Price Revelation from Insider Trading:
Evidence from Hacked Earnings News

Pat Akey, Vincent Grégoire, Charles Martineau[*]

This draft: November 25, 2021

ABSTRACT

From 2010 to 2015, a group of traders illegally accessed earnings information before their public release by hacking several newswire services. We use this scheme as a natural experiment to investigate how informed investors select among private signals and how efficiently financial markets incorporate private information contained in trades into prices. We construct a measure of qualitative information using machine learning and find that the hackers traded on both qualitative and quantitative signals. The hackers' trading caused 15% more of the earnings news to be incorporated in prices before their public release. Liquidity providers responded to the hackers' trades by widening spreads.

JEL Classification: G10, G12, G14

Keywords: cyber risks, earnings announcements, hard and soft information, informed trading, liquidity, machine learning, market microstructure, price discovery

From 2010 to 2015, a group of Russian and Ukrainian hackers illegally breached the IT systems of several large newswire companies. The hackers accessed earnings announcements' press releases several hours before their scheduled release to the public and sold them to a select group of traders. This ring of traders aggressively traded in the hours before the news was publicly released in order to exploit this private, "inside" information, which only had value for a few hours. We use this event, when nearly 9,000 earnings announcements were exposed to potential informed trading by traders with limited ability to smooth their trading, as a natural experiment to study fundamental questions in market microstructure.

First, how do the fundamental characteristics of stocks or the characteristics of private signals affect an informed investor's choice of trading strategy? Classical models of trading (e.g., Kyle, 1985, Kandel and Pearson, 1995, Kim and Verrecchia, 1997) predict that insiders trade more aggressively when the difference between the observed stock price and the fundamental value is larger and when the signal about a stock's fundamental value is more precise. Though intuitive, to our knowledge these fundamental predictions have not yet been empirically examined. Indeed, to test how an informed trader would choose among different stocks, one needs a setting such as ours, where it is possible to observe a cross-section of stocks about which the informed trader had private information and could choose to trade in different stocks. In the case of the hacking scandal, the traders had access to a large number of press releases but chose to trade only on a subsample of events. We are able to examine the exact private information that the traders used to decide how to trade (the text of the press releases). Having such information allows us to use machine learning methods to measure the strength and precision of the traders' signals and to relate the characteristics of their signals to their trading decisions.

Second, how much private information can a small amount of informed trading reveal? Existing literature suggests that trading can predict news or subsequent future returns (Kaniel, Liu, Saar, and Titman, 2012, Collin-Dufresne and Fos, 2015, Hendershott, Livdan,

1

and Schürhoff, 2015). However, quantifying *how much* information is revealed by informed trades or the speed at which this information is incorporated into prices is challenging. Indeed, one needs an ex-ante quantifiable source of fundamental news such as an earnings surprise as well as a setting in which there is plausibly exogenous variation in which traders have access to the fundamental news. The hacking scheme provides such a setting because we can observe a set of common events (earnings announcements), in which "informed retail" traders had access to "inside" information about some upcoming earnings announcements' press releases but not about others. This allows to isolate the impact of insider trading to returns relative to other factors known to impact returns prior to earnings announcements.

Finally, do liquidity providers respond to a sharp increase in informed trading? Classic models such as Copeland and Galai (1983) and Glosten and Milgrom (1985) suggest that liquidity providers (market makers) widen spreads when faced with an increased likelihood of informed trading. However, recent empirical work finds that spreads and other measures of adverse selection are negatively associated with episodes of informed trading (Collin-Dufresne and Fos, 2015, Ahern, 2020, Kacperczyk and Pagnotta, 2019). This negative association is likely influenced by informed traders' efforts to split their trades across time to avoid detection, which induces an important endogeneity problem when studying how well market makers respond to informed trading. The nature of the hacking prevented the hackers from meaningfully splitting their trades across time as they only had a few hours to profit from the information before it became public. This natural constraint on strategic timing allows us to more directly study how well market makers can respond to aggressive informed trading.

To examine these important issues, we define a set of treatment firms whose earnings announcements were exposed to the hackers' breach of a newswire at a particular point in time, and a control group of firms whose earnings announcements were not exposed at that same point in time. In publicly available court documents, the US government alleges that there were at least several hundred cases of known illegal trading occurred and that the

2

hackers sold earnings news on a black-market website accessed by more than 100 individual traders. Unfortunately, the universe of these trades is not available for study, so our primary research design focuses on *access to* material, non-public information rather than on the specific informed trades.[1] We believe that our analysis estimates a lower bound of the effect of this illegal trading on prices.

We construct measures of hard and soft information that reflect the information content of the earnings press releases to quantify the information set of the hackers. We use the conventional measure of earnings surprise (DellaVigna and Pollet, 2009) to measure "hard" information. We quantify the "soft" information of the press release using a machine learning technique (ElasticNet) on the text of the press releases that scores each word in the press release to best explain the price reaction to the earnings announcement. We find that measures of both hard and soft information explain price reactions by a similar magnitude, despite only having a correlation of about 15%. We use a sample of earnings announcements that formed the basis of the US prosecutions to identify a set of cases with known informed trading and study which stock characteristics and types of information were most correlated with the hackers' trading decisions. We find that hackers were more likely to trade when firms were larger, had larger analyst coverage, and had lower spreads, as well as for earnings news that had both hard and soft signals pointing in the same direction. Moreover, trades on earnings news that had correlated signals had the largest gap between pre- and post-announcement prices, suggesting the hackers correctly identified and traded when the gap between price and fundamental value was the largest, consistent with trade models such as Kyle (1985), Kandel and Pearson (1995), and Kim and Verrecchia (1997).

We find that informed trading impacted price discovery before the news became public during the after-hours market. Overnight returns of firms whose newswires had been hacked are less responsive to earnings surprises and the soft information contained in the text of

---

[1] We filed several Freedom of Information Act requests with the SEC to better refine our analysis and to obtain the universe of potentially illegal trades. All of these requests were denied.

Electronic copy available at: https://ssrn.com/abstract=3365024

the press releases, suggesting that prices had already incorporated some of the information. Specifically, we find that the overnight return response to a given earnings surprise and soft information after the news has been released are 15% and 16% lower, respectively, if a firm's earnings release had been exposed to hackers. We also find that price revelation was stronger when the signals of future fundamental value were more precise. More specifically, we show that the sensitivity of overnight returns to earnings surprises (soft information, respectively) is even smaller for events that were subject to a hack when the magnitude of the soft information (earnings surprise, respectively) is higher. We interpret this as evidence that the hackers traded most aggressively on more precise signals, and in so doing moved prices the most for those events. This is consistent with the predictions of models in which informed traders trade more aggressively when they can better forecast future fundamental values (e.g., Kandel and Pearson, 1995, Kim and Verrecchia, 1997).

Having established that the hackers' trading was sufficiently large to impact price discovery, we use this scheme as a laboratory to study the relationship between informed trading and spreads. We find that equity volume, equity turnover, and option volume were higher when a firm's earnings were subject to informed trading. Moreover, we find that effective and realized spreads are robustly higher, suggesting liquidity providers charged higher transaction costs in response to an aggressive increase in informed trading. Our results are consistent with the predictions of models such as Glosten and Milgrom (1985), Easley and O'Hara (1987), and Kim and Verrecchia (1994), but differ from the results of previous studies that examined spreads and informed trading when the informed traders had fewer constraints on trade splitting (Collin-Dufresne and Fos, 2015, Kacperczyk and Pagnotta, 2021).

Our study complements recent papers that examine the relationship between the distribution of private information to investors and price formation. Hu, Pan, and Wang (2017), Rogers, Skinner, and Zechman (2017), and study the market response to private information at high frequency for news (e.g., macroeconomic announcements) that was publicly available

4

for a subset of paid subscribers a few seconds before its public release. They find a fast price reaction at the time the news was disclosed to these subscribers. In contrast to these studies, our empirical tests study a large set of private information that was (illegally) made available to a limited number of retail traders. Moreover, we focus on a set of retail traders who are not capable of processing information at microsecond speed to "snipe" inefficient stale quotes, but instead traded on private information through their personal brokerage accounts over a few hours.[2] As in Hu, Pan, and Wang (2017), we examine how much of that private information is reflected in prices before announcements by examining the conditional mean response of prices to news using an ex-ante, well-defined measure of surprise. In contrast to their study, we focus on a small set of less technologically sophisticated individual traders (rather than professional funds, e.g., high-frequency traders) and study firm-level news (rather than macroeconomic news), which also allow us to examine how cross-sectional characteristics of firms' trading and information environments matter for price discovery.

Our study further complements, but is different from, several recent papers that examine how investors trade when in possession of private information. Koudijs (2015) uses data from the 18th-century London and Amsterdam markets to identify the time of arrival of private news (without observing the nature of the information) and supports the hypothesis that informed agents trade as in Kyle (1985). Kacperczyk and Pagnotta (2021) retrieve a list of 453 cases of SEC investigations about insider trading for which the authors know the personal background of the insiders and their actual trades.[3] They find that nearly 80% of insiders split their trades as in Kyle (1985) and are more likely to split when private signals are less powerful and legal risks are higher.[4] Bolandnazar, Jackson Jr, Jiang, and Mitts (2020) study the trade patterns when a subset of investors received value-relevant

---

[2]Most of the traders involved in this illegal trading scheme were retail traders who were members of the hacker's extended families or their social circles. They frequently traded using their personal brokerage accounts. This pattern of an informed network is similar to the finding in Ahern (2017) that insider information flows through strong social ties.

[3]Only 15% of these cases concern trading before earnings announcements.

[4]Several other paper papers study the role of legal risks on insider trading (Meulbroek, 1992, Garfinkel, 1997).

Electronic copy available at: https://ssrn.com/abstract=3365024

information from the SEC a few seconds earlier than financial markets at large. They find that informed investors traded more aggressively when the expected release of the news was closer and when the value-price divergence is larger. Finally, Shkilko (2018) concludes that insiders primarily focus on return timing rather than liquidity timing. Our focus relative to these papers is different because the hackers' information was short-lived; in most cases, they knew the information would become public after the end of the trading day. Thus, we do not test whether traders split their trades over several days as in Kyle (1985), but rather we study whether liquidity providers are capable of widening spreads when they are faced with a sharp increase of informed trading. Such a mechanism is central to classical models of market making (e.g., Copeland and Galai, 1983, Glosten and Milgrom, 1985) and has, to our knowledge, not been systematically documented empirically.

Our paper relates to recent literature that examines the performance of insider trading measures. Collin-Dufresne and Fos (2015) and Kacperczyk and Pagnotta (2019) both examine how measures of adverse selection respond to informed trading, either on days when activist investors take positions in stocks or on days when an insider illegally trades, respectively, and find that illiquidity measures are negatively related to stock-level trading by informed investors. They conclude that informed investors strategically time their trades so as to trade when stocks are most liquid. Ahern (2020) makes a similar point and suggests that understanding the "urgency" of informed traders to trade on their private information is crucial to detect their presence. We complement these studies by providing evidence that liquidity providers *can* respond to potential informed order flow by adjusting prices and/or spreads when the ability of informed traders to time their trades is low, thus providing empirical support for classical price formation models where traders anticipate this behavior and strategically time their trades to minimize price impact (e.g., Glosten and Milgrom, 1985, Kyle, 1985, Easley and O'Hara, 1987, 1992).

Our paper also contributes to the literature that studies the relevance of non-numerical

6

data or information for financial markets. Classic papers such as Tetlock (2007), Hoberg and Phillips (2010), Loughran and McDonald (2011), Dougal, Engelberg, Garcia, and Parsons (2012), and Boudoukh, Feldman, Kogan, and Richardson (2019) show that textual analysis has a variety of applications, such as quantifying sentiment, firms' products and innovation, and investor attention. Other recent applications of machine learning include measuring banking regulation, corporate culture, and firm innovation, among many others (e.g., Goldsmith-Pinkham, Hirtle, and Lucca, 2016, Grennan, 2019, Bellstam, Bhagat, and Cookson, 2021). Several recent papers discuss the difficulty of defining soft information but highlight that non-numerical information is important in many financial market contexts (Liberti, 2018, Liberti and Petersen, 2019, Blankespoor, deHaan, and Marinovic, 2020, Bradshaw, Lock, Wang, and Zhou, 2021, Gustafson, Ivanov, and Meisenzahl, 2021). Our paper provides another instance in which textual analysis can be used to construct a measure of soft information that has valuable information content in the context of insider trading.

Finally, our paper contributes to recent literature that examines the consequences of cyberattacks on firms and financial markets. A number of recent studies examine the causes and consequences of individual cyberattacks on publicly traded firms (e.g., Akey, Lewellen, Liskovich, and Schiller, 2020, Kamiya, Kang, Kim, Milidonis, and Stulz, 2021, Florackis, Louca, Michaely, and Weber, 2020, Crosignani, Macchiavelli, and Silva, 2021, Huang and Wang, 2021). Other studies focus on the insider trading of individual firms that are subject to cyberattacks. Mitts and Talley (2018) and Berkman, Jona, Lee, and Soderstrom (2019) discuss potential insider trading of *individual firms* that are subject to breaches in cybersecurity. In contrast to these papers, we study cyberattacks on private companies that reveal information about a wide set of *other firms*. Thus, our setting is unique because we can observe the (stolen) information set of informed traders and study how the price impact of trades differs for different types of firms or earnings.[5]

---

[5]Xie (2020) uses the same setting to study how informative earnings announcements are. In contrast to her findings, our results suggest that the hackers *did* gain informative signals about future prices from press releases and that there *is* valuable information contained in earnings press releases.

1 Empirical setting: The newswire hacks

To examine the ability of informed (retail) investors to impact price formation, we must identify one set of firms for which the informed investors have access to material non-public information and one set of firms for which they do not have access to such information. Moreover, we must be able to both observe the news and infer how the market would react to the news. Hacked newswires and earnings announcements can plausibly be used as such a setting.

Firms communicate earnings and other corporate news to markets through one of several newswire companies. They typically contract with (only) one company to release news. Since the late 2000s, a firm's earnings are typically released in the after-hours market (Michaely, Rubin, and Vedrashko, 2014), when new information about the firm's value is typically incorporated in prices very quickly (Grégoire and Martineau, 2021, Martineau, 2021). Firms provide earnings announcements to their newswire service several hours before news is to be released. They are stored in the newswire's IT systems until their prescheduled release time. This window between the earnings news being uploaded to the IT systems and their public release is the period when hackers who have access to the information from the newswire's IT systems could trade using this information. Because firms only contract with one newswire company and, as we will describe in detail below, the hackers only had access to the IT systems of some of the newswire companies, we have variation in the non-public information set that the hackers could trade on (i.e., the quantitative earnings figures and qualitative discussion in the press releases).

Between 2010 and 2015, a group of hackers and individual traders (we refer to them collectively as "the hackers" as both the hackers and traders were prosecuted for illegal trading) engaged in a series of cyberattacks against the three largest newswires: Business

8

Wire, Marketwired, and PR Newswire.[6] The hackers used sophisticated methods to attempt to penetrate the IT systems of these companies. They had varying degrees of success with different newswires, but obtained access to the IT systems of Marketwired continuously from the beginning of 2010 until the end of 2013, had intermittent access to PR Newswire from 2010 to 2012, and had access to Business Wire for half of 2015. This access allowed the hackers to obtain company press releases, including earnings announcements, after they had been sent by the client firm but before they were released to the market.

The hackers and many of their family members or members of their social circles traded on the private information using a variety of personal brokerage accounts (e.g., Interactive Brokers, Charles Schwab, and TD Ameritrade). In addition, they sold information online to investors that resided in eastern Europe. We provide an image used by the SEC to prosecute some of the hackers of a "shopping list" of earnings information on which they wanted to trade in Figure A1 of the appendix. Their trading strategies were relatively sophisticated. They took both long and short positions in equities, they occasionally transacted in options, and the traders based in Europe took positions in a derivative contract called a "contract-for-difference," which is a contract where one party agrees to pay the other party the difference between the opening and closing prices of an underlying asset at the maturity of the contract.[7] In contrast to most inside trades that typically yield small dollar profits (Cziraki and Gider, 2021), these hackers earned large profits, collectively in excess of $100 million, including several events that yielded more than a $1 million over two days.

Eventually, in 2015 the newswire companies and the SEC were able to identify and prosecute the hackers that resided in the United States. The SEC initiated a number of criminal and civil proceedings against the coders that hacked the newswire companies and the traders that orchestrated the financial transactions using the stolen information. In total,

---

[6]Our description of the event is largely taken from the SEC complaints against Arkadiy Dubovoy and co-conspirators in 2015, Vitaly Korchevsky and co-conspirators in 2015, and Evgenii Zavodchicko and co-conspirators in 2016 (cases 15-cv-06076, 1:15-cr-00381, and 16-cv-00845, respectively).

[7]Augustin, Brenner, and Subrahmanyam (2019) show that traditional insiders in the US also trade in the option markets before merger and acquisition announcements.

9

more than 20 individuals were charged, although those individuals residing in Ukraine were not extradited to the United States. Those individuals who are not still in Ukraine have either pleaded guilty or been convicted and sentenced to prison.

2   Sample description and empirical strategy

We retrieve data from several sources. Daily stock returns, overnight (4 p.m. to 9:30 a.m.) returns, trade volume, and prices are from the Center for Research in Security Prices (CRSP). Intraday transactions data (trades and quotes) are extracted from the Trade and Quote (TAQ) database following the procedure in Holden and Jacobsen (2014). Institutional ownership data are from Thomson Reuters 13-F and option turnover data are from Option-Metric. The earnings announcement sample is from I/B/E/S and newswire information is from Ravenpack. Intraday option transaction data are from the CBOE Global Markets. We provide below detailed information on the earnings announcement sample construction and how we link each announcement to its newswire provider.

Our earnings announcement sample is from January 1, 2010 to December 31, 2015—the period during which the newswires were hacked. We follow Livnat and Mendenhall (2006) and choose stocks for which the earnings announcement date is reported in Compustat, the price per share is available from Compustat as of the end of the announcement quarter and is greater than $1, the stock market capitalization is greater than $5 million, and the stock is listed on the New York Stock Exchange (NYSE), NYSE Amex Equities (rebranded as NYSE MKT in May 2012), or NASDAQ with share code 10 or 11 in the CRSP. We then select the earnings announcements for stocks that have at least one analyst forecast in I/B/E/S in the 90 days prior to the earnings announcement so as to be able to compute an earnings surprise, for a total of 62,851 earnings announcements. Because the majority of earnings are announced overnight, we exclude earnings announcements that occurred during regular trading hours (2,159 observations).

We next assign earnings announcements to the newswire company that released the earnings press releases to the public. We use Ravenpack's Press Release (PR) and Dow Jones (DJ) news edition to complete this task.[8] We merge each earnings announcement to their corresponding earnings press release. We drop earnings announcements for which we were not able to assign a press release (4,545 observations). We further drop observations with no intraday returns in the TAQ database on the trading day prior to the earnings announcement and with no opening price in the CRSP following the earnings announcement (2,140 observations). We restrict our analysis to announcements that use one of the newswires that was subject to a hack at some point to further ensure we are comparing firms that are most similar. In doing so, we identify results primarily based on the *timing* of when a firm's newswire was hacked. Our final sample consists of 43,687 earnings announcements, of which 8,980 were exposed to hacks. Figure 1 shows the number of earnings announcements exposed to hacks by newswire.

We calculate the standardized earnings surprise following DellaVigna and Pollet (2009) and Hartzmark and Shue (2018) and define earnings surprises as

$$Surprise_{i,t} = \frac{EPS_{i,t} - E_{t-1}[EPS_{i,t}]}{P_{i,t-5}}, \tag{1}$$

where $EPS_{i,t}$ is the earnings per share of earnings announcement i for a firm announcing on day t, and $E_{t-1}[EPS_{i,t}]$ is the expectation of earnings per share, measured by the consensus analyst forecast in I/B/E/S. We define the consensus analyst forecast as the median of all analyst forecasts issued over the 90 days before the earnings announcement date. If analysts revise their forecasts during this interval, we use only their most recent forecasts. We scale the surprise by the firm's stock price five trading days before the announcement. We further winsorize at the 1st and 99th percentiles to account for the outliers in the surprise distribution (the kurtosis of the raw sample is 1459.87), as is commonly done in the literature (e.g., Skinner and Sloan, 2002, Doyle, Lundholm, and Soliman, 2006).

---

[8]We use the DJ edition for press releases missing from the PR edition. Ravenpack warned us that for 2014, and 2015, Business Wire press releases are missing from the PR edition but are present in the DJ edition.

11

Panel A of Table 1 presents summary statistics for our sample. The variables are the earnings surprise, absolute value of earnings surprise, year-end log market capitalization (prior to earnings announcement), fraction of shares held by institutions, number of analyst forecasts, log Q-value,[9] the monthly share turnover (monthly trade volume in dollars divided by the market capitalization), and monthly option turnover (monthly option trade volume divided by the shares outstanding).[10]

The goal of this study is to understand the causal effect of the informed trading activity of a group of small traders on price discovery. There are several reasons why uncovering this effect is challenging empirically. The first reason is that it is difficult to identify a counterfactual trade that had the same "news content" and could have had informed trading but did not. For example, many insider trading cases that were prosecuted involve trading news related to a merger (Kacperczyk and Pagnotta, 2019, Ahern, 2020). Identifying "similar" counterfactual observations for informed trading events in a given time period is difficult (Meulbroek, 1992). The second reason is that informed traders may select the types of firms about which they become informed. They may do so for unobservable reasons that relate to either the characteristics of the firm or its trading environment, in which case naïve ordinary least squares (OLS) regressions will yield a biased estimate. Third, in equilibrium, informed traders likely strategically time their trades so as to *minimize detection* (Collin-Dufresne and Fos (2015) and Ahern (2020)), thus it is important to find news events that have an equal "urgency" to trade.

We believe this hacking and illegal trading scandal is a useful setting to help overcome many of these empirical challenges. Because firms typically only contract with one newswire company, hackers had access to material, non-public information of some but not all firms. Thus, we are able to identify variation in the information set of a group of informed traders

---

[9]Using the corresponding Compustat variables, Q-value is calculated as the ratio of firm value to the sum of physical capital $(prcc\_f \times csho + dltt + dlc - act)/ppegt$.

[10]The share turnover and option turnover are calculated on the month prior to the earnings announcement.

12

that is unlikely to be related to the underlying characteristics of the firms.[11] This, combined with the fact we only analyze firm outcomes on days before an earnings announcement, means we have a counterfactual set of observations (earnings announcements that were released by non-hacked newswires). Moreover, because the hackers only had a short window (at most a few hours) to trade on the information, concerns about strategic trade timing are less salient in our setting.

We provide evidence of our identifying assumption that the pool of firms affected by a hack of their newswire was similar to the pool of firms not affected by a hack in panels B and C of Table 1. Specifically, we regress the characteristics reported in panel A of Table 1 on a binary variable that takes the value of one if a firm's newswire has been hacked in that quarter and zero otherwise. Panel B of Table 1 presents the results of this analysis in the cross-section, while panel C presents within-firm results. Columns (1) and (2) consider the raw and absolute value of firms' earnings surprises. There are no differences between firms that have a hacked newswire and those that do not. Columns (3)–(8) present firm fundamentals and characteristics of their trading environment. In the cross-section, we find that firms subject to a newswire hack are smaller and have less institutional ownership, but this difference disappears once we introduce firm fixed effects as these variables are quite persistent. We find no evidence that the Q-value is different or that firms face different information or trading environments. Firms that are hacked have similar liquidity, share turnover, option turnover, and analyst coverage.

---

[11]Our tests that examine the speed of price discovery or market-making do not condition on the specific cases in which regulators assert that there was informed trading. Comparisons of how efficiently financial markets responded to the hackers' (endogenous) choices of particular stocks to trade could reflect underlying differences (e.g., in liquidity) between traded and untraded events. We show that such differences are not present in hacked and unhacked observations.

3 How did the traders select stocks?

We now examine our first research question: which types of stocks or signals did the hackers prefer to trade on? Recent research examines the trading behavior of traditional insiders such as managers or directors (Ahern, 2020, Kacperczyk and Pagnotta, 2021); however, these insiders likely only have infrequent access to private information and only about firms to which they have a connection. Because the hackers had access to the upcoming press releases for any firm that used an affected newswire and had sent in a press release and they presumably selected individual events for rational reasons, we can use this event as a laboratory to examine which type of private information is most "tradeable." We do so by examining the characteristics of the 724 events that formed the basis of the US government's enforcement against the US-based traders they were able to charge. We do so in several steps. We first identify all of the days on which there was an informed trading event and construct a sample of all earnings announcements released on the same day to identify the observations the traders could have traded. We then split the observations into three groups: (1) earnings events that formed part of the government's case and for which we are certain that informed trading occurred, (2) earnings events that were distributed by a newswire that was affected by a hack at that point in time but did not form part of the government's case, and (3) earnings events that were distributed by a newswire that was not affected by a hack at that point in time. We next examine whether fundamental characteristics such as size or liquidity are different between these groups to study whether the hackers preferred to trade certain types of stocks. We finally study whether the hackers preferred to trade on information with more hard or soft information. We proxy for hard information using the earnings surprise. As we outline below, we conduct textual analysis on the text of the press release the hackers accessed to quantify the soft information that they were able to read before the market at large. With measures of hard and soft information in hand, we examine whether the hackers were more likely to trade on events with more extreme realizations of

14

hard and soft information. We finally examine how much more profitable are the returns of these events.

Before proceeding, it is worth noting that we are only able to observe the subset of events described in the court documents; we cannot observe the universe of all informed trades by all of the traders. This limitation introduces several potential biases. First, the US government alleges a number of traders were based in Europe, for whom they had no jurisdiction for enforcement. Thus, there are likely to be informed trades that we cannot observe. However, we believe this would introduce measurement error and potentially bias comparisons of known events with other events that were subject to the hack downwards. On the other hand, the government chooses which cases to use as the basis of enforcement events, and their selection criteria may omit harder-to-prove events, which could bias differences between prosecuted and "potentially prosecutable" events upwards. To overcome these potential biases, we compare the prosecuted observations to both the observations that were affected by the hack and could have been traded as well as the observations that *were not* affected by the hack and could *not* have been traded. In doing so, we have two reference groups: (1) the events that the traders could have traded but seemingly did not (but potentially are subject to a difficult-to-sign bias) and (2) the events that the traders could not have traded and whose characteristics might be more representative of the underlying sample characteristics.

3.1 Firm fundamentals

We first present results examining firm fundamentals between the three groups of stocks. Figure 2 presents averages and 95% confidence intervals for all three groups of events (i.e., events that were part of the US government's enforcement, events from the same day that were not part of the enforcement but were from a newswire that was hacked on the day, and events from newswires that were not affected on that day). The stocks that were part of the prosecution were larger as measured by market capitalization, were more liquid as measured by lower quoted spreads and higher institutional ownership, and were followed by

15

more analysts. Moreover, they were more frequently traded as measured by monthly share turnover and option turnover. Finally, the prosecuted events generally involved more positive surprises, which we attribute to the greater cost associated with taking short positions (e.g., as in Diamond and Verrecchia, 1987, Marin and Olivier, 2008).

3.2 Soft information from press releases

We next quantify the role of soft information in the hackers' trading decisions using the target of the hacking scheme—the press releases themselves. A typical press release contains both quantitative information about realized sales and/or earnings, often with a reference to market expectations, as well as more qualitative information similar to the management discussion and analysis. We use machine learning techniques on the text of the press releases to quantify positive and negative signals of the expected price reaction to the news. More specifically, we download all press releases from 2010–2015 from EDGAR and use a machine learning algorithm called ElasticNet to identify which words can best explain the stock return from 12 p.m. on the afternoon before the earnings are announced to the opening of the morning after the announcement. ElasticNet is a linear regression model that blends both the linear and quadratic penalties used by LASSO and Ridge regression methods in its objective function to identify the set of words that best explains a dependent variable (i.e., earnings announcement returns). This specific model performs best in cases like ours where there is very high dimensionality in the dependent variables (i.e., words in the press releases) and relatively fewer independent variables (i.e., announcement returns) (Zou and Hastie, 2005). We run this model on the first 400 words of each press release in our sample to assign a "score" to each word so as to best explain the reaction to the earnings announcement.[12]

---

[12]In a typical earnings press release, the first few hundred words summarize the recent firm performance and indicate the key factors underlying the performance. The remaining content of the press release generally consists of financial statements. See Section 4.1.2 for robustness analysis.

16

The fitted value of the "best" model is our measure of soft information.[13] We provide word cloud figures that depict both the positive and negative words that contribute most to explaining returns in Figure 3, where larger sized words have larger explanatory power for announcement returns. The words make intuitive sense. "Rais*," "exceed," "momentum," "please*, "better," and "leverag*" are some of the strongest contributors to positive scores. "Disappoint*," "challeng*," delay," "weak," "slower", "however," "decline*," and "lower" are some of the strongest contributors to negative scores.

The text of the press releases has substantial explanatory power for the announcement returns. The univariate R^2 of the best model is 0.068, which is comparable to the R^2 of the quantitative earnings surprise. One may be concerned that the press releases are explaining in text what can be understood from the earnings surprise or other quantifiable information, however the univariate Pearson and Spearman correlations between the two measures are only 0.15 and 0.20, respectively. We present further evidence that this is not the case in Table 2. In panel A, we provide a heatmap of announcement returns after double sorting by the earnings surprise and our measure of soft information. Each row represents an ordered quintile of earnings surprise and each column represents an ordered quintile of soft information. Across both dimensions of information, returns monotonically increase as the quintile of soft or hard information increases, suggesting both measures of information capture different information. We examine this more formally by running regressions in which we separately estimate the explanatory power of the earnings surprise and our measure of soft information and then estimate the joint explanatory power. Panel B of Table 2 presents the results of these regressions. In all regressions, both the earnings surprise and soft information measure are highly statistically significant. Individually, the R^2 of the earnings surprise regression is 0.061, and the R^2 of the soft information regression is 0.068, while the adjusted R^2 of the regression that includes both variables is 0.115; this suggests combining

---

[13]We provide a more detailed discussion of the process that we used to download the data, the filtering that we used to exclude words that appear frequently in most press releases, and the implementation of the machine learning algorithm in Section A2 of the Internet Appendix.

the two measures adds incremental information when explaining returns. Collectively, these results suggest our measure of soft information captures value-relevant information that is not spanned by the most commonly used measure of quantitative earnings news.

3.2.1 Soft information and stock selection

We now examine whether the hackers' trading decisions used hard as well as soft information, and whether they traded on news that had the largest potential for profit as measured by the post-announcement return. We have shown above that the informed traders were more likely to trade on positive surprises than negative surprises; however, even given an earnings surprise there is still substantial uncertainty regarding how much prices will respond to a particular surprise. It is plausible that the hackers also relied on the qualitative news content of the press releases themselves when selecting events to trade on to minimize the uncertainty of the price reaction to the information content of the press release. As shown in Kandel and Pearson (1995), Kim and Verrecchia (1991, 1997) and Banerjee and Kremer (2010), price reaction to the news content is subject to differential interpretation by investors. Intuitively, if the content of the press releases is informative, we would expect the hackers would have been most likely to trade on events where there is a higher "agreement" of hard and soft information (i.e., when both the earnings surprise and text of the press release suggest prices should increase or decrease).

We test this hypothesis by first splitting the matched sample that we used to look at fundamental characteristics and prosecuted trades by quintiles on one dimension of information (i.e., by hard information or by soft information) and measure the average level of the second dimension of information for the three categories of observations (i.e., a prosecuted event, a same-day event impacted by a hack but not a prosecuted event, or a same-day event that was not impacted by a hack). Panels A and B of Figure 4 present the results of this analysis. In panel A, we group observations by the quintile of earnings surprise on the x-axis and present the average level of soft information for each group of observations

on the y-axis. We can see that the average values of the "prosecuted" events tend to have more extreme values of soft information than the other two "control" groups. The statistical significance of some of these comparisons is weaker because we are first sorting by surprise quintile and have only 724 prosecuted observations; however, we find the patterns striking. We see a similar pattern of more extreme earnings surprises for more positive quintiles of soft information in panel B. The hackers were more likely to select observations when both the quantitative earnings surprise and the qualitative press release were most positive.

We next examine whether these trades also corresponded to the announcements for which the gap between observed and fundamental price is the largest by looking at the announcement returns (12 p.m. to the next day open) of the three groups of events, as would be expected in models such as Kyle (1985). We present these results in panels C and D of Figure 4. Specifically, we examine the returns associated with these events sorted by the quintile of hard or soft information to examine whether the hackers' selection criteria were profitable. As both figures show, their trading decisions were indeed more profitable. The returns for the negative-information quintiles that were known to have informed trading were on average lower than the returns of events with similar "levels" of information but that were not known to have informed trading, and vice versa for the positive-information quintiles. Indeed, as shown in panels C and D, the average returns of the prosecuted events with high levels of quantitative earnings surprise (and correspondingly high levels of soft information) had average returns of 4.4% and 3.3%, which can be compared to 2.5% and 1.6% for the group of observations that were subject to the hack but not traded and 2.9% and 1.9% for the group of observations that were not subject to the hack, respectively. Collectively, we believe these results suggest the hackers were more likely to trade when they had hard and soft information that pointed in the same direction, thus reducing risk associated with differential interpretation of the news by investors at the time of the announcement and focusing on events that had the largest gap between pre-announcement price and fundamental value.

19

4 Do stock prices respond less to news when there is informed trading?

We now examine our second research question: how much private information can a small number of informed trades reveal? To address this question, we examine whether a stock's overnight return reaction to an earnings release is less responsive to the earnings surprise or our measure of soft information on events for which the hackers had access to the not-yet-public news. Grégoire and Martineau (2021) show that price discovery related to earnings happens after hours and occurs quickly.

Figure 5 graphically depicts the result. Specifically, we plot the cumulative returns around earnings announcements for firms exposed to hacks (the solid blue line) and not exposed to hacks (the dashed red line) from noon on the trading day before the announcement to close on the next trading day. Specifically, we match each hacked earnings announcement to a non-hacked earnings announcement by S&P index, earnings surprise (in panel A) or soft information (in panel B), and by total return (from 12:00 p.m. to 9:30 a.m. the next day). Panel A presents results that condition on having a positive earnings surprise (the left panel) or a negative earnings surprise (the right panel), while panel B presents results that condition on having a positive value of soft information (the left panel) or a negative value of soft information (the right panel). We observe a significant price drift towards the post-announcement price in the hours before the announcement for firms exposed to hacks, which suggests that some information regarding the unreleased news is incorporated into prices for both measures of news. By the end of the next trading day, there is no significant difference between firms have been hacked or not, which means that the total response to earnings news is the same. The effect is most pronounced in the sample of positive earnings surprises and positive values of soft information, which is consistent with our result that the hackers seemed to trade more frequently on positive news.

If the hackers' trades impacted prices, the overnight response for a given earnings surprise

20

should be smaller. To test this hypothesis, we run the following regressions:

$$Return_{i,t}^{on} = \beta_1 Surprise_{i,t} + \beta_2 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathbf{1}_{[\text{Hacked}]i,t} \quad (2)$$

$$+ \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t} \text{ and}$$

$$Return_{i,t}^{on} = \beta_1 Soft_{i,t} + \beta_2 Soft_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathbf{1}_{[\text{Hacked}]i,t} + \quad (3)$$

$$+ \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Return_{i,t}^{on}$ is the overnight (close-to-open) log return around earnings announcement i for a firm announcing on day t. $Surprise$ corresponds to earnings surprises defined in Equation (1), $Soft$ corresponds to the soft information from earnings press releases, and $\mathbf{1}_{[\text{Hacked}]i,t}$ is a binary variable that takes the value of one if the firm's newswire had been hacked at that point in time and zero otherwise. We include controls for firm size and the CAPM-predicted return calculated as the intraday CAPM beta times the intraday market return using the SPDR S&P 500 ETF as a proxy for the market. We cluster standard errors by firm and year-quarter.

Table 3 presents the results of this analysis. We begin by studying the revelation of hard information (earnings surprises) prior to the release of earnings announcement in panel A. In columns (1)–(4), we present a variety of regression models that increasingly add fixed effects and controls. Consistent with the evidence presented in Figure 5, we find a strong, positive relationship between the earnings surprise and overnight returns. In all specifications, we find (as expected) that a firm's overnight return is strongly positively related to the earnings surprise. We find that overnight returns become less responsive to earnings surprises when the hackers had access to the IT systems of the firm's newswire. The effects are highly statistically significant and economically meaningful. For example, the results presented in column (4) suggest that, *for a given firm*, the return responsiveness to its earnings news declined by 15.1% in periods when the firm's newswire had been hacked (= -0.215/1.426). The point estimates are generally stable across different regression models and are not affected by the inclusion of control variables, suggesting this effect is robust and that the empirical

setting of hacked newswires isolates variation in the ease of informed retail trading that is plausibly exogenous.

In panel B, we examine the revelation of soft information prior to the dissemination of the press release. As with the analysis in panel A and the evidence in Figure 5, there is a positive and statistically significant relationship between the soft information and the overnight return as expected. Consistent with our previous results, we find that returns are less sensitive to the soft information in the press releases when the newswire that disseminated them to the market was subject to a hack. We find similar statistical and economic results for the revelation of soft information as for hard information. For example, in column (4) we find that prices were 15.7% (= -0.210/1.334) less responsive to our measure of soft information when the firm's newswire was subject to a hack. In panel C, we combine both types of information and find similar magnitude effects for the revelation of hard and soft information, but that the reduced sensitivity of prices to soft information remains more statistically robust than the reduced sensitivity to hard information.

We verify that our main conclusions are robust to an alternative method of estimating differential price discovery. We use a regression that is similar to the unbiasedness regressions used in Biais, Hillion, and Spatt (1999), Barclay and Hendershott (2003), and Van Kervel and Menkveld (2019). Specifically, we regress the long-window log return (12 p.m. to 9:30 a.m. or 4 p.m. the next trading day) onto a short-window log return (12 p.m. to 4 p.m.) and the short-window log return interacted with a dummy "hacked". We present these results in Table 4. The interactions are 0.130 and 0.142, both statistically significant at conventional levels, which suggests that 13–14% more of the *total return* over the 21.5- or 28-hour window occurred in the time periods when the hackers were trading. We note that the economic significance of this model is essentially the same as our main empirical specification.

22

4.1 Robustness

We perform a number of tests to ensure that our main results are robust to alternative modeling or empirical choices.

4.1.1 Outliers and functional form

As described in Section 2, we winsorize the earnings surprise measure at the 1st and 99th percentiles. We next verify that our main results are robust by winsorizing at the 0.5th and 99.5th percentiles. Moreover, we use an alternative estimation approach that relaxes the assumption of a linear relationship between hard (earnings surprise) or soft information and returns and also has the advantage of ensuring our results are not driven by outliers. More specifically, we verify that our main results are robust to using a more flexible estimation approach in which we define observations as falling into various quintiles of the earnings surprise or soft information distribution, and perform a piecewise estimation to see if we observe reduced return sensitivity to high or low levels of hard or soft information as measured by the quintile of the distribution. We find that our results are robust to this alternative estimation methodology. Tables A1, A2, and A3 in the Internet Appendix present the results of all of these tests.

4.1.2 Empirical choices to measure soft information

To implement the ElasticNet algorithm, we had to make three important choices with respect to how to process the text of the press releases: how many words to keep from each press release and the maximum and minimum frequency thresholds to keep individual words (400, 0.5% of all press releases and 40% of all press releases, respectively). We verify that the conclusions about the revelation of soft information in Table 2 are robust to varying these choices by recalculating a measure of soft information using different parameters and reestimating our analysis of interest. We provide a plot that shows our estimates of the revelation of soft information are stable across a number of alternative implementation choices

that we could have made in Figure A2 of the Internet Appendix.

4.1.3 Robustness to removing large cases

US regulators allege that many traders were not located within American legal jurisdiction and, therefore, many more instances of informed trading occurred than those events that formed the basis of the enforcement actions that we used to study trading behavior. As such, our main empirical specification should capture the effect of the traders that were apprehended as well as those that were not. However, we verify that our main results remain robust when removing those 724 cases who were part of the US enforcement actions to verify who our main results are not simply capturing a relatively small number of events. As Internet Appendix Tables A4 and A5 show, our results are similar when the prosecuted cases are excluded, suggesting the scheme was indeed widespread in magnitude.

4.2 Signal strength and price discovery

We now examine whether the signal strength interacts with the strength of price revelation. While the assumption of canonical models of informed trading such as Kyle (1985) assume informed traders know the fundamental value of the asset, which causes them to trade most aggressively when the difference between the observed and fundamental prices are largest, the hackers only had a signal about the future value of the assets. Other models of trading such as Kim and Verrecchia (1991, 1997), Kandel and Pearson (1995), and Banerjee and Kremer (2010) are closest to the informed traders involved in the scheme. In such models, information precision about fundamentals further determines the degree of trade aggressiveness by insiders. Such key canonical characteristics of microstruture models have largely been untested because testing them requires the researcher to have a *cross-section* of events that investors could have traded on but did not, as well as a control group of stocks that were not subject to potential informed trading. Unfortunately, we cannot directly examine the individual trades of the hackers because our efforts to acquire this information

24

have been unsuccessful, so our tests focus on the strength of price revelation. Effectively, we are assuming that higher price revelation prior to earnings announcements reflects a higher amount of informed trading by the hackers.

To examine how the signal strength interacts with the strength of price discovery, we reestimate regressions specified in Equations (2) and (3) for announcements with high (above median) and low (below median) values of soft information and for announcements with high and low earnings surprises, respectively. That is to say, we examine whether price revelation of hard (soft) information was stronger when the magnitude of the "other" signal was larger. Panel A of Table 5 examines the revelation of the earnings surprise in the high and low soft information sample. Columns (1)–(2) suggest the return responsiveness to earnings surprises for announcements with high absolute soft information declined by 18.7% (=0.329/1.763) (statistically significant at the 5% level), while Columns (3)–(4) suggest no statistically significant change in return responsiveness to earnings surprises for events with smaller amounts of soft information.

Panel B presents the results of the revelation of soft information in the sample of high and low earnings surprises. Columns (1)–(2) show the return responsiveness to its soft information for announcements with high absolute earnings surprises declined by 20.2% (=0.33/1.63) (statistically significant at the 5% level), while columns (3)–(4) suggest no statistically significant change in return responsiveness to soft information for events with smaller earnings surprises. Collectively, these results suggest informed traders trade more aggressively when their information is a more precise signal about fundamental value, resulting in higher revelation of fundamental information.

5   Did liquidity providers react to the informed trading?

We now examine our final research question: did liquidity providers react to the increase in informed trading by charging higher spreads? A number of theoretical studies model the

25

behavior of informed investors and uninformed market makers (e.g., Glosten and Milgrom, 1985, Kyle, 1985, Easley and O'Hara, 1992, Collin-Dufresne and Fos, 2016). A common theme of these models is that if market makers were to observe an increase in informed trading activity then they would respond by widening spreads or decreasing market depth. Taking this behavior as given, informed traders strategically split or time their trades when possible to avoid revealing themselves to market makers. Existing research that "tests" models of informed trading focuses its empirical tests on how informed traders structure their trades when trying to profit from their information (e.g., Koudijs, 2015, Kacperczyk and Pagnotta, 2021, Shkilko, 2018, Bolandnazar, Jackson Jr, Jiang, and Mitts, 2020). However, to our knowledge, existing research has not examined the ability of *market makers* to respond to increased informed trading volume, a critical assumption for these models.[14] Testing this assumption is challenging because researchers require an empirical setting such as ours in which informed investors faced substantial constraints on their ability to strategically split their trades. The short window of time between when the hackers accessed upcoming press releases and their prescheduled release to financial markets provides such a setting.

We begin by examining whether market makers could plausibly have detected a sharp increase in trading activity.[15] We do so by returning to the sample of events that formed the basis of the US government's enforcements and identify the subset of events for which the legal documents provide the time of the first trade by the hackers. We can identify the time of the first trade for 170 events. We construct a matched sample of firms from the sample of firms whose newswires were not subject to a hack on the same day and compare the abnormal volume, minute-by-minute before and after the time the prosecuted hackers were known to have begun trading. We identify the three closest matches based on earnings surprise and on the stock return from 12 p.m. to the next day open. We then calculate the abnormal log equity volume and log option volume for each minute for each firm as the

---

[14]Collin-Dufresne, Fos, and Muravyev (2020) suggest that informed trading in options markets causes option spreads to widen, to our knowledge no one has shown they can also widen in the underlying market.

[15]Van Kervel and Menkveld (2019) show that high-frequency traders can detect informed orders.

difference between the log volume observed at time t minus the log of the average volume for the same minute in the past 21 days. Panel A of Figure 6 plots the average abnormal log equity volume for both groups, while panel B plots the average abnormal log option volume. In both cases, we observe a "spike" in the trade volumes of the hacked observations compared to the unhacked observations. The magnitudes are striking. In the first three minutes after this group of hackers began trading, log equity volume increased by approximately 150% relative to the unhacked observations, while log option volume increased by 125%. These increases are substantial relative to the intraday distribution of changes in log equity and log option volume. Specifically, the increase in equity volume falls at the 88.5th percentile of the intraday distribution of pre-earnings days and the 93.5th percentile for all days, while the increase in option volume falls at the 92.5th percentile of the intraday distribution of pre-earnings days and the 97.5th percentile for all days. Thus, it seems reasonable to assume that market makers could have detected such a large spike in trading volume in the afternoon before an earnings announcement and inferred an increased probability that they were transacting with informed traders.

We return to the full sample to examine whether market makers widened spreads when faced with increased informed trading. We examine whether several measures of order flow and spreads were higher when the hackers had access to earnings information. We consider four potential measures of order flow: intraday equity volume, intraday equity turnover, intraday absolute order imbalance (i.e., the difference between market buy and sell orders), and intraday option volume. We first examine which of these measures are demonstrably higher when the hackers were likely to be trading in order to verify that liquidity providers could have reacted to a measure of order flow as suggested by theory (e.g., Glosten and Milgrom, 1985, Easley and O'Hara, 1992). We then examine whether liquidity providers reacted by studying whether quoted spreads, effective spreads, realized spreads, and price impact increased when the hackers had access to non-public information. When designing

these tests, we use the fact that press releases were often uploaded in the late morning or early afternoon to the newswire servers, and the hackers typically only began trading in the afternoon, which allows us to use the morning as a second type of placebo test. We construct the measures over the time period of 9:30 a.m. to 12 p.m. and from 12 p.m. to 4 p.m. on the trading day before the earnings announcements.[16] We calculate the intraday measures of order flow and spreads using the NYSE Monthly and Daily Trades and Quotes (TAQ) database for the periods 2010 to 2013 and 2014 to 2015, respectively. We follow the recommendations in Holden and Jacobsen (2014) to reduce bias in the estimates for the liquidity measures in the monthly TAQ database.[17] We provide additional details on the construction of each measure in Appendix A1.

We estimate the following regression:

$$Outcome_{i,t} = \beta \mathbf{1}_{[\text{Hacked}]i,t} + \Gamma'Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}, \tag{4}$$

where $Outcome_{i,t}$ represents one of the eight measures that we enumerate above. $\mathbf{1}_{[\text{Hacked}]i,t}$ is a binary variable that takes the value of one if a firm's newswire company was subject to a hack at that point in time and zero otherwise. We include controls for log market capitalization, fraction of shares held by institutional investors, natural logarithm of number of analysts, natural logarithm of newswire news in the quarter leading to the announcement, daily cost of borrowing from Markit, and the inverse of the stock price.

The regression results are reported in Table 6. Panel A presents estimates for informed trading measures calculated from 12 p.m. to 4 p.m. (when the hackers were most likely to be trading), while panel B presents the results from 9:30 a.m. to 12 p.m. (when the hackers were less likely to be trading). All variables except for (log) stock and option trade volume are rescaled to have a standard deviation of one.

Columns (1)–(2) and (4) show that share volume and turnover and option volume are

---

[16]The spread measures are calculated from 9:45 a.m. because it is well known that spreads are large at the opening of markets (see Jiang and Yao, 2019).

[17]We thank Craig Holden and Stacy Jacobsen for providing their code to process TAQ data efficiently.

reliably higher when $\mathbf{1}_{[\text{Hacked}]i,t}$ is equal to one. Share turnover and volume is 4.9% and 3.5% of a standard deviation higher, respectively, and volume in option markets is 7.2% higher. In contrast, we find the absolute order imbalance is not higher for stocks exposed to hacks. We propose two reasons for why we observe no increase in the absolute order imbalance. First, as Collin-Dufresne and Fos (2015) suggest, if informed traders use both market and limit orders then it is more difficult to empirically measure informed trading using the commonly defined measure of order imbalance.[18] Second, liquidity providers could be engaging in dynamic inventory control. For example, Ho and Stoll (1981) and Hendershott and Menkveld (2014) suggest that liquidity providers generally manage inventory risk associated with large buy/sell pressure by adjusting quotes to steer volume traffic towards reverting their inventory to the desired level (Menkveld, 2013).

Columns (5)–(8) examine which measures of spreads increase when $\mathbf{1}_{[\text{Hacked}]i,t}$ is equal to one. We find that effective spreads increased by 3% of a standard deviation during periods when a firm's newswire had been hacked. We decompose this increase in effective spread into two components: a realized spread component and a price impact component. We find this increase in effective spread comes from an increase in the realized spread rather than from a direct price impact of each trade. We do not find quoted spreads are higher, though prior research has argued that effective and realized spreads are superior measures of the actual cost of executing trades (Petersen and Fialkowski, 1994).

We use the fact that the hackers did not typically have access to material, non-public information in the morning as a placebo test to further verify that unobserved, time-varying changes in the trading environment of hacked firms is not spuriously responsible for our informed trading results. Panel B of Table 6 presents the results of the same analysis as in panel A, but for measures computed from 9:30 a.m. to 12 p.m. We find little evidence that volume, turnover, or spreads are higher for firms subject to a hacked newswire *in the*

---

[18]Even if the hackers themselves were not using market orders, other traders that "lean with the wind" (see Van Kervel and Menkveld, 2019) could use limit orders when executing some of their trades, which would impact the measurement of order imbalance in the same way.

morning before an earnings announcement, when the hackers did not have access to superior information.

We present a formal test for differences of the effect of $\mathbf{1}_{[\text{Hacked}]i,t}$ for the afternoon and morning time periods in panel C. We see the increase in volume and turnover for the afternoon is highly statistically different from the "placebo" test of the morning. We find the increases in effective spreads and realized spreads in the afternoon are nearly statistically significantly different from the "placebo" tests, with p-values of 0.12 and 0.16, respectively. However, direct comparisons of the robust effect that we document in the afternoon to the "placebo" non-effect that we find in the morning are more complicated than for measures of order flow. Specifically, previous research has highlighted that spreads are substantially more volatile in the morning than in the afternoon and exhibit an intraday "seasonality" (e.g., Heston, Korajczyk, and Sadka, 2010, Bogousslavsky, 2016, Jiang and Yao, 2019)[19].

5.1 Spreads and price impact

One surprising result from Table 6 is that the price impact measure is not higher for earnings exposed to hacks. Shkilko (2018) examines the price impact at the micro level around a sample of insider trades reported by the Toronto Stock Exchange and shows that price impact does increase. The results in Table 6 do not imply that no individual trade had a large price impact, but rather that *on average* the price impact of trades in the time periods when the hackers were active was not higher. We perform further analysis to examine which quantiles of the effective spread, realized spread, and price impact distributions changed when the hackers were trading on non-public information to better understand these findings. More specifically, we compute the volume-weighted quantile for each one percent of the morning and afternoon distribution of these spread measures for each stock on each pre-earnings announcement day. We then reestimate equation (4) for each percentile for each measure.

We summarize the results of this estimation in Figure 7, where we plot the point estimate

---

[19]For early evidence on intraday seasonality, see McInish and Wood (1992) and Wood, McInish, and Ord (1985).

and confidence intervals for each percentile for the morning and the afternoon for each dependent variable. Panel A presents the results for effective spreads. The point estimates are always positive and 42% are statistically significant at the five-percent level in the afternoon. None of the point estimates are statistically significant in the morning. When we look at the two components of the effective spread in panels B and C, (the realized spread and the price impact), respectively, the average results in Table 6 become easier to understand. In the afternoon, the realized spread is statistically positive at the five-percent level above the 70th percentile and is statistically *negative* below the 6th percentile, while the price impact is only statistically positive above the 91st percentile and is negative below the 24th percentile. The negative shift in the left part of the distribution for price impact is likely either a consequence of inventory risk management by liquidity providers or of informed trading executed using limit orders, which would be measured as negative price impact when estimated with conventional measures of price impact. These results further confirm that some *individual trades* had higher price impact, which liquidity providers may have been able to observe and which may have led them to increase the spreads they charge to traders. We find no robust patterns in the morning for either variable and note that the standard errors are generally larger in the morning, particularly for the larger percentiles. Collectively, these results provide guidance for future research using these illiquidity measures to proxy the degree of insider trading. We recommend researchers further examine the change in the distribution of these measures, as the mean effect can be subsumed by the fact that insider trading can occur through market and limit orders (Kaniel and Liu, 2006, Baruch, Panayides, and Venkataraman, 2017).

6   Conclusion

We use a series of successful hacks on newswire providers when a group of traders illegally accessed press releases several hours before their pre-scheduled distribution to financial

markets to study how informed traders process financial information and how efficiently markets incorporate private information into prices. This empirical setting allows us to use a differences-in-differences methodology to identify plausibly exogenous variation in the information set of a group of traders.

We first examine which types of signal best explained the hackers' trading decisions. We quantify the soft (or qualitative) information contained in the press releases using machine learning techniques and compute the earnings surprise to quantify the hard (or quantitative) information in the press releases. We find that both types of signal were important and complementary determinants of the hacker's trading decisions—the hackers were most likely to trade on earnings announcements for which there were correlated signals that were more extreme in magnitude. These choices were profitable. The hackers' trades were associated with much larger changes in fundamental value once the news was publicly released than other trades they could have made.

We next examine how much of the information was revealed by the hackers' trading. We find that overnight returns (after the earnings releases became public) were 15% less responsive to both the hard and soft information when the hackers had early access to the press releases. We find this reduced sensitivity is strongest for earnings announcements for which the hard or soft information contained in the press release were the most extreme, suggesting the hackers traded most aggressively on signals that were the most precise.

Finally, we examine whether market makers can detect and respond to sharp increases in informed trading. We find that equity volume, equity turnover, and option volume were significantly higher in stocks for which the informed traders had access to upcoming earnings news. Moreover, we find that liquidity providers responded to this increase in informed order flow by charging higher spreads. Collectively, our results suggest that liquidity providers can efficiently respond to aggressive bursts of informed trading.

References

Ahern, Kenneth R., 2017, Information networks: Evidence from illegal insider trading tips, *Journal of Financial Economics* 125, 26–47.

———— , 2020, Do proxies for informed trading measure informed trading? Evidence from illegal insider trades, *Review of Asset Pricing Studies* 10, 397–440.

Akey, Pat, Stefan Lewellen, Inessa Liskovich, and Christoph Schiller, 2020, Hacking corporate reputations, Unpublished working paper.

Augustin, Patrick, Menachem Brenner, and Marti G. Subrahmanyam, 2019, Informed options trading prior to takeover announcements: Insider trading?, *Management Science* 65, 5697–5720.

Banerjee, Snehal, and Ilan Kremer, 2010, Disagreement and learning: Dynamic patterns of trade, *The Journal of Finance* 65, 1269–1302.

Barclay, Michael J., and Terrence Hendershott, 2003, Price discovery and trading after hours, *Review of Financial Studies* 16, 1041–1073.

Baruch, Shmuel, Marios Panayides, and Kumar Venkataraman, 2017, Informed trading and price discovery before corporate events, *Journal of Financial Economics* 125, 561–588.

Bellstam, Gustaf, Sanjai Bhagat, and J. Anthony Cookson, 2021, A text-based analysis of corporate innovation, *Management Science* 67, 4004–4031.

Berkman, Henk, Jonathan Jona, Gladys Lee, and Naomi Soderstrom, 2019, Digital insiders and informed trading before earnings announcements, *Working Paper*.

Biais, Bruno, Pierre Hillion, and Chester Spatt, 1999, Price discovery and learning during the preopening period in the Paris bourse, *Journal of Political Economy* 107, 1218–1248.

Blankespoor, Elizabeth, Ed deHaan, and Ivan Marinovic, 2020, Disclosure processing costs, investors' information choice, and equity market outcomes: A review, *Journal of Accounting and Economics* 70, 101344.

Bogousslavsky, Vincent, 2016, Infrequent rebalancing, return autocorrelation, and seasonality, *Journal of Finance* 71, 2967–3006.

Bolandnazar, Mohammadreza, Robert J. Jackson Jr, Wei Jiang, and Joshua Mitts, 2020, Trading against the random expiration of private information: A natural experiment, *The Journal of Finance* 75, 5–44.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, and Matthew Richardson, 2019, Information, trading, and volatility: Evidence from firm-specific news, *The Review of Financial Studies* 32, 992–1033.

Bradshaw, Mark T., Brandon Lock, Xue Wang, and Dexin Zhou, 2021, Soft information in the financial press and analyst revisions, *The Accounting Review* 96, 107–132.

Collin-Dufresne, Pierre, and Vyacheslav Fos, 2015, Do prices reveal the presence of informed trading?, *Journal of Finance* 70, 1555–1582.

———— , 2016, Insider trading, stochastic liquidity, and equilibrium prices, *Econometrica* 84, 1441–1475.

33

———— , and Dmitriy Muravyev, 2020, Informed trading in the stock market and option price discovery, *Journal of Financial and Quantitative Analysis* Forthcoming.

Copeland, Thomas E., and Dan Galai, 1983, Information effects on the bid-ask spread, *The Journal of Finance* 38, 1457–1469.

Crosignani, Matteo, Marco Macchiavelli, and André F. Silva, 2021, Pirates without borders: The propagation of cyberattacks through firms' supply chains, *FRB of New York Staff Report*.

Cziraki, Peter, and Jasmin Gider, 2021, The dollar profits to insider trading, *Review of Finance* Forthcoming.

DellaVigna, Stefano, and Joshua M. Pollet, 2009, Investor inattention and Friday earnings announcements, *Journal of Finance* 64, 709–749.

Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277–311.

Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher A. Parsons, 2012, Journalists and the stock market, *The Review of Financial Studies* 25, 639–679.

Doyle, Jeffrey T., Russell J. Lundholm, and Mark T. Soliman, 2006, The extreme future stock returns following I/B/E/S earnings surprises, *Journal of Accounting Research* 44, 849–887.

Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69–90.

———— , 1992, Time and the process of security price adjustment, *Journal of Finance* 47, 577–605.

Florackis, Chris, Christodoulos Louca, Roni Michaely, and Michael Weber, 2020, Cybersecurity risk, Discussion paper, National Bureau of Economic Research.

Garfinkel, Jon A, 1997, New evidence on the effects of federal regulations on insider trading: the insider trading and securities fraud enforcement act (itsfea), *Journal of Corporate Finance* 3, 89–111.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

Goldsmith-Pinkham, Paul, Beverly Hirtle, and David O. Lucca, 2016, Parsing the content of bank supervision, *Working paper*.

Grégoire, Vincent, and Charles Martineau, 2021, How is earnings news transmitted into stock prices?, *Journal of Accounting Research* Forthcoming.

Grennan, Jillian, 2019, A corporate culture channel: How increased shareholder governance reduces firm value, *Available at SSRN 2345384*.

Gustafson, Matthew T., Ivan T. Ivanov, and Ralf R. Meisenzahl, 2021, Bank monitoring: Evidence from syndicated loans, *Journal of Financial Economics* 139, 452–477.

Hartzmark, Samuel M., and Kelly Shue, 2018, A tough act to follow: Contrast effects in financial markets, *The Journal of Finance* 73, 1567–1613.

Hendershott, Terrence, Dmitry Livdan, and Norman Schürhoff, 2015, Are institutions informed about news?, *Journal of Financial Economics* 117, 249–287.

Hendershott, Terrence, and Albert J. Menkveld, 2014, Price pressures, *Journal of Financial Economics* 114, 405–423.

Heston, Steven L., Robert A. Korajczyk, and Ronnie Sadka, 2010, Intraday patterns in the cross-section of stock returns, *Journal of Finance* 65, 1369–1407.

Ho, Thomas, and Hans R. Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47–73.

Hoberg, Gerard, and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *The Review of Financial Studies* 23, 3773–3811.

Holden, Craig W., and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *Journal of Finance* 69, 1747–1785.

Hu, Grace Xing, Jun Pan, and Jiang Wang, 2017, Early peek advantage? Efficient price discovery with tiered information disclosure, *Journal of Financial Economics* 126, 399–421.

Huang, Henry He, and Chong Wang, 2021, Do banks price firms' data breaches?, *The Accounting Review* 96, 261–286.

Jiang, Wenxi, and Chen Yao, 2019, U disappeared: Indexing and the shift of diurnal patterns, *Working paper*.

Kacperczyk, Marcin, and Emiliano Pagnotta, 2019, Chasing private information, *The Review of Financial Studies* 32, 4997–5047.

——— , 2021, Becker meets Kyle: Legal risk and insider trading, *Working paper*.

Kamiya, Shinichi, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M Stulz, 2021, Risk management, firm reputation, and the impact of successful cyberattacks on target firms, *Journal of Financial Economics* 139, 719–749.

Kandel, Eugene, and Neil D. Pearson, 1995, Differential interpretation of public signals and trade in speculative markets, *Journal of Political Economy* 103, 831–872.

Kaniel, Ron, and Hong Liu, 2006, So what orders do informed traders use?, *The Journal of Business* 79, 1867–1913.

Kaniel, Ron, Shuming Liu, Gideon Saar, and Sheridan Titman, 2012, Individual investor trading and return patterns around earnings announcements, *The Journal of Finance* 67, 639–680.

Kim, Oliver, and Robert E. Verrecchia, 1991, Trading volume and price reactions to public announcements, *Journal of Accounting Research* 29, 302–321.

——— , 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41–67.

——— , 1997, Pre-announcement and event-period private information, *Journal of Accounting and Economics* 24, 395–419.

Koudijs, Peter, 2015, Those who know most: Insider trading in eighteenth-century Amsterdam, *Journal of Political Economy* 123, 1356–1409.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1335.

Lee, Charles M.C., and Mark J. Ready, 1991, Inferring trade direction from intraday data, *Journal of Finance* 46, 733–746.

Liberti, José María, 2018, Initiative, incentives, and soft information, *Management Science* 64, 3714–3734.

——— , and Mitchell A. Petersen, 2019, Information: Hard and soft, *Review of Corporate Finance Studies* 8, 1–41.

Livnat, Joshua, and Richard R. Mendenhall, 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts, *Journal of Accounting Research* 44, 177–205.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *The Journal of Finance* 66, 35–65.

Marin, Jose M., and Jacques P. Olivier, 2008, The dog that did not bark: Insider trading and crashes, *The Journal of Finance* 63, 2429–2476.

Martineau, Charles, 2021, Rest in peace post-earnings announcement drift, *Critical Finance Review* Forthcoming.

McInish, Thomas H., and Robert A. Wood, 1992, An analysis of intraday patterns in bid/ask spreads for NYSE stocks, *Journal of Finance* 47, 753–764.

Menkveld, Albert J., 2013, High frequency trading and the new market makers, *Journal of Financial Markets* 16, 712–740.

Meulbroek, Lisa K., 1992, An empirical analysis of illegal insider trading, *Journal of Finance* 47, 1661–1699.

Michaely, Roni, Amir Rubin, and Alexander Vedrashko, 2014, Corporate governance and the timing of earnings announcements, *Review of Finance* 18, 2003–2044.

Mitts, Joshua, and Eric L. Talley, 2018, Informed trading and cybersecurity breaches, *Harvard Business Law Review* 8.

Petersen, Mitchell A., and David Fialkowski, 1994, Posted versus effective spreads: Good prices or bad quotes?, *Journal of Financial Economics* 35, 269–292.

Rogers, Jonathan L., Douglas J. Skinner, and Sarah L.C. Zechman, 2017, Run EDGAR run: SEC dissemination in a high-frequency world, *Journal of Accounting Research* 55, 459–505.

Shkilko, Andriy, 2018, Insider trading under the microscope, *Working paper*.

Skinner, Douglas J., and Richard G. Sloan, 2002, Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio, *Review of Accounting Studies* 7, 289–312.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *The Journal of Finance* 62, 1139–1168.

Van Kervel, Vincent, and Albert J Menkveld, 2019, High-frequency trading around large institutional orders, *Journal of Finance*.

Wood, Robert A., Thomas H. McInish, and J. Keith Ord, 1985, An investigation of transactions data for NYSE stocks, *Journal of Finance* 40, 723–739.

Xie, Chloe, 2020, The signal quality of earnings announcements: Evidence from an informed trading cartel, Working Paper.

Zou, Hui, and Trevor Hastie, 2005, Regularization and variable selection via the elastic net, *Journal of the royal statistical society: series B (statistical methodology)* 67, 301–320.

Figure 1. Number of Earnings Announcements Exposed to Hacks

This figure shows the number of earnings announcements exposed to hacks (left y-axis, indicated with bars) per newswire and the corresponding proportion over all earnings announcements exposed to hacks (right y-axis, indicated with dots) on each quarter. The sample period is from January 2010 to December 2015.
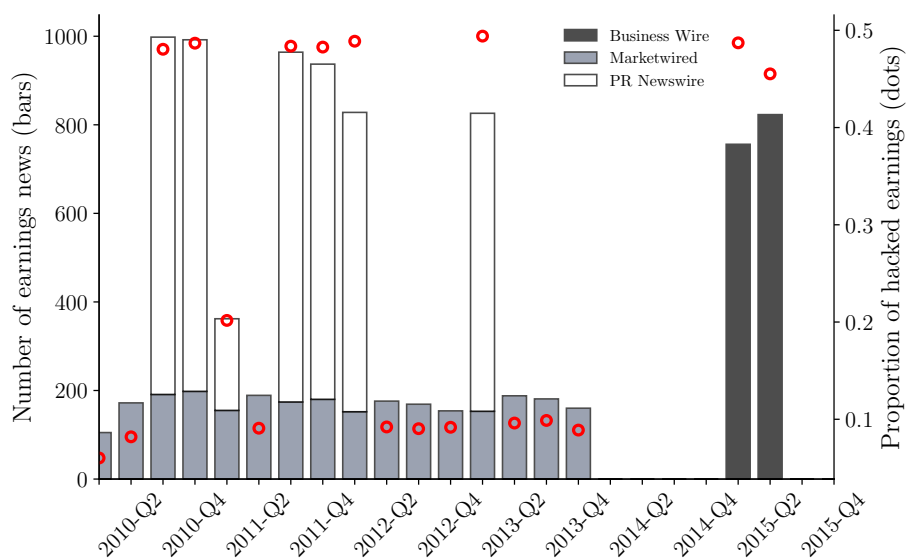
Figure 2. Stock Selection

This figure shows the average log market capitalization, number of analysts, fraction of institutional ownership (%), quoted spreads (%), monthly share turnover, monthly option turnover, and earnings surprises (%) for prosecuted, hacked (non-prosecuted), and non-hacked stocks. The sample of stocks are those with earnings announcements on the same day for which there is at least one prosecuted stock. The error bars are calculated for 95% confidence intervals.
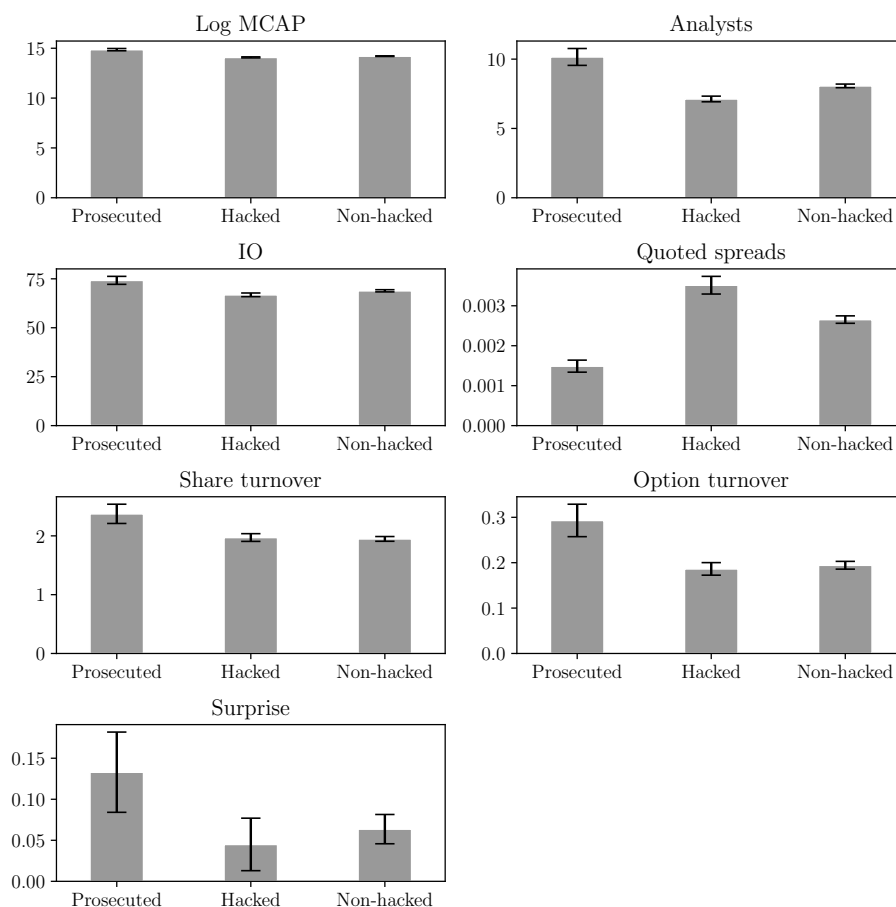
Figure 3. The Words that Most Determine Soft Information

This figure shows word clouds in which the size of each word indicates its importance at explaining positive (panel A) and negative (panel B) stock returns around earnings announcements (from 12:00 p.m. to 9:30 a.m. the next day).

Panel A. Positive weights

Panel B. Negative weights
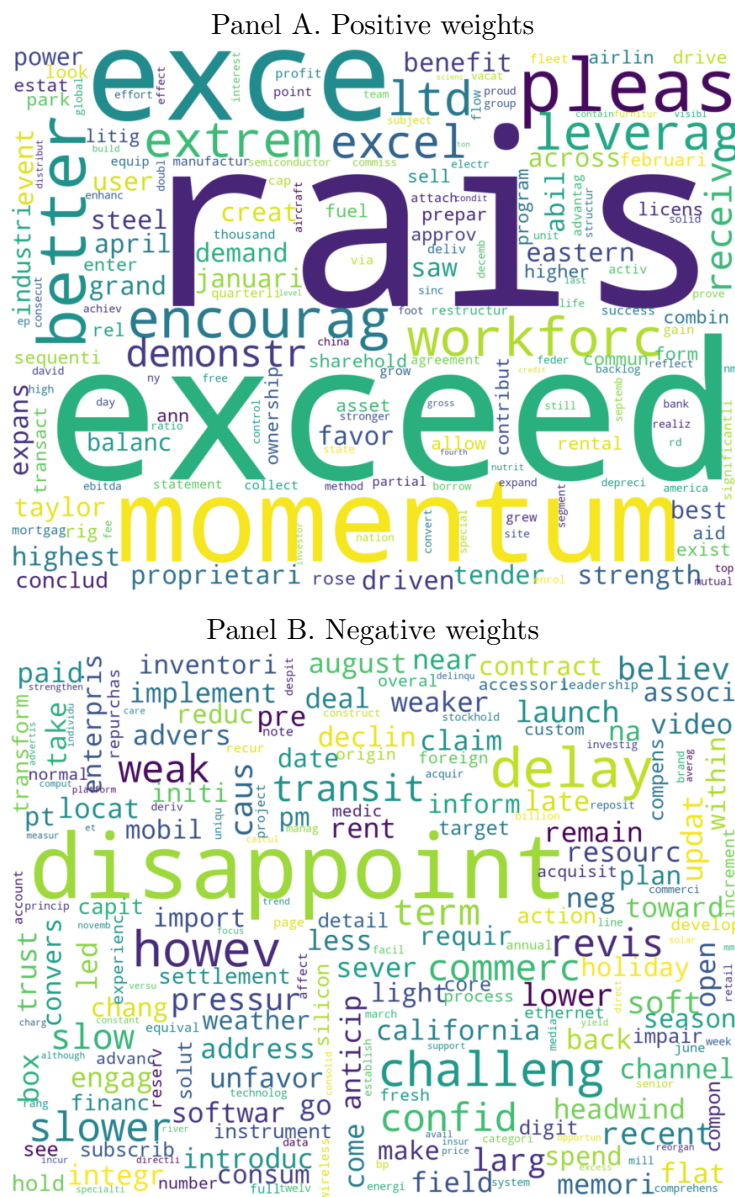
Figure 4. Surprise, Soft Information, and Returns for Prosecuted and Non-Hacked Stocks

This figure shows the average soft information by earnings surprise quintiles in panel A, average earnings surprise by soft information quintiles in panel B, and average stock returns (from 12 p.m. to next day open) by surprise quintiles in panel C and by soft information quintiles in panel D. The results are presented for non-hacked stocks, non-prosecuted hacked stocks, and prosecuted stocks. We select earnings announcements on dates for which at least one prosecuted case is reported in the SEC filings. The sample period is from January 2010 to December 2015.

Panel A. Soft by surprise quintile

Panel B. Surprise by soft quintile

Panel C. Returns by surprise quintile

Panel D. Returns by soft quintile
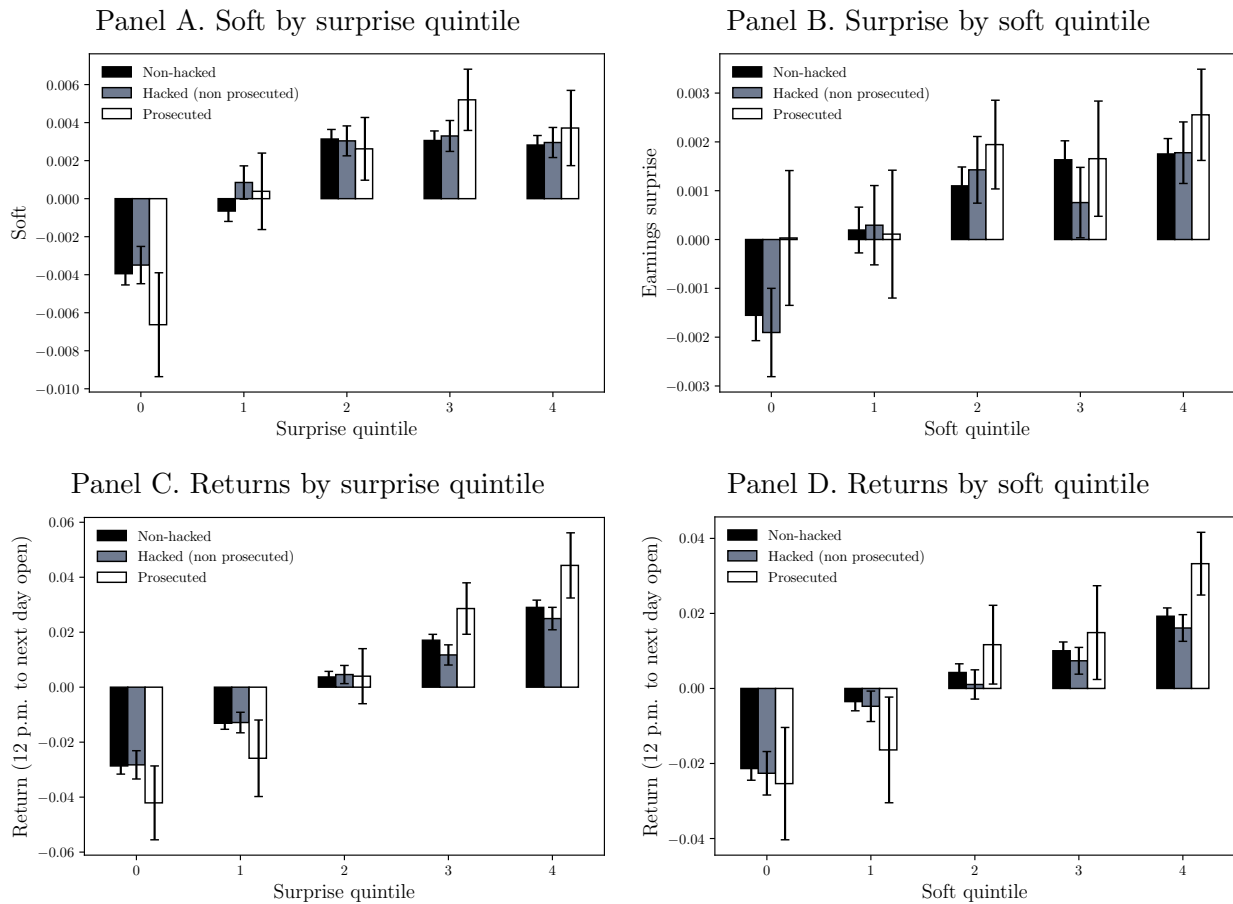
Figure 5. Cumulative Returns Around Earnings Announcements

This figure shows the average cumulative returns around earnings announcements for hacked and non-hacked matching stocks on the trading day before the announcement to the close on the next trading day. The cumulative returns for positive and negative earnings surprises and soft information are presented in panels A and B, respectively. The hacked earnings announcements are matched to one non-hacked earnings announcement for stocks belonging to the same S&P index, by earnings surprise (in panel A) or soft information (in panel B), and by total return (from 12:00 p.m. to 9:30 a.m. the next day). The shaded areas correspond to 95% confidence intervals. The gray bar indicates the overnight period and includes the first 15 minutes of trading on the next day. The sample period is from January 2010 to December 2015.

Panel A. Earnings surprise

Panel B. Soft information
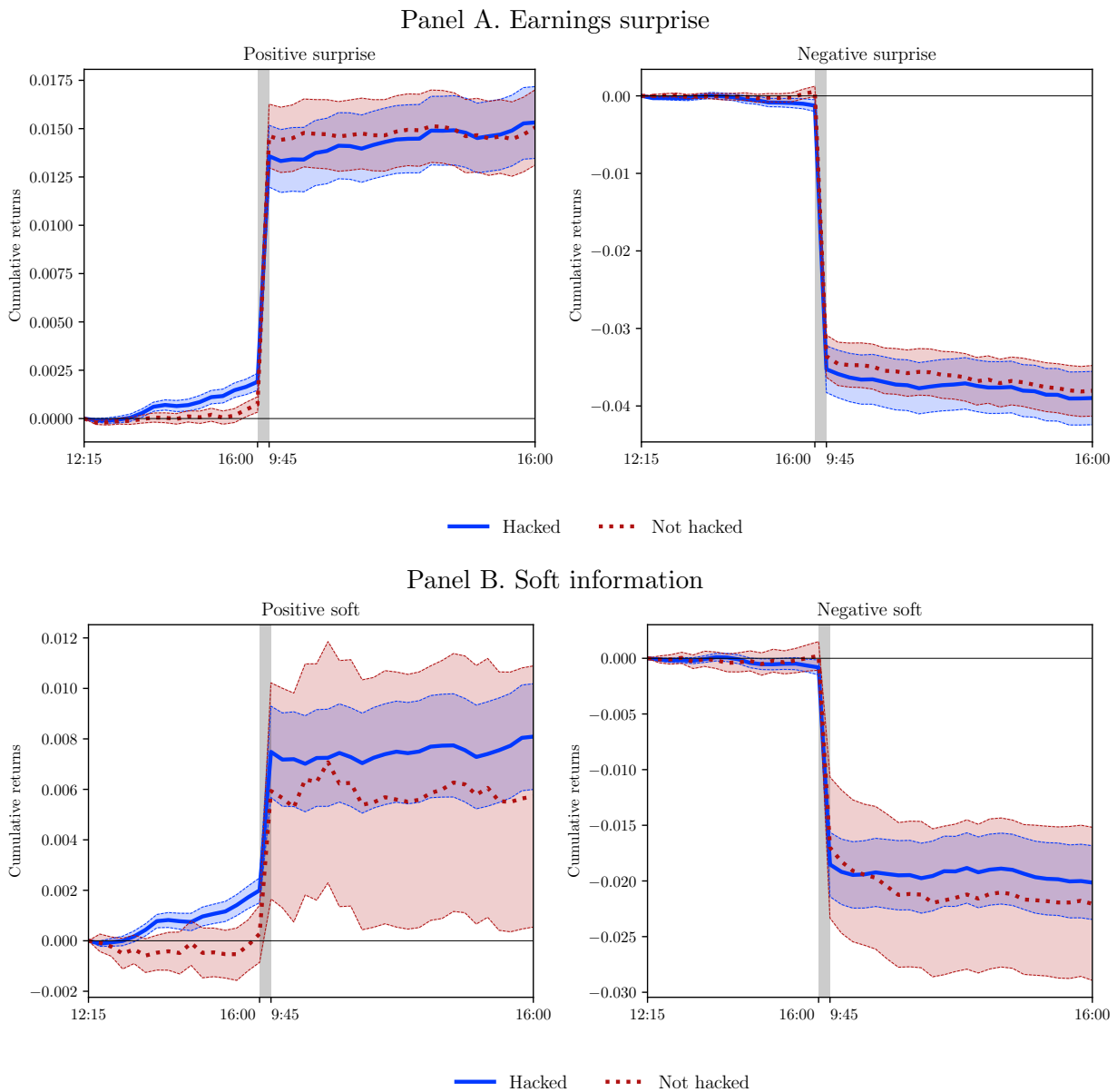
Figure 6. The Impact of Hackers' Trading on Volume

This figure shows the abnormal level in log trade volume in panel A and log option trade volume in panel B around the time of the first trade by hackers prior to earnings announcements. The x-axis corresponds to a 30-minute window around the event ($t = 0$). The abnormal measures are constructed as the difference between the log volume at time t and the log of the average volume at time t over the 21 previous trading days. The sample contains the 170 events for which we were able to identify the time of the first trade from the legal documents of the SEC prosecutions. For each event, we include three matched firms whose newswires were not subject to a hack.
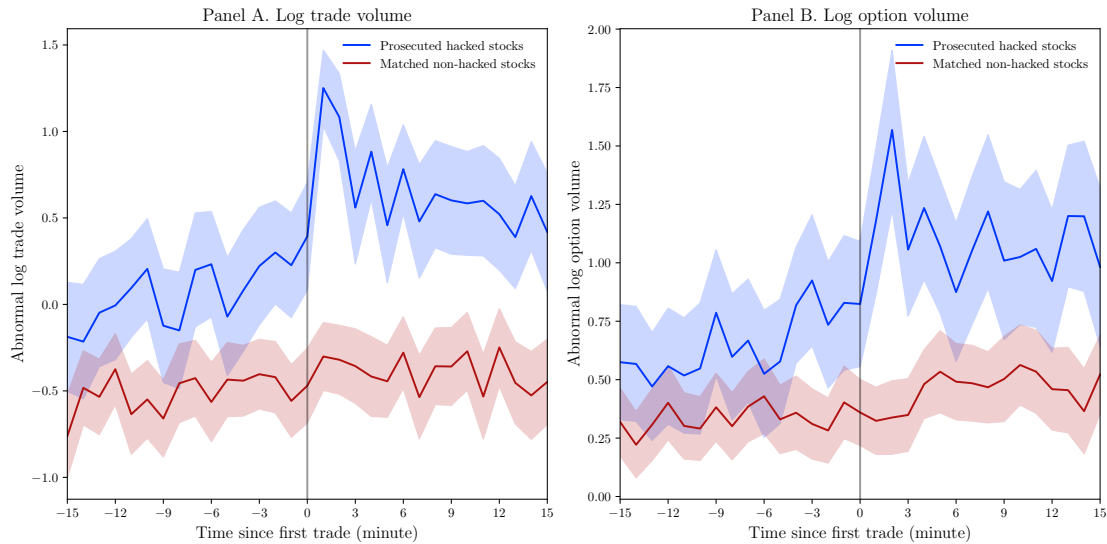
Figure 7. Changes Across the Spread Measure Distribution

This figure shows the estimated coefficients (β) of the following regression:

$$Outcome_{i,t} = \beta \mathbf{1}_{[\text{Hacked}]i,t} + \Gamma'Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.$$

$Outcome_{i,t}$ corresponds to effective spreads, realized spreads, and price impact in panels A to C, respectively. Regressions are estimated at each percentile of the distribution of the informed trading measures across all earnings announcements i for the morning (9:45 a.m. to 12 p.m.) and afternoon (12 p.m. to 4 p.m.) trading sessions. The distributions for each informed trading measure defined in Section A1 are calculated by stock-earnings announcement date across all trades and are dollar-weighted. The shaded areas correspond to the 95% confidence intervals. The sample period is from January 2010 to December 2015.

Panel A. Effective spreads

Panel B. Realized spreads

Panel C. Price impact

44
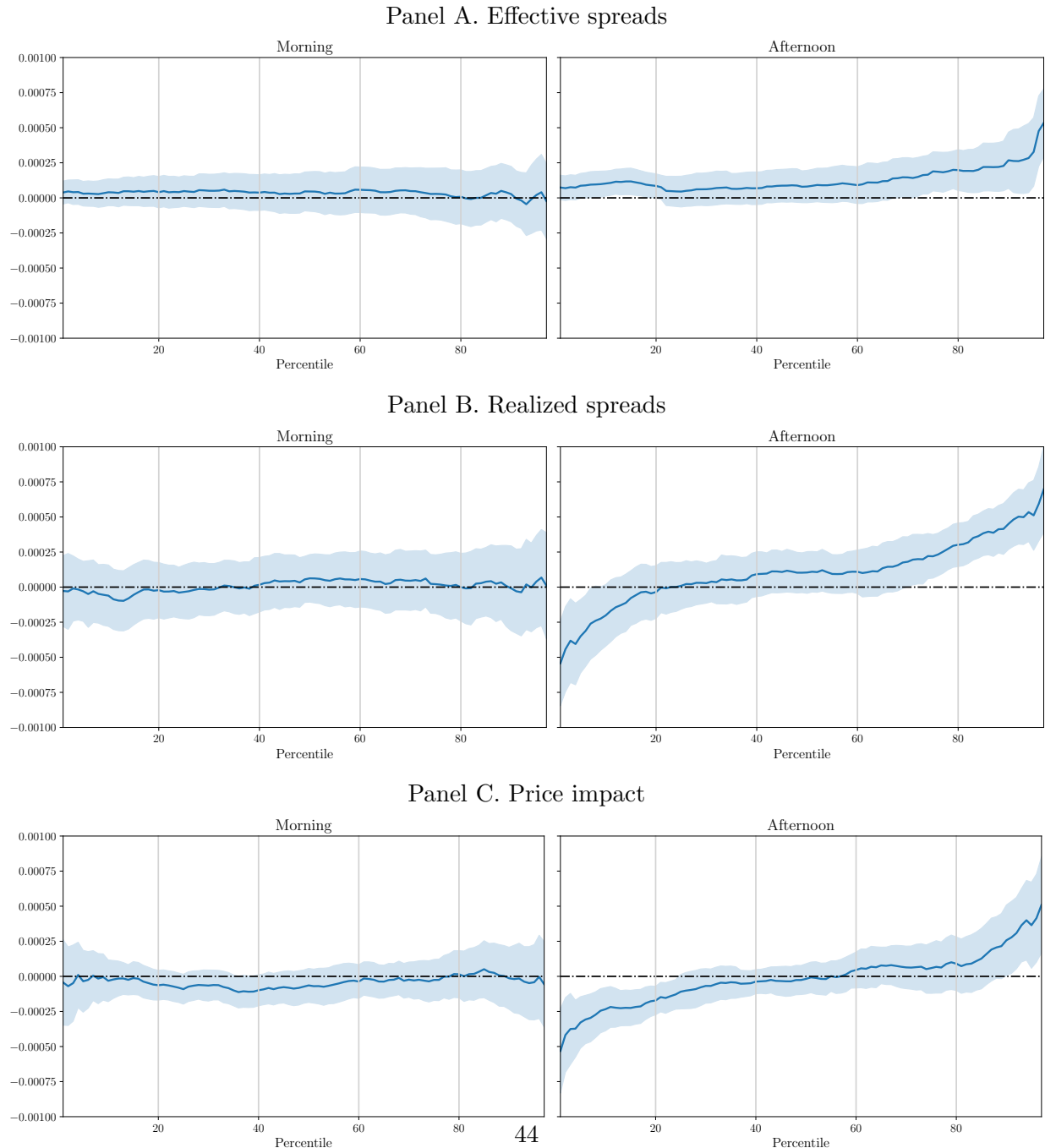
Table 1
Descriptive Statistics and Characteristic Regressions

Panel A reports the mean, median, standard deviation, and number of observations (N) for the earnings surprise, absolute earnings surprise, log market capitalization (log(MCAP)), fraction of shares held by institutions (IO, in percent), number of analysts, log Q-value of the firm, monthly share turnover, and monthly option turnover. Both turnover measures are calculated in the month prior to the earnings announcement. Panels B and C report the results of the following regression model:

$$Characteristics_{i,t} = \beta \mathbf{1}_{[\text{Hacked}]i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Characteristics_{i,t}$ corresponds to one of the variables in Panel A for earnings announcement i announced on year-quarter t. $\mathbf{1}_{[\text{Hacked}]i,t}$ is a dummy variable equal to one if announcement i is released by a hacked newswire and zero otherwise. Results are presented with year-quarter fixed-effects in panel B and firm and year-quarter fixed-effects in panel C. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Descriptive statistics

	Surprise	|Surprise|	Log(MCAP)	IO	N. analysts	Log(Q-value)	Share turn	Option turn
Mean	0.0004	0.007	14.16	68.37	7.69	0.97	1.99	0.2
Median	0.0006	0.0017	14.1	77.53	5.0	0.8	1.46	0.04
Std. dev.	0.0104	0.0373	1.77	28.82	6.92	1.53	2.11	0.47
N	43,687	43,687	43,687	43,687	43,687	35,273	43,687	43,687

Panel B: Hacked and non-hacked announcements - with year-quarter fixed-effects

	Surprise (1)	|Surprise| (2)	Log(MCAP) (3)	IO (4)	N. analysts (5)	Log(Q-value) (6)	Share turn (7)	Option turn (8)
$\mathbf{1}_{[\text{Hacked}]}$	-0.000	0.000	-0.195**	-3.119**	-0.153	0.045	0.051	0.003
	(0.000)	(0.001)	(0.079)	(1.336)	(0.221)	(0.055)	(0.061)	(0.014)
N	43,687	43,687	43,687	43,687	43,687	35,273	43,687	43,687
Adjusted R^2	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000
Year-Quarter F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Firm F.E.	N	N	N	N	N	N	N	N

Panel C: Hacked and non-hacked announcements - with firm and year-quarter fixed-effects

	Surprise (1)	|Surprise| (2)	Log(MCAP) (3)	IO (4)	N. analysts (5)	Log(Q-value) (6)	Share turn (7)	Option turn (8)
$\mathbf{1}_{[\text{Hacked}]}$	0.000	-0.000	0.002	0.084	0.088	-0.021	-0.010	-0.004
	(0.000)	(0.001)	(0.011)	(0.293)	(0.078)	(0.019)	(0.032)	(0.006)
N	43,687	43,687	43,687	43,687	43,687	35,273	43,687	43,687
Adjusted R^2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year-Quarter F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y	Y	Y	Y	Y

Table 2
Earnings Surprises and Soft Information Heatmap

Panel A shows the average after-hours announcement return by earnings surprise and soft information quintiles in a heatmap. Panel B reports the regression of after-hours returns onto earnings surprise and soft information. The sample period is from January 2010 to December 2015.

Panel A. Heatmap

Earnings surprise	Soft information 1	2	3	4	5
5	.011	.017	.023	.030	.039
4	-.002	.008	.013	.015	.022
3	-.011	-.002	.003	.005	.009
2	-.028	-.017	-.014	-.011	-.004
1	-.051	-.025	-.022	-.017	-.010
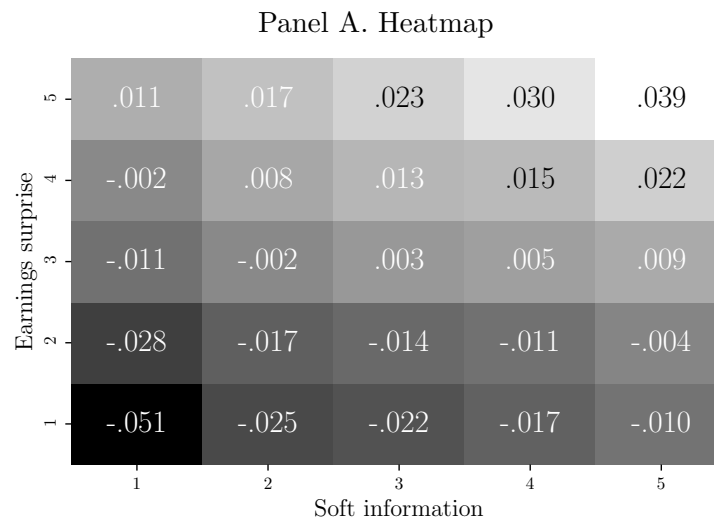
Panel B. Regression

	(1)	(2)	(3)
Surprise	1.313***		1.184***
	(0.048)		(0.051)
Soft		1.201***	1.047***
		(0.036)	(0.035)
Intercept	-0.002***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)
N	43,687	36,750	36,750
Adjusted R^2	0.061	0.068	0.115

46

Table 3
Informed Trading and Price Formation

This table reports the coefficients of the following regressions:

$$Return_{i,t}^{on} = \beta_1 Surprise_{i,t} + \beta_2 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \text{ in panel A},$$
$$Return_{i,t}^{on} = \beta_1 Soft_{i,t} + \beta_2 Soft_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \text{ in panel B},$$
$$Return_{i,t}^{on} = \beta_1 Surprise_{i,t} + \beta_2 Soft_{i,t} + \beta_3 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\beta_4 Soft_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \text{ in panel C}.$$

$Return_{i,t}^{on}$ corresponds to the log overnight return calculated using stock prices at 4 p.m. to 9:30 a.m. on the following day, $Surprise$ corresponds to the earnings surprise, $Soft$ corresponds to the soft information contained in the earnings press release, and $\mathbf{1}_{[\text{Hacked}]i,t}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPDR S&P 500 ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 2010 to December 31, 2015.

Panel A. Earnings surprises				
	(1)	(2)	(3)	(4)
$Surprise$	1.364***	1.436***	1.443***	1.426***
	(0.073)	(0.079)	(0.079)	(0.062)
$Surprise \times \mathbf{1}_{[\text{Hacked}]}$	-0.212**	-0.234**	-0.241**	-0.215**
	(0.103)	(0.103)	(0.104)	(0.101)
$\mathbf{1}_{[\text{Hacked}]}$	-0.001	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	43,687	43,687	43,687	43,687
Adjusted R^2	0.062	0.060	0.071	0.067
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

Table 3
Informed Trading and Price Formation (cont.)

Panel B. Soft information

	(1)	(2)	(3)	(4)
$Soft$	1.246***	1.352***	1.345***	1.334***
	(0.062)	(0.065)	(0.063)	(0.045)
$Soft \times \mathbf{1}_{[Hacked]}$	-0.184*	-0.220**	-0.222**	-0.210**
	(0.097)	(0.099)	(0.100)	(0.083)
$\mathbf{1}_{[Hacked]}$	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
N	36,750	36,750	36,750	36,750
Adjusted R^2	0.069	0.062	0.072	0.068
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

Panel C. Earnings surprises and soft information

	(1)	(2)	(3)	(4)
$Surprise$	1.225***	1.326***	1.336***	1.308***
	(0.071)	(0.081)	(0.081)	(0.064)
$Soft$	1.090***	1.192***	1.183***	1.176***
	(0.060)	(0.060)	(0.059)	(0.042)
$Surprise \times \mathbf{1}_{[Hacked]}$	-0.174	-0.199*	-0.207*	-0.144
	(0.111)	(0.111)	(0.112)	(0.113)
$Soft \times \mathbf{1}_{[Hacked]}$	-0.179**	-0.202**	-0.202**	-0.193**
	(0.088)	(0.091)	(0.092)	(0.080)
$\mathbf{1}_{[Hacked]}$	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
N	36,750	36,750	36,750	36,750
Adjusted R^2	0.115	0.110	0.120	0.116
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

Table 4
Unbiasedness Regressions

This table reports the coefficients of the following regression:

$$Return_{i,t\rightarrow t+1} = \beta_1 Return_{i,t}^{12-4PM} + \beta_2 Return_{i,t}^{12-4PM} \times \mathbf{1}_{[Hacked]i,t} +$$
$$\beta_3 \mathbf{1}_{[Hacked]i,t} + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where the dependent variable $Return_{i,t\rightarrow t+1}$ corresponds to the log return from 12 p.m. on day t to 9:30 a.m. on the following trading day in column (1) and from 12 p.m. on day t to 4 p.m. on the following trading day $t+1$ in column (2) of announcement i. All earnings announcements occur outside of regular trading hours, between 4 p.m. on day t and 9:30 a.m. on the following trading day. $\mathbf{1}_{[Hacked]i,t}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPDR S&P 500 ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. The sample period is from January 2010 to December 31, 2015.

	Dependent variable return window	
	12 p.m.(t)-9:30 a.m.(t+1)	12 p.m.(t)-4 p.m.(t+1)
	(1)	(2)
$Return^{12-4PM}$	0.776***	0.591***
	(0.029)	(0.038)
$Return^{12-4PM} \times \mathbf{1}_{[Hacked]}$	0.130***	0.142**
	(0.044)	(0.059)
$\mathbf{1}_{[Hacked]}$	-0.001	-0.001
	(0.001)	(0.001)
N	43,687	43,687
Adjusted R^2	0.082	0.057
Controls	Y	Y
Year-Quarter F.E.	Y	Y
Firm F.E.	Y	Y

Table 5
Signal Strength and Price Formation

This table reports the coefficients of the following regressions:

$$Return_{i,t}^{on} = \beta_1 Surprise_{i,t} + \beta_2 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \text{ in panel A and}$$
$$Return_{i,t}^{on} = \beta_1 Soft_{i,t} + \beta_2 Soft_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \text{ in panel B,}$$

where $Return_{i,t}^{on}$ corresponds to the log overnight log return calculated using stock prices at 4 p.m. to 9:30 a.m. on the following day, $Surprise$ corresponds to the earnings surprise, $Soft$ corresponds to the soft information contained in the earnings press release, and $\mathbf{1}_{[\text{Hacked}]i,t}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPDR S&P 500 ETF. We present the results for high (above median) and low (below median) absolute soft information in panel A and high and low absolute surprises in panel B. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 2010 to December 31, 2015.

Panel A. Price formation of surprise for high and low absolute soft

	High |soft|		Low |soft|	
	(1)	(2)	(3)	(4)
$Surprise$	1.759***	1.763***	1.291***	1.314***
	(0.119)	(0.119)	(0.108)	(0.107)
$Surprise \times \mathbf{1}_{[\text{Hacked}]}$	-0.334**	-0.329**	-0.215	-0.245
	(0.164)	(0.162)	(0.150)	(0.153)
$\mathbf{1}_{[\text{Hacked}]}$	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.001)	(0.001)	(0.001)
N	18,375	18,375	18,375	18,375
Adjusted R^2	0.076	0.087	0.048	0.061
Controls	N	Y	N	Y
Year-Quarter F.E.	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y

Panel B. Price formation of soft for high and low absolute earnings surprises

	High |surprise|		Low |surprise|	
	(1)	(2)	(3)	(4)
$Soft$	1.638***	1.630***	0.934***	0.925***
	(0.091)	(0.090)	(0.058)	(0.058)
$Soft \times \mathbf{1}_{[\text{Hacked}]}$	-0.336**	-0.333**	-0.126	-0.134
	(0.131)	(0.131)	(0.122)	(0.123)
$\mathbf{1}_{[\text{Hacked}]}$	-0.001	-0.001	0.000	0.000
	(0.002)	(0.002)	(0.001)	(0.001)
N	18,375	18,375	18,375	18,375
Adjusted R^2	0.079	0.084	0.037	0.064
Controls	N	Y	N	Y
Year-Quarter F.E.	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y

50

Table 6
Informed Trading, Order Flow, and Spreads

This table reports the coefficients of the following regression:

$$Outcome_{i,t} = \beta\mathbf{1}_{[Hacked]i,t} + \Gamma'Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.$$

$Outcome_{i,t}$ corresponds to one of the dependent variables represented in each column and $\mathbf{1}_{[Hacked]i,t}$ is a dummy equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. See section A1 in the appendix for the dependent variable definitions. All variables except for (log) stock and option volume are rescaled to have a standard deviation of one. Panels A and B present the results for the afternoon (12 p.m. to 4 p.m.) and morning (9:30 a.m. to 12 p.m.) trading sessions, respectively. Panel C presents tests of differences for the morning and afternoon coefficients in panels A and B. All regressions include controls for log market capitalization, fraction of shares held by institutional investors, natural logarithm of number of analysts, natural logarithm of newswire stories about a firm in the quarter leading to the announcement, daily cost of borrowing from Markit, and inverse of the stock price. The sample period is from January 2010 to December 31, 2015. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Afternoon (12 p.m. to 4 p.m.)

	Order flow measures				Spread measures			
	Share turn	Log(share vol)	|OI|	Log(option vol)	Effective spread	Realized spread	Price impact	Quoted spread
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbf{1}_{[Hacked]}$	0.0490***	0.0350**	0.0080	0.0721**	0.0312***	0.0445***	0.0153	-0.0140
	(0.02)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.02)
N	43,687	43,687	43,687	43,687	43,687	43,687	43,687	43,687
R^2	0.021	0.038	0.013	0.054	0.153	0.034	0.128	0.022
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Morning (9:30 a.m. to 12 p.m.)

	Order flow measures				Spread measures			
	Share turn	Log(share vol)	|OI|	Log(option vol)	Effective spread	Realized spread	Price impact	Quoted spread
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbf{1}_{[Hacked]}$	-0.0092	-0.0176	-0.0099	0.0331	0.0063	0.0127	-0.0137	0.0178*
	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
N	43,687	43,687	43,687	43,687	43,687	43,687	43,687	43,687
R^2	0.017	0.033	0.011	0.046	0.100	0.030	0.024	0.101
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Year-Quarter F.E.	Y	Y	Y	Y	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y	Y	Y	Y	Y

Panel C: Difference Afternoon-Morning

	Order flow measures				Spread measures			
	Share turn	Log(share vol)	|OI|	Log(option vol)	Effective spread	Realized spread	Price impact	Quoted spread
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Difference	0.0582	0.0526	0.0179	0.0390	0.0249	0.0318	-0.0003	-0.0026
P-value	0.0049	0.0123	0.3537	0.3533	0.1201	0.1594	0.9915	0.8545

51

Internet Appendix to
Price Revelation from Insider Trading: Evidence from Hacked Earnings News

Intended for online publication.

A1 Variable Definitions

We describe below the variables used in the manuscript.

- **Absolute order imbalance ($|OI|$):** We define the absolute order imbalance as $\left|\frac{B-S}{B+S}\right|$, where B and S correspond to the number of traded shares that are buy and sell market orders, respectively. The trades are signed using the Lee and Ready (1991) algorithm.

- **Effective spread:** The percent effective spread is calculated per trade, k, as $2d_k(log(P_k) - log(M_k))$. d_k a trade indicator that equals +1 if trade k is a market order buy and -1 if it is a market order sell. P_k is the trade price of trade k, and M_k is the midpoint of the prevailing NBBO quotes when trade k occurs. Following Holden and Jacobsen (2014), we aggregate the effective spread to the daily level by taking the dollar-volume weighted average of the effective spread across all trades for the morning (9:45 a.m. to 12 p.m.) and afternoon (12 p.m. to 4 p.m.) periods.

- **Institutional ownership (IO):** The fraction of shares held by institutional investors.

- **Log MCAP:** Natural logarithm of the firms' market capitalization.

- **Log option volume:** Natural logarithm of one plus the number of traded options.

- **Log Q-value:** Natural logarithm of the firms' Q-value. Using the corresponding Compustat variables, Q-value is calculated as the ratio of firm value to the sum of physical capital ($prcc\_f \times csho + dltt + dlc - act)/ppegt$.

- **Log share volume:** Natural logarithm of the number of traded shares.

- **Number of analysts:** The number of analysts issuing quarterly earnings forecasts in I/B/E/S.

- **Option turnover:** The number of options traded divided by the number of shares outstanding.

- **Price impact:** The percent price impact is calculated per trade, k, as $2d_k(log(M_{k+5}) - log(M_k))$. d_k a trade indicator that equals +1 if trade k is a market order buy and -1 if it is a market order sell. M_k is the midquote of the prevailing NBBO quotes when trade k occurs and M_{k+5} is the prevailing midquote 5 minutes after trade k. Following Holden and Jacobsen (2014), we aggregate the price impact to the daily level by taking the dollar-volume weighted average of the price impact across all trades for the morning (9:45 a.m. to 12 p.m.) and afternoon (12 p.m. to 4 p.m.) periods. The price impact presented in the Internet Appendix uses a prevailing midquote 1 minute after trade k.

- **Quoted spread:** The percent quoted bid-ask spread is $log(A_t) - log(B_t)$, where A_t is the National Best Ask quoted at time t and B_t is the National Best Bid quoted at time t. Following Holden and Jacobsen (2014), we aggregate the quoted spread to the daily level for the morning (9:45 a.m. to 12 p.m.) and afternoon (12 p.m. to 4 p.m.) periods by taking the weighted average of the intraday quoted spreads, where the weights on each observed intraday spread are the amount of time the spread is in play.

A1

- **Realized spread**: The percent realized spread is calculated per trade, k, as $2d_k(log(P_k)-log(M_{k+5}))$. d_k a trade indicator that equals +1 if trade k is a market order buy and -1 if it is a market order sell. P_k is the trade price of trade k, M_k is the midquote of the prevailing NBBO quotes when trade k occurs, and M_{k+5} is the prevailing midquote 5 minutes after trade k. Following Holden and Jacobsen (2014), we aggregate the realized spread to the daily level by taking the dollar-volume weighted average of the realized spread across all trades for the morning (9:45 a.m. to 12 p.m.) and afternoon (12 p.m. to 4 p.m.) periods. The realized spread presented in the Internet Appendix uses a prevailing midquote 1 minute after trade k.

- **Share turnover**: The number of shares traded divided by the number of shares outstanding.

- **Soft**: A measure of soft information computed using textual analysis of earnings press releases. See Section 3.2 for details.

- **Surprise**: The earnings surprise measure is computed as $\frac{EPS_{i,t}-E_{t-1}[EPS_{i,t}]}{P_{i,t-5}}$, where $EPS_{i,t}$ is the earnings per share of earnings announcement i for a firm announcing on day t, and $E_{t-1}[EPS_{i,t}]$ is the expectation of earnings per share, measured by the consensus analyst forecast in I/B/E/S. We define the consensus analyst forecast as the median of all analyst forecasts issued over the 90 days before the earnings announcement date.

A2 Construction of soft information measure

Hackers had access not only to quantitative information but also to the full press releases that make up earnings announcements. We call the information contained in the text of the press releases *soft information*, as opposed to the *hard information* that is the earnings surprise. To assess the impact of the soft information on hackers' actions and on their impact on price discovery, we build a measure of soft information using an off-the-shelf machine learning algorithm called ElasticNet, which is a form of regularized linear regression. We describe below how we retrieved the earnings press releases and the machine learning technique used to measure the soft information content of the press releases.

A2.1 Press releases

We obtain press releases from Edgar by scraping all Exhibits 99 and 99.1 to form 8-K filed within one day of an earnings announcement. We merge our announcement data with Edgar based on CIK code from Compustat. To filter documents that are not earnings press releases, we exclude those missing the word "quarter" and those that include statements linked to other corporate filings:

- "conference call script",

- "conference call transcript",

- "conference call participants",

- "corporate participants",

- "financial results conference call",

- "conference call script",

- "conference call notes",

- "transcript of earnings release conference call",

- "public offering common stock",

- "names new director".

Finally, we manually review all events for which we are left with more than one document. Our final sample comprises 36,750 press releases. We next clean the resulting press releases by applying the following steps:[20]

1. remove all HTML tags,

2. tokenize (convert everything to lower cap, keeping only words, i.e., drop numbers and punctuation),

3. remove English stop words (frequent words such as "a" and "the"),

4. stem the words (remove the ending) using the Porter Stemming Algorithm,

5. of the remaining words, keep only the first 400,

6. remove words that appear in less than 0.5% or more than 40% of the sample.

Overall, we are left with 2,393 unique "tokens", i.e., word parts after stemming.

A2.2 Machine learning

Our measure is estimated using an ElasticNet regression, which is a regularized linear regression that combines both L_1 (i.e., LASSO) and L_2 (i.e., Ridge) penalties. Similar to OLS, an ElasticNet regression aims to estimate the β coefficient of a linear system of equations:

$$y = X\beta. \tag{5}$$

Our objective is to predict market returns to an announcement, and thus we use announcement returns from 12 p.m. in the trading session before the announcement to the opening of the following trading day as the dependent variable, y. X is a matrix with each column corresponding to one token, and each entry is the number of occurrences of that token in the corresponding text.

The difference with OLS lies in the objective function used to estimate $\hat{\beta}$:

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}} \left(\sum_{i=1}^{N} (y_i - X_i\beta)^2 + \lambda_1 \sum_{j=1}^{M} |\beta_j| + \lambda_2 \sum_{j=1}^{M} \beta_j^2 \right) \tag{6}$$

The first term corresponds to the sum of squared residuals used in OLS, the second term is the L_1 penalty based on the absolute value of each coefficient, and the third term is the L_2 penalty based on the square of

---

[20]These steps were done in Python using the following packages: html, BeautifulSoup, NLTK, and scikit-learn.

A3

each coefficient. We set $\lambda_1 = \lambda_2$ such that both penalties have the same weight. Regularization is necessary because the number of columns in the X matrix is very large (2,393; one for each token).

While regularization helps with over-fitting, we also use 5-fold cross-validation. This process splits our sample into five sub-samples of equal size. For each sub-sample, we evaluate the performance of the $\hat{\beta}$ estimated using the other four sub-samples and then pick the model that performs best "out-of-sample". We then use this model to compute the fitted value for all announcements, and the result is our measure of soft information.

The final $\hat{\beta}$ estimates can be interpreted as in an OLS regression; in this case, they correspond to the association between a specific token and announcement returns. Negative coefficients correspond to "negative words", and vice versa for positive. To illustrate this point, Figure 3 in the main text shows word clouds for tokens with negative and positive weights for which the size of the token is proportional to the magnitude of its weight.

A3   Supplementary Figures and Tables

Figure A1. Evidence of Hackers "Shopping" List

This figure shows an example taken from SEC court filings of a "shopping list" for press releases sent to the hackers by traders in regards to upcoming earnings announcements.

Figure A2. Soft Information Measure: Robustness Tests

This figure shows the coefficient β_3 in Panel A and the R^2 in Panel B of the following regression:

$$Return_{i,t}^{ON} = \beta_1 Soft_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Soft_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Return_{i,t}^{on}$ corresponds to the log after-hours return calculated using stock prices from 4 p.m. on day t to 9:30 a.m. on the following trading day, $Soft$ corresponds to the soft information contained in the earnings press release, and $\mathbf{1}_{[\text{Hacked}]i,t}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The gray bars correspond to the 95% confidence intervals. The soft information is computed using machine learning (ElasticNet) and the results are presented for different choices of parameters to demonstrate the robustness. Specifically, the choice of parameters includes the number of words chosen from the earnings press releases (starting at the beginning of the press releases; see x-axis) and the minimum and maximum frequencies of word appearance across all press releases (see the legend).

Panel A. β_3

Panel B. R^2

A6
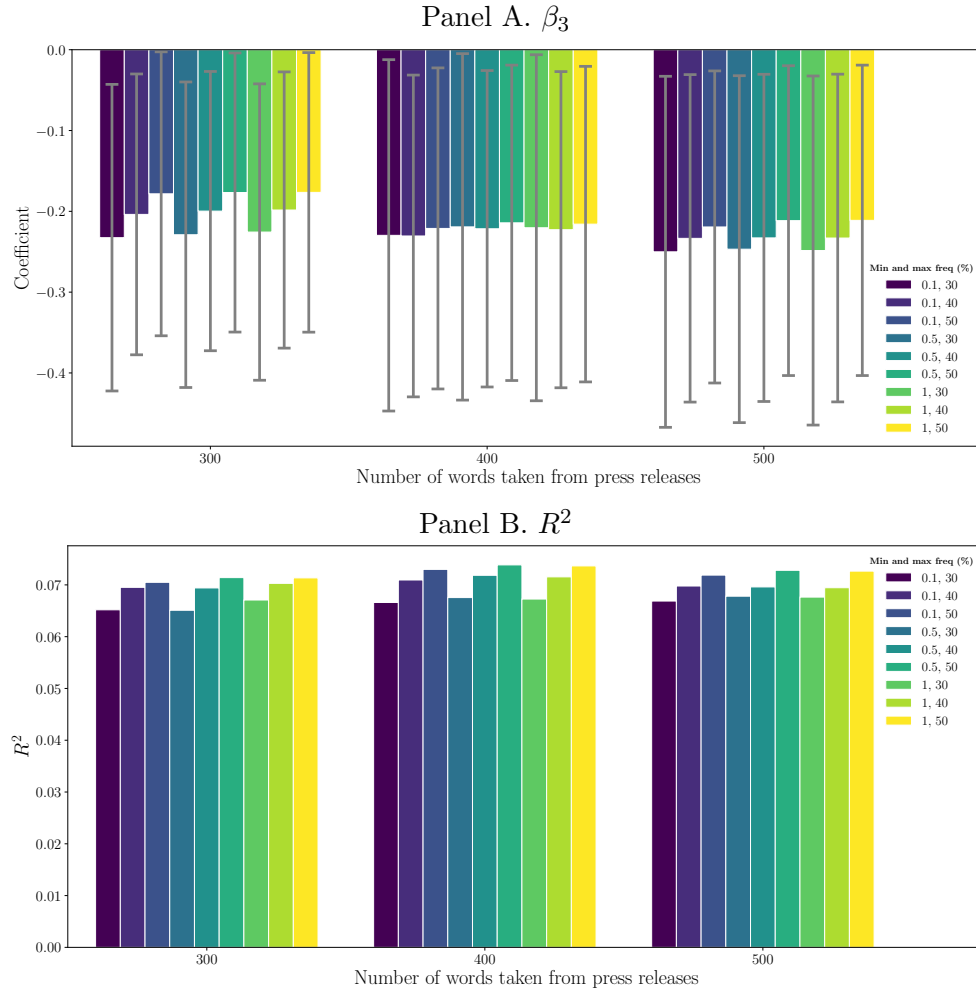
Table A1
Informed Trading and Overnight Price Formation:
Winsorize Earnings Surprise at 0.5%

This table reports the coefficients of the following regression in panel A:

$$Return_{i,t}^{on} = \beta_1 Surprise_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

where $Return_{i,t}^{on}$ corresponds to the log overnight return calculated using stock prices at 4 p.m. to 9:30 a.m. on the following day, $Surprise$ corresponds to the earnings surprise winsorized at 0.5%, and $\mathbf{1}_{[\text{Hacked}]i,t}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPDR S&P 500 ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively. The sample period is from January 2010 to December 31, 2015.

	(1)	(2)	(3)	(4)
Surprise	0.988***	1.034***	1.046***	1.032***
	(0.060)	(0.065)	(0.065)	(0.052)
Surprise $\times \mathbf{1}_{[\text{Hacked}]}$	-0.195**	-0.207**	-0.211**	-0.184**
	(0.082)	(0.084)	(0.085)	(0.083)
$\mathbf{1}_{[\text{Hacked}]}$	-0.001	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	43,687	43,687	43,687	43,687
Adjusted R^2	0.050	0.048	0.060	0.055
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

Table A2
Informed Trading and After Hours Price Formation
Piecewise Regression – Earnings Surprises

This table reports the coefficients of the following regression:

$$Return_{i,t}^{on} = \beta_1 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\beta_4 \mathbf{1}_{[\text{Surprise qnt} \in [2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.$$

$Return_{i,t}^{on}$ corresponds to the log after-hours return calculated using stock prices from 4 p.m. on day t to 9:30 a.m. on the following trading day. $\mathbf{1}_{[\text{Surprise qnt=5}]i,t}$, $\mathbf{1}_{[\text{Surprise qnt} \in [2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Surprise qnt=1}]i,t}$ are dummy variables equal to one if the earnings surprise belongs to the top, second to fourth, and bottom earnings surprise quintiles, respectively, and zero otherwise. $\mathbf{1}_{[\text{Hacked}_{i,t}]}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPY ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. 724 cases of actual informed trading before earnings announcements reported in SEC court filings are excluded from this analysis. The sample period is from January 2010 to December 31, 2015.

	(1)	(2)	(3)	(4)
$\mathbf{1}_{[\text{Surprise qnt=5}]}$	0.026***	0.028***	0.027***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Surprise qnt=1}]}$	-0.030***	-0.030***	-0.031***	-0.030***
	(0.002)	(0.002)	(0.002)	(0.001)
$\mathbf{1}_{[\text{Surprise qnt=5}]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.004**	-0.004*	-0.004*	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
$\mathbf{1}_{[\text{Surprise qnt} \in [2,3,4]]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.001*	-0.001*	-0.002**	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Surprise qnt=1}]} \times \mathbf{1}_{[\text{Hacked}]}$	0.003	0.004*	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
N	43,687	43,687	43,687	43,687
Adjusted R^2	0.097	0.099	0.107	0.103
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

A8

Informed Trading and After Hours Price Formation
Piecewise Regression – Soft Information

This table reports the coefficients of the following regression:

$$Return_{i,t}^{on} = \beta_1 \mathbf{1}_{[\text{Soft qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Soft qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Soft qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\beta_4 \mathbf{1}_{[\text{Soft qnt}\in[2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Soft qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma'Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.$$

$Return_{i,t}^{on}$ corresponds to the log after-hours return calculated using stock prices from 4 p.m. on day t to 9:30 a.m. on the following trading day. $\mathbf{1}_{[\text{Soft qnt=5}]i,t}$, $\mathbf{1}_{[\text{Soft qnt}\in[2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Soft qnt=1}]i,t}$ are dummy variables equal to one if the soft information belongs to the top, second to fourth, and bottom soft information quintiles, respectively, and zero otherwise. $\mathbf{1}_{[\text{Hacked}_{i,t}]}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPY ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. 724 cases of actual informed trading before earnings announcements reported in SEC court filings are excluded from this analysis. The sample period is from January 2010 to December 31, 2015.

	(1)	(2)	(3)	(4)
$\mathbf{1}_{[\text{Soft qnt=5}]}$	0.015***	0.015***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Soft qnt=1}]}$	-0.025***	-0.026***	-0.025***	-0.025***
	(0.002)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Soft qnt=5}]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.004***	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.002)	(0.002)
$\mathbf{1}_{[\text{Soft qnt}\in[2,3,4]]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.005***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Soft qnt=1}]} \times \mathbf{1}_{[\text{Hacked}]}$	0.002	0.003	0.003	0.002
	(0.003)	(0.003)	(0.003)	(0.002)
N	36,750	36,750	36,750	36,750
Adjusted R^2	0.052	0.043	0.053	0.050
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

Table A4
Informed Trading and After Hours Price Formation with Exclusion of Prosecuted Hacks – Earnings Surprises

This table reports the coefficients of the following regression in panel A:

$$Return_{i,t}^{on} = \beta_1 Surprise_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

and the coefficients of the following regression in panel B:

$$Return_{i,t}^{on} = \beta_1 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\beta_4 \mathbf{1}_{[\text{Surprise qnt} \in [2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.$$

$Return_{i,t}^{on}$ corresponds to the log after-hours return calculated using stock prices from 4 p.m. on day t to 9:30 a.m. on the following trading day. $Surprise_{i,t}$ corresponds to the earnings surprise i occurring after the trading session on day t, $\mathbf{1}_{[\text{Hacked}_{i,t}]}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise and $\mathbf{1}_{[\text{Surprise qnt=5}]i,t}$, $\mathbf{1}_{[\text{Surprise qnt} \in [2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Surprise qnt=1}]i,t}$ are dummy variables equal to one if the earnings surprise belongs to the top, second to fourth, and bottom earnings surprise quintiles, respectively, and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPY ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. 724 cases of actual informed trading before earnings announcements reported in SEC court filings are excluded from this analysis. The sample period is from January 2010 to December 31, 2015.

	Panel A. Linear estimation			
	(1)	(2)	(3)	(4)
Surprise	1.364***	1.437***	1.443***	1.427***
	(0.073)	(0.079)	(0.079)	(0.062)
Surprise $\times \mathbf{1}_{[\text{Hacked}]}$	-0.243**	-0.267**	-0.270***	-0.245**
	(0.100)	(0.105)	(0.105)	(0.100)
$\mathbf{1}_{[\text{Hacked}]}$	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
N	42,963	42,963	42,963	42,963
Adjusted R^2	0.062	0.061	0.072	0.068
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

A10

Table A4

Informed Trading and After Hours Price Formation with Exclusion of Prosecuted Hacks – Earnings Surprises (cont.)

Panel B. Piecewise estimation				
	(1)	(2)	(3)	(4)
$\mathbf{1}_{[\text{Surprise qnt}=5]}$	0.026***	0.028***	0.027***	0.027***
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Surprise qnt}=1]}$	-0.030***	-0.030***	-0.031***	-0.030***
	(0.002)	(0.002)	(0.002)	(0.001)
$\mathbf{1}_{[\text{Surprise qnt}=5]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.005***	-0.005**	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
$\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.001**	-0.002***	-0.002***	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Surprise qnt}=1]} \times \mathbf{1}_{[\text{Hacked}]}$	0.003	0.004**	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
N	42,963	42,963	42,963	42,963
Adjusted R^2	0.097	0.099	0.107	0.104
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

A11

Informed Trading and After Hours Price Formation with Exclusion of Prosecuted Hacks – Soft Information

This table reports the coefficients of the following regression in panel A:

$$Return_{i,t}^{on} = \beta_1 Soft_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Soft_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},$$

and the coefficients of the following regression in panel B:

$$Return_{i,t}^{on} = \beta_1 \mathbf{1}_{[\text{Soft qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Soft qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Soft qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\beta_4 \mathbf{1}_{[\text{Soft qnt} \in [2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Soft qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} +$$
$$\Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.$$

$Return_{i,t}^{on}$ corresponds to the log after-hours return calculated using stock prices from 4 p.m. on day t to 9:30 a.m. on the following trading day, $Soft_{i,t}$ corresponds to the press release i soft information occurring after the trading session on day t, $\mathbf{1}_{[\text{Hacked}_{i,t}]}$ is a dummy variable equal to one if announcement i is released by a newswire that was hacked at time t and zero otherwise, and $\mathbf{1}_{[\text{Soft qnt=5}]i,t}$, $\mathbf{1}_{[\text{Soft qnt} \in [2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Soft qnt=1}]i,t}$ are dummy variables equal to one if the soft information belongs to the top, second to fourth, and bottom earnings soft quintiles, respectively, and zero otherwise. The control variables are the log market capitalization and the product of the stock intraday return CAPM beta and the market return proxied by the SPY ETF. Fixed-effects are indicated in the table. Robust standard errors clustered by firm and year-quarter are presented in parentheses and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels respectively. 724 cases of actual informed trading before earnings announcements reported in SEC court filings are excluded from this analysis. The sample period is from January 2010 to December 31, 2015.

		Panel A. Linear estimation		
	(1)	(2)	(3)	(4)
$Soft$	1.246***	1.349***	1.341***	1.330***
	(0.062)	(0.065)	(0.063)	(0.045)
$Soft \times \mathbf{1}_{[\text{Hacked}]}$	-0.215**	-0.254***	-0.256***	-0.249***
	(0.096)	(0.097)	(0.098)	(0.085)
$\mathbf{1}_{[\text{Hacked}]}$	-0.001*	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
N	36,131	36,131	36,131	36,131
Adjusted R^2	0.069	0.062	0.072	0.068
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

Informed Trading and After Hours Price Formation with Exclusion of Prosecuted Hacks – Soft Information (cont.)

Panel B. Piecewise estimation				
	(1)	(2)	(3)	(4)
$\mathbf{1}_{[\text{Soft qnt=5}]}$	0.016***	0.015***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Soft qnt=1}]}$	-0.024***	-0.025***	-0.025***	-0.024***
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Soft qnt=5}]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.004***	-0.004***	-0.004***	-0.004**
	(0.001)	(0.001)	(0.001)	(0.002)
$\mathbf{1}_{[\text{Soft qnt} \in [2,3,4]]} \times \mathbf{1}_{[\text{Hacked}]}$	-0.001**	-0.002**	-0.002***	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
$\mathbf{1}_{[\text{Soft qnt=1}]} \times \mathbf{1}_{[\text{Hacked}]}$	0.002	0.003	0.003	0.003
	(0.002)	(0.003)	(0.002)	(0.002)
N	42,963	42,963	42,963	42,963
Adjusted R^2	0.043	0.035	0.046	0.042
Controls	N	N	Y	Y
Year-Quarter F.E.	Y	Y	Y	N
Firm F.E.	N	Y	Y	Y
Date F.E.	N	N	N	Y

A13