References

Davydiuk, T., Gupta, D., and Rosen, S. (2022). "De-crypto-ing signals in initial coin offerings: Evidence of rational token retention". Management Science, forthcoming.

Fahlenbrach, R. and Frattaroli, M. (2020). "ICO investors". Financial Markets and Portfolio Management.

Williams, A., Nangia, N., and Bowman, S. (2017). "A broad-coverage challenge corpus for sentence understanding through inference". arXiv preprint arXiv:1704.05426.

Yin, W., Hay, J., and Roth, D. (2019). "Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach". arXiv preprint arXiv:1909.00161.