The Dark Web and Capital Markets

Brigham Brau
University of North Carolina
Brigham_Brau@kenan-flagler.unc.edu

Juuso Itkonen
Cyber Intelligence House and Macquarie University
juuso@cyberintelligencehouse.com

Junyoung Jeong
University of North Carolina
Junyoung_Jeong@kenan-flagler.unc.edu

Mark Lang
University of North Carolina
Mark_Lang@kenan-flagler.unc.edu

Mikko Niemelä
Cyber Intelligence House
mikko@cyberintelligencehouse.com

June 2025

ABSTRACT: We use granular dark web data to infer cyber incidents and link them to capital market outcomes. We find that dark web activity increases before cyber breaches are disclosed, suggesting that many breaches occur substantially earlier than reported. Consistent with capital market awareness, information asymmetry and informed trade rise during periods of elevated dark web activity prior to cyber breach disclosure. Next, we identify firm-level periods of heightened dark web activity with no subsequent breach disclosure, which we term "shadow breaches." These episodes exhibit similar characteristics to disclosed breaches, are associated with increased information asymmetry and informed trading and are followed by negative stock returns. Using dark web activity to identify an adjusted event date, we find a negative market reaction around the adjusted breach date, but not around the disclosure date, suggesting that dark web activity preempts breach disclosure. Lastly, we find that breach activity is more likely to be disclosed in high-tech industries and among firms monitoring the dark web. Our findings support the usefulness of dark web data to detect cyber breaches and suggest cyber breaches are more prevalent than disclosed, disclosure is often delayed, and firms monitoring the dark web are more likely to detect breaches.

Keywords: Dark web, informed trading, cybercrime, cyber breach
JEL classification: G14, G18, D82

This paper benefited from helpful comments from workshop participants at the A. B. Freeman School of Business at Tulane University, Columbia Business School, the Kenan-Flagler Business School at the University of North Carolina at Chapel Hill, and the Naveen Jindal School of Management at the University of Texas at Dallas.

1. Introduction

In 2019, the World Economic Forum's Executive Opinion Survey identified cyberattacks as a top risk for CEOs (World Economic Forum, 2019). Survey evidence suggests that about 50% of companies have experienced direct cyber breaches (Kron, 2022), and research by two cybersecurity firms estimates that 98% of organizations use at least one third-party vendor that has experienced a breach in the last two years (Cyentia Institute and SecurityScorecard Research, 2023). Firms are also concerned about cybercrime. An analysis of Fortune 100 firms finds that 95% disclosed a focus on cybersecurity in the risk oversight section of their proxy statements filed in the year ending May 2022, and 99% disclosed efforts to mitigate cybersecurity risk (Ernst & Young, 2022). Despite widespread concern over cyber risk and evidence that breaches are common, formal disclosure of cyber incidents remains relatively rare.

Because cybercrime typically involves a firm's internal systems and data, formal disclosure of cyberbreaches is frequently the only verifiable public source of information about the extent and implications of such incidents. The SEC has recently increased its focus on cybercrime, disclosure practices, and associated risks amid growing evidence that formal disclosures are relatively rare and are delayed relative to cybercrime activity. In 2022, the SEC documented only 35 Form 8-K filings reporting material cybersecurity incidents, despite examining 71,505 Form 8-K filings from 7,416 filers. Among 977 disclosed cybersecurity breaches recorded by Audit Analytics (2019–2025), only 385 (39.4%) included information on when the firm first became aware of the incident. The companies that disclosed the timing of breaches took an average of 42 days to discover a breach and 89 days to disclose it after discovery. This timeline does not fully capture the period of exposure, as industry estimates suggest breaches may go undetected for months (CYFOR Secure, n.d.; IBM, 2022). Further, as discussed later, most disclosed breaches are reported through State

Attorneys General or other outlets rather than being included in SEC filings.

Detecting cyber breaches is inherently difficult for both firms and researchers because cybercriminals operate under conditions of secrecy and anonymity. Breaches are often designed to remain hidden, making it challenging to observe when and how a firm has been compromised. Moreover, firms have an incentive to limit disclosure to avoid reputational damage. Regulators, governments and firms worry that requiring detailed disclosure of breaches risks laying out a roadmap for cyber criminals by highlighting mechanisms underlying successful breaches, as well as detection and mitigation strategies. However, because there are gains to specialization and trade, cybercriminals who access privileged data (e.g., login credentials, account identifiers, passwords, or credit card data) have strong incentives to connect with buyers who specialize in exploiting stolen information, either directly or in combination with other personal data they possess.

As a result, unindexed forums, websites, and chat rooms, collectively referred to as the dark web, serve as hubs for the sale of data stolen in cyber breaches.[1] The dark web is a continuous source of anonymized communication, including discussions and offers to transact in stolen data. Potential sellers typically advertise breached data on dark web forums and discussion boards and move to more secure forums (e.g., encrypted messaging apps) to complete transactions. Given the fact that the value of illicit data often decreases rapidly (e.g., firms reinforce security and change credentials once they realize a breach has occurred and customers change passwords), the dark web potentially provides a continuous and timely lens into cyber risk exposure. While anonymized from the perspective of individual buyers and sellers, discussions and proposed transactions on the dark web are observable using specialized software configurations.

[1] "The dark web is the World Wide Web content that exists on darknets (overlay networks) that use the Internet but require specific software, configurations, or authorization to access. Through the dark web, private computer networks can communicate and conduct business anonymously without divulging identifying information, such as a user's location." Wikipedia, accessed 2/13/2025.

A growing industry of companies such as Cyber Intelligence House (CIH), specializes in continuously scraping the dark web to provide clients with real-time data to provide insight into potential breaches and risk exposure.[2] CIH uses a variety of techniques to gain access to a wide range of marketplaces, chatrooms and discussion boards on which cyber breaches are discussed and transactions are proposed. Some companies use CIH data directly to track their cyber exposure, while others outsource to cybersecurity consulting and service firms who integrate this data into broader risk assessments, ongoing monitoring, and mitigation strategies.

For our main empirical analyses, we use weekly firm-level dark web activity recorded and categorized by CIH algorithms. We focus on dark web selling and transactions offers ("*Blackmarkets*") for NYSE-listed firms over the period 2020 to early 2025. Our descriptive statistics indicate that nearly all of our NYSE sample firms have some level of dark web exposure during our sample period. While offers to sell firm information, measured by *Blackmarkets*, are a rarer event, they are still prevalent in our sample.[3] More than 75% of our sample firms have *Blackmarkets* activity at some point during our sample period and more than 20% have some *Blackmarkets* activity in a given week.

Industry reports suggest that cybercriminals quietly access systems to steal and quickly sell data, as exposed breaches reduce the value of stolen information due to heightened security measures. Our empirical analysis supports this notion, showing that *Blackmarkets* (dark web sales activity) rises sharply in the months before a cyber breach is disclosed and returns to normal levels following breach disclosure.

[2] Cyber Intelligence House "detects and monitors cyber exposure via dark web, deep web and data breaches alerting individuals and businesses about potential cyber threats." Wikipedia, accessed 2/13/25.

[3] We observe offers to sell rather than actual transactions. The dark web serves primarily as a forum to advertise breached information, while the actual sales take place over more secure platforms (e.g., instant messaging apps like Jabber). We follow CIHs convention of focusing on sustained and abnormally high levels of dark web activity.

The SEC's recent rule requiring prompt cyber breach disclosure stems from concerns that many firms delay disclosure, limit detail or avoid disclosure entirely, potentially to limit reputational damage or avoid disclosing sensitive information to cybercriminals.[4] The SEC worries that delayed or incomplete disclosure may permit informed traders, insiders, and even hackers to exploit cyber breaches by trading in capital markets.[5]

To contextualize the delay in disclosure, and illustrate the potential role of dark web data in providing insight, we begin by presenting anecdotal evidence from a major cyber breach, involving the MOVEit file transfer company and compromising data for 30 NYSE firms. This case study is informative because MOVEit disclosed the timing of the breach, and we can observe dark web transaction data around the breach date. In addition, we can observe disclosure lags by the 30 NYSE breached firms, whose data was all breached at the same time. We find clear evidence of a surge in dark market transaction activity around the breach date. Consistent with SEC concerns, disclosure was significantly delayed and variable across breached firms, with the first firms disclosing 4-5 months after the breach and the plurality of firms disclosing more than a year after the breach.

We then turn to our primary empirical analysis to evaluate dark web activity for disclosed breaches more broadly. We focus initially on 215 disclosed breaches on Audit Analytics of which 44 were disclosed in SEC filings, 127 in reports to Attorneys General and the remainder through other sources. An advantage of this sample is that we know that a breach occurred. For the

[4] The SEC's final rule states: "Several commenters indicated both that investors look for information regarding registrants' cybersecurity incidents and that current disclosure levels are inadequate to their needs in making investment decisions. In addition, we note […] evidence showing that delayed reporting of cybersecurity incidents can result in mispricing of securities, and that such mispricing can be exploited by threat actors, employees, related third parties, and others through trades made before an incident becomes public. Accordingly, we believe it is necessary to adopt a requirement for uniform current reporting of material cybersecurity incidents." (Securities and Exchange Commission, 2023, p. [28-29]).

[5] See Piccotti and Wang (2023); Mitts and Talley (2019); Amir et al. (2018); Amir et al. (2019); Lin et al. (2020); Chen et al. (2024); Motoki and Pinto (2024).

disclosed breaches, we find clear evidence of elevated dark web transaction activity prior to the disclosure date, peaking 3-4 months prior to disclosure. To assess whether the stock market was aware of the breaches prior to disclosure, we examine measures of stock market activity around the dark web activity and find evidence of significant increases in information asymmetry and informed trading in weeks with abnormal black market transaction activity. Overall, the results suggest that dark web activity is elevated around breaches, disclosure is delayed and capital market participants seem aware of the activity.

We then turn to "shadow breaches," cases of abnormal sustained dark web transaction activity without subsequent breach disclosure.[6] We find that shadow breaches are as common as disclosed breaches and descriptively resemble disclosed breaches, with similar industry distribution and dark web activity patterns. Consistent with our earlier analysis of disclosed breaches, we find significant increases in information asymmetry and informed trading during shadow breaches, with larger effect sizes when we apply stricter definitions of shadow breaches.

Next, we turn to stock price reactions to shadow breaches. Consistent with dark web activity conveying negative information, we find significant negative returns following shadow breaches. In some sense, this runs counter to a stream of literature that finds only mixed evidence of stock market reaction around breach disclosure dates.[7] This difference may reflect the lag between dark web activity and breach disclosure. To test this possibility, we estimate breach dates for disclosed breaches based on dark web transaction activity (similar to our approach for shadow breaches). On average, this adjusted breach date occurs 4-5 months prior to formal disclosure. Consistent with prior research, we find limited market reaction at the official disclosure date. However, we

[6] The cyber disclosure regulation adopts the usual securities law definition of materiality; i.e., "a substantial likelihood that the disclosure of the omitted fact would have been viewed by a reasonable investor as having significantly altered the total mix of information made available", TSC Industries, Inc. v. Northway, Inc., 426 U.S. 438 (1976).

[7] Kvochko and Pant, 2015; Campbell et al., 2003; Kannan et al., 2007; Goel and Shawky, 2009.

document a significant negative return following the estimated breach date, suggesting that investors incorporate breach information well before it is formally disclosed.

Lastly, we compare disclosed breaches with shadow breaches to provide descriptive evidence on determinants of a firm's disclosure decision. A potential reason that a firm with dark web activity suggesting a breach might not disclose a breach is that they may not monitor dark web activity as part of their cyber risk mitigation activity. We examine Form 10-K cyber risk disclosure for evidence on whether firms monitor the dark web as part of their risk assessment process.[8] We find that larger and more technology-oriented firms are more likely to disclose breaches. Potentially more telling, firms that disclose in their SEC filings that they actively monitor the dark web as part of their cyber risk efforts are more likely to disclose breaches. While we cannot infer causality, the results suggest that firms that more closely monitor the dark web are more likely to disclose.

We make three primary contributions. First, we provide initial evidence on the potential for dark web data to offer important insights into breaches and cyber risks for firms, investors, and regulators. Cyber risk is increasingly recognized as a significant issue with considerable economic implications for firms and their stakeholders, as breaches can lead to severe negative consequences (Kamiya et al., 2021; Ashraf, 2022; Crosignani et al., 2023). Our evidence suggests that dark web data can potentially serve as an early-warning mechanism and valuable resource for proactive risk management, and that dark web activity is reflected in overall capital markets outcomes. In addition, our results highlight the potential usefulness of dark web data to capital markets researchers in assessing and measuring cyber risk and breach incidents.

Second, we document significant capital market implications associated with cyber events. Our

[8] The SEC requires that, "[r]egistrants must describe their processes, if any, for the assessment, identification, and management of material risks from cybersecurity threats."

results suggest that information asymmetry, illiquidity and informed trading respond to undisclosed cyber breaches (Lin et al., 2020; Mitts & Talley, 2019; Akey et al., 2022). In addition, our results highlight the potential usefulness of dark web data for identifying cyber breach dates to infer stock market reactions (Kannan et al., 2007; Campbell et al., 2003).

Finally, our results speak to the recent SEC disclosure requirements on cyber risk and firm-level cyber monitoring. Prior research highlights the challenges inherent in timely identification of cybersecurity incidents and the ambiguous disclosure materiality thresholds (Ali et al., 2019; Foerderer & Schuetz, 2022). Our results suggest that dark web tracking may offer firms and regulators actionable sources of information for evaluating and managing cyber threats, which is particularly relevant given recent regulatory shifts mandating prompt cyber risk disclosure.

2. Institutional Background

2.1 Deep and Dark Web

The dark web is a concealed layer of the internet that is inaccessible via conventional search engines and requires specialized software to access.[9] Although multiple darknet protocols exist (e.g., I2P, Freenet, and Tor), the Tor network is the most widely adopted, and the scope of this discussion is limited to dark web content hosted on Tor hidden services. These services employ layered encryption and use non-standard domain extensions (e.g., *.onion), which are only resolvable through Tor-enabled browsers. While originally developed to safeguard privacy and anonymity, the Tor-based dark web has evolved into a central platform for illicit activity. Cybercriminals use dark web forums, encrypted messaging channels, and darknet marketplaces to

[9] The "clear web" (or surface web) refers to publicly accessible websites that are indexed by standard search engines. The "deep web" comprises online content not indexed by search engines, including private databases, academic repositories, intranet systems, and content behind paywalls or authentication barriers. The "dark web" is a subset of the deep web that is intentionally hidden and accessible only via specialized anonymity-preserving protocols, such as Tor or I2P. While the deep web includes both benign and secure resources (e.g., banking portals or medical records), the dark web is commonly associated with anonymized communication and, in many cases, illicit activity.

disseminate, trade, and monetize compromised assets—including corporate credentials, personally identifiable information (PII), financial account data, and proprietary intellectual property.

The anonymity of dark web platforms facilitates candid discussion among malicious actors, and fresh data continuously becomes available with the rapid turnover of active transactions. Cybersecurity companies such as CIH continuously monitor and scrape content from the dark and deep web, collecting breached data, infostealer logs, and social channel discussions for activity related to firms' cyber risk. CIH data coverage includes dark web forums, deep web marketplaces, malware ecosystems, and hacker chatter. CIH classifies activity into discrete categories including *Blackmarkets* activity (offers to sell firm-specific data) and *Databreach Information* (instances where a firm is mentioned in the context of a data breach) allowing us to construct a continuous measure of firm-specific cyber breach activity on the dark web.

Dark web data has the potential to serve at least two purposes for capital markets research. First, the dark web serves as a real-time repository of cyber threat information, offering potential insights into both the timing and magnitude of data breaches. Investors, recognizing the potential value of real-time information flow on cyber risk, have started including dark web monitoring tools as part of their alternative data stack (Kolanovic and Smith, 2019). Second, because firms may not immediately detect or publicly disclose breaches, dark web activity can serve as an early indicator of "shadow breaches," periods of sustained abnormal activity that hint at data leakage without formal disclosure. As a result, the dark web provides a complementary data source to official disclosures, enabling researchers and regulators to assess the timing, nature, and extent of cyber risks.

That said, we view the ability of our dark web measures to capture cyber breach activity and provide information to capital market participants to be an open empirical question for several reasons. First, the dark web is vast and unregulated so it is possible that, while cybercriminals with specific knowledge may be able to parse the data, a broader algorithmic approach may be less effective. Second, despite the fact that CIH uses a wide range of approaches to access a comprehensive range of data sources including marketplaces, chatrooms and discussion boards, it is possible that CIH is only able to access a subset of activity given the inherent secrecy of the dark web. By virtue of being unindexed and continuously evolving, it is infeasible to exhaustively document all relevant dark web activity. Third, given the dark web is unregulated, it is likely the case that some actors advertise nonexistent or stale data in an attempt to attract unsophisticated buyers, reducing the information content of advertised data.[10]

2.2 SEC Mandates and Cyber Breach Disclosure

The SEC recently enacted requirements for publicly traded firms to disclose material cyber incidents. In particular, "Public companies must provide the required cybersecurity incident disclosure within four business days after the company determines the incident to be material." Prior to the changes, cyber risk was covered by the more generic requirement to disclose significant events relevant to shareholders within four business days.

Despite these regulatory safeguards, anecdotal evidence discussed earlier suggests that only a small fraction of cyber breaches are formally disclosed. The discrepancy likely stems from some combination of lack of knowledge of breaches and concerns over reputational damage, legal liability, and corporate secrecy. In addition, the definition of "material" in the context of cyber risk is potentially ambiguous. Although firms are required to disclose events that "a reasonable shareholder would consider important," the materiality threshold remains subjective, permitting

[10] While some offers to sell breached data may be illegitimate (e.g., the seller is hoping to attract a naïve buyer), buyers are relatively sophisticated in requiring proof that data is legitimate (e.g., access to part of the dataset before agreeing to purchase or escrowing funds until data can be verified), and sellers may rely on reputation to reassure potential buyers. We follow CIH's approach of focusing on abnormally high sustained levels of dark web activity.

discretion in what is reported. This discretion creates the potential for selective disclosure and informational gaps that may affect capital market participants. Disclosure patterns support this view, with a majority of firms submitting cyber breach disclosures to State Attorneys General rather than filing with the SEC.[11]

3. Data

Our main data source is CIH, a cyber intelligence agency founded in 2015 that specializes in tracking cyber exposure on the dark web. CIH continuously monitors and scrapes the dark web for activity, using algorithms to classify this activity for individual firms based on the nature of the dark web activity. A key advantage of CIH is that, in addition to real-time monitoring, it maintains an archive of the dark web activity, enabling ex-post investigation of breaches, data leaks, and other cyber threats. CIH sells its data directly to individual firms seeking to assess their own cyber exposure, cybersecurity service providers that firms hire to manage and mitigate risks, investors, governments and intergovernmental agencies such as Interpol and the UNODC.

3.1 Dark Web Data

Using CIH's dark web data, we construct a weekly panel of dark web activity for NYSE firms from 2020 to early 2025. CIH uses curated firm domains, emails, and keywords to identify if dark web activity is related to a firm and then uses a proprietary algorithm to classify the activity. We are primarily interested in the category *Blackmarkets* provided by CIH. *Blackmarkets* is categorized as dark web activity involving a sales offering. Because many of the transactions are not finalized on the dark web for security reasons (e.g., using private chats), *Blackmarkets* primarily comprises messages and listings offering firm data for sale. The typical transaction involves a listing of information for sale on a dark web forum directing the potential buyer to an

[11] The discrepancy likely reflects the fact that many states mandate notification when a breach affects a certain number of residents, independent of the breach’s materiality under federal securities laws.

anonymous platform (e.g., a private chat) where the transaction can be completed.

Appendix B provides examples of dark web activity that is classified as *Blackmarkets*. We provide detailed examples because ours is the first paper, of which we are aware, to examine dark web activity and firm breaches. All the examples relate to a single disclosed breach involving a mortgage-related financial firm which we anonymize as “Financial Firm”. The examples in Appendix B illustrate the variety of platforms and types of data offered for sale from the breach.

Much of the sales activity that is classified as *Blackmarkets* comes from three types of dark web activity. The first common dark web sales activities relate to the sale of logs, also known as "cloud logs" or "stealer logs”. Logs have become one of the most common types of information sold and are often among the first information provided for sale on the dark web following a breach. Unlike traditional credential dumps, which contain static usernames and passwords, logs offer real-time access by including active session tokens and authentication data, allowing cybercriminals to bypass login security. These logs are typically harvested through infostealer malware, which extracts credentials and session data from infected devices.[12]

Unlike stolen credentials, which can trigger security alerts or require password resets, logs let attackers blend in with normal traffic, move across accounts undetected, and maintain access until the session expires or is manually revoked. This makes them particularly dangerous for companies, as attackers can stealthily extract data, escalate privileges, or deploy ransomware without setting off immediate alarms. Logs are sold as active access and are regularly refreshed with new stolen

[12] Infostealer malware is a type of malicious software designed to extract stored credentials, session cookies, authentication tokens, browser fingerprints, and other sensitive data from infected devices. Common infostealers, such as RedLine, Raccoon Stealer, and Vidar, silently run in the background, collecting login details from browsers, VPN clients, cryptocurrency wallets, and financial applications before transmitting the stolen data to a command-and-control (C2) server controlled by the attacker. These logs provide real-time access to compromised accounts, allowing attackers to bypass security measures like multi-factor authentication (MFA) by hijacking active sessions. Infostealers are often distributed via phishing emails, malicious attachments, drive-by downloads, and trojanized software, making them an efficient and scalable method for cybercriminals to gather data at scale that can be used to breach systems.

data, ensuring that buyers maintain continuous entry into compromised systems. Because logs are particularly time sensitive, they tend to be sold shortly after the breach, while the systems are still likely to be compromised and especially vulnerable. A surge in newly listed logs on underground marketplaces can signal recent successful infections, with "fresh" logs indicating that the stolen credentials or sessions are still valid and more likely to bypass detection or multi-factor authentication controls.

Logs are advertised and sold on dark web marketplaces, forums, and encrypted messaging groups, where sellers highlight key details such as the type of access available, the platform or service compromised, and any corporate or financial accounts included. Listings often specify the domain or company associated with the stolen credentials, allowing buyers to target specific firms, while sellers may provide sample data or redacted login details to verify authenticity before purchase. Logs are often labeled to indicate they contain recently stolen, active session data, ensuring buyers can access accounts before credentials are reset or sessions expire.

Examples 1 and 2 in Appendix B illustrate how a forum was used to notify potential buyers when new logs became available for the Financial Firm.[13] We focus our discussion on Example 1, which is from a Russian language forum known for advertising log files. The site is offering fresh logs from a breach and potential buyers can communicate with the seller on Jabber, an encrypted messaging protocol widely used by cybercriminals, with payment to be made in Bitcoin.[14] To mitigate concerns about trustworthiness, the seller indicates a willingness to escrow, typically through a neutral third party paid for by the buyer as a percentage of the transaction price. The

[13] These examples were accessed from links to still-existing locations provided in the CIH data. Since CIH scrapes the dark web as of a point in time, the associated links may no longer be accessible. We have masked potentially sensitive data in the examples such as names and other personal identifying information.

[14] Jabber is an instant messaging protocol based on XMPP (Extensible Messaging and Presence Protocol), which enables decentralized communication across a network of servers. Jabber allows users to connect via private, self-hosted servers and can be paired with Off-the-Record (OTR) encryption for enhanced security. Jabber is considered one of the most popular among Russian-speaking cybercriminals.

escrow agent holds the payment pending confirmation from the buyer.

A second common type of activity occurs on forums, message boards, and private messaging groups, typically involving specific breached information. These platforms are more limited in terms of potential buyers because they require both buyers and sellers to be connected within the same underground network groups which screen members rather than relying on an open marketplace. Because they are more secure, these platforms are particularly appealing to cybercriminals dealing in high-value or sensitive information. Transactions on these platforms often begin with general discussions, with specific details and negotiations shifting to more secure private chats, making monitoring more challenging. These platforms are more limited in terms of the range of potential buyers because participants are typically vetted before being granted access to sites. The advantage of these platforms is that these forums and groups lack a single point of failure, making them harder to infiltrate or shut down. CIH quietly joins these forums, messaging groups, and private groups to monitor activity.

A widely used platform for these exchanges is Telegram, due to its encryption and anonymity. Users have the ability to form private groups where cybercriminals can congregate to communicate and trade illicit data with reduced risk of detection. Example 3 in Appendix B provides an example of a message related to the Financial Firm breach. An identical version circulated across at least six private Telegram groups dedicated to buying and selling stolen data, with names like "Bank Lords", "Gang Market" and "FraudStars". The listing is from a cybercriminal who is advertising stolen information and notifying potential buyers that the seller has access to a specific firm's logins. Buyers interested in purchasing the credentials would move to a private messaging service to indicate interest and complete the transaction.

A third common medium for transaction offers is marketplace listings (loosely analogous to legitimate marketplaces like Amazon, Etsy, or eBay). Marketplace listings often offer stolen personal information related to a small group or single individuals, and mostly represent personal information being sold after the firm is compromised and the cybercriminal has exfiltrated all possible information. These listings typically include personally identifiable data such as names, account credentials, Social Security numbers, and bank account details that have been extracted in a breach. Appendix B, Example 4, provides listings from the breach of the Financial Firm for individual people and includes their full name, address, accounts, credit limits, and account balances. Upon purchase, the complete information is provided to the buyer. Market-based transactions are relatively structured and are attractive to sellers because they can attract a wide range of potential buyers. However, because these platforms are less restrictive in access and are more visible, they are also more susceptible to law enforcement action and may be shut down. For example, in 2023 the Genesis Market, a well-known illicit marketplace, was seized by the FBI in a coordinated international operation (U.S. Department of Justice, 2023).

While CIH tracks *Blackmarkets* activity, which indicates a specific offer to transact in a company's data, the dark web also hosts a wide range of activity beyond direct offers to sell access and information. CIH also tracks a broader category of discussion, *Databreach Information*, which, like *Blackmarkets*, tracks mentions of firms in dark web activity related to breaches. However, *Databreach Information* captures broader instances in which a firm is mentioned in posted leaks, discussions, or data dumps that are not directly linked to specific sale offers. For example, posters might discuss past leaks or post free versions of datasets from past breaches that others could find useful to aggregate with other data to create more complete profiles. While these discussions are identified by CIHs algorithms as being related to breaches, they are more general than specific offers and may be less timely indicators of breach activity.

While it might seem that data breach evidence on the dark web would be relatively rare, Figure 1 suggests otherwise. Figure 1 Panel A plots the frequency distribution of *Blackmarkets* and *Databreach Information* for all NYSE firms from 2020-2025. *Databreach Information* (discussion of breaches) is present for virtually all of our sample firms, while *Blackmarkets* (offers to transact) are present for more than 75% of our firms. While the proportions for *Blackmarkets* may seem high, they are consistent with the evidence noted earlier that 98% of organizations use at least one third-party vendor that has been breached in the last two years and 95% note a focus on cybersecurity risk in the risk oversight section of their proxy statements. Figure 1 Panel B shows the distribution by firm-week for our sample firms. *Blackmarkets* offers are present in 21.9% of weeks, although weekly numbers need not correspond to individual breaches since a single breach may lead to transactions spanning multiple weeks. Overall, our evidence suggests that dark web activity related to NYSE firms is widespread, with breach-related discussion and transaction offers frequently circulating across forums and marketplaces.

3.2 Building the Panel

To build our firm-week panel, we incorporate cyber breach disclosures from Audit Analytics for our sample of NYSE firms. Audit Analytics collects disclosures of cyber breaches from SEC filings, state documents, press releases and other sources. Table 1, Panel A reports the number of disclosed breaches per firm in our sample of firms with disclosed breaches. Most sample firms did not disclose a breach during the sample period. Of those that did, 131 only disclosed one breach while 32 disclosed more than one breach. Table 1 Panel B reports breaches by disclosure source. Only 44 of the 215 disclosures initially took place through SEC filings, of which 29 were "timely filings" through Form 8-K as opposed to periodic quarterly or annual filings (untabulated). Most of the disclosures (127) occurred through filings with State Attorneys General, while 21 were through news reports and 23 were through other outlets. Even counting cases in which initial disclosure was through other sources, only a total of 59 of the 215 total were eventually included in SEC filings.

Table 2, reports descriptive variables by firm-week, including capital markets variables computed over two-year windows around disclosed breaches and shadow breaches to match our empirical specifications. We use three measures to examine information asymmetry and informed trading, averaged at the firm-week level. Variable definitions are included in Appendix A. From CRSP we calculate *Amihud Illiquidity* (Amihud, 2002), the average daily absolute stock return divided by trading volume, multiplied by one million. Table 2, Panel A shows the average value of *Amihud Illiquidity* is 0.172. Higher values suggest reduced liquidity, as a proxy for the price impact of informed trades. We also examine the *Bid-Ask Spread*, calculated as the difference between the closing ask and bid prices, divided by their midpoint and multiplied by one thousand. The average *Bid-Ask Spread* in our sample is 2.780, where higher values suggest increased trading costs and greater information asymmetry. Lastly, we construct the Conditional Probability of an Information Event (*CPIE OWR*) following Odders-White and Ready (2008). *CPIE OWR* reflects the model-implied probability that private information arrives on a given trading day, based on both order imbalance and price movements. Among the models they evaluate, Duarte et al. (2019) find that *CPIE OWR* performs particularly well in identifying economically meaningful information events and exhibits limited mechanical correlation with trading volume.[15] *CPIE OWR*

[15] We estimate the structural OWR model following Odders-White and Ready (2008), which identifies private information arrival using both order imbalance and returns. The model incorporates daily intra-day and overnight returns, along with order flow imbalance, to generate a time-varying measure of the conditional probability of private information arrival. This probability, known as the Conditional Probability of an Information Event (CPIE), reflects the likelihood that an asset's trading activity on a given day is driven by private information. Following Duarte, Hu, and Young (2019), we estimate the model parameters at the stock-year level using all available trading days in a given calendar year and then apply those parameters to compute daily CPIE values within the same year. This approach allows us to construct a daily measure of information risk while holding model parameters fixed within each annual

ranges from 0 to 1, with an average of 0.393, with higher values suggesting a greater likelihood of private information.

Table 2 Panel A shows for the Disclosed Breaches *Blackmarkets* has an average value of 12, and the 75th percentile is 5. This pattern aligns with Figure 1, where *Blackmarkets* activity clusters in a minority of firm weeks, suggesting that breaches are episodic rather than continuous, reflecting concentrated periods of heightened illicit trade rather than a steady flow of transactions. As illustrated in our earlier example, an individual breach typically results in multiple offers to sell across a range of platforms so, for example, the 5 offers at the 75th percentile may well be linked to a single breach. We winsorize continuous variables at the 1st and 99th percentiles and count variables at the 99th percentile (since they are bounded below by zero). In our empirical analysis, we focus on *Blackmarkets* because our interest is in offers to transact, which generally precede broader discussions of dark web activity that may be triggered by data breach disclosure.

4. Empirical approach and results

Our analyses fall into two broad categories. First, we examine patterns in dark web activity and capital markets behavior around breaches that are eventually disclosed. Second, we use dark web activity to identify likely breaches that are never formally disclosed ("shadow breaches") and examine if they provide information to capital markets.

4.1 Cyber Breach Disclosure

Our analysis of disclosed cyber breaches serves three main purposes. First, it provides a validation of dark web activity as a signal of breach activity. Because we know when these breaches were eventually disclosed, we can examine whether dark web activity rises beforehand, offering evidence on both the presence and timing of breach-related signals on the dark web.

estimation window. Our implementation closely follows the publicly available code from Duarte, Hu, and Young (2019), available at https://github.com/edwinhu/pin-code (accessed February 14, 2025).

Second, by comparing the timing of dark web activity with disclosure dates, we can estimate how long it takes firms to detect breaches and initiate formal reporting. Third, given the substantial lag between dark web activity and public disclosure, we can compare the capital market's response to the dark web activity relative to the eventual disclosure.

Figure 2 plots *Blackmarkets* trends for the two-year window around the cyber breach disclosure date, orthogonalized with firm fixed effects to control for background levels of dark web activity. There is clear evidence of an increase in *Blackmarkets* activity starting about six months before the breach disclosure date and it remains elevated until about one month before the disclosure. It is difficult to interpret the lags in Figure 2 as firm-level averages because the graph pools dark web activity across firms. As an alternative, we identify the peak week of *Blackmarkets* activity before disclosure for each firm and compute the gap to the disclosure date. The mean lag is 17.4 weeks, with the median firm disclosing 16 weeks after the peak in dark web activity, consistent with the aggregate pattern in Figure 2.[16] Overall, the results suggest that *Blackmarkets* activity data reflect cyber breach activity, and the timing of this activity varies substantially across breaches. Moreover, the fact that elevated activity occurs six months on average before disclosure points to a significant lag between dark web activity and disclosure, consistent with the SEC's concerns about the timeliness of breach reporting.

The preceding analyses suggest that dark web transaction activity often precedes disclosed breaches, consistent with offers to sell firm data emerging soon after a breach occurs. To more formally test this pattern, we estimate whether there is a significant increase in *Blackmarkets* activity during the 24 weeks leading up to a firm's breach disclosure, relative to the two-year window around a breach. We include firm fixed effects to account for average levels of dark web

[16] We restrict the sample to firms with a unique weekly maximum for *Blackmarkets* (i.e., no ties across weeks) and require that the peak value exceeds one. Applying these criteria yields a sample of 148 breaches.

activity for the firm, and year-week fixed effects to control for common shocks affecting all firms in a given week. This structure ensures that the estimated effects reflect changes in a firm's dark web activity relative to its own historical baseline and relative to overall dark web activity. We use fixed-effects Poisson regressions because the dependent variables are counts, which are non-negative and right-skewed. As noted by Cohn, Liu, and Wardlaw (2022), Poisson models are appropriate for count data and provide consistent estimates without requiring strong assumptions about the distribution of the errors. Results (untabulated) are robust when using OLS instead of Poisson regression. We cluster standard errors at the firm level to account for within-firm autocorrelation. Table 3, Panel A, presents the results. *Blackmarkets* activity increases significantly during the 24 weeks prior to a breach disclosure, with a coefficient of 0.304 ($p <$ 0.05). This finding confirms that offers to sell firm data on the dark web often emerge before formal disclosure. In terms of economic magnitude, the estimated effect in Column (1) implies an increase in *Blackmarkets* activity of roughly 36 percent during the pre-disclosure period, relative to a firm’s own baseline (calculated as: $36\% = (e^{0.304} - 1) * 100)$).

Next, we conduct a similar analysis focused on the period immediately following a breach disclosure. We define a *Post Disclosure* indicator equal to one for the eight weeks after the public announcement of a breach and estimate a regression using the same two-year window centered on the disclosure date, with firm and year-week fixed effects included, using a fixed-effects Poisson model. In Table 3, Panel B, the coefficient on *Post Disclosure* is –0.074 (p = 0.62), indicating no statistically significant change in *Blackmarkets* activity after disclosure. Consistent with Figure 2, these results suggest that dark web transaction activity increases substantially before cyber breach disclosure and then reverts to normal levels following disclosure, likely because the data loses its value after detection.

The increase in *Blackmarkets* prior to a disclosed breach suggests that dark web transaction activity may serve as an early signal of breach events. We next examine whether capital markets respond to the dark web activity that occurs prior to breach disclosure. Specifically, we test whether there are higher levels of *Bid-Ask Spread*, *Amihud Illiquidity*, and *CPIE OWR* during weeks in which dark web activity is elevated prior to the breach disclosure. For each breach, we define the *Blackmarkets Period* as equal to one during the four-week window surrounding the peak in dark web activity. To avoid contaminating dark web activity with breach disclosure, we exclude observations where the dark web activity window overlaps with the disclosure date. We include firm-breach fixed effects to focus on within-breach variation, comparing capital market behavior during the weeks of peak dark web transaction activity to other pre-disclosure weeks for the same firm. Because the window precedes public disclosure of the breaches, the capital markets activity reflects the awareness by capital markets participants of breach activity prior to the breach disclosure (e.g., they observe the dark web activity directly or are aware of the breach through other nonpublic sources).

Table 4 reports the association between dark web activity and capital market measures in the year leading up to a breach. In Column (1), where the dependent variable is *Bid–Ask Spread*, the coefficient on *Blackmarkets Period* is significant and positive, 0.068 ($p < 0.05$), corresponding to a 4.9% increase relative to the within-breach average in non-indicator weeks.[17] In Column (2), where the dependent variable is *Amihud Illiquidity*, the coefficient is 0.0075 ($p < 0.05$), an 18.6% increase relative to the within-breach average, consistent with greater price impact and reduced liquidity. In Column (3), where the dependent variable is *CPIE OWR*, the coefficient is 0.0084 (p

[17] Economic magnitudes are calculated by dividing the coefficient on *Blackmarkets Indicator* by the regression constant from each specification. Since the regressions include breach fixed effects, the constant captures the mean of the dependent variable during non-indicator weeks within each breach. This scaling expresses the estimated effect as a percentage change relative to the within-breach baseline.

< 0.05), representing a 2.3% increase from the within-breach average, a magnitude consistent with known insider trades (Duarte et al., 2019).[18]

Taken together, the results suggest that dark web data can reveal breach activity well before public disclosure, and that capital markets begin to respond during this pre-disclosure period. Although the timing of dark web activity varies across breaches, it typically precedes disclosure by a meaningful margin. These findings align with SEC concerns with delays and lack of timing detail in breach disclosures and suggest that the dark web may offer a potential alternative signal.

4.2 Shadow Breaches

Thus far, we have examined cyber breaches publicly disclosed by firms. However, it is possible that some cyber breaches reflected in dark web activity remain undisclosed. In particular, the earlier anecdotes suggest that although breach disclosures are relatively rare, most firms have been exposed to breach activity either directly or through breaches at third party vendors. Breaches may go undisclosed because the firm was unaware of the incident, determined it was not material under the SEC's or State Attorney General's definitions, or made a strategic decision to withhold disclosure. We next examine whether patterns in dark web activity indicate breaches that were not disclosed, and whether these events are reflected in capital markets.

We use the preceding evidence from disclosing firms, which shows that increases in *Blackmarkets* activity can signal periods of active breach, to identify potential breaches among firms that do not disclose. In particular, we use extended abnormal dark web transaction activity to construct a measure of "shadow breaches", suggesting periods in which breaches were likely but were not disclosed. We define a period as abnormal when the within-firm standardized

[18] The magnitude of increase in *CPIE OWR* is comparable to that in Figure 9 of Duarte, Hu, and Young (2020), which shows that *CPIE OWR* rises by about 0.006 (≈1.6%) around known insider trades, increasing from roughly 0.364 to 0.370 in event time.

Blackmarkets value exceeds three standard deviations above the firm mean in at least three of four consecutive weeks. We treat the first of these consecutive weeks as the start of a shadow breach and exclude any cases that occur within one year of a disclosed breach.[19] This process yields 245 shadow breaches, slightly more than the number of disclosed breaches.

Descriptively, the industry distribution for firms with shadow breaches is similar to that for disclosed breaches based on the Fama-French 12 classification. The top three industries are Money and Finance, Other Industries, and Retail/Wholesale/Services, which together account for the majority of breaches in both groups. Money and Finance represent the largest share, with 26% of shadow breaches and 24% of disclosed breaches. Other Industries comprise 15% and 18%, respectively, and Retail/Wholesale/Services account for 10% and 12%.

In Figure 3, we plot *Blackmarkets* activity, orthogonalized using firm fixed effects, relative to the timing of each shadow breach. Unlike the disclosed breaches in Figure 2, we cannot align these events to formal disclosure dates because no disclosure was made. In general, *Blackmarkets* activity follows a similar descriptive pattern relative to disclosed breaches in Figure 2. Dark web activity rises over several months, peaks, and then drops quickly.

The preceding evidence suggests that shadow breaches resemble disclosed breaches in their dark web transaction patterns. However, it is unclear whether these breaches are associated with capital market activity. Recall our earlier findings that market participants respond to dark web transaction activity before formal disclosure. If shadow breaches reflect meaningful, though undisclosed, events, we would expect to see similar patterns in capital market behavior around the time of the elevated dark web activity, even without subsequent disclosure.

[19] Results are robust to other thresholds such as one, two and three consecutive weeks. We focus on a more conservative threshold to balance identifying activity that is truly abnormal with ensuring a reasonable sample size. Results are also robust to alternative event windows around actual disclosures, including [-10, +10], [-20, +20], [-25, 0], and [-52, 0].

To provide initial descriptive evidence, Figure 6 plots average trends in *Bid-Ask Spread, Amihud Illiquidity* and *CPIE OWR* relative to the start of a shadow breach, orthogonalized with firm fixed effects. *CPIE OWR* begins to rise first, consistent with an increase in informed trading activity. *Amihud Illiquidity* increases shortly thereafter, indicating that trades begin to move prices more as liquidity becomes more limited. *Bid-Ask Spread* also increases, suggesting that quotes widen, likely due to increased uncertainty. The sequence of responses across these measures is consistent with private information influencing informed trading and liquidity at the firm level.

To formally test the capital markets activity variables, we define the *Shadow Breach Period* as an indicator for four weeks after the start of the shadow breach.[20] For comparison, we also construct a stricter measure of "high-intensity" shadow breaches, which are characterized by four (rather than three) consecutive weeks of abnormal *Blackmarket* activity. This stricter criterion reduces the number of identified breaches from 245 to 148. *Shadow Breach Period – High Intensity* is an indicator for the period associated with these high-intensity shadow breaches. We examine a two-year window surrounding the start of each shadow breach and estimate regressions using *Bid-Ask Spread*, *Amihud Illiquidity*, and *CPIE OWR* as dependent variables. We include fixed effects for each firm-shadow breach to control for firm and breach-specific factors, allowing us to test for whether market behavior changes in response to abnormal dark web activity controlling for normal levels of informed trading and liquidity in the period around the breach episode.[21]

Table 5, Panel A shows that the coefficient on *Shadow Breach Period* is positive and statistically significant across all three measures: 0.332 ($p < 0.05$) for *Bid-Ask Spread*, 0.043 ($p <$

[20] Results are robust if the window is shortened to one month.

[21] This design mirrors our earlier analysis that correlates capital market activity with *Blackmarkets* prior to disclosure, with two differences. First, rather than limiting the window to the year before disclosure, we include both the year before and after the shadow breach. Since shadow breaches are undisclosed, there are no concerns about confounding effects from the disclosure itself. Second, instead of using the single-week maximum to define abnormal dark web activity (*Blackmarkets Period*), we use *Shadow Breach Period*, which applies a more conservative threshold based on sustained abnormal *Blackmarkets* activity over multiple consecutive weeks.

0.05) for *Amihud Illiquidity*, and 0.018 (p < 0.01) for *CPIE OWR*. The high-intensity shadow breach coefficient estimates in Panel B are larger for *Bid-Ask Spread* (0.836; p < 0.01) and *Amihud Illiquidity* (0.090; p < 0.05), although not for *CPIE OWR* (0.008; p < 0.10). These results are consistent with the idea that abnormal dark web sales activity is associated with greater trading frictions, reduced liquidity, and increased informed trading risk, even absent formal disclosure. Based on the standard deviation of the dependent variables, the estimated effects are economically meaningful. The *Bid–Ask Spread* coefficient (0.332) corresponds to an 11.9% increase, the *Amihud Illiquidity* coefficient (0.043) corresponds to a 17.2% increase, and the CPIE OWR coefficient (0.018) corresponds to a 4.6% increase. Overall, the evidence on shadow breaches suggests that, similar to disclosed breaches, sustained abnormal dark web activity is associated with increased illiquidity, information asymmetry and informed trading.

4.3 Shadow Breaches and Stock Returns

Despite the potential negative monetary and reputational costs of cybercrime, prior research finds that market reactions to cyber breach disclosure are typically small and often insignificant. These mixed findings have been attributed to sample size limitations, selection bias, variation in breach types, and differences in the information environment at the time of disclosure (e.g., Amir et al. (2018), Foerderer and Schuetz (2022), Kamiya et al. (2021), and Yayla and Hu (2011). Our previous findings suggest that informed trading and illiquidity respond in advance of cyber breach disclosures, particularly in periods of increased dark web transaction activity. This pattern suggests that investors may become aware of breaches before they are publicly disclosed. If this is the case, dark web activity may serve as an information source even absent formal disclosure by firms.

To assess whether investors view elevated dark web activity as an indicator of a potential breach, we examine market-adjusted stock returns around *Shadow Breach* dates. Abnormal returns are estimated over a 100-day window ending 70 trading days before the breach, and cumulative abnormal returns (CARs) are computed over the 35-day (5-week) period following the breach onset.[22] Figure 4 Panel A presents the results, splitting the sample between the full *Shadow Breach* sample and *Shadow Breach-High Intensity* subsample. The results indicate substantial negative returns following shadow breaches for both samples, with more pronounced negative returns for the *Shadow Breach-High Intensity* subsamples. Table 6, Panel A reports statistical tests for the CARs. The CAR for the overall sample of *Shadow Breach* is -3.53%, while the CAR for the *Shadow Breach-High Intensity* subsample is -5.82%; both are statistically significant ($p < .01$).[23]

4.4 Disclosed Breaches and Stock Returns

A possible implication of the preceding test is that at least part of the reason prior research finds limited evidence of negative returns associated with cyber breach disclosure dates may be because the lag between the breach and disclosure means that the market has already impounded much of the negative news based on dark web activity prior to disclosure. To examine that possibility, we compare two return windows around alternative event dates for firms disclosing cyber breaches. First, we consider the actual reported cyber breach disclosure date using the disclosure date reported by Audit Analytics as the first public announcement of the breach. Second, we compare the actual breach disclosure date to an "*Adjusted Breach Date*" based on the dark market activity. We use the same algorithm we used to determine *Shadow Breach* to infer the timing of the dark web activity preceding cyber breach disclosure and adjust the event date to

[22] Market-adjusted returns are benchmarked against the CRSP equally weighted index. Results are qualitatively unchanged when we drop the top 5% or 10% of events with the most negative CARs; extend the estimation window to 150 or 200 trading days (or shift its end to 30 days before the event date); or use a simple market model or Fama–French three-factor model, either with or without a momentum factor, to estimate abnormal returns.

[23] While the returns trend appears to begin after 2 weeks, recall that the *Shadow Breach* measure is computed over a four-week window and Week 0 is the beginning of the window. As a result, the pattern is consistent with investors learning the seriousness of the breach over the *Shadow Breach* measurement window. Put another way, if the returns window started in the middle of the *Shadow Breach* window, the return reaction would be almost immediate.

reflect the first week of the shadow breach period. We limit our sample of disclosed breaches to 81 firms with enough black web activity to trigger the *Shadow Breach* criteria.[24] On average, the *Shadow Breach* dark web activity begins 29 weeks before the cyber breach is disclosed (untabulated), which aligns with the IBM data security report estimate that it takes about 277 days on average to identify and report a breach (CYFOR Secure, n.d.; IBM, 2022).

Figure 4, Panel B plots CARs for the disclosing firms around the cyber breach disclosure date. As with prior literature, the returns are negative but muted. Table 6, Panel B presents statistical tests for the CARs associated with the *Disclosed Breach Date*. The mean CAR following disclosure is -0.75% ($p > 0.10$) consistent with prior research suggesting a muted stock price response associated with cyberbreach disclosure dates.

Figure 4, Panel B also plots CARs for the *Adjusted Breach Date* based on dark web activity. The negative returns following the *Adjusted Breach Date* are substantially larger than for the *Disclosed Breach Date*. The results are tabulated in Table 6, Panel B. The CAR for the *Adjusted Breach Date* based on the timing of dark web activity totals -5.25% ($p < 0.05$), suggesting a significant negative response of similar magnitude to that following a *Shadow Breach* but much larger than that following the *Disclosed Breach Date*. Taken together, the results suggest that the delay in cyber breach disclosure means that much of the information is incorporated into share price before the announcement, consistent with the SEC's concerns over the timeliness of cyberbreach disclosure. Coupled with the earlier analyses, the results suggest that capital markets respond similarly in terms of informed trading, illiquidity and negative returns to elevated dark web transaction activity irrespective of whether or not a breach is ultimately disclosed. However,

[24] In untabulated tests, we relax the threshold to include cases with only one week of abnormal activity, yielding 148 events with an average adjustment of 21 weeks—consistent with Figure 3. The 5-week abnormal CARs remain significantly negative, and our main inferences remain unchanged.

the negative returns are concentrated around periods of elevated dark web transaction activity rather than when a breach is actually disclosed, consistent with dark web market activity pre-empting the disclosure.

4.4 Why do firms not disclose breaches?

In our final analysis, we examine cross-sectional variation in the decision to disclose a breach, recognizing that the evidence is descriptive rather than causal. Our previous analyses suggest that disclosed breaches are similar to shadow breaches in terms of patterns of dark web activity and capital market responses. That raises the question of what factors might differentiate between disclosed breaches and shadow breaches. There are several reasons a firm might not disclose a breach (e.g., lack of knowledge, lack of materiality, or strategic reasons). One possibility is that some firms may not monitor the dark web and, therefore, may not be aware of elevated dark web activity suggesting breaches. Limited monitoring of the dark web aligns with concerns that firms underinvest in cybersecurity relative to its risks (Blau, 2017; Gordon et al., 2014).

To examine factors differentiating between disclosed and shadow breaches, we construct a panel of firms that: (1) disclosed a breach or (2) were classified as having a shadow breach. We define an indicator variable, *Disclosed*, equal to one if the breach was publicly disclosed and zero otherwise. We also control for firm characteristics measured in the quarter of the breach: *Leverage*, *Assets*, *Book to Market*, *Cash Holdings*, and *Q* (Tobin's Q). Our primary independent variable of interest is *Cyber Monitoring 10-K*, an indicator equal to one if the firm reported engaging in dark web monitoring or broader cyber risk oversight in its 10-K filing prior to the breach.[25] We also

[25] We classify firms as disclosing cyber monitoring activity in their 10-K filings if the text contains references to terms associated with dark web surveillance (e.g., "dark web," "darknet," "onion sites," "Tor network," "hacker forums") or broader cybersecurity monitoring practices (e.g., "cyber monitoring," "threat intelligence," "security operations center," "SIEM systems," "endpoint security," "anomaly detection," "data leak monitoring," and "compromised account monitoring"). These keywords are precompiled using regular expressions to capture minor linguistic variations (e.g., plural forms, hyphenation, or spacing).

include a variable that measures the extent of elevated dark web activity, *Databreach Information*, to control for the intensity of the firm's exposure to dark web activity, and indicator variables for whether the firm belongs to a technology or financial industry since those industries may be particularly sophisticated in monitoring cyber risk.[26]

Results are presented in Table 6. In Column (1), which includes only *Cyber Monitoring 10-K* without controls, we find a positive and statistically significant association with cyber breach disclosure. In Column (2), we include additional firm-level controls. The estimated effect remains of similar magnitude and statistical significance with or without controls. In terms of economic significance, the predicted probability of disclosure is 14.2% for firms that do not mention cyber monitoring in their 10-K filings versus 41.9% for those that do, a difference of almost 28 percentage points. *Assets* are also positively associated with disclosure, suggesting that larger firms are more likely to disclose breaches. *Tech Firm* is also positively associated with disclosure (p < 0.10), and *Leverage* is significant at the 10% level, perhaps consistent with heightened external monitoring (Huang and Wang, 2021). We emphasize that the evidence is intended to be descriptive but not causal because the choice of whether to monitor the dark web is endogenous. That said, the results suggest that, for a given level of dark web activity, firms that disclose that they track the dark web as part of their cyber risk mitigation strategy are substantially more likely to disclose cyber breaches.

5. Conclusion

Our results suggest that elevated dark web activity often precedes a firm's disclosure of known

[26] Financial firms are identified as those with SIC codes between 6000 and 6999. Technology firms are classified using the Fama-French 12 industry definition, which includes firms in the following SIC codes and sectors: 3570–3579 (Computer and Office Equipment), 3660–3669 (Communications Equipment), 3670–3679 (Electronic Components and Accessories), 3810–3819 (Engineering, Laboratory, and Scientific Instruments), 7370–7379 (Computer Programming and Data Processing), 4810–4819 (Telephone Communications), 4820–4829 (Telegraph and Other Communications), 4830–4839 (Radio and TV Broadcasting), and 4890–4899 (Other Communication Services). Results are robust to alternative classifications used in prior work (e.g., Loughran and Ritter, 2004).

breaches and may, in some cases, be useful in identifying “shadow breaches” that remain unreported. We document correlations between capital market measures of information asymmetry and dark web transaction activity, suggesting that elevated dark web activity is associated with capital market outcomes before the disclosure of a breach. Further, the associations between elevated dark web activity, informed trading and illiquidity are consistent even in cases of shadow breaches in which there is no public disclosure of a breach.

We document negative abnormal stock returns associated with elevated dark web activity for both disclosed and shadow breaches. For disclosed breaches, comparing the actual disclosure date to an estimated breach date based on dark web transaction activity, we document a significant negative market response around the estimated breach date, but not the disclosure date. Lastly, our determinants analysis suggests that firms actively monitoring the dark web for cyber risk are more likely to disclose breaches.

Our analyses support the potential usefulness of dark web data for researchers, firms, investors and regulators in understanding and responding to the increasing prevalence of cybercrime. From a regulatory perspective, these results support the SEC’s recent push for more timely cyber breach disclosure and emphasize the potential utility of dark web intelligence to complement firms’ internal controls and traditional cybersecurity defenses. More broadly, while we focus on corporate cyber breaches, our results suggest the potential usefulness of dark web data in investigating a wide range of illegal activity.

References

Akey, P., Grégoire, V., & Martineau, C. (2022). Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics*, 143(3), 1162–1184.

Ali, A., Li, N., & Zhang, W. (2019). Restrictions on managers' outside employment opportunities and asymmetric disclosure of bad versus good news. *The Accounting Review*, 94(5), 1–25.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31-56.

Amir, E., Levi, S., & Livne, T. (2018). Do firms underreport information on cyber-attacks? Evidence from capital markets. *Review of Accounting Studies*, 23, 1177-1206.

Amir, E., Levi, S., & Livne, T. (2019). Insider trading and disclosure: The case of cyberattacks. Available at SSRN 3355978.

Ashraf, Musaib. (2022). The Role of Peer Events in Corporate Governance: Evidence from Data Breaches. *The Accounting Review* 97(2): 1–24.

Blau, A. (2017). The behavioral economics of why executives underinvest in cybersecurity. *Harvard Business Review*. https://hbr.org/2017/06/the-behavioral-economics-of-why-executives-underinvest-in-cybersecurity

Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security*, 11(3), 431–448.

Chen, J., Henry, E., & Jiang, X. (2023). Is cybersecurity risk factor disclosure informative? Evidence from disclosures following a data breach. *Journal of Business Ethics*, *187*(1), 199-224.

Chen, Xi, Gilles Hilary, and Xiaoli Tian. "Mandatory Data Breach Disclosure and Insider Trading." *Journal of Business Finance & Accounting* (2024).

Cohn, J. B., Liu, Z., & Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, *146*(2), 529-551.

Crosignani, Matteo, Marco Macchiavelli, and Antonio F. Silva. 2023. Pirates without Borders: The Propagation of Cyberattacks through Firms' Supply Chains. *Journal of Financial Economics* 147(2): 432–448.

Cyentia Institute and SecurityScorecard Research. (2023).

CYFOR Secure. (n.d.). How long does it take to detect a cyber attack*? CYFOR Secure*. Last accessed March 6, 2025, from https://cyforsecure.co.uk/how-long-does-it-take-to-detect-a-cyber-attack/

Duarte, J., Hu, E., & Young, L. (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics*, 135(3), 795–815. https://doi.org/10.1016/j.jfineco.2019.08.005

Easley, D., Kiefer, N. M., O'Hara, M., & Paperman, J. B. (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance*, 51(4), 1405–1436.

Ernst & Young. (2022). How cyber governance and disclosures are closing the gaps in 2022. *EY Center for Board Matters*. https://www.ey.com/content/dam/ey-unified-site/ey-com/en-ca/campaigns/board-matters/documents/ey-how-cyber-gov-and-disclosures-are-closing-gaps-in-2022.pdf

Florackis, Chris, Christodoulos Louca, Roni Michaely, and Michael Weber. 2023. Cybersecurity Risk. *The Review of Financial Studies* 36(1): 351–407.

Foerderer, J., & Schuetz, S. W. (2022). Data breach announcements and stock market reactions: A matter of timing? *Management Science*, 68(10), 7298–7322.

Foley, Sean, Jonathan R. Karlsen, and Tālis J. Putniņš. 2019. Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies? *The Review of Financial Studies* 32(5): 1798–1853.

Gordon, L. A., Loeb, M. P., & Zhou, L. (2016). Investing in cybersecurity: Insights from the Gordon-Loeb model. *Journal of Information Security*, *7*(02), 49.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., & Zhou, L. (2014). Externalities and the magnitude of cyber security underinvestment by private sector firms: a modification of the Gordon-Loeb model. *Journal of Information Security*, *6*(1), 24-30.

Goel, S., & Shawky, H. A. (2009). Estimating the market impact of security breach announcements on firm values. *Information & Management*, 46(7), 404–410. https://doi.org/10.1016/j.im.2009.06.005

Hernandez, J. C. (2024). Preventing credential stuffing: Strategies and tips. *Prey Project*. Last accessed March 6, 2025, from https://preyproject.com/blog/credential-stuffing-attacks

Huang, H. H., & Wang, C. (2021). Do banks price firms' data breaches?. *The Accounting Review*, *96*(3), 261-286.

IBM. (2022, July 27). *IBM report: Consumers pay the price as data breach costs reach all-time high. IBM Newsroom.* Last accessed March 6, 2025, from https://newsroom.ibm.com/2022-07-27-IBM-Report-Consumers-Pay-the-Price-as-Data-Breach-Costs-Reach-All-Time-High

Kamiya, S., Kang, J. K., Kim, J., Milidonis, A., & Stulz, R. M. (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, *139*(3), 719-749.

Kannan, K., Rees, J., & Sridhar, S. (2007). Market reactions to information security breach announcements: An empirical analysis. *International Journal of Electronic Commerce*, 12(1), 69–91.

Kolanovic, M., & Smith, R. (2019). 2019 Alternative Data Handbook. *J.P. Morgan Big Data and AI Strategies*, October 2019

Kron. (2022, May 29). Cyber attacks throw one-fifth of businesses in Europe and the US into bankruptcy. Retrieved from https://www.kron.com/blog/cyber-attacks-throw-one-fifth-of-businesses-into-bankruptcy

Kvochko, E., & Pant, R. (2015). Why data breaches don't hurt stock prices. *Harvard Business Review*.

Lin, Z., Sapp, T. R., Ulmer, J. R., & Parsa, R. (2020). Insider trading ahead of cyber breach announcements. *Journal of Financial Markets*, 50, 100527.

Mitts, J., & Talley, E. (2019). Informed trading and cybersecurity breaches. Harvard Business Law Review, 9, 1.

Morgan, S. (2024, February 5). Top 10 cybersecurity predictions and statistics for 2024. *Cybersecurity Ventures.* Retrieved January 13, 2025, from https://cybersecurityventures.com/research/

Motoki, F. & Pinto, J. (2024). Regulating data: Evidence from corporate America. *Journal of Business Finance & Accounting*, 52(1), 541-568.

Odders-White, E. R., & Ready, M. J. (2008). The probability and magnitude of information events. *Journal of Financial Economics*, *87*(1), 227-248.

Piccotti, L. R., & Wang, H. (2023). Informed trading in the options market surrounding data breaches. *Global Finance Journal*, 56, 100774.

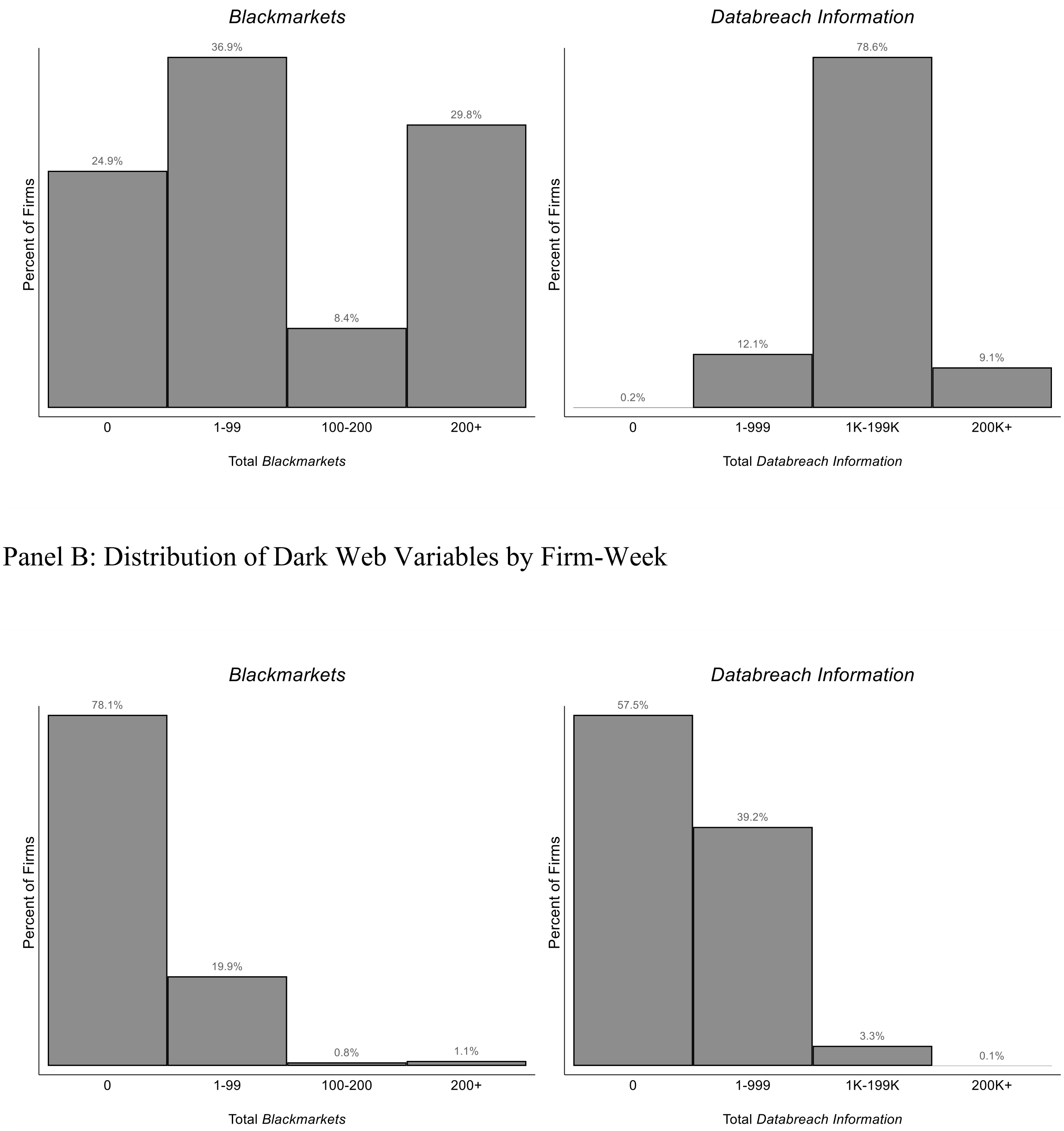
Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 39(4), 1127–1139. https://doi.org/10.1111/j.1540-6261.1984.tb03897.

Securities and Exchange Commission. (2023). *Cybersecurity risk management, strategy, governance, and incident disclosure* (Release No. 33-11216; 34-97989). https://www.sec.gov/files/rules/final/2023/33-11216.pdf

U.S. Department of Justice. (2023). Criminal marketplace disrupted in international cyber operation. *U.S. Department of Justice.* Last accessed March 6, 2025, from https://www.justice.gov/archives/opa/pr/criminal-marketplace-disrupted-international-cyber-operation

World Economic Forum.(2019). *The Global Risks Report 2019* (14th ed.). https://www.weforum.org/publications/the-global-risks-report-2019/

Figure 1. Distribution of Dark Web Variables for NYSE Firms

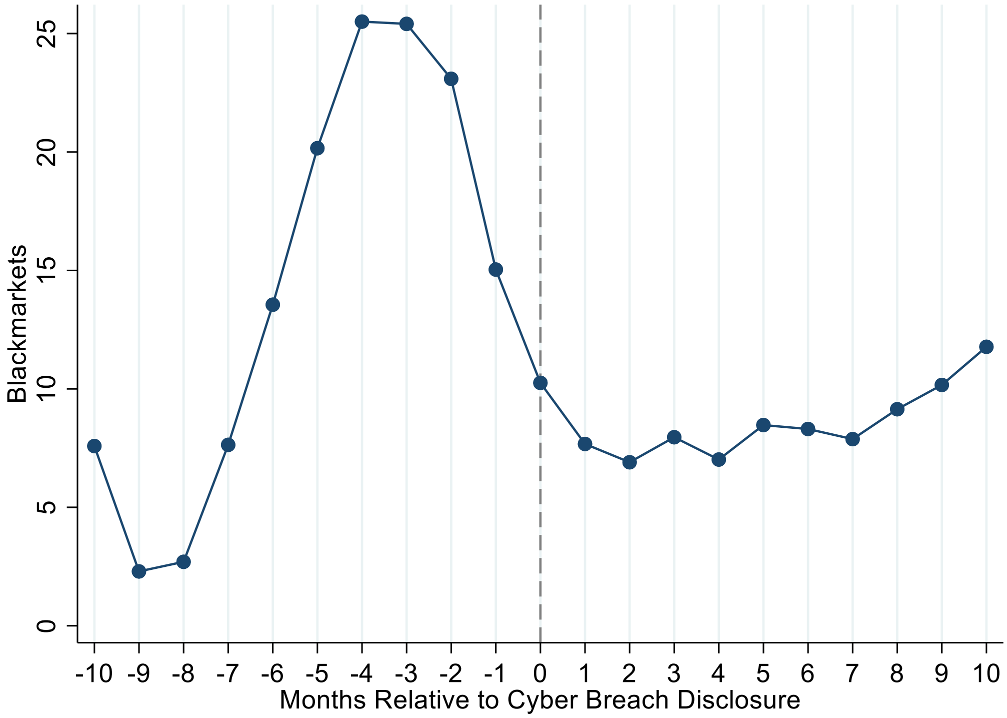


Panel A: Distribution of Dark Web Variables by Firm

Panel B: Distribution of Dark Web Variables by Firm-Week

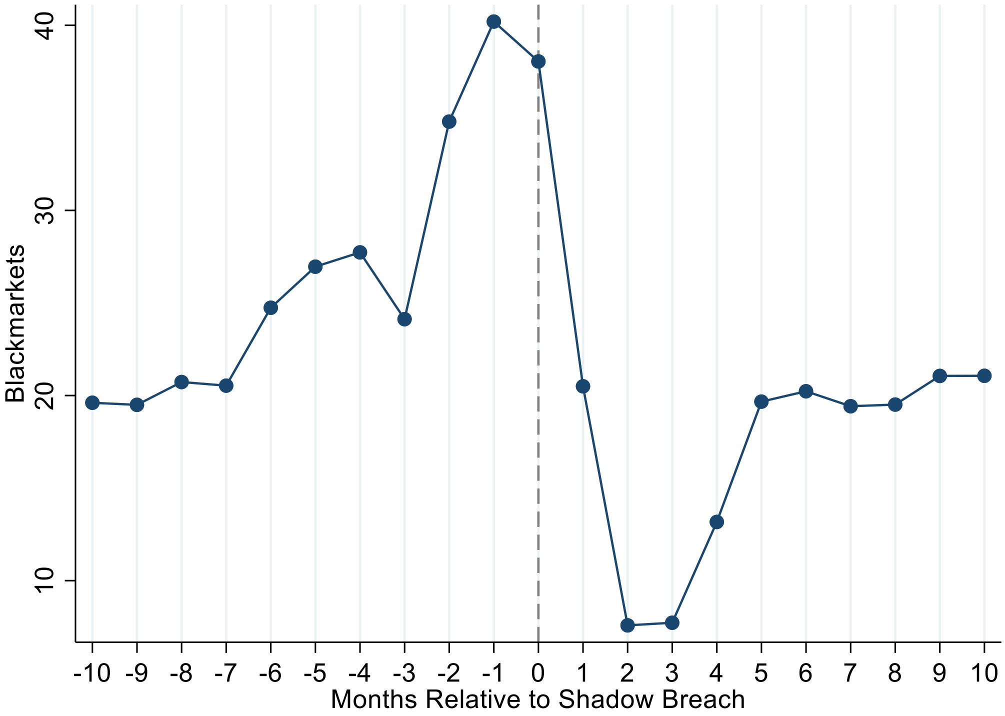
Notes: This figure shows the distribution of *Blackmarkets* and *Databreach Information* for NYSE firms from 2020-2025. Panel A presents the firm-level distribution, while Panel B displays the firm-week distribution of these dark web measures.

Figure 2. *Blackmarkets* Around Cyber Breach Disclosure



Notes: Figure 2 shows the monthly trends in *Blackmarkets* activity, orthogonalized using firm fixed effects, in the 20-month window around a disclosed cyber breach. Trends are plotted for 215 disclosed breaches among NYSE-listed firms in our sample from 2020 to 2025. *Blackmarkets* activity refers to the frequency of dark web activity related to the sale of firm information.

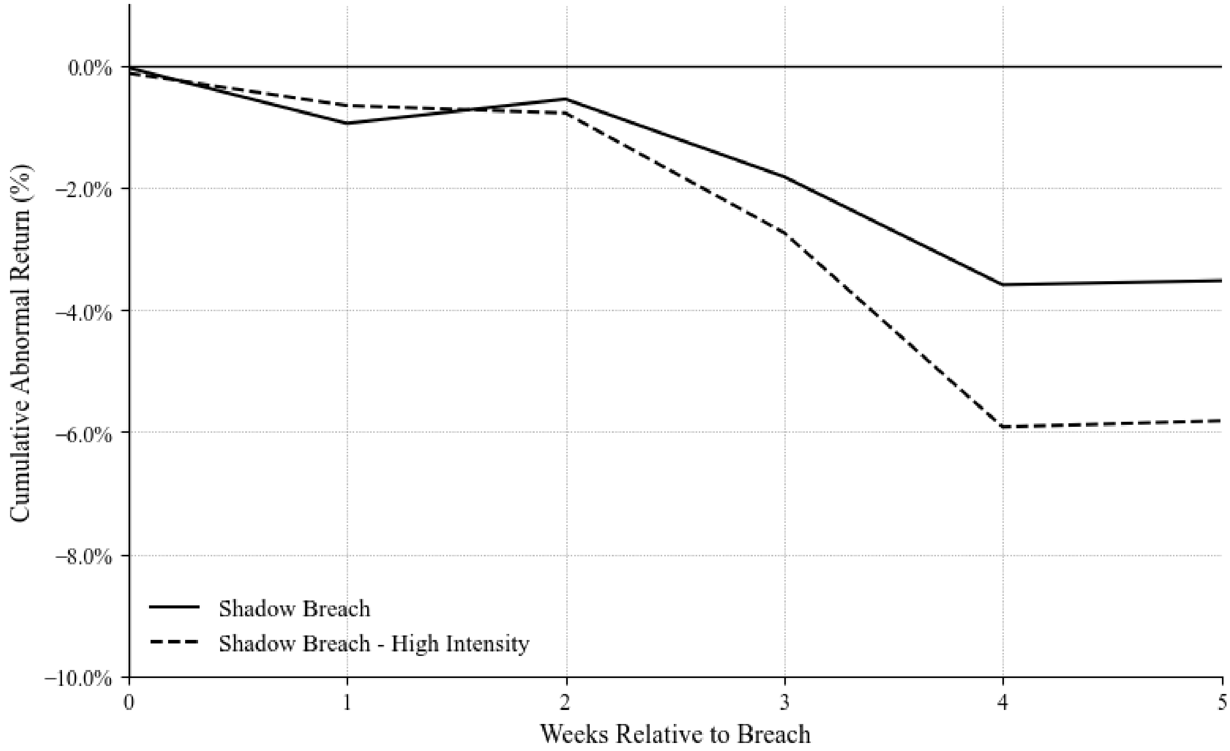
Figure 3. ***Blackmarkets*** **Around Shadow Breach**



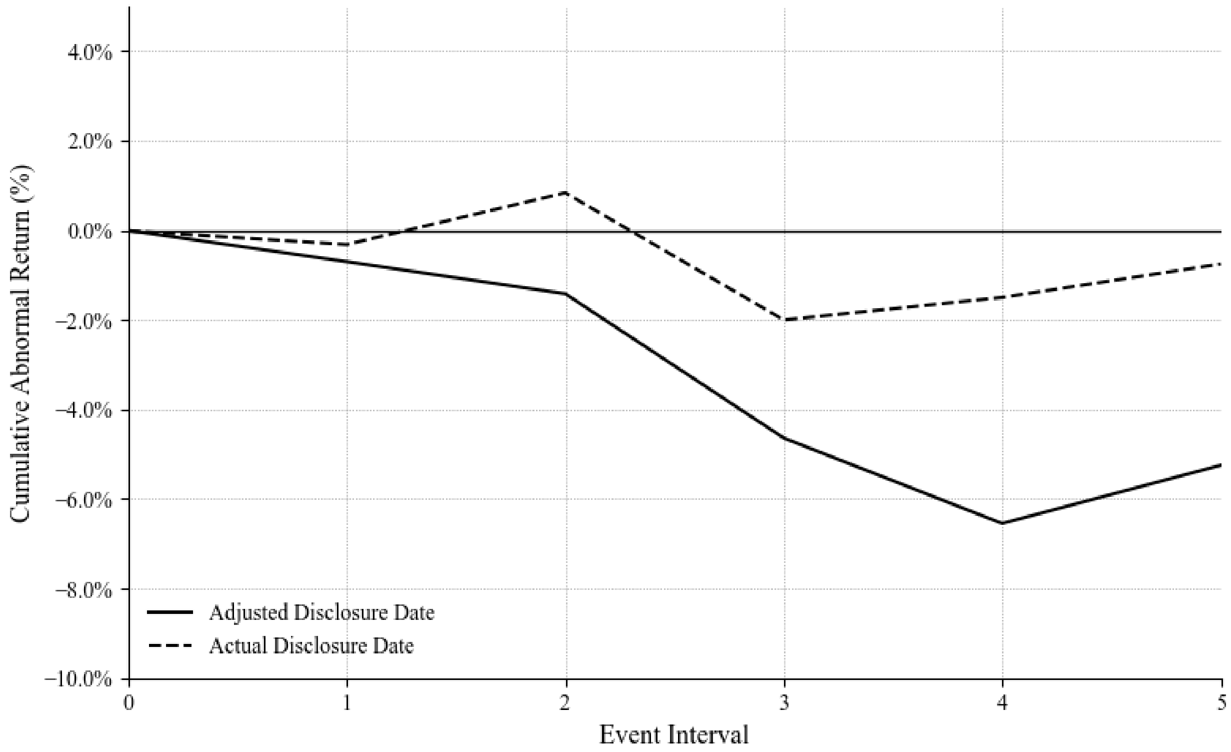
Notes: Figure 3 shows the monthly trends in *Blackmarkets* activity, orthogonalized using firm fixed effects, in the 20-month window around a shadow cyber breach. *Blackmarkets* is the frequency of activity related to the sale of firm information on the dark web.

Figure 4. Cumulative Abnormal Returns

Panel A: Cumulative Abnormal Returns for *Shadow Breach*

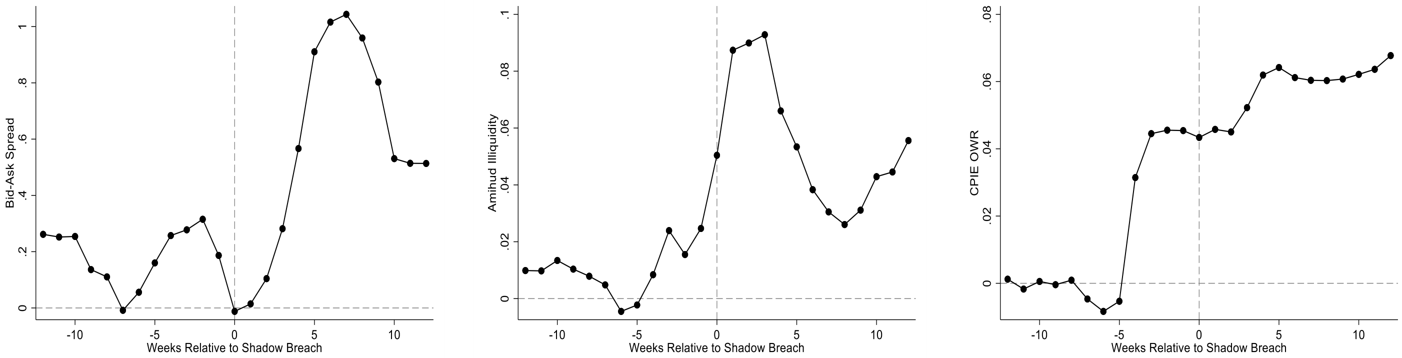


Panel B: Cumulative Abnormal Returns for Adjusted Disclosure Date



Notes: This figure presents cumulative abnormal returns (*CAR*) from week 0 to +5 for breach-related events. Panel A reports *CAR* trends for *Shadow Breach Period* and *Shadow Breach Period – High Intensity*, using the first week of sustained abnormal dark web sales activity as the event date. Panel B presents *CAR* trends for 81 breaches for both their actual disclosure dates and adjusted disclosure dates, where the adjusted dates are based on abnormal dark web activity in the year prior to disclosure. Abnormal returns are estimated using a market-adjusted model. Expected returns are calculated using a 100-day estimation window, ending 70 trading days before the event window. This figure corresponds to the cumulative abnormal returns reported in Table 6.

Figure 5. Informed Trading Around Shadow Breach



Panel A *Bid-Ask Spread*

Panel B *Amihud Illiquidity*

Panel C *CPIE OWR*

Notes: Panel A-C of Figure 5 plot informed trading measures, each orthogonalized by firm fixed effects, over a 24-week window surrounding a shadow cyber breach. *Bid-Ask Spread*, *Amihud Illiquidity*, and *CPIE OWR* are shown in Panels A, B, and C, respectively.

Table 1. Sources of Disclosure

Panel A: Number of firms reporting data breach

# of Disclosed Breaches	# of Firms
0	1,751
1	131
2	22
3	4
4+	6

Panel B: Sources of public disclosure

Source	Frequency	# of Breaches eventually reported to SEC
Attorney General	127	6
SEC Filing	44	7
News Report	21	2
Other	16	44
Total	215	59

Notes: This table reports sources of the 215 cyber breach disclosures in the sample. Panel A shows the distribution of firms by the number of breaches they disclosed. Panel B shows where the first public disclosure was made, and the number of those breaches that were eventually reported to the Securities and Exchange Commission (SEC).

Table 2. Descriptive Statistics

Panel A: Shadow breaches and disclosed breaches

	Variable	*Obs*	*Mean*	*Std. dev.*	*P25*	*P75*
Shadow Breach	*Amihud Illiquidity*	22,274	0.172	0.789	0.002	0.031
	Bid-Ask Spread	22,274	2.780	6.727	0.495	1.886
	Blackmarkets	22,274	11	38	0	2
	CPIE OWR	22,274	0.393	0.165	0.341	0.453
Disclosed Breach	*Amihud Illiquidity*	15,625	0.038	0.151	0.001	0.011
	Bid-Ask Spread	15,625	1.362	2.636	0.378	1.151
	Blackmarkets	15,625	12	29	0	5
	CPIE OWR	15,625	0.375	0.126	0.337	0.443

Panel B: Firm level

Variable	*Obs*	*Mean*	*Std. dev.*	*P25*	*P75*
BTM	465	0.378	0.315	0.173	0.542
Cash Holding	465	0.107	0.103	0.034	0.148
Dark Web 10K	465	0.862	0.345	1.000	1.000
Disclosed	465	0.314	0.465	0.000	1.000
High Tech	465	0.215	0.411	0.000	0.000
Leverage	465	0.288	0.172	0.167	0.396
Loss	465	0.206	0.405	0.000	0.000
Q	465	1.021	0.382	0.749	1.181
ROA	465	0.025	0.031	0.013	0.037
Size	465	9.243	1.798	7.984	10.207
Tangible Assets	465	0.210	0.170	0.070	0.318

Notes: This table reports descriptive statistics. Panel A reports firm-week descriptives for the two years around a cyber breach disclosure and for the two years around or a shadow breach. Panel B reports firm-level descriptives for firms that have a disclosed breach or a shadow breach. Detailed variable descriptions can be found in Appendix A.

Table 3. Dark Web Activity Around Cyber Breach Disclosure

Panel A: Dark Web Activity Before Cyber Breach Disclosure

	(1)
VARIABLES	*Blackmarkets*
Pre Disclosure	0.304**
	(0.137)
Observations	15,625
Pseudo R-Squared	0.903
Firm FE	Y
Year-Week FE	Y

Panel B: Dark Web Activity Before Cyber Breach Disclosure

	(1)
VARIABLES	*Blackmarkets*
Post Disclosure	-0.074
	(0.126)
Observations	15,625
Pseudo R-Squared	0.909
Firm FE	Y
Year-Week FE	Y

Notes: This table reports the results of fixed-effects Poisson regressions of *Blackmarkets* activity, the dependent variable, on *Pre Disclosure* indicator, which equals one for up to 24 weeks prior to disclosure and *Post Disclosure* indicator, which equals one for up to 8 weeks post disclosure. The regressions use a [-52, +52] week window around cyber breach disclosures. Firm and Year-Week fixed effects are included, and standard errors are clustered at the firm level. The intercept is not shown in the table for simplicity. Coefficients are reported with standard errors in parentheses.*** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 4. Informed Trading and *Blackmarkets* (Cyber Breach Disclosure)

	(1)	(2)	(3)
VARIABLES	*Bid-Ask Spread*	*Amihud Illiquidity*	*CPIE OWR*
Blackmarkets Period	0.068**	0.007**	0.008**
	(0.028)	(0.003)	(0.003)
Observations	7,000	7,000	7,000
Adjusted R-Squared	0.908	0.889	0.547
Firm-Breach FE	Y	Y	Y

Notes: This table reports the results of fixed-effects regressions of *Bid-Ask Spread*, *Amihud Illiquidity* and *CPIE OWR*, the dependent variables, on *Blackmarkets Period*. *Blackmarkets Period* is equal to one for the four weeks around the highest level of *Blackmarkets* activity. The regressions use a [-52,0] week window around cyber breach disclosures. Firm-Breach fixed effects are included, and standard errors are clustered at the firm level. The intercept is not shown in the table for simplicity. Coefficients are reported with standard errors in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 5. Informed Trading and *Blackmarkets* (Shadow Breach)

Panel A: Shadow Breach

	(1)	(2)	(3)
VARIABLES	*Bid-Ask Spread*	*Amihud Illiquidity*	*CPIE OWR*
Shadow Breach Period	0.332** (0.163)	0.043** (0.017)	0.018*** (0.005)
Observations	22,274	22,274	22,274
Adjusted R-Squared	0.875	0.821	0.525
Firm-Breach FE	Y	Y	Y

Panel B: Shadow Breach – High Intensity

	(1)	(2)	(3)
VARIABLES	*Bid-Ask Spread*	*Amihud Illiquidity*	*CPIE OWR*
Shadow Breach Period - High Intensity	0.836*** (0.306)	0.090** (0.035)	0.008* (0.005)
Observations	22,274	22,274	22,274
Adjusted R-Squared	0.853	0.803	0.465
Firm-Breach FE	Y	Y	Y

Notes: This table reports the results of fixed-effects regressions of *Bid-Ask Spread*, *Amihud Illiquidity* and *CPIE OWR* on Shadow *Breach Period* and *Shadow Breach Period – High Intensity*. *Shadow Breach Period – High Intensity* is a stricter version of the *Shadow Breach Period* measure and represents months in which all four weeks exceed this threshold. The regressions use a [-52, 52] week window around the shadow breach. Firm-Breach fixed effects are included, and standard errors are clustered at the firm level. The intercept is not shown in the table for simplicity. Coefficients are reported with standard errors in parentheses. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 6. Cumulative Abnormal Returns

Panel A: Shadow Breaches

	CAR [0,5] (Weeks)	t-stat	Number of Breaches
Shadow Breach Period	-3.53%***	-2.98	245
Shadow Breach Period - High Intensity	-5.82%***	-3.09	148

Panel B: Adjusted Disclosure Dates

	CAR [0,5] (Weeks)	t-stat	Number of Breaches
Adjusted Breach Date	-5.25%**	-2.07	81
Disclosed Breach Date	-0.75%	-0.34	81

Notes: This table reports cumulative abnormal returns (*CAR*) from week 0 to +5 for breach related events. Abnormal returns are estimated using a market-adjusted model. Expected returns are computed over a 100-day estimation window, ending 70 trading days before the event window. Panel A reports CAR for *Shadow Breach Period* and *Shadow Breach Period – High Intensity*, using the Monday of the first week in the Shadow Breach period as the event date. Panel B presents the distribution of the number of weeks between the actual disclosure date and the adjusted disclosure date, where the adjusted disclosure date is based on sustained abnormal dark web sales activity in the year prior to disclosure. This table corresponds to the cumulative abnormal returns reported in Figure 4.

Table 7. Determinants of Cyber Breach Disclosure

	(1)	(2)
VARIABLES	*Disclosed*	*Disclosed*
Cyber Monitoring 10-K	1.430***	1.582***
	(0.343)	(0.373)
Tech Firm		0.467*
		(0.280)
Leverage		1.265*
		(0.687)
Assets		0.307***
		(0.069)
Financial Firm		-0.190
		(0.318)
Book to Market		-0.113
		(0.338)
Cash Holdings		0.244
		(1.136)
Q		0.420
		(0.377)
Databreach Information		-0.000
		(0.000)
Observations	453	441
Pseudo R-Squared	0.037	0.091

Notes: This table reports the results of a logistic regression examining which firm-level characteristics predict whether cyber breach is publicly disclosed. Each observation represents a firm that experienced either a publicly disclosed breach or a "shadow breach" (an undisclosed breach identified through dark web selling activity). The dependent variable, *Disclosed*, is an indicator equal to 1 if the breach was publicly disclosed. *Cyber Monitoring 10K*, is an indicator equal to 1 if the firm's 10-K filings contain keywords related to cybersecurity monitoring. Additional explanatory variables include *Tech Industry* (an indicator for technology firms), *Financial Firm* (an indicator for financial firms) *Leverage*, *Assets*, *Book to Market*, *Cash Holdings*, *Q*, and *Databreach Information*. The intercept is not shown in the table for simplicity. Coefficients are reported with standard errors in parentheses. Pseudo R-squared values are reported at the bottom. *** indicates significance at the 1% level, ** at the 5% level, and * at the 10% level.

Appendix A: Variable definitions

Variable:	Definition:
Amihud Illiquidity	An illiquidity measure from Amihud (2002), calculated as the average of the absolute daily stock return divided by its daily dollar trading volume, multiplied by one million.
Bid-Ask Spread	Calculated as the difference between the closing ask and bid prices, divided by their midpoint, multiplied by one thousand.
Blackmarkets	The count of firm-specific dark web activity that involves a sales offering or transaction.
Blackmarkets Period	An indicator equal to one for the four-week window surrounding the maximum *Blackmarkets* activity in the year before a cyber breach disclosure. Observations on or after the disclosure date are excluded.
BTM	The book-to-market ratio, calculated as Book Value of Equity divided by Market Value of Equity.
Cash Holding	The ratio of the firm's cash and cash equivalents to its total assets.
CPIE OWR	The average weekly probability (bounded between 0 and 1) that private information arrived on a given trading day, estimated using the structural model of Odders-White and Ready (2008). The model jointly incorporates order flow imbalance and both intra-day and overnight returns to distinguish information-driven trades from noise. Parameters are estimated annually at the stock level using the full year of data, then applied to generate daily CPIE values. Higher CPIE OWR indicates a greater likelihood that trading activity reflects informed rather than uninformed behavior.
Dark Web 10K	An indicator variable if a firm mentions monitoring cyber security or the dark web in their 10-K. For example, "dark web", "black market", or "cyber monitoring".
Databreach Information	The count of firm-specific mentions on the dark web that explicitly reference a data breach or include leaked data associated with the company.
Disclosed	An indicator equal to one if a firm discloses a cyber breach and zero otherwise.
High Tech	An indicator equal to one if the firm operates in a technology or information-technology sector using Fama French SIC classifications, and zero otherwise.
Leverage	The ratio of the firm's total debt to its total assets.
Post Disclosure	An indicator equal to one for the month and a half period immediately following a firm's official disclosure of a cyber breach, and zero otherwise.
Pre Disclosure	An indicator set equal to one for the six-month period prior to a firm's official disclosure of a cyber breach, and zero otherwise.
Q	Tobin's Q, measured as Market Value of Equity + Book Value of Debt divided by Book Value of Total Assets
ROA	Return on assets, calculated as net income divided by total assets.

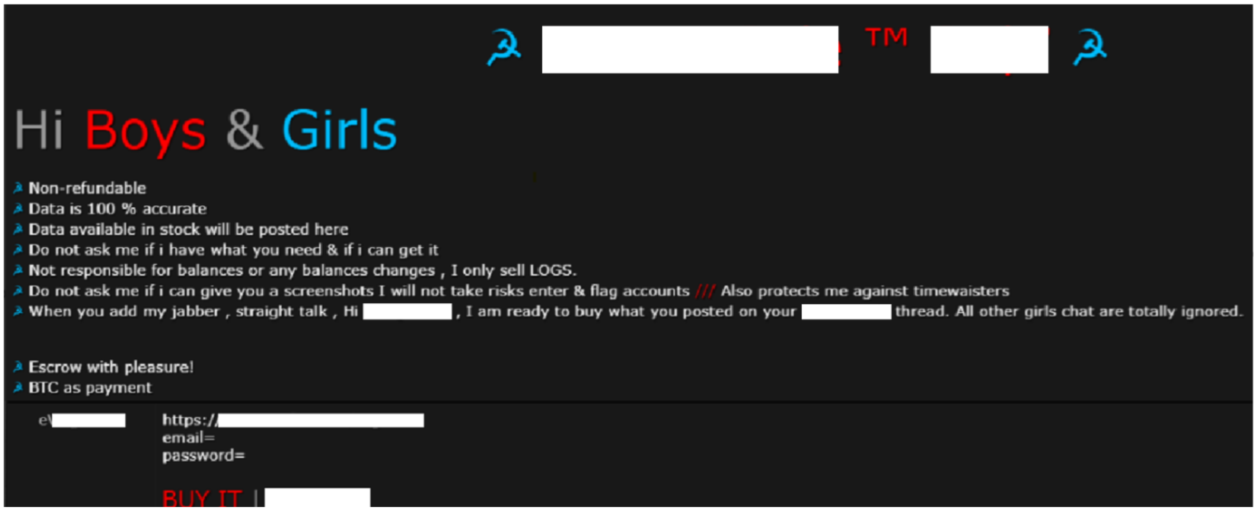
Shadow Breach Period	An indicator equal to one for the two months following a shadow breach. Shadow breach is calculated as the period where at least three weeks have abnormally high (>3SD) within-firm dark web sales activity (*Blackmarkets*) and that is not followed by any official disclosure.
Shadow Breach Period – High Intensity	An indicator equal to one for the two months following a high-intensity shadow breach. High-Intensity Shadow Breach is calculated as the period where four consecutive weeks have abnormally high (>3SD) within-firm dark web sales activity (*Blackmarkets*) and that is not followed by any official disclosure.
Size	The firm's total assets.
Tangible Assets	The ratio of the firm's property, plant, and equipment (PPE) to its total assets.

Appendix B: Dark Web Black Markets and Forums

Below, we present various examples of dark web activity that CIH classified as our primary variable, *Blackmarkets*. Dark web activity is classified as *Blackmarkets* if it involves sales activity, including forums, online marketplaces, and private channels. The term black markets is also used to describe a specific type of site on the dark web, however, the variable *Blackmarkets* when used by CIH captures selling activity more generally than on just these sites. In addition to *Blackmarkets*, CIH classifies activity into other categories which may not be mutually exclusive. The examples provided here are drawn from dark web activity in the year leading up to a disclosed data breach at a major financial institution in our sample, referred to as the "Financial Firm." To preserve confidentiality and protect sensitive data, all identifying information related to the firm and individuals has been redacted. When sites are accessible, we provide direct screenshots. When they are not, we rely on archived records maintained by CIH.

Example 1: New posted logs

A Russian-language forum known for distributing stolen data advertised a new log for sale with the Financial Firm's primary domain being the only advertised organization in the logs. This forum operates through member-only sections, where participants buy, sell, and discuss methods for accessing data illegally. The posting also includes details about the transaction, such as expecting Bitcoin as payment, an escrow being available, and that the actual transaction will take place with interested customers initiating contact through a more private communication platform. The seller also makes it clear that they are selling logs.



[redacted] TM [redacted]

Hi Boys & Girls

- Non-refundable
- Data is 100 % accurate
- Data available in stock will be posted here
- Do not ask me if i have what you need & if i can get it
- Not responsible for balances or any balances changes , I only sell LOGS.
- Do not ask me if i can give you a screenshots I will not take risks enter & flag accounts /// Also protects me against timewaisters
- When you add my jabber , straight talk , Hi [redacted] , I am ready to buy what you posted on your [redacted] thread. All other girls chat are totally ignored.

- Escrow with pleasure!
- BTC as payment

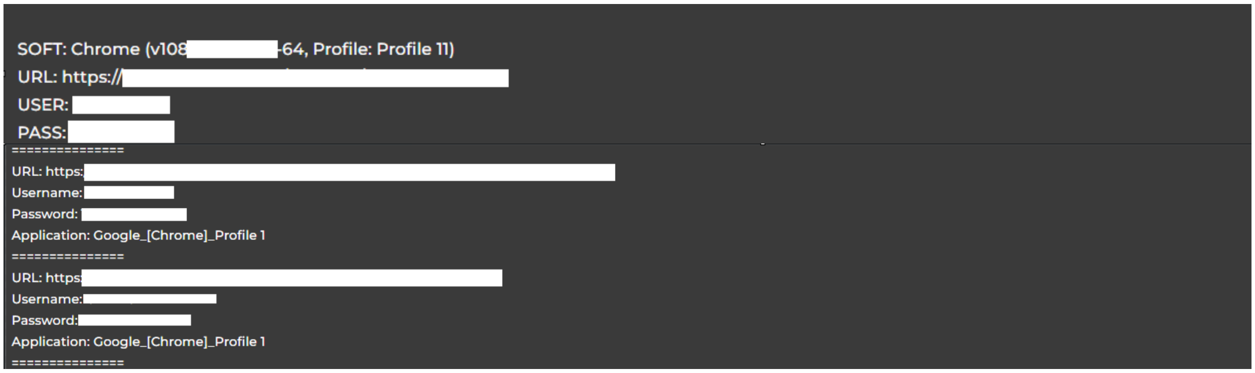
```
e\[redacted]      https://[redacted]
                  email=
                  password=

                  BUY IT | [redacted]
```

(Pictures from dark web site)

Example 2: Logs for Sale

On the dark web, *logs* are data packages harvested from compromised machines, typically via information-stealing malware. The logs are then sold to other parties interested in using the data in the logs to access compromised systems. Logs vary in complexity, ranging from basic keylogger captures of typed passwords to more advanced browser-exfiltration packages that include session tokens, cookies, and handshake artifacts—usually bundled together to enable attackers to bypass authentication steps entirely. A surge in newly listed logs on underground marketplaces often signals a recent wave of successful infections, with "fresh" logs indicating that the stolen credentials or sessions are still valid and more likely to bypass detection or multi-factor authentication controls.

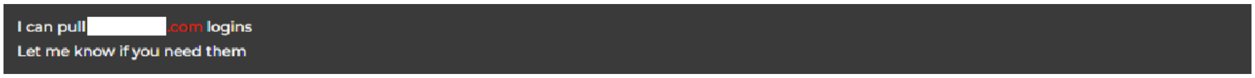
Sellers typically display truncated or partial login details to demonstrate authenticity while withholding critical information until a purchase is completed. In some instances, the logs highlight only the systems that remain actively accessible, indicating the possibility of expanded unauthorized access once buyers acquire the full dataset. Anecdotally, logs appear to be one of the more common data types sold before a breach is disclosed. Prior to the official breach disclosure of the Financial Firm, CIH recorded an increase in dark web listings offering logs linked to the Financial Firm's credentials.



```
SOFT: Chrome (v108[redacted]-64, Profile: Profile 11)
URL: https://[redacted]
USER: [redacted]
PASS:[redacted]
==============
URL: https:[redacted]
Username:[redacted]
Password: [redacted]
Application: Google_[Chrome]_Profile 1
==============
URL: https:[redacted]
Username:[redacted]
Password:[redacted]
Application: Google_[Chrome]_Profile 1
==============
```

(Pictures from CIH archive)

Example 3 - Telegram:
Telegram is a popular messaging platform that offers end-to-end encryption, making it attractive for cybercriminals seeking anonymous communication and data exchange. These platforms often rely on closed or invite-only messaging groups; similar channels include Discord, Signal, Jabber (XMPP), Wickr, Tox, ICQ, and private IRC networks.[27] CIH maintains long-standing digital personas that have built trust and credibility over time, allowing analysts to monitor closed-circle communities and access sensitive threat intelligence that would be inaccessible to outsiders. Before the breach was publicly disclosed, CIH observed discussions in private Telegram groups—dedicated to buying and selling data—about stolen information linked to the Financial Firm.[28] The following message was posted identically in at least six private Telegram groups on the same day:



I can pull[redacted].com logins
Let me know if you need them

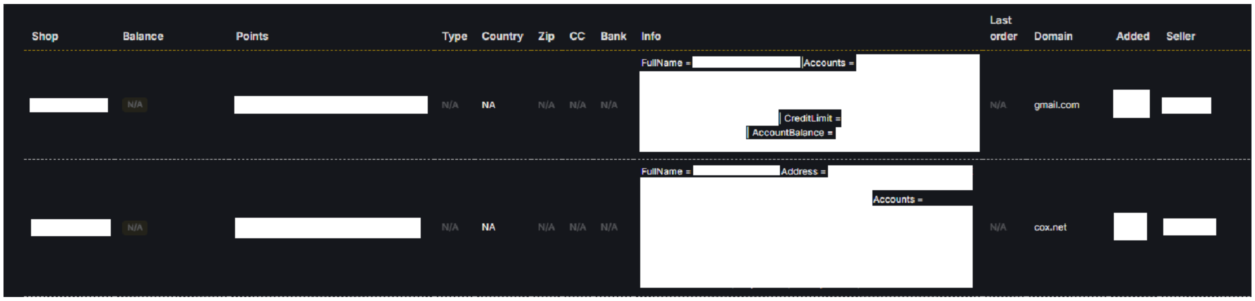
(Picture from CIH archive)

Financial firms are identified as those with SIC codes between 6000 and 6999. Technology firms are classified using the Fama-French 12 industry definition, which includes firms in the following SIC codes and sectors: 3570–3579 (Computer and Office Equipment), 3660–3669 (Communications Equipment), 3670–3679 (Electronic Components and Accessories), 3810–3819 (Engineering, Laboratory, and Scientific Instruments), 7370–7379 (Computer Programming and Data Processing), 4810–4819 (Telephone Communications), 4820–4829 (Telegraph and Other Communications), 4830–4839 (Radio and TV Bro

Example 4 – Markets:

Some underground marketplaces for stolen data closely resemble mainstream e-commerce platforms, where they primarily sell individual user profiles. Each listing typically contains login credentials, personal information, and sometimes session tokens or financial data—packaged and priced per identity. These platforms feature searchable catalogs, filters by data type or geography, vendor ratings, and even limited customer support. Payment is usually made in cryptocurrency, and some markets offer premium access or bulk discounts. While illicit, their design mirrors legitimate storefronts, making it easy for buyers to locate and purchase specific types of compromised data. Anecdotally, this type of *Blackmarket* activity is less tied to initial large system breaches and becomes more common as time progresses and individual profile information is verified and listed.

Around the time of the breach disclosure, CIH observed a small uptick in the number of individual profiles, credentials, and other personal information associated with the Financial Firm being offered for sale on these structured marketplaces. Listings appeared across at least six distinct black market sites, each with its own interface, vendor base, and user community—some of which have since gone offline. Anecdotally, this type of activity often lags well behind the initial breach, appearing in smaller, less concentrated volumes as stolen data is slowly validated, broken into individual records, and gradually released for sale.



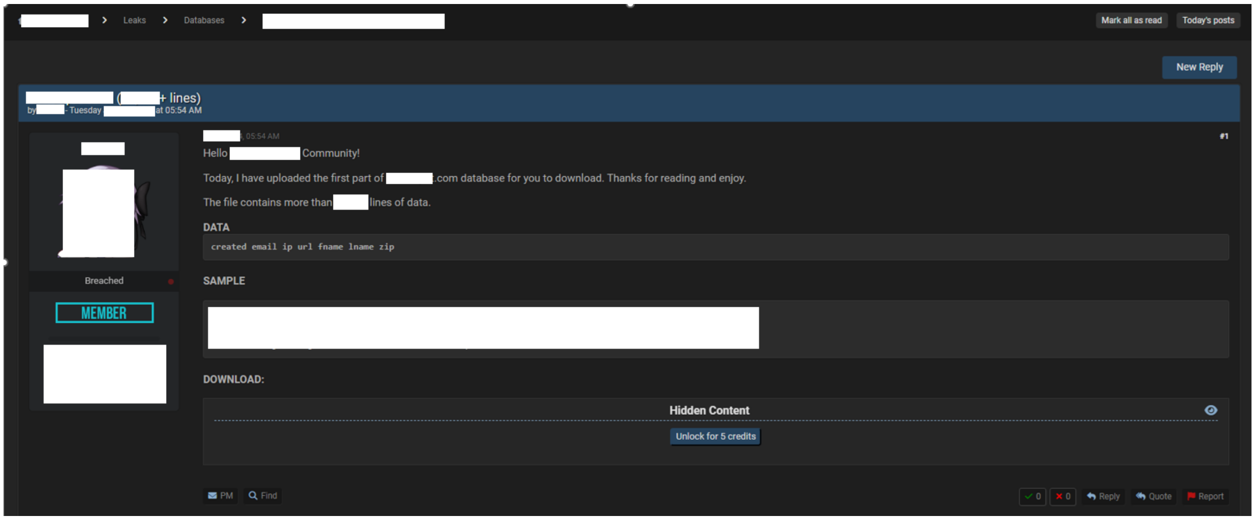
Shop	Balance	Points	Type	Country	Zip	CC	Bank	Info	Last order	Domain	Added	Seller
[redacted]	N/A	[redacted]	N/A	NA	N/A	N/A	N/A	FullName = [redacted] Accounts = [redacted] CreditLimit = [redacted] AccountBalance = [redacted]	N/A	gmail.com	[redacted]	[redacted]
[redacted]	N/A	[redacted]	N/A	NA	N/A	N/A	N/A	FullName = [redacted] Address = [redacted] Accounts = [redacted]	N/A	cox.net	[redacted]	[redacted]

(Picture from dark web site)

Example 5: Data Dumps

On breach-focused forums, it's common for stolen data to be released in stages—either to generate attention, drive traffic, or support resale through tiered access. Partial dumps can be posted behind paywalls, with users required to purchase site credits to download samples. These forums also typically serve as redistribution hubs, where data from earlier marketplace listings, data dumps, or malware logs are frequently repackaged and re-listed, or released for free.

Several weeks after the initial wave of credential sales and Telegram chatter, one such forum featured a partial dump of over a hundred thousand user profiles allegedly linked to the Financial Firm. Although this dataset represented only a fraction of the broader breach, its appearance signaled that the data had begun to circulate more widely. Subsequent user comments suggested the dump was being resold and repurposed across multiple channels.



[redacted] ([redacted]+ lines)
by [redacted] - Tuesday [redacted] at 05:54 AM

[redacted], 05:54 AM #1

Hello [redacted] Community!

Today, I have uploaded the first part of [redacted].com database for you to download. Thanks for reading and enjoy.

The file contains more than [redacted] lines of data.

DATA

```
created email ip url fname lname zip
```

SAMPLE

[redacted]

DOWNLOAD:

Hidden Content

(Picture from dark web site)

Appendix C: Contextualizing Individual Breaches

Figure 1: MOVEit Breach - 30 NYSE firms simultaneously breached

The MOVEit breach was one of the largest cyberattacks in recent years, compromising data from hundreds of organizations and exposing millions of records. The attack targeted a zero-day vulnerability in Progress Software's MOVEit Transfer, a managed file transfer (MFT) application widely used for securely exchanging sensitive data. Many organizations relied on MOVEit for automated file transfers, often integrating it with vendors and service providers. As a result, the breach extended beyond direct users to companies that had outsourced data processing or relied on third-party firms using MOVEit. Evidence suggests that attackers began exploiting the vulnerability as early as March 2023 (Remacle, 2023). The breach became public on May 31, 2023, when Progress Software issued a security advisory.[29]

We identify firms in the NYSE that were exposed to the MOVEit breach through regulatory filings, corporate disclosures, attorney general notifications, and attacker leak sites. Some companies disclosed the breach proactively, while others were identified through leaked data or vendor disclosures before issuing any public statements. In all, we were able to identified 30 firms from our sample who were directly exposed and experienced a breach of data in the MOVEit breach.[30]

We analyze dark web activity over a six-month window centered on the MOVEit breach, with relative week 0 defined as the week of March 6, 2023—the earliest known evidence of compromise. This window excludes the first public disclosure on May 31, 2023, allowing us to observe dark web activity linked to multiple affected firms without the confounding influence of media attention. Dark web activity is measured as a percentage of *Blackmarkets* relative to the first week in the window (t -12). We classify the 30 firms that lost data in connection with the MOVEit breach as Breached Firms, while all other NYSE-listed firms serve as the Non-Breached Firms in our analysis. Panel A shows that firms affected by the breach experienced a sharp rise in black market activity, peaking at roughly 70% above baseline two weeks after the earliest known vulnerability. Firms that are unaffected by the breach show very little if any, increase. Since the window is prior to any information about the breach being publicized, the activity is unlikely to be driven by publicity or market attention. We then calculate how many weeks after March 6, 2023, it takes news of the breach to become public for the affected firms, either through disclosure or news reports in Panel B. The mean is about 50 weeks and the median is about 36.

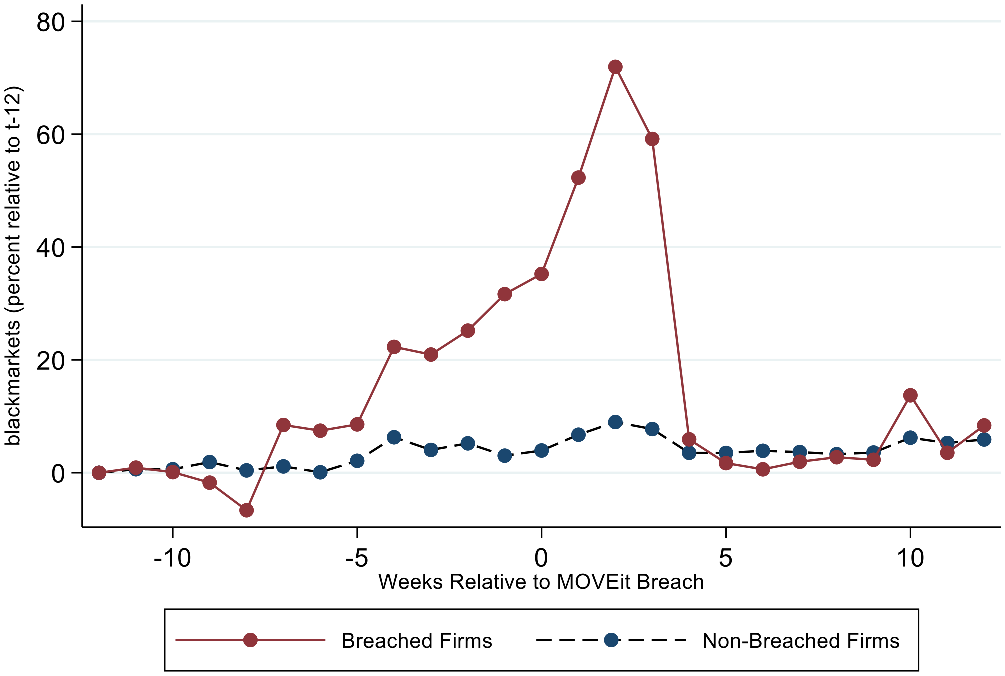
This example provides institutional context of how breaches occur and roll out. It also provides some suggestive evidence in favor of dark web activity capturing breach activity prior to disclosure, albeit limited by the small sample size. It is noteworthy that there is considerable variation across weeks in terms of when a cyber breach becomes publicly known. However, the underlying black market activity and date when firms become exposed (when identifiable) exhibit a predictable pattern, with an increase in black market activity following the breach.

adcasting), and 4890–4899 (Other Communication Services). Results are robust to alternative classifications used in prior work (e.g., Loughran and Ritter, 2004).

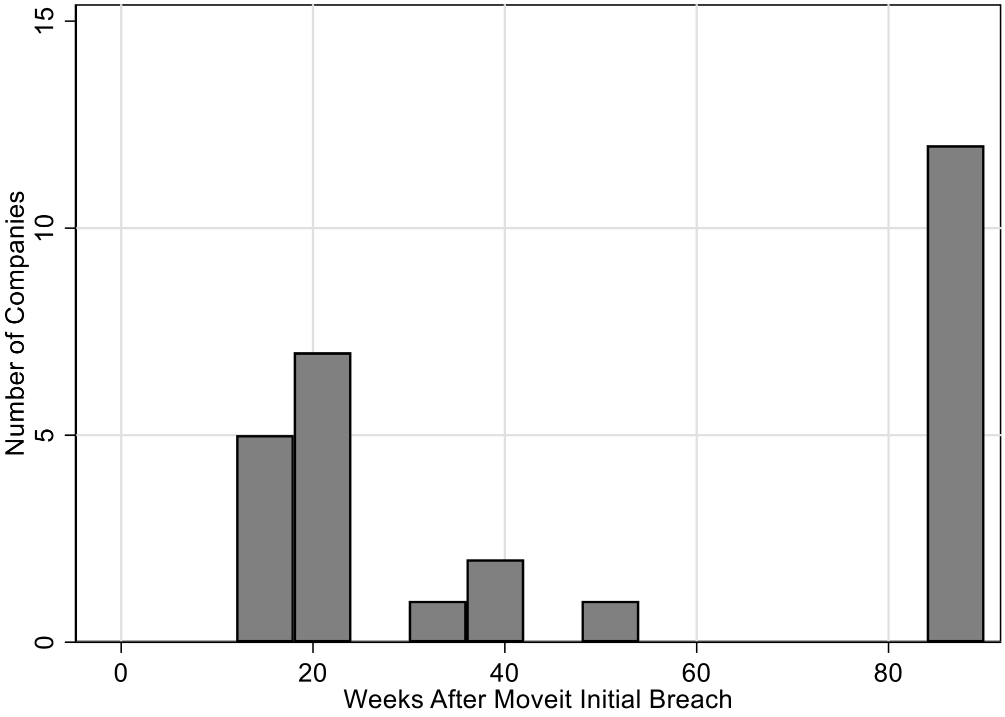
systems being exposed and data exfiltration.

[30] An important caveat is that it is possible that not all affected firms were identified, given the range of channels through which breach-related information was disclosed. Any omissions likely bias against the trends presented here. Nevertheless, the patterns described in this appendix are intended to offer institutional context rather than precise firm-level attribution.

Panel A: Dark Web Around MoveIt Initial Breach



Panel B: Number of Weeks Until Disclosure After MoveIt Initial Breach



Appendix References:

Remacle, M. (2023). Progress' MOVEit Transfer Critical Vulnerability: CVE-2023-34362. *GreyNoise Blog.* https://www.greynoise.io/blog/progress-moveit-transfer-critical-vulnerability