# Price Revelation from Insider Trading: Evidence from Hacked Earnings News

Pat Akey, Vincent Grégoire, Charles Martineau*

This draft: December 31, 2019

ABSTRACT

From 2010–2015, a group of convicted traders accessed earnings information hours before their public release by hacking several major newswire services. We use their "insider" trading as a natural experiment to investigate how efficiently markets incorporate private information in prices. 15% of a firms earnings surprise was incorporated into its stock price prior to its public release when the hackers had access to non-public information. Volume and spread-based measures of informed trading detect this activity, but order flow-based measures do not. We find evidence that uninformed professional traders traded in the same direction, amplifying the impact of informed trading.

*JEL Classification*: G10, G12, G14

*Keywords*: cyber risks, earnings announcements, insider trading, market price efficiency

*We thank Kenneth Ahern, Jaewon Choi, Peter Cziraki, Denis Gromb, Toshiki Honda, Vincent van Kervel, Sonya Lim, Tom McCurdy, Kazuhiko Ohashi, Andreas Park, Ioanid Rosu, Mike Simutin, Marius Zoican, conference participants at ESSFM Gerzensee (Informal Session), LBS Alumni Conference, NFA, Colorado Finance Summit, along with seminar participants at Dauphine University, Bank of Canada, Banque de France, the Financial Service Agency of Japan (Kinyuucho) HEC Paris, Hitotsubashi University, Tilburg University, University of Toronto, University of Illinois at Chicago, and the TD Management Data and Analytics Lab Research Roundtable Conference for comments and suggestions. We thank Jay Cao for his excellent research assistance. We acknowledge the financial support from the FinHub and the University of Toronto. Akey: University of Toronto, 105 St. George Street, Toronto ON, Canada, M5S 3E6 (pat.akey@rotman.utoronto.ca). Grégoire: HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, Montréal, QC H3T 2A7 (vincent.3.gregoire@hec.ca). Martineau: University of Toronto, 105 St. George Street, Toronto ON, Canada, M5S 3E6 (charles.martineau@rotman.utoronto.ca).

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 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. We use this event as a natural experiment to understand price formation in the presence of aggressive trading on illegally acquired private information by a small group of traders.

Fama (1970) defines strong-form efficiency as a market where private information is fully reflected in prices. This cyberattack episode provides a unique setting to study how efficiently markets detect the presence of private information and how quickly private information can be incorporated into prices prior to its public release. This setting allows us to isolate variation in the information set of a group of informed traders that is plausibly unrelated to firm fundamentals or firm-specific determinants of the efficiency of price discovery. More specifically, we compare the price discovery dynamics of a group of "treated" firms whose earnings were subject to potential informed trading by the cyberattack of their newswire to the price discovery of a "control" group of firms whose earnings were not subject to such potential informed trading.

We use this setting to study two questions. First, how much private information can a small amount of trade by retail traders reveal?[1] Existing literature has suggested that institutional trading can predict news or subsequent future returns (Hendershott, Livdan, and Schürhoff, 2015, Collin-Dufresne and Fos, 2015). However, the existing literature is silent on the question of *how much* information is revealed by informed trades and the speed at which this information is incorporated into prices. Our setting allows us to use a common set of events, earnings announcements, where "informed retail" traders had access to "inside"

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

information about some earnings announcements (e.g., earnings surprises) but not others.

We next ask how well the commonly used empirical measures of adverse selection in financial markets detect the presence of this informed trading. Collin-Dufresne and Fos (2015), Ahern (2018), Garriott and Riordan (2019), and Kacperczyk and Pagnotta (2018b) conclude that when informed traders (in a Kyle (1985) or Collin-Dufresne and Fos (2016) setting) can select when and how to trade, standard measures of adverse selection may fail to capture the presence of informed trading. However, in our setting, informed traders faced a short-time interval, at most a few hours, to execute their trades. Using the standard measures of adverse selection that are volume-based, order imbalance-based, and liquidity-based, we document which of these measures captures the risk of informed trading before scheduled public announcements.

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 point in time. In publicly available court documents, the U.S. government alleges that there were at least several hundred cases of known illegal trading and that the hackers sold earnings news broadly on a black-market website that was accessed by more than 100 individual traders.[2] Given that the universe of illegal trades is currently unavailable for study, we consider all earnings announcements between 2010 and 2015 with at least one analyst forecasts in I/B/E/S publicly released by one of the three hacked newswires for our analysis as potentially treated. We believe that our analysis captures a lower bound estimate of the effect of this illegal informed trading on prices.

We first find that informed trading in the hours prior to an earnings release predicts the earnings surprise. Specifically, we find that afternoon returns in the hours after press releases had been uploaded into a hacked newswire's IT system but before the news has been released

[2] We have 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.

to the public has a stronger ability to predict the quintile of the subsequent earnings surprise than pre-announcement afternoon returns for firms whose newswire had not been hacked. The effect is sizable; the increase in earnings surprise predictability is three times larger than the baseline predictability. Such increased predictability is not present in treatment firms in the *morning* returns on the day before an announcement return since firms generally tended to upload their earnings reports to newswire servers in the late morning or early afternoon, which further demonstrates that the hackers (as opposed to all traders) were specifically trading on non-public information about earnings surprises.

We next document that this informed trading impacted price discovery once the news became public, typically during the after-hours market. We find that overnight returns of firms whose newswires had been hacked are less responsive to earnings news. Specifically, we find that the overnight return sensitivity for a given earnings surprise is 15% lower if a firm's earnings release had been exposed to hackers. We find that this effect is strongest for large earnings surprises, both for large positive and negative earnings surprises, which is consistent with the informed traders trading most aggressively on the surprises that were most likely to earn them large profits.

We next examine how trading costs and signal precision interact with informed trading and the impact on overnight price discovery. We hypothesize that traders in possession of hacked press releases will choose to trade stocks that are "easier" to trade, i.e., more liquid and with fewer short-selling constraints. We use institutional ownership as a proxy of liquidity (Nagel, 2005) and show that the effect of insider trading is only found in stocks with high institutional ownership. We also find that the effect of hacked press releases on price formation is present only for earnings with at least five analyst forecasts. This suggests that the hackers traded on signals that were the most precise as higher analyst coverage tends to indicate lower information uncertainty (Zhang, 2006).

Having established that the hackers' trading was sufficiently large to impact price discovery, we use this scheme as a laboratory to test how well different measures of informed trading capture this behavior. We use numerous different measures of potential informed trading based on order flow and illiquidity. Specifically, we examine whether volatility, volume, turnover, absolute order imbalance, Volume-Synchronized Probability of Informed Trading (VPIN), the Amihud illiquidity measure, several measures of spreads or price impact, and the Kyle $\lambda$ are higher when a firm's newswire was hacked. We find that (log) volume, volatility, share turnover, effective spreads, and realized spreads are robustly higher when a firm's earnings were subject to informed trading. In contrast, we do not find that order imbalance measures, the Amihud measure, price impact, or Kyle's $\lambda$ capture this behavior.

Finally, we conclude by providing some evidence on *how* the hackers' information made its way into prices during afternoon trading. We examine whether sophisticated traders, e.g., high-frequency traders (HFTs), seem to infer the presence of increased trading and "lean with the wind" (Van Kervel and Menkveld (2019)) by increasing their own trading volume.[3] Specifically, we document that professional share turnover is more strongly associated with retail share turnover for earnings announcements that had been exposed to the hackers.

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 Bolandnazar, Jackson Jr, Jiang, and Mitts (2018) study the market response to private information at high-frequency for news (e.g., macroeconomic announcements and corporate SEC filings) that were publicly available for a subset of paid subscribers few seconds before their public releases. They find 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 that are not capable of

[3] Van Kervel and Menkveld (2019) show that following large institutional orders, HFTs initially lean against these orders and eventually take positions in the same direction for the most informed institutional orders. As documented in Barbon, Di Maggio, Franzoni, and Landier (2019), brokers might have played a role in spreading informed order flow information in the stock market.

processing information at microsecond speed to "snipe" inefficient stale quotes, but rather traded on private information using market orders and derivatives over few hours. As in Hu, Pan, and Wang (2017), we examine how much of that private information is reflected into prices before the announcements by examining the conditional mean response of prices to news using an ex-ante, well-defined measure of surprise. In contrast with this study, we focus on firm-level news rather than macroeconomic news.

Our study further complements but is distinct 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 (2018a) retrieves a list of 453 cases of SEC investigations about insider trading[4] for which the authors know the personal background of the insiders and their actual trades and find almost 70% of insiders do not split their trades as in Kyle but are more likely to split when private signals are less powerful and when legal risks increases. The findings of Shkilko (2018) suggest that insiders primarily focus on return timing rather than liquidity timing. Our setting differs because our insiders acquired private information for a large number of firms, which only had "value" for few hours, until the closing of markets. Finally, a recent literature looks at the insider trading of individual firms that are subject to cyber attacks. Mitts and Talley (2018) examines insider trading before the public disclosure of cyberattacks of *individual firms*, while Berkman, Jona, Lee, and Soderstrom (2019) examine potential informed trading before earnings announcements for firms that do not discuss cyber security practices in their 10-K filings. 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 since we can observe the (stolen) information set of the informed traders and study how price impact differs for different types of firms or earnings.

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

Finally, our paper relates to a recent literature that examines the performance of insider trading measures. Collin-Dufresne and Fos (2015) and Kacperczyk and Pagnotta (2018b) 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 (2018) makes a similar point and suggests that understanding the "urgency" of informed traders to trade on their private information is crucial to detecting their presence. We complement these studies by providing evidence that liquidity providers *can* respond to potential informed trading 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 (e.g., Kyle (1985)) where traders anticipate this behavior and strategically time their trades so as to minimize price impact.

## 1 Empirical setting: The newswire hacks

In order 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 (2019) and Martineau (2019)). We use the (standardized) difference between realized earnings and analysts' earnings forecasts for a given firm as a source of material news to the market.

Firms provide earnings announcements to newswire services several hours before news is to be released. These announcements are typically uploaded to a newswire's IT systems at least several hours before the news is released publicly. 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. Since 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.

Between 2010 and 2015 a group of hackers and individual traders (we refer to them collectively as "the hackers" as both the hackers and the traders were prosecuted for illegal trading) engaged in a series of cyberattacks against the three largest newswires, Business Wire, Marketwired, and PR Newswire.[5] The hackers used sophisticated methods to attempt to penetrate the IT systems of the 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

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

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 that they wanted to trade on 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 that were 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.[6] Figure 1 depicts the parties to the scheme and their relationship to each other. Figure 2 presents a timeline of the successful hacks and number of earnings news that was potentially exposed to the hackers based on our earnings announcement sample.

Using publicly available court documents, we are able to identify the specific trades for a small subset of the informed trades that the group of hackers made.[7] Table 1 provides some examples of the information and trading strategies that the hackers used. The evidence provided in the table shows that firms typically provided earnings news at least several hours before the close of markets and that the hackers typically began trading in the afternoon. In these examples, the hackers always transacted in the equity and frequently transacted using derivative contracts. In Figure 3, we graphically illustrate two examples of their trading strategies on the day when the earnings news was uploaded but had not yet been released and their two-day dollar profits. The solid blue line plots the stock price over the day. The vertical solid gray bars indicate trading in the stock by the traders, while the dotted gray bars indicate trading in derivatives by the traders. The dotted red line tabulates the cumulative trading profits associated with taking a position at a given time and holding it until the close of the next day. These figures illustrate that the traders profited on both positive and negative news by taking appropriately bullish or bearish positions in both equities and

[6] Augustin, Brenner, and Subrahmanyam (2019) shows that traditional insiders the U.S. also trade in option markets before merger and acquisition announcements.

[7] The full list of trades that the government used in their prosecutions against the hackers is not publicly available. We filed several Freedom of Information Act requests that have been denied. Our analysis is based on publicly available documents from PACER.

derivatives. In contrast with most inside trades that typically yield small dollar profits (Cziraki and Gider, 2018), these hackers earned large profits, collectively in excess of 100 million dollars, including several events that yielded more than a million dollars 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 programmers that hacked the newswire companies and the traders that orchestrated the financial transactions using the stolen information. In total, 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

## 2.1 Data sources

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) is extracted from the Trade and Quote (TAQ) database following the procedure in Holden and Jacobsen (2014). Institutional ownership data is from Thomson Reuters 13-F and option turnover data is from OptionMetric. The earnings announcement sample is from I/B/E/S and newswire information is from Ravenpack. We provide below detailed information on the earnings announcement sample construction and how we link each announcement to its newswire provider.

## 2.2 Earnings announcement sample and newswires

Our earnings announcement sample is from January 1, 2010 to December 31, 2015, which corresponds to the period that 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 quarter $q$ 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 CRSP. We then select the earnings announcements for stocks that have at least one analyst forecasts in I/B/E/S in the 90 days prior to the earnings announcement, 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).

An important step into our analysis is to correctly assign for each earnings announcement the corresponding newswire company responsible for disseminating 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 TAQ on the trading day prior to the earnings announcement and with no opening price in CRSP following the earnings announcement (2,140 observations).

Panel A of Figure 4 shows the proportion of earnings announcements by newswire. The three newswires that are the largest by market share, Business Wire, PR Newswire, and Marketwired, were subject to hacking, although at different time periods and for different lengths of time. We restrict our analysis to observations that use one of these newswires to ensure that our results are not influenced by firms that have selected to use a newswire that was not affected by a hack. In doing so, we identify results primarily off of the *timing* of when a firm's newswire is hacked. Our final sample consists of 43,991 earnings announcements.

Panel B of Figure 4 presents the proportion of earnings releases that were exposed by the different newswires. PR Newswire represents the largest fraction of earnings exposed

[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.

to hacks with 52.3% followed by Marketwired with 30.2% and Business Wire with 17.5%. Despite the fact that Marketwired is only the third largest newswire by market share, its earnings releases account for a large proportion of treatment observations because its IT systems were hacked for the longest time.

### 2.3 Descriptive statistics

The magnitude of the earnings surprise is key to our analysis. We make the assumption that the earnings surprise is the main source of information determining the likelihood and direction of trading for the hackers on these events. We follow Hartzmark and Shue (2017) 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. 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 and winsorize earnings surprises at the 1st and 99th percentiles.

Figure 5 and 6 show a probability plot of the earnings surprise against the quantile of a normal distribution (i.e., a quantile-quantile plot) and the average earnings surprise by earning surprise quintile and S&P index membership. These figures show that the distribution of earnings surprises has high kurtosis relative to a normal or t-distribution for both the pooled sample of firms and for several subgroups of firms. We estimate our main effects in a variety of ways that allow us to account for the leptokurtic nature of the distribution.

Table 2 presents summary statistics for our sample. The variables are the earnings surprise, the absolute value of the earnings surprise, year-end log market capitalization (prior to earnings announcement), the fraction of shares held by institutions, the number of analyst forecasts, the monthly Amihud illiquidity ratio[9], the log Q-value[10], the monthly share turnover (monthly trade volume in dollars divided by the market capitalization), and the monthly option turnover (monthly option trade volume divided by the shares outstanding).[11]

### 2.4 Evidence of identification

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 cases involve trading news related to a merger (Kacperczyk and Pagnotta (2018b), Ahern (2018)). 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 of its trading environment, in which case naive OLS regressions will yield a biased estimate. Third, in equilibrium, informed traders likely engage in strategic trading behavior so as to *minimize detection* (Collin-Dufresne and Fos (2015) and Ahern (2018)), thus it is important to find news events that have an equal "urgency" to trade. Finally, traders can choose to trade via market or limit orders, which will impact the ability of different classical measures of informed trading to detect this behavior.[12]

We believe that this hacking and illegal trading scandal is a useful empirical setting to

[9]The Amihud illiquidity ratio is calculated as $\frac{1}{D_{imy}} \sum_{d=1}^{D_{imy}} \frac{|R_{idmy}|}{VOLD_{idmy}} \times 10^6$, where $D_{i,m,y}$ is the number of $d$ days for which data are available for stock $i$ in month $m$ for year $y$, $|R_{idmy}|$ is daily absolute stock return and $VOLD$ is the total trade volume in dollars.

[10]Using the corresponding Compustat variables, the Q-value is calculated as the ratio of firm value to the sum of physical capital (prcc_f*csho+dltt+dlc-act)/ppegt.

[11]The Amihud illiquidity ratio, the share turnover, and option turnover are calculated on the month prior to the earnings announcement.

[12]Ahern (2018) makes a similar point about the difficulty of credibly identifying the effect of informed trading.

help overcome many of these empirical challenges. Since 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. This fact, combined with the fact that we only analyze firm outcomes on days before an earnings announcement means that we have counterfactual set of observations (earnings announcements that were released by non-hacked newswires). This clear treatment and control set of common events (earnings announcements) addresses the first empirical concern we raise above.

In order for our analysis to have a causal interpretation, these two sets of firms must not have been selected because of underlying differences between the two groups of firms. For example, if firms that used services that were hacked generally had a different distribution of earnings surprises or if these firms' price formation differed, an analysis of how price discovery changed in periods when the newswires were hacked could be confounded by these underlying differences. We provide evidence that the characteristics of firms whose information was subject to a hack do not systematically differ from those firms whose information was not subject to a hack, which addresses the second point we raise above.

Concerns about strategic trade timing, the third point that we raise above, are less salient in our setting. The information that the hackers had access to was scheduled to become public only hours after they acquired it and thus they did not have the ability to strategically time their trading as in many models of informed trading (e.g., Kyle (1985) and Collin-Dufresne and Fos (2016)). Finally, we discuss how the use of market and/or limit orders impacts our analysis of informed trading measures when we introduce those measures.

We now provide evidence that our setting is likely to have a causal interpretation. We first examine the industries of firms that use different newswire services. Figure 7 presents the proportion of earnings releases by different newswire companies split by industry. Aside from a higher proportion of observations of IT firms earnings in Marketwired, the distribution of industries appears relatively similar.

We next examine whether there are systematic differences between firm or earnings characteristics when their newswire is hacked. Specifically, we regress a variety of characteristics 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 A of Table 3 presents the results of this analysis in the cross section, while Panel B 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 since these variables are quite persistent. We find no evidence that Tobin's q is different and no evidence that firms face different information or trading environments. Firms that are hacked have similar liquidity, share turnover, option turnover, and analyst coverage.

# 3 Results

## 3.1 Does hackers' trading predict earnings surprises?

We begin our analysis by documenting whether potential informed trading by hackers can predict earnings surprises. As we described above, earnings releases were typically uploaded to newswire companies' servers several hours before they were to be released publicly in the after-hours market. The hackers, therefore, had the opportunity to trade on material, non-public information for a limited amount of time, typically in the afternoon before and often just before the close of the market. Indeed, the anecdotal examples that we provide in Table 1 suggests that the hackers typically began trading after 1 p.m. We test this hypothesis by examining whether the intraday return on the day before the earnings release could better predict the quintile of a firm's earnings surprise in time periods when its newswire had been hacked. Specifically, we run the following regression

$$\begin{aligned} \mathit{Surprise\,Quintile}_{i,t} =& \beta_1 \mathit{Log\,Return}_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 \mathit{Log\,Return}_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} \\ & + \Gamma' \mathit{Controls}_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \end{aligned} \tag{2}$$

where $\mathit{Surprise\,Quintile}_{i,t}$ is the (ordinal) quintile of earnings announcement $i$ for a firm announcing on day $t$. We prefer using a quantile of earnings surprise for this analysis since the distribution of earnings surprises exhibits significant kurtosis, as shown in Figure 5. We use earnings surprise quintiles in this analysis as it closely mirrors the empirical approach we will use in the rest of our analysis. In unreported results, we verify that we obtain similar patterns when predicting earnings surprise deciles instead of quintiles. $\mathit{Log\,Return}_{i,t}$ is the stock return of the firm computed over various windows on the day before the earnings news is released, 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. In some specifications we include firm size and the CAPM-predicted intraday return for the firm.[13]

Table 4 presents the results of this estimation. Columns (1) and (2) present estimates where $\mathit{Log\,Return}_{i,t}$ is computed over the entire day, columns (3) and (4) present estimates when $\mathit{Log\,Return}_{i,t}$ is computed for the morning return only (9:30 – 12:00), and columns (5) and (6) present estimates when the $\mathit{Log\,Return}_{i,t}$ is computed for the afternoon return only (12:00 – 16:00). We find across most specifications that higher daily returns have explanatory power for the soon-to-be-released earnings surprise. Moreover, we find that this relationship is *stronger* for days when the hackers had access to the IT systems of a firm's newswire and could, therefore, trade on this information. We find the effect is present when we consider returns over the entire day (columns (1) and (2)), as well as in the afternoon (columns (5) and (6)). The incremental effect is substantial. For example, in the column (2) the baseline coefficient is 1.976 and the interaction term is 2.661, suggesting that the ability of returns

[13]The CAPM-predicted return is the estimated CAPM beta of the stock times the market return proxied by the SPDR S&P 500 ETF. The CAPM beta is estimated using the previous 250 trading days.

to explain future earnings surprises increases by 135%. This increase is even stronger when we consider afternoon returns, the periods when the hackers were most actively trading. In column (6), for example, we see that the baseline effect is 2.684, and the interaction is 5.158. The stronger relationship is *not present* for the morning returns, the period in which the hackers often did not have access to the earnings surprises as they had not yet been uploaded. Thus the increased return predictability *only occurs* when the hackers were (ex-post) known to be trading on large earnings surprises. It seems from columns (5) and (6) that the baseline predictability is higher in the afternoon than the morning. We verify that this pattern exists even if we do not introduce variables related to the hacked newswires in columns (7) and (8) and confirm that afternoon predictability seems to be higher generally.

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

### 3.2.1 Baseline results

We now examine our main research question: how much private information can a small amount of trade by retail traders reveal? In order to examine this question, we examine whether a stock's overnight return reaction to an earnings surprise is less sensitive to the earnings surprise on events for which the hackers had access to the not-yet-public news. Existing literature suggests that price discovery related to earnings happens in the after-hours markets and occurs quickly (Grégoire and Martineau, 2019).

Figure 8 graphically depicts our main empirical 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 the close on the next trading day for S&P 1500 and non-S&P 1500 stocks in Panels A and B, respectively. The left panel shows returns for the top quintile of earnings surprises while the right panel shows returns for the bottom quintile of earnings surprises. For S&P 1500 stocks, 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. By the end of the next trading day, there is no significant difference between firms that have been hacked or not, which means that the total response to earnings news is the same. For non-S&P 1500 stocks, we observe a significant price drifts towards the post-announcement price in the hours before the announcement but only for positive earnings surprises. Shorting these stocks might be difficult which could explain why we observe no price drifts for negative earnings surprises.[14]

If the hackers' trades impacted prices, the overnight response for a given earnings surprise should be smaller. To test this hypothesis, we run the following regression:

$$\begin{aligned} Log\,Return\,ON_{i,t} =& \beta_1 Surprise_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} \\ &+ \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \end{aligned} \tag{3}$$

where $Log\,Return\,ON_{i,t}$ is the overnight (close-to-open) return around earnings announcement $i$ for a firm announcing on day $t$. $Surprise_{i,t}$ captures the unexpected earnings surprise, and $\mathbf{1}_{[\text{Hacked}]i,t}$ is, as previously defined, 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.

Panel A of Table 5 presents the results of this analysis. 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 8, we find a strong, positive relationship between the

[14] We note that there is a small price drift for hacked stocks following the opening of markets for firms outside of the S&P 1500, the least liquid stocks. Grégoire and Martineau (2019) show that price drifts following the opening of markets is largely due to price pressure generated by order flow going in the opposite direction of earnings surprises (e.g., pressure from marketable sell orders following positive earnings surprises). Consistent with the findings of Hendershott and Menkveld (2014), price pressure dampens price efficiency where liquidity providers accommodate order flow through a price concession (e.g., charge a lower bid price to incoming sell orders) before gradually relaxing the price concession (e.g., increasing the bid price) resulting in a price drift. We believe this price pressure explains the price drift in Panel B of Figure 8.

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 sensitive 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 (3) suggest that *for a given firm*, the return sensitivity to its earnings news declined by 16.4% in periods when the firm's newswire had been hacked (= -0.236/1.436). The point estimates are generally stable across different regression models and are not affected by the inclusion of control variables, suggesting that 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.[15]

We next perform a secondary analysis using a more flexible estimation approach. Equation 3 imposes a linear response to both the earnings surprise and the *differential* impact of informed trading on the surprise. However, we have previously shown that the underlying distribution of earnings news has a substantial mass in both the right and left tails of the distribution. Moreover, as informed trading is most profitable when the news will have the most impact, it is likely that the hackers more actively traded on larger news surprises. We perform a second estimation where we allow for the earnings response to vary by quintile of the earnings surprise. This estimation approach has two advantages. First, it minimizes the likelihood that outlier observations drive our results and second, it allows us to examine whether the informed trading impacted price formation differently for positive or negative

[15]In Table A1 in the Appendix, we present the results of an alternative price discovery regression similar to an unbiasedness regression. 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". The interactive coefficient is close to 0.15 and statistically significant.

news. More formally, we estimate the following regression:

$$\begin{aligned} Log\,Return\,ON_{i,t} =& \beta_1 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} + \\ & \beta_3 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_4 \mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \beta_5 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}, \end{aligned} \tag{4}$$

where $\mathbf{1}_{[\text{Surprise qnt=5}]i,t}$ is a binary variable that takes the value of one if the earnings surprise is in the top quintile and zero otherwise, $\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t}$ is a binary variable that takes the value of one if the earnings surprise is in the second, third, or fourth quintile and zero otherwise, $\mathbf{1}_{[\text{Surprise qnt=1}]i,t}$ is a binary variable that takes the value of one if the earnings surprise is in the bottom quintile and zero otherwise, and other variables are defined as before. The interpretation of these coefficients merits a brief explanation. $\beta_1$ and $\beta_2$ capture the average overnight return for an earnings surprise in the top and bottom quintile in a non-hacked time period, relative to an earnings surprise in the second to fourth quintile, which serves as the reference group. $\beta_3$ captures the *change* in return sensitivity for earnings surprises in the top quintile when a firm's earnings were subject to informed trading by the hackers. $\beta_4$ captures the same (differential) effect for earnings surprises in the middle three quintiles, and $\beta_5$ captures the effect for earnings surprises in the bottom quintile.

Panel B of Table 5 presents the results of this estimation. As in Panel A, columns (1) – (4) present estimates with increasing numbers of fixed effects and controls using the full sample. As expected, we find that there is a positive (negative) relationship between the top (bottom) quintile of earnings surprises and overnight returns. Across all specifications, we find that $\beta_3$ is negative and generally statistically significant. For example, in specification (3) we find that the coefficient is -0.004, which indicates that stock return sensitivity to the most positive quintile of news was 15.4% less sensitive to earnings surprises ( = -0.004/0.026) when the hackers had access to the IT systems of a firm's newswire. We similarly find that $\beta_5$ is positive and statistically significant across most specifications. For example, in specification

(3), we find that the coefficient is 0.004, which suggests that stock return sensitivity is 12.9% lower for the lowest quintile of news surprises ( = 0.004/-0.031) when the hackers had access to the IT systems of a firm's newswire. We find somewhat mixed evidence for the middle three quintiles of news surprises. The coefficients are negative across all specifications although the estimates are measured with less precision. The negative coefficients indicate that stock return sensitivity to the middle three news quintiles was somewhat smaller since both the mean and the median earnings surprise for this set of observations is positive (see Figure 6).[16]

### 3.3 Examples from prosecuted SEC cases

Our analysis so far has focused on isolating variation in the *ability* of traders to profit from non-public information by using variation in when a firm's newswire had been hacked. We have not used any of the events that formed the basis of the SEC's legal proceedings. It is reasonable to assume that the SEC's legal case consisted of the trading events that were the most likely to lead to a favorable legal outcome. We believe that inference focusing on variation in the ability of traders to profit off of non-public information is likely to give a lower bound on this effect similar to an intention-to-treat estimation, while focusing on the cases that formed the basis of the SEC's legal proceedings would most likely yield an estimation that would be the upper bound of potential pricing effects.

We verify this conjecture by examining the cases that formed the basis of the SEC's legal case. In order to do so, we retreived a large number of court documents from PACER related to the legal proceedings of these cases. We are able to identify 385 cases of actual trading by the hackers before earnings announcements, for which we can get all of the data necessary

[16] We repeat the same analysis of this section, separately for S&P 500, S&P 1500 (excluding S&P 500), and non-S&P 1500 stocks and present the results in Table IA.1 of Internet Appendix. We further redo the same analysis by excluding the 385 prosecuted SEC cases documented in the court documents and report the results in Table IA.2 of the Internet Appendix. We report similar effects.

for our analysis.[17] We match each known event of illegal trading to five earnings events from the set of earnings in the same quarter that were exposed to the hack but did not form part of the SEC's legal case and to five earnings events that were not exposed to the hack in the same quarter. We match using the total return over the event window within S&P index to identify events of similar news content and trading environments.

Figure 9 presents the cumulative returns from 12 p.m. on the trading session before the earnings announcement until the closing of markets on the following trading day for the top and bottom earnings surprises. The solid blue line ("Traded") plots the returns for firms that formed the basis of the SEC's case, the dotted red line ("Hacked") plots the returns of firms that were subject to informed trading but did not form the basis of this case, while the dashed black line ("Not hacked") plots the returns of firms that were not subject to potential informed trading. These results confirm that indeed, the price discovery effects in the cases that formed the basis of the SEC's case are substantially larger than what we previously documented. Moreover, there appears to be price discovery effects on some of the cases where there could have been informed trading (shown by the "Hacked" line) but was not part of the SEC's case.[18] These findings highlight that analysis made using ex-post prosecuted effects are likely to yield larger estimates that may not be representative of an average effect, a point made by Ahern (2017) and Ahern (2018).

### 3.4 Trading costs

We examine how the impact of informed trading on price formation varies with the level of institutional ownership. Higher levels of institutional ownership indicate better stock liquidity (Nagel, 2005). We rerun Equation 4 on subsamples that are split above and below the

[17] We attempted to get a fuller picture of the events that the SEC believed were associated with this scheme via several Freedom of Information Act requests with the SEC. Our requests were summarily denied. The SEC was unwilling to even provide the list of cases that we ultimately identified in publicly available documents.

[18] Media reports related to this case suggest that a number of traders that profited from this scheme are not easily prosecuted by U.S. law enforcement. These non-prosecuted cases that exhibit similar price discovery effects could be from these traders or others that the SEC found more difficult to prosecute.

median of institutional ownership (76% of shares held by institutions). Table 6 presents the results of this estimation.[19] Columns (1) and (2) present results for firms with high institutional ownership, while columns (3) and (4) present results for firms with low institutional ownership. We find that our results are strongest in the subsample of firms with higher levels of liquidity as measured by institutional ownership. For example, the results in column (2) are largely consistent with our previous results. We find that price discovery is 20% less responsive to positive news and 23% less responsive to negative news in periods when hackers had access to the IT systems of a firm's newswire, comparable to our baseline estimates (= -0.006/0.030 and = 0.009/-0.039, respectively). In contrast, we find little evidence that price discovery was impacted by potential insider trading in firms with lower institutional ownership. The coefficients in specifications (3) and (4) are insignificant, both statistically and economically.

### 3.4.1 Number of analysts

We examine how the effect of informed trading on price discovery varies by analyst coverage. We split the sample by the median number of analysts per event (5) and repeat our estimation of equation 4 in the samples of high and low analyst coverage. We conjecture that the impact of informed trading would be stronger for earnings news that is a surprise relative to an estimate made by more analysts. To the extent that professional analysts provide useful information, private information about an earnings surprise based off of a stock with higher analyst coverage would likely be more precise information about future stock prices than an earnings surprise based off of a stock with lower analyst coverage because of lower information uncertainty (Zhang, 2006). Moreover, pre-event trading on events with more analyst coverage might also be more revealing to uninformed but sophisticated traders.

Table 7 presents the results of this analysis. Columns (1) and (2) present estimates for

[19] In non-tabulated results, we find that stocks with above median institutional ownership have quoted, effective, and realized spreads that are three to four times smaller than stocks with below median institutional ownership during the regular-hour trading session prior to the earnings announcement.

earnings surprises with high analyst coverage, while columns (3) and (4) present estimates for earnings with low analyst coverage. We find evidence consistent with our conjecture. The negative effect of informed trading on price discovery is concentrated in earnings surprises with higher analyst following. For example, the estimates in column (2) suggest that price discovery is 25.9% less responsive to positive news and 21.6% less responsive to negative news in periods when hackers had access to the IT systems of a firm's newswire ( = -0.007/0.027 and = 0.008/-0.037, respectively). We find little evidence that there was a price impact of potentially informed trading for earnings surprises with low analyst coverage.

# 4 Did measures of informed trading detect the hackers' activity?

Recent studies have investigated the ability of informed trading measures to detect informed trading (Collin-Dufresne and Fos (2015), Kacperczyk and Pagnotta (2018b), Ahern (2018)). These studies suggest that many informed trading measures, particularly measures based on spreads, are *inversely* related to informed trading events, because informed traders strategically time their trades. Collectively, these studies find that volume and volatility are robustly correlated with informed trading activity, but that other measures are not.[20] We revisit the question of whether illiquidity and order flow measures are able to detect the presence of informed traders in the hours leading up to earnings announcements. As we have shown, we have plausibly exogenous variation in the information set of the hackers that had measurable price impact in a short window. We believe that our tests have higher power to test the ability of informed trading measures to detect informed trading prior to public announcements, holding the strategic element constant.

We consider several measures of informed trading that are based on order flow: volatility, volume, turnover, the absolute order imbalance, and the Volume-Synchronized Probability of Informed Trading (VPIN) of Easley, López de Prado, and O'Hara (2012). Additionally, we

[20] Ahern (2018) finds that absolute order imbalance is reliably correlated as well.

consider several measures that are based on liquidity: the Amihud illiquidity measure, quoted spreads, effective spreads, realized spreads, price impact, and the Kyle $\lambda$.[21] 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. We calculate the intraday measures of illiquidity using the NYSE Monthly and Daily Trades and Quotes (TAQ) database for the periods of 2010 to 2013 and of 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.[22] We further divide each variables by the cross-sectional standard deviation (except the log trading volume). We provide additional details on the construction of each measure in Appendix A1.

We estimate the following regression:

$$Informed_{i,t} = \beta \mathbf{1}_{[\text{Hacked}]i,t} + \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}, \tag{5}$$

where $Informed_{i,t}$ represents one of the 11 measures of informed trading 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.

The regression results are reported in Table 8. 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) volume have been standardized. Columns (1) – (3) show that trading or volume-based measures are reliably higher when $\mathbf{1}_{[\text{Hacked}]}$ is equal to one. Trading volume was roughly 3% higher, and volatility and turnover were roughly 4% of a standard deviation higher. Column (4) – (6) suggests that the Amihud measure of illiquidity, the absolute order imbalance, and the VPIN do not reliably detect

[21] The effective spread compares the execution (transaction) price to the midquote at the time of the trade. Realized spread can be thought of as the theoretical profits of a liquidity provider. It is the effective spread minus its price impact component.

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

this behavior.[23] Columns (7) – (9) suggest that some spread-based measures of illiquidity can detect this behavior. Specifically, we find that effective spreads increased by 3.1% 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 that this increase in effective spread comes from an increase in the realized spread rather than from a direct price impact of each trade. As we noted earlier, the ability of traders to choose either to trade with market orders or with limit orders can make the detection of informed trading difficult. If informed traders systematically trade with limit orders, or if liquidity providers manage their inventory risk, order imbalance may not reflect informed trading, whereas realized spreads will increase, which is what we document.[24] The results presented in Table 8 for realized spread and price impact uses a prevailing midquote 5 minute after a trade. Conrad and Wahal (2019) show that in today's financial markets the majority of the price impact of trades in large and small-capitalization stocks takes place within 60 seconds. We report in Table IA.3 of the Internet Appendix the results for effective and realized spreads and price impact using a prevailing midquote 1 minute after a trade and find similar results.

We use the fact that the hackers did not typically have access to material, non-public information 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 8 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 informed trading measures were higher for firms subject to a hacked newswire *in the morning* before

[23]We also constructed the VPIN using the signing of trades as in Ellis, Michaely, and O'hara (2000) and the results remain similar.

[24]Our results do not imply that individual trades did not have price impact, but rather that *on average* price impact of trades in the time periods when the hackers were active was not higher. Given that realized spreads *were* on average higher, liquidity providers seem able to detect the presence of informed trading and widen spreads over that time period. Indeed, we show in Figure A2 that that price impact does increase at the time of the hackers trades for two examples of cases that were prosecuted by the SEC for which we know when the hackers did trade (at the minute-frequency).

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

A surprising result 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 exchange and shows that price impact does increase. Consistent with Shkilko (2018), we show in Figure A2 of the Appendix that price impact does increase at the time of insider trading for two examples of prosecuted case by the SEC for which we know when the hackers did trade (at the minute-frequency). The figure also shows that price impact is generally negative, indicating an increase in the realized spreads.

A second surprising result from Panel A of Table 8 is that the absolute order imbalance is not significant. Kacperczyk and Pagnotta (2018b) and Ahern (2018) find that a positive association between absolute order imbalance and informed trading. To better understand our result, we examine the order imbalance for the seven cases for which we know when insider did trade from the SEC filings. Figure 10 shows the cumulative return and trade volume (right pane), the cumulative trade volume, cumulative insider trade volume, cumulative order imbalance from TAQ, and the cumulative order imbalance for trades executed on Nasdaq ITCH (middle pane), and the quoted spreads (right panel) at a one minute frequency from 9:30 a.m. to 4 p.m. for Align Technologies on 2013/10/17 and Edward Life Sciences on 2013/04/23 in Panel A and B respectively. These two cases are the same cases that we show in Figure 3. We examine the order imbalance using both TAQ and Nasdaq ITCH because while TAQ includes trades executed on all U.S. venues, we must infer trade signs using the Lee and Ready (1991) algorithm, which is not 100% accurate. Nasdaq ITCH includes only trades executed on the Nasdaq platform, but trades are perfectly signed.[25] The use of the Lee and Ready (1991) algorithm appears reasonable in our context as the cumulative OI computed from both data sources are very close. We present the other 5 case studies in the Internet Appendix. We observe similar findings.

Two important findings emerge from these figures. First, we note that there is a large

[25] For more detail on Nasdaq ITCH, see Grégoire and Martineau (2019).

increase in the overall trade volume in the stock following insider trades spikes, however, this increase in total volume is larger than what a ring of hacker-affiliated traders (who were potentially capital constrained) could have themselves traded. Indeed in these cases, the cumulative trade volume of the stock is substantially larger than the cumulative volume of the hackers that were prosecuted. Second, the cumulative order imbalance computed using both using TAQ and Nasdaq ITCH end the day near 0, which suggests that liquidity providers efficiently manage inventory so that in aggregate they end the day with close to zero exposure. We believe that the ability of liquidity providers to manage their inventory can explain why we find no significant differences in order imbalance on days when insiders trade.

Recent studies document that informed traders prefer to trade when liquidity is highest (Collin-Dufresne and Fos, 2015) and by splitting orders (Kacperczyk and Pagnotta, 2018a) to avoid detection. Our results complement these recent studies by providing first evidence that establishes that liquidity providers can plausibly detect informed trading in real time when informed traders trade deviate from the equilibrium (e.g., à la Kyle, 1985). We feel that these results provide evidence that rationalize the existing focus on strategic behavior to avoid detection by liquidity providers. We use these findings to motivate the remainder of our empirical analysis.

# 5 How was the information reflected in prices?

The analysis so far has shown that there was a substantial increase in trading volume when hackers traded. In the few cases where we are able to identify the specific trades of the hackers that were prosecuted, it seems that there was a large increase in professional trade volume that followed the hackers' trades. A recent paper by Van Kervel and Menkveld (2019) finds that high-frequency traders can detect the presence of large institutional orders and trade in the same direction as the institutional order. It is thus possible that the substantial volume that we observe is a result of sophisticated traders detecting the presence of informed traders and trading on their information. If indeed that is the case, the price discovery before earnings announcements is not solely due to the hackers but also to other market participants that "lean with the wind" and trade in the same direction as the hackers. We explore this possibility more systematically by examining the relationship between retail trading activity and professional trading activity in a different period.

Specifically, we analyze the sensitivity of professional trading activity to retail activity in the morning and the afternoon for hacked and non-hacked earnings announcements. We classify trades as either retail trades or professional trades using the procedure described in Boehmer, Jones, and Zhang (2017).[26] We compute the turnover in the morning or afternoon of each day as the number of retail or non-retail trades scaled by the number of shares outstanding and run the following regression:

$$\begin{aligned} Non\ Retail_{i,t} = &\beta_1 Retail_{i,t} + \beta_2 Retail_{i,t} \times \mathbf{1}_{[\text{PM}]i,t} + \beta_3 Retail_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ &\beta_4 Retail_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} \times \mathbf{1}_{[\text{PM}]i,t} + \beta_5 \mathbf{1}_{[\text{PM}]i,t} + \beta_6 \mathbf{1}_{[\text{Hacked}]i,t} + \\ &\beta_7 \mathbf{1}_{[\text{Hacked}]i,t} \times \mathbf{1}_{[\text{PM}]i,t} + \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}. \end{aligned} \tag{6}$$

where $Non\ Retail$ represents professional turnover on a given day, for a given stock, for a given morning or afternoon time period. *Retail* is defined analogously for a trading activity that is classified as retail trading. $\mathbf{1}_{[\text{PM}]}$ is a binary variable that takes the value of one if the trading activity occurs after 12 p.m. and is zero otherwise. $\mathbf{1}_{[\text{Hacked}]}$ is a binary variable that takes the value of one if the firm's newswire was subject to a hack at that particular time period and is zero otherwise. The primary coefficient of interest is $\beta_4$, which captures the differential sensitivity of professional trading to retail trading in the afternoon (when the hackers were trading) as compared to the morning (when they typically were not trading)

[26] Following Boehmer, Jones, and Zhang (2017), we define a trade as retail if the transaction price suggest it was filled by a wholesaler and reported to a FINRA Trade Reporting Facility (TRF) (exchange code 'D' in TAQ). All the trades that are classified according to this method are market (and not limit) orders. It is fair to assume that insiders preferred market over limit orders due to the limited time window available to trade on their private information.

for observations that were exposed to hacks, compared to those that were not. A coefficient that is positive would be consistent with a theory where professional traders "lean with the wind" and amplify abnormal (informed) trading activity conducted by the hackers.

Table 9 presents the results of this analysis. We find that there is a strong, positive relationship between retail turnover and professional turnover. Indeed, the coefficient on *Retail* is positive and very statistically significant. Moreover, both higher professional turnover in the afternoon and professional turnover is more strongly associated with retail turnover in the afternoon since both of the coefficients on $\mathbf{1}_{[\text{PM}]}$ and $Retail \times \mathbf{1}_{[\text{PM}]}$ are positive and significant. Consistent with our conjecture, we find that $\beta_4$, the coefficient associated with $Retail_{i,t} \times \mathbf{1}_{[\text{Hacked}]} \times \mathbf{1}_{[\text{PM}]}$, is positive and statistically significant, which suggests that retail share turnover more strongly predicts professional trade volume in the afternoon. The economic magnitude is large. If we compare the increase in afternoon sensitivity of professional turnover to retail turnover on hacked days to the general increase in afternoon sensitivity of professional turnover to retail turnover (i.e., by comparing $\beta_4$ to $\beta_2$) we find that the afternoon sensitivity is 92% larger in column (2) (= 0.2735/0.2961). We do not find that retail turnover in the morning of hacked days exhibits any increased association with professional turnover, as the coefficient on $Retail \times \mathbf{1}_{[\text{Hacked}]}$ is insignificant. Moreover, there do not seem to be any non-retail turnover related effects of the hack on professional trading turnover since the coefficients on $\mathbf{1}_{[\text{Hacked}]} \times \mathbf{1}_{[\text{PM}]}$ and $\mathbf{1}_{[\text{Hacked}]}$ are insignificant, both statistically and economically.

The results of this section suggest that liquidity providers and/or HFTs potentially play an important role in the price discovery process. While we are unable to precisely show which non-retail traders increased their buying or selling activity, the results of this analysis do suggest that professional market participants at least observed abnormal retail trading activity and traded in the same direction, thus amplifying the effect of a modest increase in retail trading activity and likely speeding up the process of price discovery of the illegally acquired information.

## 6 Conclusion

In this paper, we use a sophisticated newswire hacking scandal as a natural experiment to study how efficiently markets incorporate private information into prices. This natural experiment allows us to isolate variation in the information set of a small group of traders related to upcoming earnings announcements and to test how quickly their (illegally obtained) "inside" information was reflected in prices. The staggered nature of the cyberattacks on different newswire companies over time allows us to use a difference-in-difference methodology to overcome many limitations inherent in empirical research about insider trading.

We find that roughly 15% of a firms earnings surprise was incorporated its stock price when before its public release when the hackers had access to non-public information. We find that this effect is concentrated in firms that have higher institutional ownership and better analyst coverage, suggesting that the hackers traded in stocks that were more liquid and on signals that were more precise. We find that some, but not all, measures of informed trading detect this activity. Specifically, volume and spread-based measures reliably detect this activity, but order flow-based measures do not. Finally, we find suggestive evidence that uninformed, professional traders traded in the same direction, amplifying the impact of informed trading. Our results suggest that liquidity providers do adjust prices and quotes in response to informed trading.

Cyber risk presents a new form of systematic risk. In our setting, companies relying on a newswire provider to disseminate corporate press releases are exposed to cyber intrusion into the newswire providers. The cyber intrusion or hacks that we have explored in this paper resulted in a wide range of illegal insider trading over a period of six years. Some have argued that information risk is a determinant of asset returns (Easley, Hvidkjaer, and O'hara, 2002). Increasingly, there seem to be examples of cyberattacks such as in 2017, when the same hackers responsible for our event were able to hack into the SEC platform EDGAR and illegally access information. Thus we believe that information risk may be increasing as the cyberattacks become more sophisticated.

## References


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

———, 2018, Do proxies for informed trading measure informed trading? Evidence from illegal insider trades, .

Augustin, Patrick, Menachem Brenner, and Marti G. Subrahmanyam, 2019, Informed options trading prior to m&a announcements: Insider trading?, *Managment Science* Forthcoming.

Barbon, Andrea, Marco Di Maggio, Francesco Franzoni, and Augustin Landier, 2019, Brokers and order flow leakage: evidence from fire sales, *Journal of Finance, Forthcoming.*

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

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, 2017, Tracking retail investor activity, *Working paper.*

Bolandnazar, Mohammadreza, Robert J. Jackson Jr, Wei Jiang, and Joshua Mitts, 2018, Trading against the random expiration of private information: A natural experiment, *Journal of Finance, Forthcoming.*

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

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

Conrad, Jennifer, and Sunil Wahal, 2019, The term structure of liquidity provision, *Journal of Financial Economics, Forthcoming.*

Cziraki, Peter, and Jasmin Gider, 2018, The dollar profits to insider trading, *Working Paper.*

Easley, David, Soeren Hvidkjaer, and Maureen O'hara, 2002, Is information risk a determinant of asset returns?, *The Journal of Finance* 57, 2185–2221.

Easley, David, Marcos M López de Prado, and Maureen O'Hara, 2012, Flow toxicity and liquidity in a high-frequency world, *The Review of Financial Studies* 25, 1457–1493.

Ellis, Katrina, Roni Michaely, and Maureen O'hara, 2000, When the underwriter is the market maker: An examination of trading in the IPO aftermarket, *The Journal of Finance* 55, 1039–1074.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance* 25, 383–417.

Garriott, Corey, and Ryan Riordan, 2019, Trading on long-term information, *Working paper.*

Grégoire, Vincent, and Charles Martineau, 2019, How is earnings news transmitted into stock prices?, *Working Paper.*

Hartzmark, Samuel M., and Kelly Shue, 2017, A tough act to follow: Contrast effects in financial markets, *Working Paper.*

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.

Holden, Craig W., and Stacey Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *The 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.

Kacperczyk, Marcin T., and Emiliano Pagnotta, 2018a, Becker meets kyle: Inside insider trading, *Working paper.*

——— , 2018b, Chasing private information, *Review of Financial Studies, Forthcoming.*

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* pp. 1315–1335.

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

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.

Martineau, Charles, 2019, The evolution of financial market efficiency: Evidence from earnings announcements, *Available at SSRN 3111607.*

Meulbroek, Lisa K., 1992, An empirical analysis of illegal insider trading, *The 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.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics* 78, 277–309.

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.*

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

Zhang, X. Frank, 2006, Information uncertainty and stock returns, *The Journal of Finance* 61, 105–137.

**Figure 1.** Hackers and Traders Relationship

This figure depicts the parties to the hacking scheme of the press releases and the relationship between the hackers, the traders, and the behavior for which they were charged by the U.S. government. Profits correspond to the share of profits shared between the hackers, the middlemen, and traders for trading on advance press release information ahead of earnings announcements.

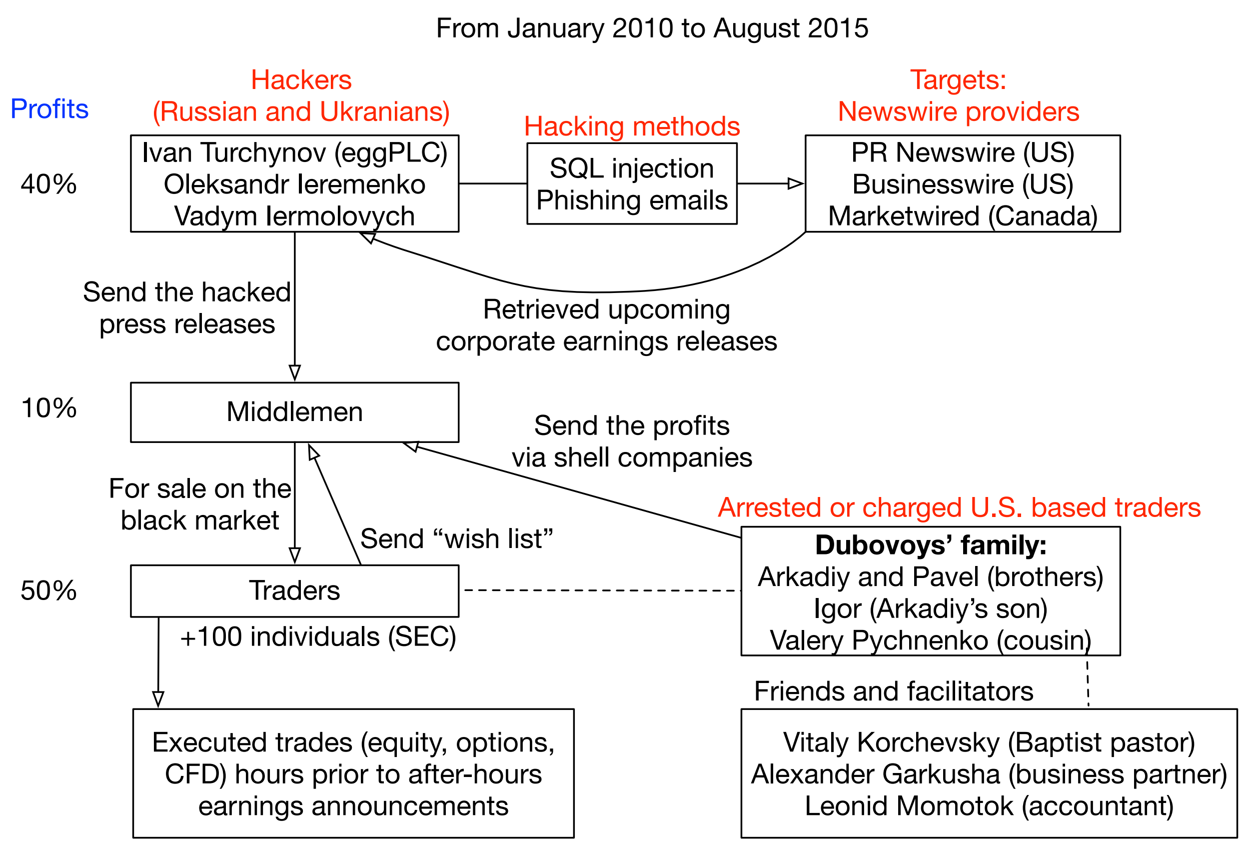


**Figure 2.** Number of Earnings Announcements Exposed to Hacks

This figure shows the number of earnings announcements exposed to hacks (left y-label) per newswire and the corresponding proportion over all earnings announcements exposed to hacks (right y-label) on each quarter. The sample period is from January 2010 to December 2015.

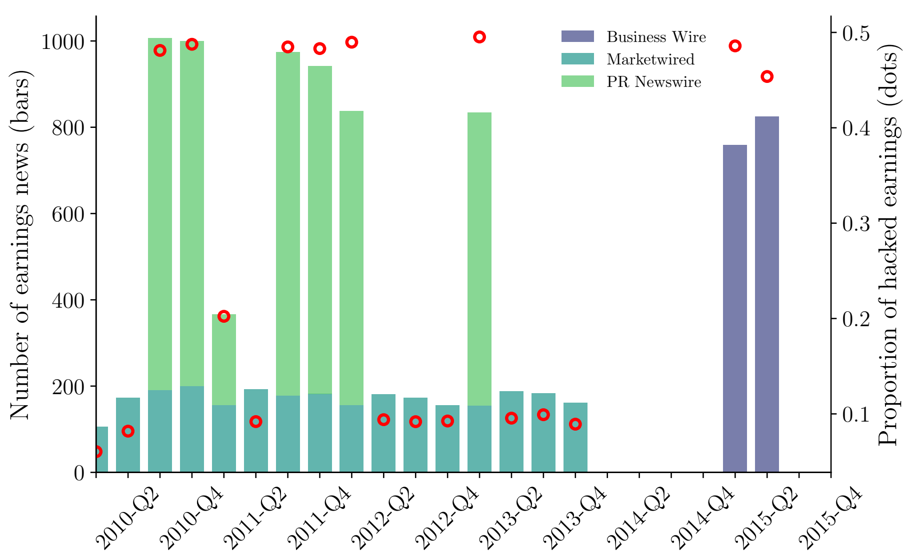
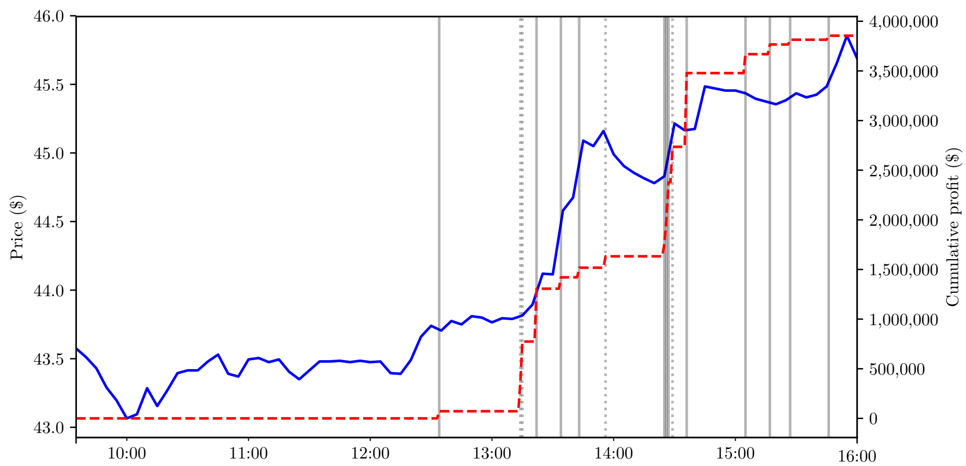


**Figure 3.** Trading Anecdotes

This figure shows two examples of price changes around trading by the insiders during regular trading hours (9:30 a.m. to 4 p.m.) leading to the earnings announcements for Align Technology on October 17, 2013 and Edward Life Science on April 23, 2013 in Panels A and B, respectively. Earnings for both companies were announced right after the closing of markets at 4 p.m. The solid blue lines correspond to the stock price and the red dashed lines correspond to the cumulative profits based on each trade according to the SEC. Solid vertical bars indicate trades in the stock while dotted vertical bars indicate trades using derivative instruments (i.e., options and contract for difference).

Panel A. Align Technology (2013/10/17), next day closing price: $57.98




Panel B. Edward Life Science (2013/04/23), next day closing price: $64.60

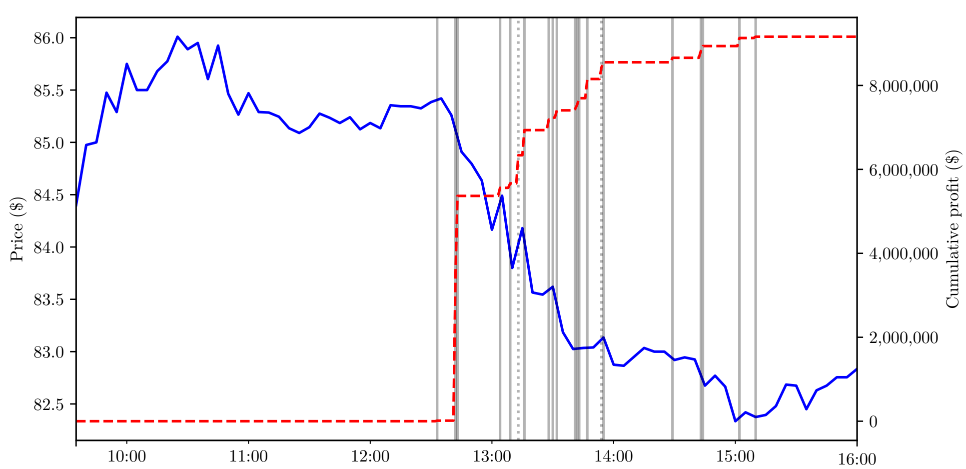
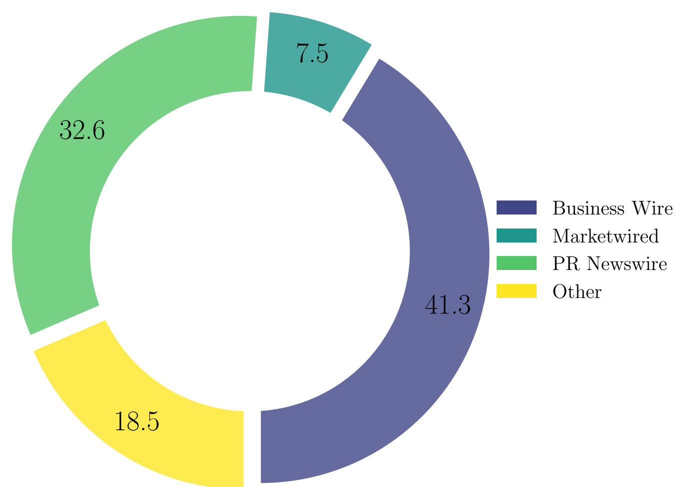


**Figure 4.** Fraction of Observations per Newswire

This figure shows the fraction of earnings announcements per newswire in Panel A and the fraction of earnings announcements exposed to hacks per newswire in Panel B.

Panel A. The fraction of earnings announcements per newswire




Panel B. The fraction of earnings announcements exposed to hacks per newswire

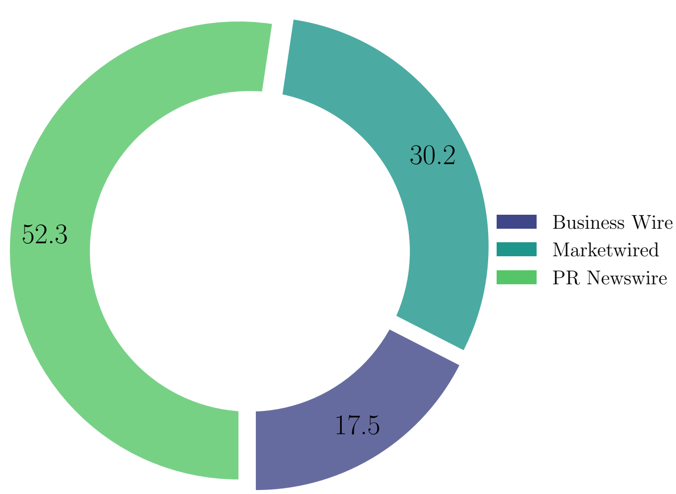


**Figure 5.** Earnings Surprise Quantile-Quantile Plot

This figure shows a probability plot of the earnings surprise data against the quantiles of a normal distribution. The red line corresponds to the best fit line. Earnings surprises below the 1% percentile and above the 99% percentile are excluded from the quantile-quantile plot.

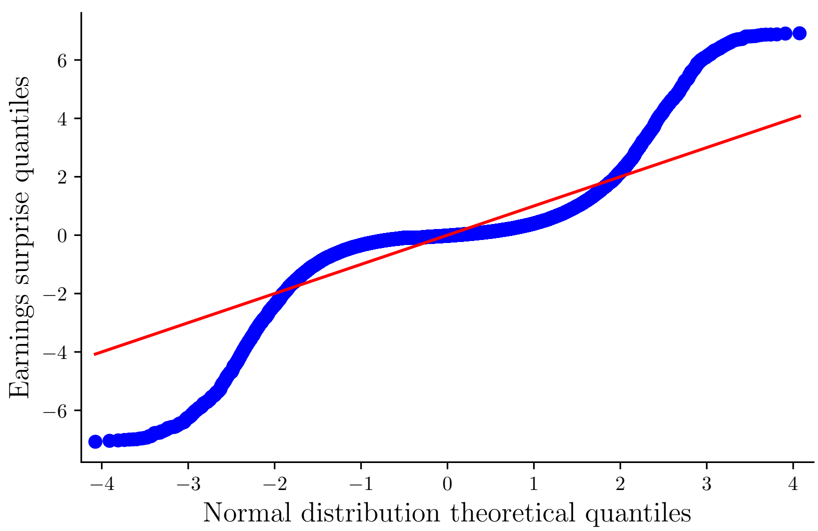


**Figure 6.** The Average Earnings Surprise Distribution by Earnings Surprise Quintiles

This figure shows the average earnings surprise by earnings surprise quintiles and by hacked newswire.

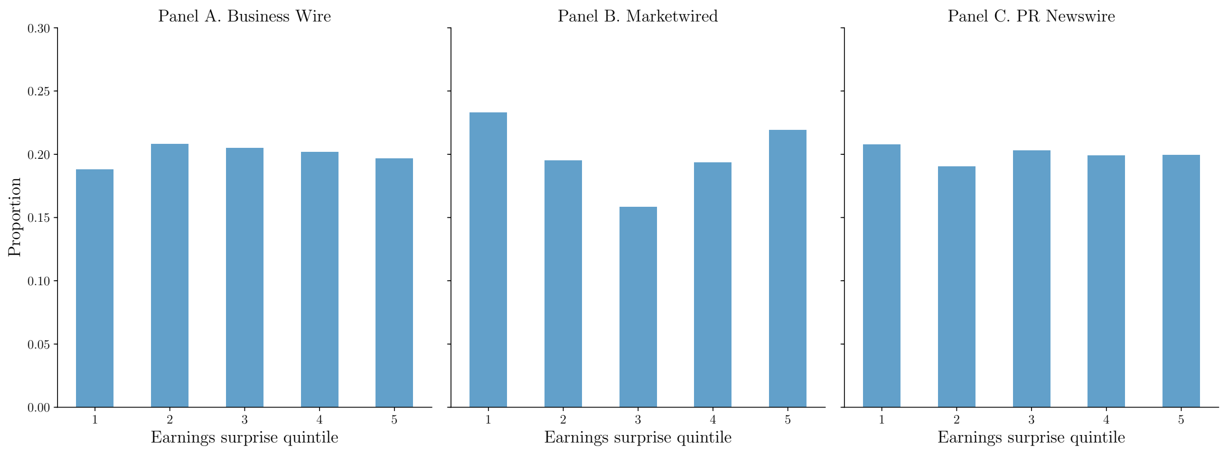



**Figure 7.** Industry Composition by Newswire

This figure shows the proportion of earnings announcements by industry using the 11 Global Industry Classification (GIC) codes by hacked newswire.

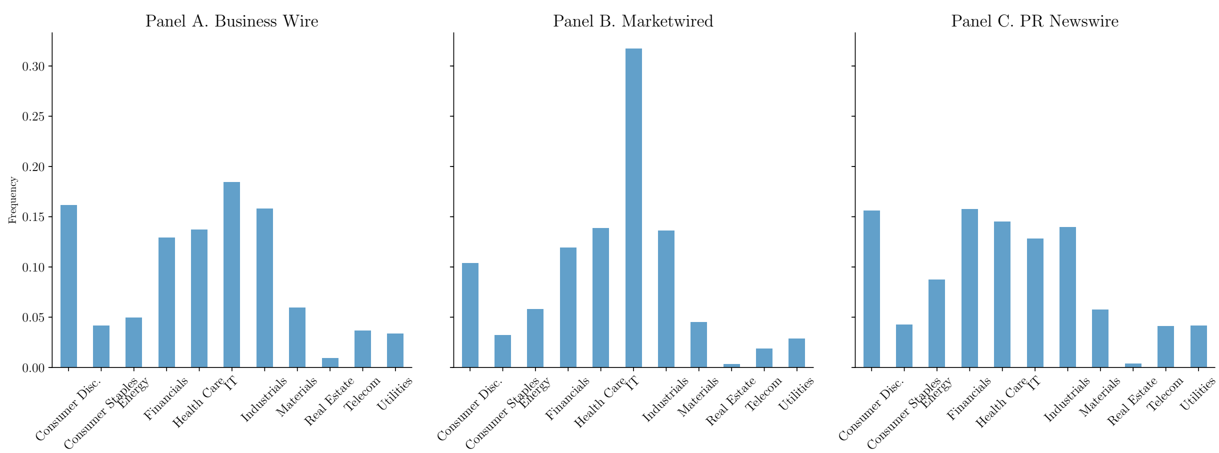
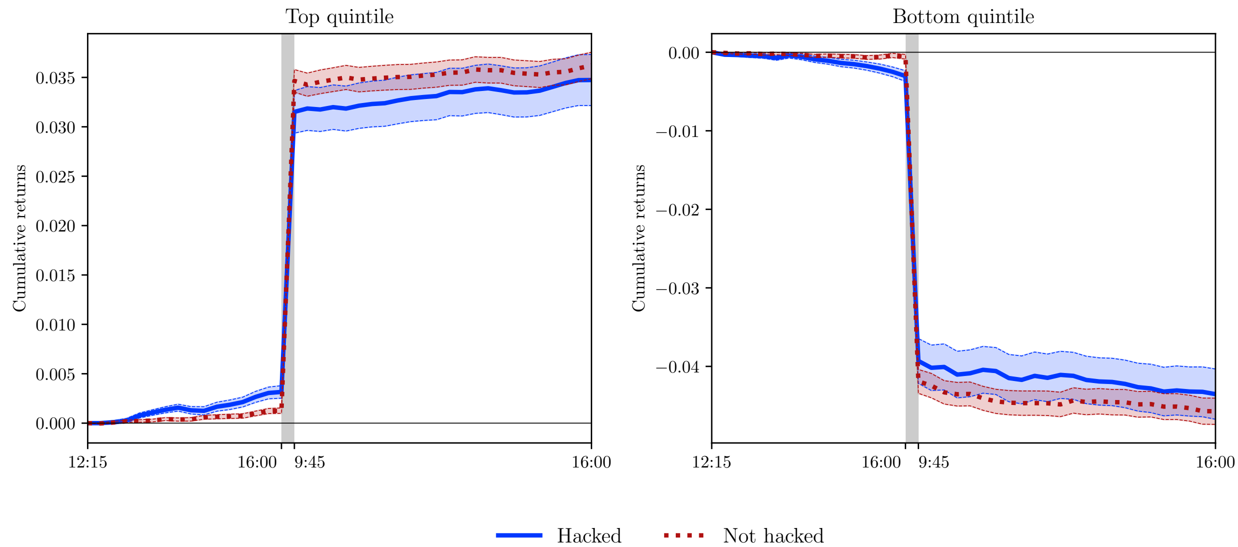


**Figure 8.** Cumulative Returns Around Earnings Announcements

This figure shows the average cumulative returns, in percent, 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 the close on the next trading day for S&P 1500 and non-S&P 1500 stocks in Panels A and B, respectively. The left panel shows returns for the top quintile of earnings surprises while the right panel shows returns for the bottom quintile of earnings surprises. The shaded areas correspond to the 95% confidence intervals. The gray bar in Panel A 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. S&P 1500 stocks




Panel B. Non-S&P 1500 stocks

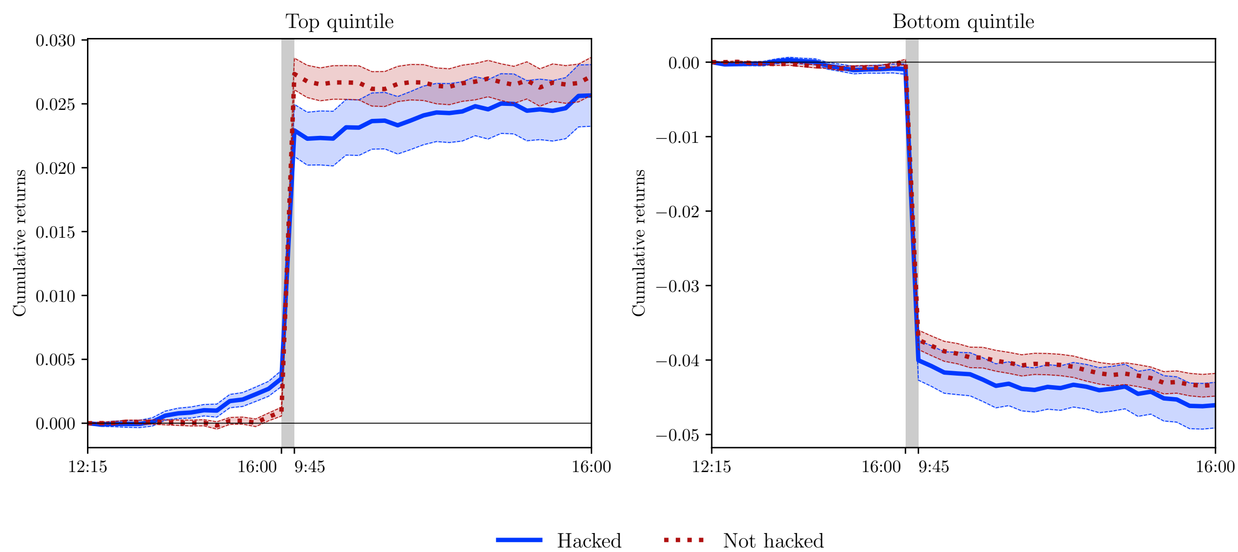


**Figure 9.** Cumulative Returns Around Earnings Announcements for Prosecuted Cases

This figure shows the average cumulative returns, in percent, around earnings announcements for firms that were mentioned in the SEC prosecution (Traded) and for two matching groups: exposed to hacks (Hacked) but not mentioned in the SEC prosecution, and not exposed to hacks (Not hacked) from noon on the trading day before the announcement to the close on the next trading day. Each earnings announcements mentioned in the SEC documentation is matched with five earnings announcements for firms belonging to the same S&P index occurring in the same earnings year-quarter and by total return (from 12:00 to 16:00 the next day). The left panel shows returns for the top quintile of earnings surprises while the right panel shows returns for the bottom quintile of earnings surprises. The shaded areas correspond to the 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.

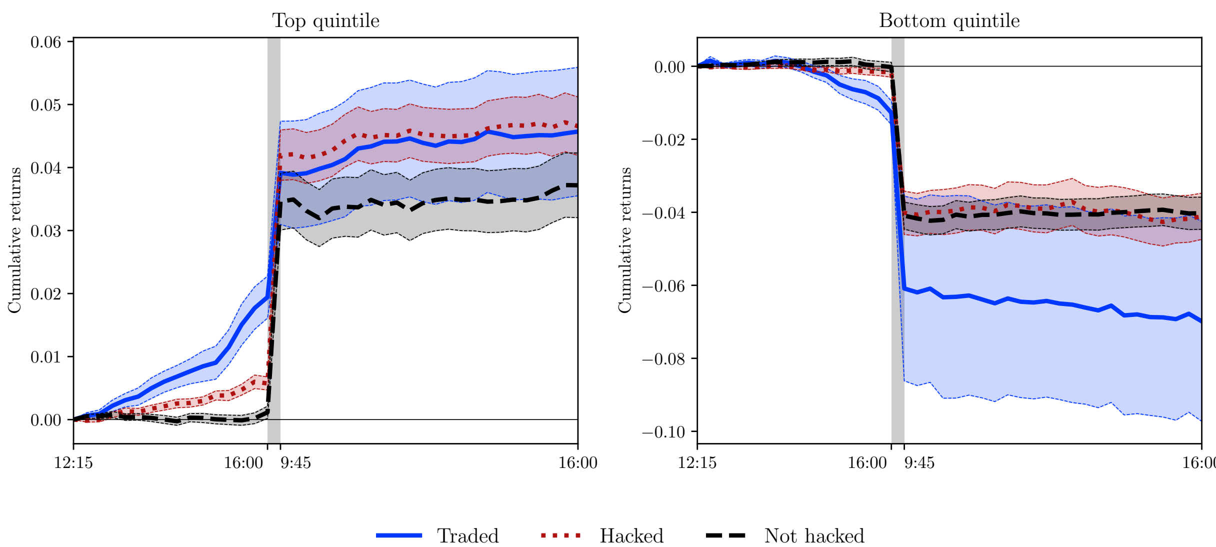


**Figure 10.** Prosecuted SEC Case Examples

This figure shows in the right panel the cumulative return and trade volume and the left panel shows the cumulative trade volume, cumulative insider trade volume, cumulative order imbalance, and the cumulative order imbalance for trades executed on Nasdaq ITCH at a one minute frequency from 9:30 a.m. to 4 p.m. for Align Technologies on October 17, 2013 and Edward Life Sciences on April 23, 2013 in Panels A and B, respectively. Earnings for both companies were announced right after the closing of markets at 4 p.m. The gray vertical lines represent the time of an insider trade (buying or shorting a the stock).

Panel A. Align Technology (2013/10/17)

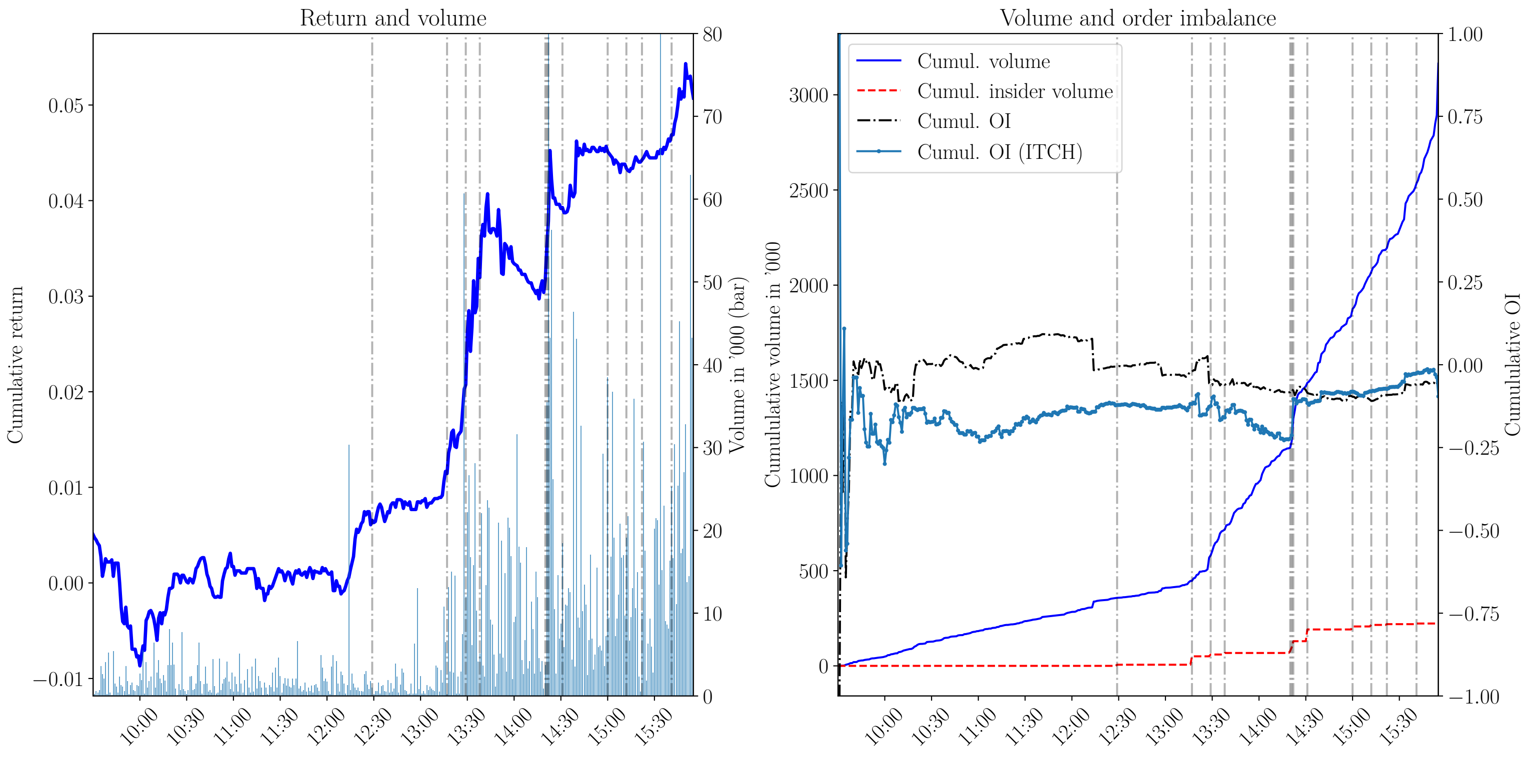


**Figure 10.** Prosecuted SEC Case Examples (continued)

Panel B. Edward Life Sciences (2013/04/23)

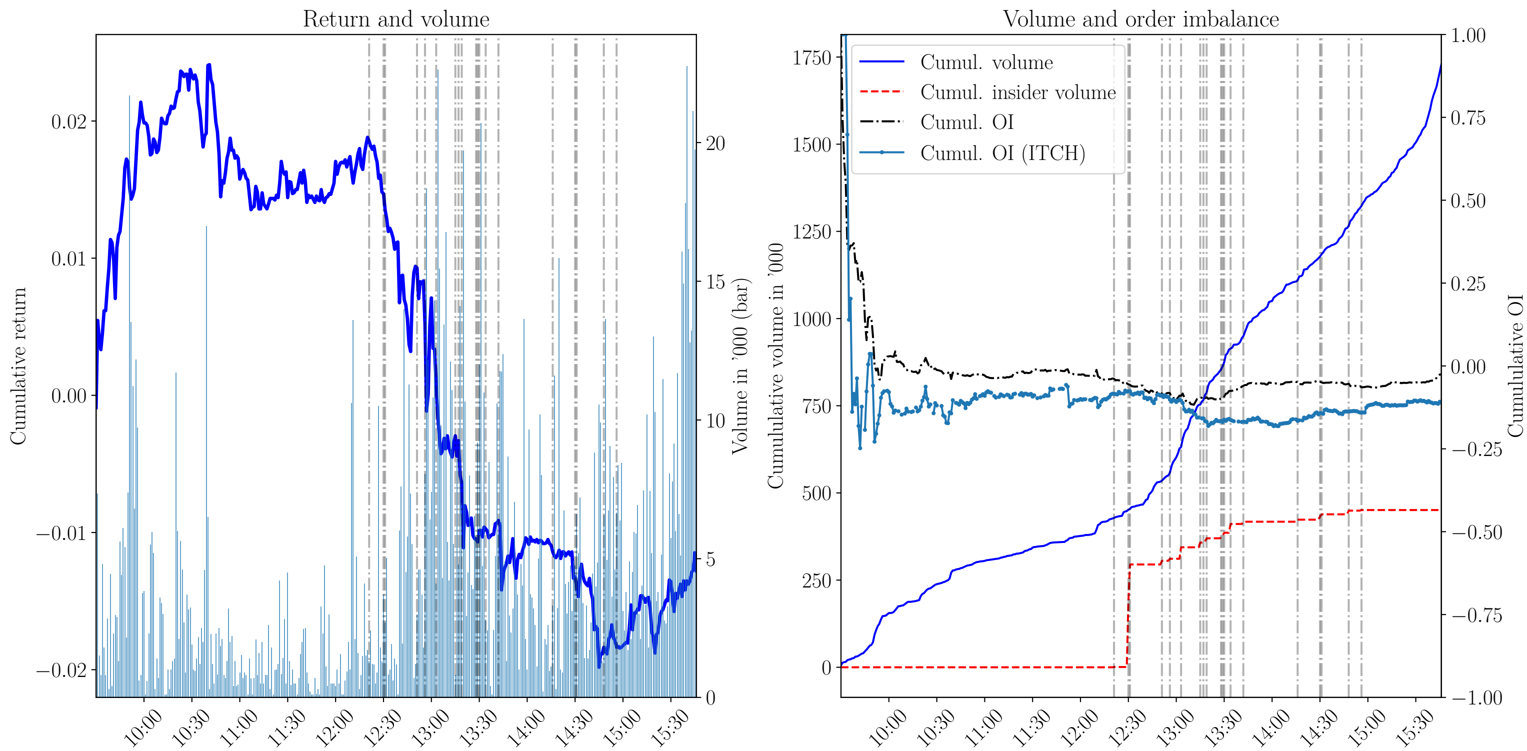


Table 1: Anecdotal Evidence

This table presents examples of trading based off of illegally hacked material non-public information. The examples are taken from SEC court filings.

| **Firm** | **Trading Date** | **Time News Uploaded** | **Time of First Trade** | **Used Equity** | **Used Options** | **Used Contract for Difference** | **Earnings Surprise (%)** | **Newswire** |
|---|---|---|---|---|---|---|---|---|
| Caterpillar Inc, | 10/21/2011 | 9:40 PM | 1:30 PM | Yes | No | Yes | 8.92 | PR Newswire |
| Las Vegas Sands | 7/25/2012 | 11:08 PM | 9:47 AM | Yes | No | Yes | -27.87 | MarketWired |
| BroadSoft Inc. | 8/3/2012 | 11:38 AM | 1:24 PM | Yes | Yes | Yes | 65.00 | MarketWired |
| NVIDIA Corp. | 8/9/2012 | 7:48 PM | 1:31 PM | Yes | Yes | Yes | 57.14 | MarketWired |
| Lexmark Intl. | 1/28/2013 | 1:08 PM | 2:48 PM | Yes | No | No | -3.91 | PR Newswire |
| Edwards Life Sciences | 4/23/2013 | 11:29 AM | 12:23 PM | Yes | Yes | Yes | -5.26 | MarketWired |
| Solar Winds | 4/30/2013 | 1:03 PM | 2:09 PM | Yes | No | No | 10.81 | MarketWired |
| Panera Bread Co. | 7/23/2013 | 10:00 AM | 12:15 PM | Yes | Yes | Yes | -1.69 | MarketWired |
| VMware Inc | 7/23/2013 | 12:09 PM | 1:23 PM | Yes | No | Yes | 2.60 | MarketWired |
| TIBCO Software Inc | 9/19/2013 | 1:47 PM | 2:35 PM | Yes | No | Yes | 27.27 | MarketWired |
| Align Technology | 10/17/2013 | 1:28 AM | 12:34 PM | Yes | No | Yes | 40.00 | MarketWired |

**Table 2**
**Descriptive Statistics**

This table reports the mean, median, standard deviation (Std. dev.) and the number of observations (N) for the earnings surprise, the absolute earnings surprise, the log market capitalization, the fraction of shares held by institutions (in percent), the number of analysts, Amihud illiquidity measure, the 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.

| | Surprise | \|Surprise\| | Ln MCAP | IO | N. analysts | Ln Q-value | Share turn. | Option turn. |
|---|---|---|---|---|---|---|---|---|
| Mean | 0.0004 | 0.007 | 14.16 | 68.36 | 7.71 | 0.9689 | 2 | 0.21 |
| Median | 0.0005 | 0.0017 | 14.11 | 77.48 | 5 | 0.7927 | 1.46 | 0.04 |
| Std. dev. | 0.0104 | 0.0373 | 1.77 | 28.81 | 6.93 | 1.5317 | 2.13 | 0.48 |
| N | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 35,556 | 43,991 | 43,991 |

**Table 3**
**Characteristic Regressions**

This table reports 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 the earnings surprise, the absolute earnings surprise, the log market capitalization, the number of analysts, Amihud illiquidity measure, the log Q-value of the firm, monthly share turnover, and monthly option turnover for earnings announcement $i$ announced on year-quarter $t$. Both turnover measures are calculated in the month prior to the earnings announcement. $\mathbf{1}_{[\text{Hacked}]}$ is a dummy variable equal to one if announcement $i$ is released by a hacked newswire, zero otherwise. Results are presented with year fixed-effects in Panel A and firm and year-quarter fixed-effects in Panel B. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively.

Panel A. With year-quarter fixed-effects

| | Surprise | \|Surprise\| | Ln MCAP | IO | N. analysts | Ln Q-value | Share turn. | Option turn. |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\mathbf{1}_{[\text{Hacked}]}$ | -0.000 | 0.000 | -0.193** | -3.101** | -0.125 | 0.039 | 0.070 | 0.006 |
| | (0.000) | (0.001) | (0.078) | (1.323) | (0.224) | (0.055) | (0.070) | (0.017) |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 35,556 | 43,991 | 43,991 |
| Adjusted $R^2$ | 0.000 | 0.000 | 0.001 | 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 B. With firm and year-quarter fixed-effects

| | Surprise | \|Surprise\| | Ln MCAP | IO | N. analysts | Ln Q-value | Share turn. | Option turn. |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\mathbf{1}_{[\text{Hacked}]}$ | 0.000 | -0.000 | 0.005 | 0.124 | 0.098 | -0.018 | -0.014 | -0.004 |
| | (0.000) | (0.001) | (0.012) | (0.290) | (0.082) | (0.020) | (0.032) | (0.006) |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 35,556 | 43,991 | 43,991 |
| 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 4**
**Predicting Earnings Surprise**

This table reports the coefficients of the following regression:

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

where $Surprise\,Quintile_{i,t}$ corresponds to the ordinal earnings surprise quintiles for earnings announcement of firm $i$ announced in year-quarter $t$. Log return corresponds to different window length of intraday returns from 9:30 to 16:00, 9:30 to 12:00, and from 12:00 to 16:00, during the trading day prior to the earnings announcement. $\mathbf{1}_{[\text{Hacked}]}$ is a dummy variable equal to one if announcement $i$ is released by a hacked newswire 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. Results are presented with firm and year-quarter fixed-effects. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The sample period is from January 2010 to December 31, 2015.

| | Return window | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 9:30-16:00 | | 9:30-12:00 | | 12:00-16:00 | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| *Log Return* | 1.346*** | 1.976*** | 0.689 | 1.117** | 2.049*** | 2.684*** | 3.365*** | 3.948*** |
| | (0.384) | (0.404) | (0.468) | (0.478) | (0.524) | (0.574) | (0.531) | (0.614) |
| *Log Return* $\times \mathbf{1}_{[\text{Hacked}]}$ | 2.647*** | 2.661*** | -0.339 | -0.336 | 5.109*** | 5.158*** | | |
| | (0.872) | (0.824) | (1.023) | (0.984) | (1.011) | (0.953) | | |
| $\mathbf{1}_{[\text{Hacked}]}$ | 0.012 | 0.014 | 0.009 | 0.012 | -0.001 | 0.000 | | |
| | (0.019) | (0.019) | (0.020) | (0.020) | (0.018) | (0.018) | | |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 |
| Adjusted $R^2$ | 0.002 | 0.010 | 0.000 | 0.008 | 0.002 | 0.010 | 0.002 | 0.009 |
| Controls | N | Y | N | Y | N | Y | N | Y |
| Year-Quarter F.E. | Y | Y | Y | Y | Y | Y | Y | Y |
| Firm F.E. | Y | Y | Y | Y | Y | Y | Y | Y |

**Table 5**
**Informed Trading and Overnight Price Formation**

This table reports the coefficients of the following regression in Panel A

$$\begin{aligned} Log\,Return\,ON_{i,t} =& \beta_1 Surprise_{i,t} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Surprise_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t}, \end{aligned}$$

and the coefficients of the following regression in Panel B

$$\begin{aligned} Log\,Return\,ON_{i,t} =& \beta_1 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \beta_4 \mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}. \end{aligned}$$

$Log\,Return\,ON_{i,t}$ 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, $\mathbf{1}_{[\text{Hacked}]}$ is a dummy variable equal to one if announcement $i$ is released by a hacked newswire and zero otherwise, $\mathbf{1}_{[\text{Surprise qnt=5}]i,t}$, $\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Surprise qnt=1}]i,t}$ are dummy variables equals to one if the earnings surprise belongs to the top, second to fourth, and bottom earnings surprise quintiles 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 parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The sample period is from January 2010 to December 31, 2015.

Panel A. Linear estimation

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Surprise* | 1.357*** | 1.428*** | 1.436*** | 1.419*** |
| | (0.073) | (0.078) | (0.078) | (0.061) |
| $Surprise \times \mathbf{1}_{[\text{Hacked}]}$ | -0.213** | -0.228** | -0.236** | -0.211** |
| | (0.100) | (0.103) | (0.103) | (0.100) |
| $\mathbf{1}_{[\text{Hacked}]}$ | -0.001 | -0.001 | -0.001 | -0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 |
| 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 5**
**Informed Trading and Overnight Price Formation (cont.)**

Panel B. Piecewise estimation

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| $\mathbf{1}_{[\text{Surprise qnt}=5]}$ | 0.026*** | 0.028*** | 0.026*** | 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) |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 |
| Adjusted $R^2$ | 0.097 | 0.098 | 0.106 | 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 |

**Table 6**
**Regression: By Institutional Ownership**

This table reports the coefficients of the following regression

$$\begin{aligned} Log\,Return\,ON_{i,t} =& \beta_1 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \beta_4 \mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}. \end{aligned}$$

$Log\,Return\,ON_{i,t}$ corresponds to the log overnight return calculated using stock prices at 4 p.m. to 9:30 a.m. on the following day, $\mathbf{1}_{[\text{Hacked}]}$ is a dummy variable equal to one if announcement $i$ is released by a hacked newswire and zero otherwise, $\mathbf{1}_{[\text{Surprise qnt=5}]i,t}$, $\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Surprise qnt=1}]i,t}$ are dummy variables equals to one if the earnings surprise belongs to the top, second to fourth, and bottom earnings surprise quintiles 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. The regression is estimated for stocks with institutional ownership (IO) that is greater or equal to the median (76%) and less than the median. Results are presented with firm and year-quarter fixed-effects. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The sample period is from January 2010 to December 31, 2015.

| | High IO | | Low IO | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\mathbf{1}_{[\text{Surprise qnt=5}]}$ | 0.033*** | 0.030*** | 0.025*** | 0.024*** |
| | (0.002) | (0.002) | (0.001) | (0.001) |
| $\mathbf{1}_{[\text{Surprise qnt=1}]}$ | -0.038*** | -0.039*** | -0.024*** | -0.024*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| $\mathbf{1}_{[\text{Surprise qnt=5}]} \times \mathbf{1}_{[\text{Hacked}]}$ | -0.006** | -0.006** | -0.002 | -0.002 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| $\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]} \times \mathbf{1}_{[\text{Hacked}]}$ | -0.002 | -0.002* | -0.000 | -0.001 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $\mathbf{1}_{[\text{Surprise qnt=1}]} \times \mathbf{1}_{[\text{Hacked}]}$ | 0.009** | 0.009** | -0.000 | -0.000 |
| | (0.004) | (0.004) | (0.002) | (0.002) |
| $N$ | 21,996 | 21,996 | 21,995 | 21,995 |
| Adjusted $R^2$ | 0.106 | 0.127 | 0.094 | 0.099 |
| Controls | N | Y | N | Y |
| Year-Quarter F.E. | Y | Y | Y | Y |
| Firm F.E. | Y | Y | Y | Y |

**Table 7**
**Regression: By Analyst Coverage**

This table reports the coefficients of the following regression

$$
\begin{aligned}
Log\,Return\,ON_{i,t} = & \beta_1 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} + \beta_2 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} + \beta_3 \mathbf{1}_{[\text{Surprise qnt=5}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\
& \beta_4 \mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \beta_5 \mathbf{1}_{[\text{Surprise qnt=1}]i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\
& \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}.
\end{aligned}
$$

$Log\,Return\,ON_{i,t}$ corresponds to the log overnight return calculated using stock prices at 4 p.m. to 9:30 a.m. on the following day,$\mathbf{1}_{[\text{Hacked}]}$ is a dummy variable equal to one if announcement $i$ is released by a hacked newswire and zero otherwise, $\mathbf{1}_{[\text{Surprise qnt=5}]i,t}$, $\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]i,t}$, and $\mathbf{1}_{[\text{Surprise qnt=1}]i,t}$ are dummy variables equals to one if the earnings surprise belongs to the top, second to fourth, and bottom earnings surprise quintiles 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. The regression is estimated for stocks with the number of analysts in IBES that is greater or equal to the median (5 analysts) and less than the median. Results are presented with firm and year-quarter fixed-effects. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The sample period is from January 2010 to December 31, 2015.

| | N. analysts $\geq$ 5 | | N. analysts $<$ 5 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\mathbf{1}_{[\text{Surprise qnt=5}]}$ | 0.029*** | 0.027*** | 0.028*** | 0.026*** |
| | (0.001) | (0.001) | (0.002) | (0.001) |
| $\mathbf{1}_{[\text{Surprise qnt=1}]}$ | -0.036*** | -0.037*** | -0.023*** | -0.024*** |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| $\mathbf{1}_{[\text{Surprise qnt=5}]} \times \mathbf{1}_{[\text{Hacked}]}$ | -0.007** | -0.007** | -0.002 | -0.001 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| $\mathbf{1}_{[\text{Surprise qnt}\in[2,3,4]]} \times \mathbf{1}_{[\text{Hacked}]}$ | -0.002* | -0.002* | -0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| $\mathbf{1}_{[\text{Surprise qnt=1}]} \times \mathbf{1}_{[\text{Hacked}]}$ | 0.008** | 0.008*** | 0.000 | 0.000 |
| | (0.003) | (0.003) | (0.002) | (0.002) |
| $N$ | 24,947 | 24,947 | 19,044 | 19,044 |
| Adjusted $R^2$ | 0.096 | 0.106 | 0.104 | 0.112 |
| Controls | N | Y | N | Y |
| Year-Quarter F.E. | Y | Y | Y | Y |
| Firm F.E. | Y | Y | Y | Y |

## Table 8
## Insider Trading, Order Flow, and Illiquidity Measures

This table reports the coefficients of the following regression:

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

$Informed$ corresponds to one of the dependent variables represented in each columns and $\mathbf{1}_{[\text{Hacked}]i,t}$ is a dummy equal to one if the earnings announcement $i$ is exposed to a hack and zero otherwise. See section A1 in the appendix for the dependent variable definitions. 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 session, respectively. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The control variables are the log market capitalization, fraction of shares held by institutional investors, realized volatility of the S&P 500 proxied by the SPDR S&P 500 ETF, and the inverse of the stock price. The sample period is from January 2010 to December 31, 2015.

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

| | Volatility | Log(volume) | Turnover | Amihud | $\lvert OI \rvert$ | VPIN | Quoted spread | Effective spread | Realized spread | Price impact | Kyle's $\lambda$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| $\mathbf{1}_{[\text{Hacked}]}$ | 0.0421*** | 0.0328** | 0.0458*** | 0.0108 | 0.0086 | 0.0021 | 0.0163* | 0.0316*** | 0.0449*** | -0.0136 | 0.0140 |
| | (0.01) | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) | (0.01) |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 |
| $R^2$ | 0.180 | 0.033 | 0.016 | 0.121 | 0.012 | 0.044 | 0.133 | 0.159 | 0.035 | 0.024 | 0.062 |
| Controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Year-Quarter F.E. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Firm F.E. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

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

| | Volatility | Log(volume) | Turnover | Amihud | $\lvert OI \rvert$ | VPIN | Quoted spread | Effective spread | Realized spread | Price impact | Kyle's $\lambda$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| $\mathbf{1}_{[\text{Hacked}]}$ | -0.0140 | -0.0191 | -0.0131 | 0.0171 | -0.0091 | -0.0047 | 0.0186* | 0.0138 | 0.0222 | -0.0024 | 0.0146 |
| | (0.01) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.02) | (0.01) |
| $N$ | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 | 43,991 |
| $R^2$ | 0.044 | 0.029 | 0.010 | 0.102 | 0.011 | 0.033 | 0.108 | 0.073 | 0.029 | 0.018 | 0.035 |
| Controls | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Year-Quarter F.E. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Firm F.E. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

**Table 9**
**Retail Trading**

This table reports the coefficients of the following regression

$$\begin{aligned} Non\ Retail_{i,t} =& \beta_1 Retail_{i,t} + \beta_2 Retail_{i,t} \times \mathbf{1}_{[\text{PM}]i,t} + \beta_3 Retail_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \beta_4 Retail_{i,t} \times \mathbf{1}_{[\text{Hacked}]i,t} \times \mathbf{1}_{[\text{PM}]i,t} + \beta_5 \mathbf{1}_{[\text{PM}]i,t} + \beta_6 \mathbf{1}_{[\text{Hacked}]i,t} + \\ & \beta_7 \mathbf{1}_{[\text{Hacked}]i,t} \times \mathbf{1}_{[\text{PM}]i,t} + \Gamma' Controls_{i,t} + \alpha_t + \alpha_i + \varepsilon_{i,t}. \end{aligned}$$

*Non Retail Turn* and *Retail Turn* corresponds to the share turnover by non-retail and retail traders, respectively. $\mathbf{1}_{[\text{Hacked}]i,t}$ and $\mathbf{1}_{[\text{PM}]i,t}$ are dummies equal to one if the earnings announcement $i$ is exposed to a hack and the trade volume is after 12 p.m. and zero otherwise. The control variables are the log market capitalization, realized volatility of the S&P 500 proxied by the SPDR S&P 500 ETF, and the inverse of the stock price. We classify trades into retail and non-retail as in Boehmer, Jones, and Zhang (2017). Results are presented with firm and year-quarter fixed-effects. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The sample period is from January 2010 to December 31, 2015.

| | (1) | (2) |
|---|---|---|
| *Retail* | 4.6769*** | 4.6990*** |
| | (0.10) | (0.09) |
| *Retail* $\times \mathbf{1}_{[\text{PM}]}$ | 0.3059*** | 0.2939*** |
| | (0.08) | (0.08) |
| *Retail* $\times \mathbf{1}_{[\text{Hacked}]}$ | 0.0643 | -0.0135 |
| | (0.14) | (0.14) |
| *Retail* $\times \mathbf{1}_{[\text{Hacked}]} \times \mathbf{1}_{[\text{PM}]}$ | 0.2501* | 0.2750** |
| | (0.13) | (0.13) |
| $\mathbf{1}_{[\text{PM}]}$ | 0.0015*** | 0.0015*** |
| | (0.00) | (0.00) |
| $\mathbf{1}_{[\text{Hacked}]}$ | 0.0000 | 0.0001 |
| | (0.00) | (0.00) |
| $\mathbf{1}_{[\text{Hacked}]} \times \mathbf{1}_{[\text{PM}]}$ | 0.0001 | 0.0001 |
| | (0.00) | (0.00) |
| $N$ | 87,982 | 87,982 |
| $R^2$ | 0.483 | 0.493 |
| Controls | N | Y |
| Year-Quarter F.E. | Y | Y |
| Firm F.E. | Y | Y |

# A1 Variable Definitions

We describe below the variables used in Table 8. These variables are constructed on the trading day preceeding an earnings announcement for two trading periods, from 9:30 a.m. to 12 p.m. (morning) and from 12 p.m. to 4 p.m. (afternoon). We further divide each variables by the cross-sectional standard deviation (except the log trading volume).

- **Volatility**: $\sqrt{\sum_t^T r_t^2}$. $r_t$ is the 5 minute returns. The returns are calculated using midquotes (the midpoint between the national best ask and bid price).

- **Trading volume**: Log(volume) where volume is the number of traded shares.

- **Turnover**: $\frac{\text{Volume}}{\text{Share outstanding}}$

- **Amihud**: $\frac{|r_t|}{\text{Dollar traded volume}} \times 10,000$ where $|r_t|$ is the absolute return for the morning or afternoon period.

- **Absolute order imbalance** ($|OI|$): $\left|\frac{B-S}{B+S}\right|$, where $B$ and $S$ corresponds 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.

- **Volume imbalance and trade intensity (the VPIN toxicity metric):** We follow Easley, López de Prado, and O'Hara (2012) and compute the VPIN measure as $\sum_{i=1}^{n} \frac{|B_i - S_i|}{nV}$, where $B_i$ and $S_i$ denotes the buying volume and selling volume (market orders) for trade bin $i$, respectively. The trades are signed using the Lee and Ready (1991) algorithm. To obtain a daily VPIN measure, we construct $n = 50$ trade bins of equal volume $V$, where $V$ is equal to a fraction $1/n$ of the daily volume.

- **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 for the morning and afternoon trading sessions by taking the weighted average of the intraday quoted spreads, where the weights on each observed intraday spread is the amount of time the spread is in play.

- **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 the trade 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 NBBO quotes prevailing 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 and afternoon trading session.

- **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 the trade 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 NBBO quotes prevailing 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 and afternoon trading session. The realized spread presented in the Internet Appendix uses a prevailing midquote 1 minute after trade $k$.

- **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 the trade is a market order buy and -1 if it is a market order sell. $M_k$ is the midquote of the NBBO quotes prevailing 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 and afternoon trading session. The price impact presented in the Internet Appendix uses a prevailing midquote 1 minute after trade $k$.

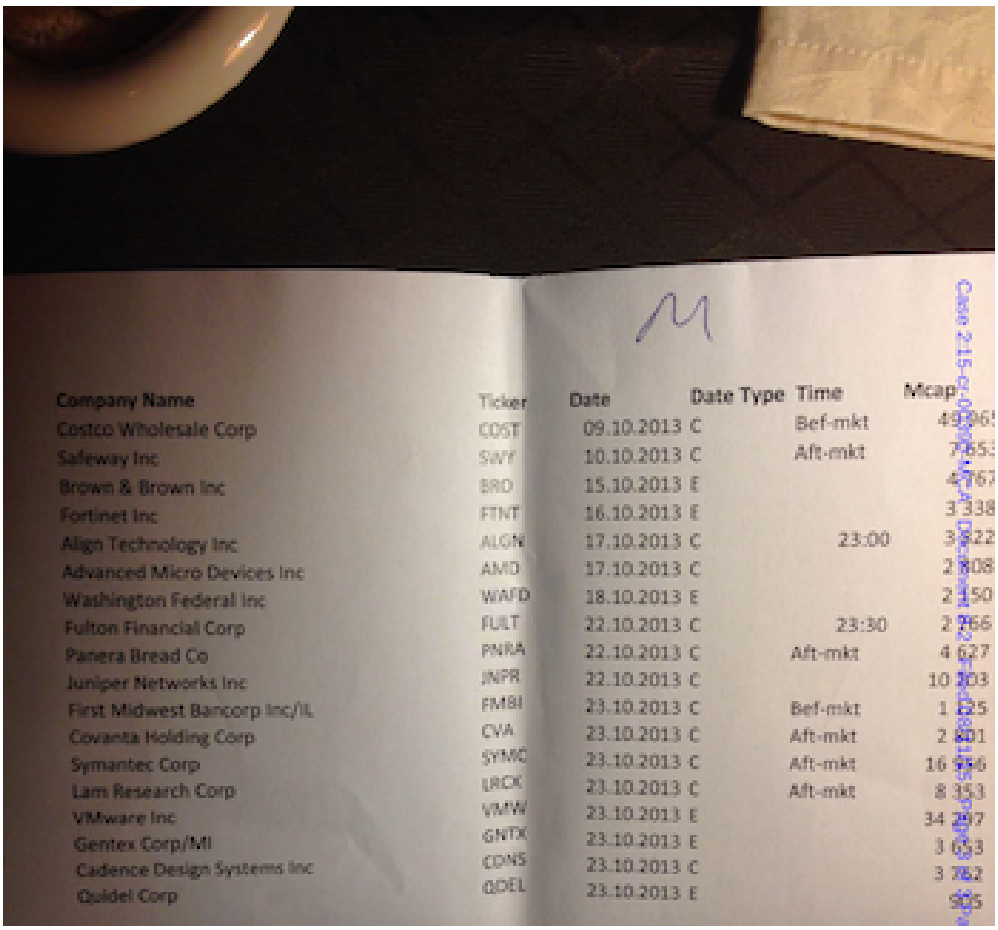
- **Kyle's $\lambda$**: We follow Ahern (2018) and estimate Kyle (1985) lambda as the coefficient $\lambda$ (times 1000) in the following regression:

$$\Delta p_k = \lambda S_k + u_k,$$

  where $\Delta p_k$ is the change in the transaction price, $S_k = d_k \sqrt{DollarVolume}$ is the signed dollar volume of the trade, and $u_k$ is an unobserved error term. This regression is estimated across all trades for the morning and afternoon trading sessions.

**Figure A1.** Evidence of Hackers "Shopping" List

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



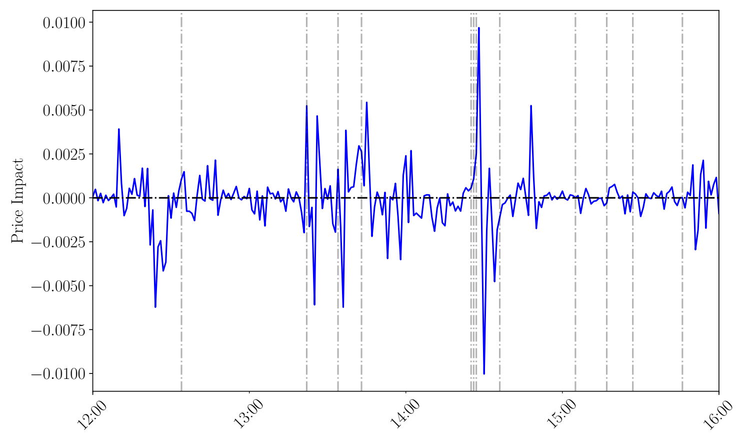
M

| Company Name | Ticker | Date | Date Type | Time | Mcap |
|---|---|---|---|---|---|
| Costco Wholesale Corp | COST | 09.10.2013 | C | Bef-mkt | 49 965 |
| Safeway Inc | SWY | 10.10.2013 | C | Aft-mkt | 7 653 |
| Brown & Brown Inc | BRO | 15.10.2013 | E | | 4 767 |
| Fortinet Inc | FTNT | 16.10.2013 | E | | 3 338 |
| Align Technology Inc | ALGN | 17.10.2013 | C | 23:00 | 3 822 |
| Advanced Micro Devices Inc | AMD | 17.10.2013 | C | | 2 808 |
| Washington Federal Inc | WAFD | 18.10.2013 | E | | 2 150 |
| Fulton Financial Corp | FULT | 22.10.2013 | C | 23:30 | 2 766 |
| Panera Bread Co | PNRA | 22.10.2013 | C | Aft-mkt | 4 627 |
| Juniper Networks Inc | JNPR | 22.10.2013 | C | | 10 203 |
| First Midwest Bancorp Inc/IL | FMBI | 23.10.2013 | C | Bef-mkt | 1 125 |
| Covanta Holding Corp | CVA | 23.10.2013 | C | Aft-mkt | 2 801 |
| Symantec Corp | SYMC | 23.10.2013 | C | Aft-mkt | 16 956 |
| Lam Research Corp | LRCX | 23.10.2013 | C | Aft-mkt | 8 353 |
| VMware Inc | VMW | 23.10.2013 | E | | 34 207 |
| Gentex Corp/MI | GNTX | 23.10.2013 | E | | 3 653 |
| Cadence Design Systems Inc | CDNS | 23.10.2013 | C | | 3 782 |
| Quidel Corp | QDEL | 23.10.2013 | E | | 905 |

**Figure A2.** Real-Time Price Impact from Prosecuted Cases

This figure shows the average one-minute price impact from 12:00 p.m. to 4 p.m. for Align Technologies on October 17, 2013 and Edward Life Sciences on April 23, 2013 in Panels A and B, respectively. Earnings for both companies were announced right after the closing of markets at 4 p.m. Price impact is defined as $2d_k(M_{k+5} - M_k)$. $d_k$ a trade indicator that equals +1 if the trade is a market order buy and -1 if it is a market order sell. $M_k$ is the midquote of the NBBO quotes prevailing when trade $k$ occurs and $M_{k+5}$ is the prevailing midquote 5 minutes after trade $k$. The gray vertical lines represent the time of an insider trade (buying or shorting a the stock).

Panel A. Align Technology (2013/10/17)




Panel B. Edward Life Sciences (2013/04/23)

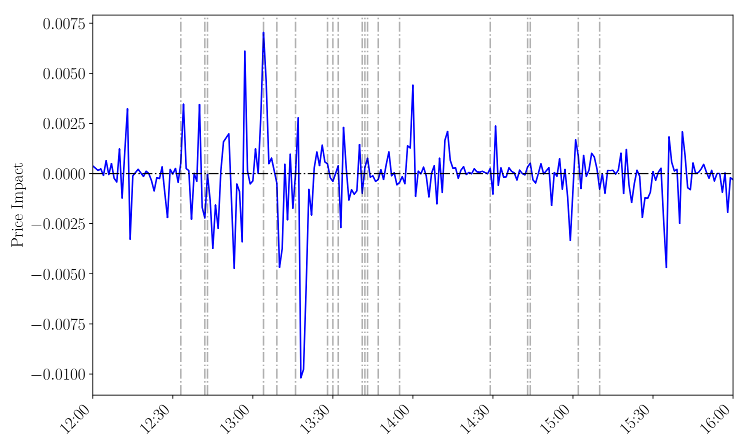


**Table A1**
**Unbiasedness regressions**

This table reports the coefficients of the following regression:

$$
\begin{aligned}
LogReturn_{i,t\to t+1} =& \beta_1 Log\,Return_{i,t,12-4PM} + \beta_2 \mathbf{1}_{[\text{Hacked}]i,t} + \beta_3 Log\,Return_{i,t,12-4PM} \times \mathbf{1}_{[\text{Hacked}]i,t} \\
& + \Gamma' Controls_{i,t} + \alpha_i + \alpha_t + \varepsilon_{i,t},,
\end{aligned}
$$

where the dependent variable $LogReturn_{i,t\to t+1}$ corresponds to the log return from 12 p.m. on day $t$ to 9:30 a.m. the following trading day $t+1$ in column (1) and from 12 p.m. on day $t$ to 4 p.m. on the following trading day $t+1$ of firm $i$ earnings announcement. An earnings announcement occurs during after hours (between 4 p.m. on day $t$ to 9:30 a.m. $t+1$ ). $\mathbf{1}_{[\text{Hacked}]}$ is a dummy variable equal to one if announcement $i$ is released by a hacked newswire 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. Results are presented with firm and year-quarter fixed-effects. Robust standard errors clustered by firm and year-quarter are presented in parenthesis and *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level respectively. The sample period is from January 2010 to December 31, 2015.

| | Dependant variable return window | |
|---|---|---|
| | 12 p.m.(t)-9:30 a.m.(t+1) | 12 p.m.(t)-4 p.m.(t+1) |
| | (1) | (2) |
| $Log\ Return_{12-4PM}$ | 0.776*** | 0.589*** |
| | (0.028) | (0.038) |
| $Log\ Return_{12-4PM} \times \mathbf{1}_{[\text{Hacked}]}$ | 0.133*** | 0.148** |
| | (0.046) | (0.060) |
| $\mathbf{1}_{[\text{Hacked}]}$ | -0.001 | -0.001 |
| | (0.001) | (0.001) |
| $N$ | 43,991 | 43,991 |
| Adjusted $R^2$ | 0.082 | 0.057 |
| Controls | Y | Y |
| Year-Quarter F.E. | Y | Y |
| Firm F.E. | Y | Y |