

# First to “Read” the News: News Analytics and Algorithmic Trading

von Beschwitz, Bastian, Donald B. Keim, and Massimo Massa

Please cite paper as:  
von Beschwitz, Bastian, Donald B. Keim, and Massimo Massa  
(2018). First to “Read” the News: News Analytics and  
Algorithmic Trading. International Finance Discussion Papers  
1233.

<https://doi.org/10.17016/IFDP.2018.1233>



## International Finance Discussion Papers

Board of Governors of the Federal Reserve System

Number 1233

July 2018

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 1233

July 2018

**First to “Read” the News: New Analytics and Algorithmic Trading**

Bastian von Beschwitz, Donald B. Keim, and Massimo Massa

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at [www.federalreserve.gov/pubs/ifdp/](http://www.federalreserve.gov/pubs/ifdp/). This paper can be downloaded without charge from the Social Science Research Network electronic library at [www.ssrn.com](http://www.ssrn.com).

# First to “Read” the News: News Analytics and Algorithmic Trading

Bastian von Beschwitz\*  
Federal Reserve Board

Donald B. Keim\*\*  
Wharton School

Massimo Massa\*\*\*  
INSEAD

May 16, 2018

## Abstract

Exploiting a unique identification strategy based on inaccurate news analytics, we document a causal effect of news analytics on the market irrespective of the informational content of the news. We show that news analytics speed up the stock price and trading volume response to articles, but reduce liquidity. Inaccurate news analytics lead to small price distortions that are corrected quickly. The market impact of news analytics is greatest for press releases, which are timelier and easier to interpret algorithmically. Furthermore, we provide evidence that high frequency traders rely on the information from news analytics for directional trading on company-specific news.

**JEL classification: G10, G12, G14**

**Keywords: Stock Price Reaction, News Analytics, High Frequency Trading, Press Releases.**

---

\* Bastian von Beschwitz, Federal Reserve Board, International Finance Division, 20th Street and Constitution Avenue N.W., Washington, D.C. 20551, tel. +1 202 475 6330, e-mail: [bastian.vonbeschwitz@frb.gov](mailto:bastian.vonbeschwitz@frb.gov) (corresponding author).

\*\* Donald B. Keim, Wharton School, University of Pennsylvania, Philadelphia, PA 19104; [keim@wharton.upenn.edu](mailto:keim@wharton.upenn.edu)

\*\*\* Massimo Massa, INSEAD, Finance Department, Bd de Constance, 77305 Fontainebleau Cedex, France, tel. + 33-(0)160-724-481, email: [massimo.massa@insead.edu](mailto:massimo.massa@insead.edu)

An earlier version of this paper was titled "Media-Driven High Frequency Trading : Evidence from News Analytics". We are grateful to RavenPack for providing their data, and Malcolm Bain in particular for his expertise on different RavenPack releases. Thanks also to the technical personnel at WRDS, especially Mark Keintz, for making the construction of the intraday-market indexes possible. We thank Joseph Engelberg, Nicholas Hirschey, Todd Gormley, Markus Leippold, Joel Peress, Ryan Riordan, Paul Tetlock, Sarah Zhang and conference participants at the NBER Microstructure Meeting, European Winter Finance Summit, FIRS, and DGF for valuable comments. We acknowledge the financial support of the Wharton-INSEAD Center for Global Research and Education. All remaining errors are our responsibility. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

## **Introduction**

A major purpose of financial markets is the assimilation of information into prices. Since the advent of securities trading, informationally-relevant news has been read and processed by humans, first directly from newspapers, then from news wires such as Dow Jones, Reuters, and Bloomberg. However, in the last two decades, computer algorithms have increasingly been used to read and interpret financial news. Given the importance of news for financial markets, it is crucial to understand how the algorithmic processing of news releases by computers (“news analytics”) affects financial markets. In particular, in what ways do news analytics affect stock returns and trading volume? Who are the users of news analytics? And for which type of articles are news analytics most important?

We address these questions using news analytics provided by RavenPack, the leading provider of news analytics in the market. RavenPack uses computer algorithms to determine for each article in the Dow Jones Newswire its relevance to each company mentioned in it, and whether the news is positive or negative. This processed content is then electronically delivered to RavenPack’s subscribers within a third of a second, allowing them to react to the news faster than humans possibly could.

We address three broad questions in this paper. The first is whether news analytics have a causal effect on the stock market, making prices and trading volumes react faster to news wire articles and thereby increasing market efficiency. The second asks for which articles is the causal effect of news analytics most important. In particular, we compare press releases that are directly released by companies to other articles that are written by the journalists of Dow Jones. The third asks whether news analytics are used only for directional trading or also to avoid adverse selection.

We study this question by focussing on high frequency traders (HFTs), which we argue are the type of trader most likely to use news analytics to avoid adverse selection.

These questions are difficult to address in practice because the *response to news analytics normally cannot be distinguished from the reaction to the news itself*. We are able to address this distinction by exploiting a unique identification strategy based on inaccuracies in news analytics that are revealed by comparing older and newer versions of RavenPack. We use the back-filled analytics of increasingly more sophisticated versions of RavenPack to identify inaccuracies in the old version that was released to the market. Finding evidence that markets react to such inaccuracies would suggest a causal impact of RavenPack on the stock market.

To identify inaccuracies in news analytics, we focus on differences in RavenPack’s “relevance score”, which measures the importance of an article for a certain company. The relevance score is very important: highly relevant articles that are positive (negative) are on average followed by positive (negative) stock returns, while there is almost no reaction to articles with a low relevance score. Differences in relevance scores between the old and new RavenPack versions are due to improvements in the algorithm when identifying companies in the article and determining the article’s relevance to the company.

We use these differences in relevance scores to define three categories of articles: High-relevance articles Released as High-relevance articles (HRH) are articles that were correctly released to the market; Low-relevance articles Released as High-relevance articles (LRH) are false positives, i.e. articles that are wrongly attributed to a company; and High-relevance articles Released as Low-relevance articles (HRL) are false negatives, i.e. articles that the old version of RavenPack failed to attribute to the correct company.

To assess the causal effect of Ravenpack, we start by focussing on LRH articles. We find that the market indeed reacts to such false positives, but the effect does not persist. The market initially overreacts to the incorrect information, realizes the inaccuracy, and quickly corrects after 30 seconds. This finding confirms the causal effect of RavenPack on stock prices but also suggests that the market is quite resilient against disturbances from inaccurate news analytics.

To reinforce our finding of a causal effect, we examine the difference in the market's reaction to HRH and HRL articles. These two article types are of similar relevance according to the most recent version of RavenPack, but only HRH articles were released to the market as highly relevant. Because HRL articles were incorrectly released as not relevant, they should not trigger a causal effect on stock prices. Thus, comparing the difference in market responses between HRH and HRL articles allows us to assess the causal effect of RavenPack.

We find that the share of stock price reaction concentrated in the first 5 seconds after an article, compared to the total reaction over 120 seconds, is significantly greater for HRH articles than for HRL articles. This speed of the stock price response is 1.3 percentage points higher for HRH articles, or *10% relative to the mean*. The market not only reacts faster to HRH articles, but it also reacts in the sentiment direction indicated by RavenPack. The RavenPack sentiment direction of an article predicts the stock price reaction to HRH articles better than to HRL articles. This implies that traders use RavenPack to trade in the direction of the sentiment indicator provided by the news analytics.

In addition to the faster stock price response, we also document an increase in the share of trade volume concentrated in the first 5 seconds compared to the two minutes after an article. This increase in the speed of trade volume response is consistent with the theoretical prediction that investors with a speed advantage trade aggressively on signals that they can exploit before other

traders (e.g., Foucault, Hombert, and Rosu (2016)). Taken together, these findings confirm that RavenPack has a causal effect on the stock market, resulting in both prices and trading volume reacting more quickly to the information delivered by news analytics and, thereby, improving market efficiency.

Having established the baseline finding that news analytics affect the stock market, we ask for which article types news analytics have the largest causal effect. We distinguish between press releases that are directly released by companies and articles written by Dow Jones' journalists, and find that RavenPack has a statistically significantly larger effect for press releases. The speed of stock price response increases 2.6% for press releases that are HRH, while it increases only 0.8% for other HRH articles. This difference is even starker for trading volume: Being correctly covered in RavenPack increases the speed of trade volume response by 1.4% for press releases but only by an insignificant 0.2% for other articles. Taken together, these results confirm that the effect of RavenPack is mainly concentrated in press releases.

Why do news analytics have a larger effect for press releases than for other articles? We show that press releases are timelier: they are 8% more likely to be the first article of the day for the company and 17% more likely to be a new news story rather than a reprint of an earlier story. In addition, RavenPack sentiment is more accurate for press releases, correctly predicting the direction of the stock price reaction in the 2 minutes around the article more often.

These findings are consistent with the notion that traders view RavenPack as being more reliable for press releases. We extend this idea to the time series, by asking whether users of RavenPack learn dynamically about its signal quality. We find that they do: the causal effect of RavenPack on the 5-second return is stronger if RavenPack has been more informative in the past 6 months, measured by whether sentiment scores accurately predicted 2-minute returns following

the article. This finding suggests that algorithmic traders learn dynamically about the precision of RavenPack, and that they rely more heavily on RavenPack's sentiment scores if these scores have been more informative in the past. Such learning could be programmed into their algorithms (machine learning) or can come from manually updating their algorithms over time.

Next, we focus on the two ways in which traders can use news analytics. They can either use them to get an informational edge to conduct directional trades, or they can use them to learn when to get out of the market to avoid adverse selection or elevated order execution costs. The causal effects of RavenPack on returns and trading volume clearly suggest that RavenPack is used for directional trading. But is it also used to avoid adverse selection?

To examine this question, we focus on high frequency traders (HFTs). HFTs are a subset of algorithmic traders that have invested heavily to gain a speed advantage, for example through co-location at an exchange or hyper-fast connections between different exchanges (such as microwave towers). Common trading strategies associated with HFTs include market making and cross-venue arbitrage (Boehmer, Li, and Saar (2017), Zhang (2017)). While executing these strategies, HFTs submit limit orders that are at risk of being picked off when new fundamental information reaches the market. Therefore, we believe that HFTs are the class of trader most likely to use RavenPack to avoid adverse selection. They would hold their usual algorithms and cancel their outstanding orders whenever a new article about the firm is released.

To study this question, we use the NASDAQ high frequency trading data first used in Brogaard, Hendershott and Riordan (2014) which identifies the traders that NASDAQ knows are HFTs. Because this sample is limited to 120 stocks and just two years of data (2008-2009), a comparison between HRL and HRH articles is not feasible. Instead we conduct a simple time-series comparison on how the release of RavenPack to the market affects HFT trading for all



relevant articles. We focus on the fraction of HFT trading in the 5 seconds after an article, standardized by the fraction of HFT trading in the 120 seconds after the article. If HFTs use RavenPack to avoid adverse selection, we would expect this measure to decrease. Instead, it increases by 1.8% after the release of RavenPack, indicating that HFTs make up a larger fraction of trading in the 5 seconds after an article once RavenPack is live. We also find, in line with our previous results, that this effect is much stronger for press releases. Moreover, we find that HFTs mainly increase their liquidity demanding trades after an article and that these trades are predominantly in the direction of the article sentiment. Taken together, these results suggest that HFTs do not use RavenPack to avoid adverse selection but rather to place directional bets.

Given that even HFTs mainly engage in liquidity demanding trades after the release of an article, we ask whether RavenPack causes a faster decline in liquidity following articles. The idea is that the directional trades triggered by RavenPack hit existing quotes and cause liquidity to decline following an article. We find that this is indeed the case. Both effective spreads and Amihud illiquidity increase in the five seconds following an HRH article (compared to HRL articles).

A series of robustness checks confirm our results. One potential concern is that HRH articles may be systematically different from HRL articles. We address this concern in two ways. First, we show that HRH and HRL articles are similar in terms of long-run stock price reactions and several other characteristics. Second, we use the fact that RavenPack has back-filled the data of all versions to February 2004 and conduct placebo tests during the time before RavenPack went live. If our results are driven by actual differences between the two article types, rather than a causal impact of RavenPack, then we should find significant differences in price reactions before RavenPack went live. However, for all tests before RavenPack went live we find insignificant differences

(between HRH and HRL articles). Moreover, the stock price reactions to HRH and HRL articles start to diverge precisely when RavenPack went live, and the resulting increase in the difference between HRH and HRL articles is significant. All of this suggests that our results are robust.

In this paper we show that many algorithmic traders, including HFTs, use RavenPack for directional trading. This results in RavenPack having a significant impact on the market in terms of returns, trading volume, and liquidity. This effect goes beyond the underlying influence of the news itself. While our study can only detect the effect of RavenPack, there are other providers of news analytics, and traders may conduct algorithmic news processing in house. Thus, the total effect of algorithmic news processing is likely much larger than the effect of RavenPack measured in this paper. Also, given that RavenPack is the leading provider of news analytics, we expect the results to carry through to wide-subscription news analytics services more generally.

Our results contribute to four major strands of literature. First, we contribute to the literature on the causal effect of media on the stock market.<sup>1</sup> Methods to address the endogeneity of media coverage include exogenous scheduling of journalists (Dougal, Engelberg, Garcia, and Parsons, 2011), local media coverage and its delay due to extreme weather (Engelberg and Parsons, 2011) and newspaper strikes (Peress, 2014). We add to this literature in three ways. First, we study news analytics, rather than news articles themselves. News analytics are special in that they are a derivative of news articles that contain *less* information than the article itself. Their only advantage is that they are easier to process algorithmically. Our results show that in the age of algorithmic trading, processability is just as important as informational content. Second, we study the effect of

---

<sup>1</sup> There is a wider literature on media and stock markets including for example Chan (2003), Tetlock (2007, 2011), Fang and Peress (2009), Griffin, Hirschey, and Kelly (2011), Boudoukh et al. (2016), Loughran and McDonald (2013), Garcia (2013), Ferguson et al. (2015), Hu, Pan, and Wang (2017)). For a review on textual analysis in finance see Kearney and Liu (2014).

such news analytics on algorithmic traders rather than private investors. This focus increases the policy relevance of our findings in a regulatory environment that is increasingly focused on news analytics. Third, we show that the impact of news analytics on prices are particularly important for press releases, which are not the subject of the prior studies.

Second, we contribute to the literature on news analytics. Prior papers in this literature study the correlation between the market and news analytics without passing judgment on whether there is a causal impact of news analytics on the market (e.g. Dzielinski and Hasseltoft (2017), Riordan, Storkenmaier, Wagener, and Zhang (2013), Gross-Klugmann and Hautsch (2011), Sinha (2016), Heston and Sinha (2016)). In contrast, our paper is the first to show the *causal* impact of news analytics on stock markets.

Third, we contribute to the growing empirical literature on algorithmic and high frequency trading.<sup>2</sup> Several papers show that high frequency traders use information from order flow (e.g. Hirschey (2018) or information from related asset prices (e.g., Chaboud et al. (2014), Boehmer, Li, and Saar (2017), Zhang (2017)). In contrast to these studies, we show that HFTs do not only trade on market information, but also enter directional bets based on news analytics, which contain new, company-specific information that is not yet reflected in any market prices.

Fourth, our results are consistent with recent models of high frequency trading in which some traders have an informational advantage. For example, Foucault, Hombert, and Rosu (2016) model a situation in which a speculator receives information one period ahead of the market maker in a

---

<sup>2</sup> Examples of this literature include Brogaard, Hendershott and Riordan (2014), Boehmer, Fong and Wu (2015), Hendershott and Riordan (2013), Hendershott, Jones, and Menkveld (2011), Baron, Brogaard, Hagströmer, and Kirilenko (2017), Menkveld (2013), Jovanovic and Menkveld (2010), Riordan and Storkenmaier (2012), Boehmer, Fong, and Wu (2015), Hasbrouck and Saar (2013), Benos and Sagade (2016), Clark-Joseph (2013), Hirschey (2018), Brogaard et al. (2014), Chordia, Green, and Kottimukkalur (2017). A survey of this literature is provided by Jones (2013).

set-up similar to Kyle (1985); in Martinez and Rosu (2013) some agents have a short-lived informational advantage; and in Dugast and Foucault (2017), speculators face a trade-off between processing a signal faster or more accurately. Faster traders in these models make markets more informationally efficient, but also more unstable. We find support for both effects.

## **2. Test design, identification strategy, and data sources**

In this section we first describe the RavenPack news analytics data and how it is used in our identification strategy and tests. After briefly describing our stock market data, we then present summary statistics for the variables used in our tests. Variable definitions are in Appendix 1.

### *2.1 RavenPack*

RavenPack provides real-time news analytics based on the Dow Jones (DJ) Newswire. This service analyzes all the articles on the DJ Newswire with a computer algorithm and delivers article-level relevance and sentiment metrics to its users. It determines which companies are mentioned in the article, how relevant the article is to the company and reports different sentiment indicators about whether the article is good or bad news for the company. The latency – i.e. the time from the release of the DJ Newswire to the release of the RavenPack metrics – is approximately 300 milliseconds. RavenPack claims it has the “timeliest company sentiment indicators in the marketplace.”<sup>3</sup> As such, RavenPack is ideally suited for the use of traders engaging in algorithmic news trading.

#### *2.1.1 Ravenpack – definition of variables*

We extract from RavenPack the following variables. *Article Category* is a variable determining the topic of the article and the role played by the company in the article. For example, *Article*

---

<sup>3</sup> “RavenPack Enables Trading Programs with Sentiment on 10,000 Global Equities,” RavenPack press release from May 28, 2009.

*Category* might be “acquisition – completed – acquirer” for a company announcing the completion of an acquisition of another company or “rating – change – negative – rater” for a rating company that just downgraded another company. The identification of the news topic is based on a purely algorithmic approach, and a large percentage of articles cannot be classified in this way. *Article Category Identified* is a dummy variable equal to 1 if *Article Category* is identified by RavenPack, and zero otherwise.

There are two major sentiment scores in RavenPack. The *Composite Sentiment Score* (CSS) is based on several individual RavenPack sentiment measures. It takes a value ranging from 100 (positive) to 0 (negative), where 50 is a neutral article. It is available for each article. The *Event Sentiment Score* (ESS) is coded in the same way as CSS, but available only if the category of the article can be identified. We aggregate these two scores into a single sentiment variable called *Sentiment Direction*, which is primarily based on ESS and uses CSS only if ESS is either missing or equal to 50 (neutral).

*Relevance* is an index provided by RavenPack that indicates the relevance of an article to the company. *Relevance* takes values ranging from 0 (least relevant) to 100 (most relevant). If the type of the article can be identified and the company plays an important role in the main context of the story – e.g. is an acquirer or announces a buyback – then *Relevance* is 100. If the company is mentioned in the title, but the type of article cannot be identified, then *Relevance* ranges between 90 and 100. If the company is mentioned, but plays an unimportant role, then it gets a low *Relevance* score – e.g., a bank advising an acquisition might get a score of 20. We would not expect such articles to affect the bank’s stock price very much.

In line with this, RavenPack recommends “filtering for *Relevance* greater than or equal to 90 as this helps reduce noise in the signal”. To examine this claim, Figure 1 plots the market reaction

to news as a function of *Relevance*. We plot the cumulative returns relative to news events from April 1, 2009 to September 10, 2012. We multiply returns by the article's sentiment direction. The articles with *Relevance* greater than 90 do indeed have an important effect on stock prices, but there is no reaction to articles with *Relevance* below 90. Thus, we will refer to articles with *Relevance* below 90 as low relevance. This analysis suggests that RavenPack is good at identifying both relevance and sentiment of an article.

That the reaction to high relevance articles starts about 60 seconds before the article suggests that *some* of the news events are covered in other news sources before they are covered in the DJ Newswire (used by RavenPack). Cases where the DJ Newswire is not the first to report an event should only work against us by making it more difficult to find a causal impact of RavenPack. We have no reason to believe that this issue should bias the results because it should be unrelated to whether RavenPack makes a mistake interpreting the article. While some trades in the 5 seconds after a RavenPack article may be due to human traders reacting to earlier coverage of the news elsewhere, this trading should affect both HRL and HRH articles. Thus the *additional* trading following HRH articles (relative to HRL articles) should only be due to algorithmic traders reacting to the coverage in RavenPack itself.

### *2.1.2 Ravenpack – test design using different product versions*

RavenPack released its first version (v. 1.0) to the market on April 1, 2009,<sup>4,5</sup> and a revised version of the service (v. 2.0) with additional features on June 6, 2011. The most recent version we use (v.

---

<sup>4</sup> Even though the official release date of the RavenPack service was May 2009, some customers had access to the service as early as from April 1, 2009. Thus, we refer to April 1, 2009 as the introduction of RavenPack. Before April 2009 RavenPack had a pre-existing service that also released sentiment information on the Dow Jones News Wire. However, this service was meant for longer term news analysis, such as charting sentiment over several days. The prior service was not provided timely enough to be used at high frequency.

<sup>5</sup> RavenPack 1.0 was actually released on Sept 6, 2010. A predecessor to v.1.0, that was similar to v.1.0, is the version that was released on April 1, 2009. This predecessor version was not made available to us, but RavenPack confirmed that it was very similar to RavenPack 1.0.

3.0) was released on September 10, 2012. RavenPack has provided us with data from each of the release-specific algorithms, each having been back-filled to February 2004. RavenPack does not continuously update its algorithm, so as not to distort its customers' trading strategies which might be based on specific variable definitions. Rather, RavenPack rolls out any changes to its algorithm when releasing a new version, meaning that stock-specific metrics from the three releases can sometimes differ.<sup>6</sup> These differences are often related to the way companies are identified in an article and how the relevance of an article to a company is determined.<sup>7</sup> Thus, there are articles that might be associated with a particular company in one RavenPack release, but not in another. Such differences in the relevance of articles to companies in different versions provide the basis for our tests. Assuming the most recent version of RavenPack (v. 3.0, hereafter New RavenPack) is the most accurate, we can identify inaccuracies in RavenPack 1.0 and RavenPack 2.0 (hereafter Old RavenPack) that were released to the market. If the market reacts to these inaccuracies, it is an indication of a causal effect of RavenPack on the stock market.

Our analysis can be thought of as assuming two types of traders: Algorithmic traders that subscribe to RavenPack and human traders that manually read the article to determine its content. Further, we assume that human traders can more precisely derive the relevant signal from the article, while algorithmic traders have an advantage in terms of speed (a setting modelled by Dugast and Foucault (2017)). This means that RavenPack allows its subscribers to trade faster on a possibly less precise signal. In the short run, when only algorithmic traders can react to news,

---

<sup>6</sup> Because the algorithm is proprietary, we do not know exactly what changes RavenPack implemented but some examples of articles where the two versions disagree are provided in the Internet Appendix.

<sup>7</sup> In addition, the number of companies covered by RavenPack has also increased between releases. There are 156 companies (3%), which are only covered in New RavenPack. We ensure by using company fixed effects that this difference in coverage is not driving our results.

RavenPack will have the largest impact; while in the long run human traders determine the price reaction because their signal is more precise.

In the empirical implementation we choose specific time intervals to constitute the short and long run. We define the short run to be 5 seconds, because this is long enough to capture the full reaction of algorithmic traders (and accommodates slower algorithmic traders that are not co-located and not trading within milliseconds), but is too short for a human trader to read an article, process it and make a trading decision based on it. We choose two minutes as the long run because this permits enough time to read an article and trade on it, whereas longer time windows will be more affected by noise. In the Internet Appendix, we provide robustness checks in which we use both 1 and 10 seconds for the short run and 5 minutes for the long run.

We define the following article types that we also list in Panel A of Table 1. *High relevance article Released as High relevance article* (HRH) is defined as an article classified as relevant in both Old and New RavenPack. We predict that such a correctly released article creates a fast and persistent market reaction. *High relevance article Released as Low relevance article* (HRL) is defined as an article with high relevance in New RavenPack, but incorrectly assigned low relevance in Old RavenPack. Low relevance means either the article was not assigned to the company or the relevance score was below 90. We expect an HRL article to have a similar long-run market reaction as an HRH article because they are of similar relevance according to New RavenPack. However, we would expect a slower market reaction to an HRL article as it was not released originally as relevant. *Low relevance article Released as High relevance article* (LRH) is an article that was incorrectly released to investors as having high relevance but has low relevance in New RavenPack. For these articles we expect an initial overreaction of algorithmic traders, which might later be reversed by human traders. Examples of all three article groups are provided



in the Internet Appendix. A fourth article category is *Low relevance articles Released as Low relevance articles* (LRL); these articles have a relevance score below 90 in both versions.<sup>8</sup> We do not expect much market reaction to LRL articles.

These predictions allow for two possible empirical set-ups using the two types of mistakes. First, we could study overreaction to *false positives* by examining LRH articles or we could study underreaction to *false negatives* by examining HRL articles. In both cases, it would be desirable to have a control group of articles where RavenPack determined the relevance accurately. Having a control group would enable us to include time and firm fixed effects and a host of control variables. The ideal control group would consist of articles that are similar in terms of importance and content but have been released to the market with a different relevance. The obvious candidates are articles that have the same relevance (according to New RavenPack) but were released to the market as having a different relevance (in Old RavenPack). Thus, the candidate control group for HRL articles are HRH articles and the candidate control group for LRH articles are LRL articles.

For these control groups to be appropriate, we would require HRL and HRH articles to be of similar relevance and LRH and LRL articles to be of similar relevance. This is the case if New RavenPack determines the relevance of an article accurately and if Old RavenPack contains no information on the relevance of the article over and above the information contained in New RavenPack. This is a fairly strong assumption, but it is testable. Because we have data before RavenPack went “live” in 2009, we can compare the market impact to the different types of articles during the time period when RavenPack could not have had any causal market impact, because it was not yet “live”. To examine the market impact of the different article types, we regress absolute

---

<sup>8</sup> LRL articles also include articles that have a relevance score below 90 in either Old RavenPack or New RavenPack and are not assigned to the company in the other version.

return and turnover in the two minutes after the article's release on dummy variables equal to 1 for HRH, HRL and LRH (with LRL being the omitted category). To control for firm- and time-specific effects, we include firm, date and hour-of-the-day fixed effects.

The results are presented in Panel B of Table 1. When we test the difference between LRH and LRL articles, we find a large and significant difference for both absolute returns and turnover, suggesting LRH articles are significantly more important than LRL articles. Thus, a test of overreaction comparing LRH and LRL articles is not possible because the two article types are fundamentally different. Thus, instead of comparing LRH and LRL articles, we rely on graphical evidence and also compare the reaction to LRH articles before and after RavenPack went "live".

When we test the significance of the difference between the coefficients on HRH and HRL, however, we find the difference is small and insignificant for both turnover and absolute returns. This finding suggests that HRH and HRL articles have very similar relevance. Thus, any difference in their market reaction can be attributed to the causal effect of RavenPack. Therefore, most of our tests are based on the comparison between these two article types. To ensure further that differences in importance between HRH and HRL are not driving our results, we control for article characteristics and study the speed of market reaction, i.e. the size of the short run reaction relative to the long-run reaction. We also conduct placebo checks for our tests showing that our results are not driven by differences between HRH and HRL articles, but by the causal effect of RavenPack.

## *2.2 Stock market data*

We use intraday quotes and trade data from TAQ.<sup>9</sup> We use the TAQ National Best Bid and Offer (NBBO) file provided by WRDS for quotes. As a first step, we aggregate trading volume at the

---

<sup>9</sup> We use the usual filters of excluding all trades with zero size, negative prices, correction code different from 0 and bid ask quotes where the bid is above the asked.

frequency of one second, and compute second-by-second returns based on end-of-second bid-ask midpoints. We use bid-ask midpoints rather than trading prices to avoid bid-ask bounce effects. Even after this aggregation, the data for *all* stocks in our 8-year sample is far too large to be used in a second-by-second panel set-up. But we are interested only in the market reaction around specific company news events, so we limit our analysis to a few minutes around these events. This simplification allows us to study all US common stocks over the full 8-year sample period.

To control for the overall market movements taking place during this period, we compute a second-by-second intraday market index from the total TAQ universe. We compute second-by-second returns, turnover, and value-weighted volatility for the market index. We also compute returns for industry-specific indices for the 12 Fama French industries. The details of the index construction are explained in Internet Appendix 1. To control for stock-specific information, we use the CRSP daily stock file and compute the prior month's return, volatility, turnover, Amihud (2002) illiquidity measure, and market capitalization.

We employ the following filters: To be included in our sample, a stock must be covered in CRSP and TAQ, must have SHRCD 10 or 11, must have a beginning of the day stock price of at least \$1 and must have a beginning of the day percentage bid-ask spread of less than 10%. We exclude articles that occur outside trading hours or in the first or last 20 minutes of trading in the day. To avoid distortions from overlapping windows around articles, we exclude articles for which the company had an article in the prior 15 minutes. We also exclude four companies that appear in articles mainly as information providers: McGraw-Hill, NASDAQ, CME and Moody's. Because we need an initial bid-ask midpoint to compute a first return and because we want to avoid a stock's turnover influencing the stock price we measure, we use seconds  $t-480$  to  $t-1$  as a burn-

in period. Only articles for which the stock has a quote in those 8 minutes before the article are included in our analysis.

### *2.3 Summary statistics and comparison between HRH and HRL*

The final sample consists of 321,912 article-firm combinations, starting with the release of RavenPack 1.0, over the period April 1, 2009 to September 10, 2012. In Panel A of Table 2, we report descriptive statistics for all our variables for the combined sample of articles classified as HRH and articles classified as HRL.

As alluded to previously, a concern is that the information content of HRH and HRL articles might be different. Therefore, we compare their difference in terms of observable variables in Panel B. For this purpose, we regress each article characteristic on a dummy variable,  $D(\text{HRH})$  equal to 1 if the article is HRH, and relevance, category, hour and date fixed effects. We report the coefficient of  $D(\text{HRH})$  as well as a t-statistic two-way clustered at the firm and date level. There is no statistical significant difference between the HRH and HRL articles in terms of firm size, sentiment scores, time since the last article, turnover, and illiquidity. Most importantly, we find no evidence that HRH are more important than HRL articles: the absolute returns both over the 2 minutes following an article and on the full trading day of the article are actually (insignificantly) lower for HRH articles. The only significant difference is that companies in HRH articles have a slightly lower return (0.03%) in the prior month than those associated with HRL articles, and that HRH articles cover fewer firms per article. However, these differences are small in economic terms (0.05 and 0.22 standard deviations) and we account for these differences with control variables in all our subsequent tests. In addition, we show in the Internet Appendix that our results are robust to including fixed effects for the number of firms mentioned in the article. That the characteristics of HRH and HRL articles are similar alleviates worries that our results are driven by differences

in the article types. In addition, we run placebo tests to confirm that unobservable differences are not driving our results.

### **3. News analytics and impact on the market**

#### *3.1 News analytics and temporary price distortions*

Here we examine whether inaccuracies in news analytics lead to temporary price distortions, i.e. to an overreaction in stock price that is afterwards reversed. As explained in Section 2.1.2, we expect the market to overreact to LRH articles, i.e. articles that New RavenPack identifies as having low relevance, but that were incorrectly released as having high relevance in Old RavenPack.

We first consider graphical evidence. In Figure 2, we compare the market reaction of articles HRH and LRH articles. We focus on the cumulative returns from  $t-60$  to  $t+120$  seconds around the article (measured relative to  $t=0$ ). We focus on signed returns—i.e. returns multiplied with the sentiment direction of the article—to be able to combine positive and negative news in one analysis. We exclude articles with neutral sentiment. Figure 2 shows that the market overreacts to LRH articles. In the short-run these articles have a price reaction that is very similar to HRH articles. However, after approximately 30 seconds – a reasonable time for a fast human trader to verify that no important news for the stock has arrived – the stock price reaction to LRH articles starts to retreat. After approximately 2 minutes, a large part of the short-run reaction to these articles has reversed.<sup>10</sup> In contrast, articles classified as HRH have a longer-term effect on price, lasting more than two minutes. This finding is consistent with a causal effect of RavenPack that leads algorithmic traders to trigger an initial overreaction to the article that is then corrected by

---

<sup>10</sup> We would never expect full mean reversion given that LRH articles contain at least some information (see Table 1 Panel B).

other traders. While the magnitude of this effect is only about 1 basis point on average, this is due to the fact that the average price reaction to company-specific news is fairly small. The average absolute return reaction in the first 2 minutes after an article is only 11.4 basis points and the average signed return is only 1.9 basis points (see Table 2).

Next, we provide a multivariate analysis. The problem in studying LRH articles in a regression set-up is that we do not have an appropriate control group for these articles as they are more relevant than LRL articles, but less relevant than HRH articles (see Table 1 Panel B). Therefore, we use LRH articles from the period before RavenPack went “live” as the control group.<sup>11</sup> During this period, the fact that these articles would have been marked as relevant by RavenPack should not affect stock prices. However, after the introduction of RavenPack, LRH articles can have a causal effect on the market, which should lead to a short-run overreaction to the article. Thus, we study whether LRH articles have a stronger short run stock price impact and a larger reversal after RavenPack goes “live”, as compared to before. We estimate the following article-level regression including only LRH articles:

$$Return(t-1, t+5)_a = \beta_1 * Sentiment\ Direction_a * RavenPack\ Release_t + \beta_2 * Sentiment\ Direction_a + \beta_3 * RavenPack\ Release_t + \gamma * Controls + \varepsilon_a$$

where *Sentiment Direction* indicates whether the article is positive or negative news and *RavenPack Release* is a dummy variable equal to 1 after RavenPack went “live” in April 2009. The coefficient of interest is the interaction between *Sentiment Direction* and *RavenPack Release*. In addition, we include various combinations of control variables and fixed effects. To control for stock-specific information, we use its market capitalization, return, volatility, and turnover

---

<sup>11</sup> We confirm in IA-Table 1 in the Internet Appendix that the relevance of LRH articles does not change after the introduction of RavenPack.

measured over the prior month, and our illiquidity measure based on Amihud (2002). For brevity, the coefficients on these control variables are reported only in Internet Appendix 3. To control for characteristics of the news announcement, we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score, and the hour during the day in which the article was released. In regressions 3 and 6, we add as additional controls the absolute return, turnover, and volatility for each industry and the market over the horizon from  $t-1$  to  $t+5$  seconds around the article. All standard errors are two-way clustered at the firm and date level. Appendix 1 contains a detailed description of all variables.

We present the results in Table 3. In regressions 1-3, the dependent variable is the short-run stock return from 1 second before to 5 seconds after the article. We find that the short-run return associated with LRH articles is significantly more positively correlated with the sentiment of the article after RavenPack went “live” in 2009. Because the relevance of LRH articles was the same both before and after the introduction of RavenPack, this finding indicates an overreaction to these articles. Indeed, it seems plausible that algorithms would trade in the direction of the article’s sentiment, because RavenPack (incorrectly) labelled such articles as highly relevant.

Next, we ask whether this overreaction is subsequently reversed. In regressions 4-6, we use the stock price reaction from 6 to 120 seconds after the LRH article as the dependent variable. We find that it is more *negatively* correlated with the article sentiment after RavenPack went “live”, consistent with a reversal. While this result is not statistically significant, the negative magnitude of this coefficient is about the same as the positive magnitude of the coefficient in regressions 1-3, implying that almost all of the short run overreaction is reversed in the two minutes after the article. This result is not significant because two-minute returns are much noisier than the five-second returns studied in regressions 1 to 3.

Taken together, our graphical and regression analyses of LRH articles results provide evidence that inaccuracies in news analytics can cause short term overreaction that is afterwards reversed.

### 3.2 News analytics and speed of stock price response

In this section, we study whether news analytics improve market efficiency by increasing the speed with which stock prices and traders react to news. For this purpose, we focus on false negatives, i.e. HRL articles that are highly relevant according to New RavenPack, but were released as having low relevance in Old RavenPack. For these articles we have a good control group in the form of articles that have been reported as having high relevance in both versions (HRH). Comparing the market reaction to those two article groups allows us to see whether the market underreacts to relevant news when RavenPack does not classify it as relevant.

#### 3.2.1 Regression analysis – speed of stock price response

We consider two alternative analyses for market reaction. First, we examine whether stock prices respond faster to HRH articles irrespective of the direction of the reaction. Then we study whether the sentiment of HRH articles predicts the directional stock price response better than the sentiment of HRL articles. For the first analysis, we define *Speed of Stock Price Response* as:

$$\frac{\text{Abs}(\text{Return } t-1, t+5)}{\text{Abs}(\text{Return } t-1, t+5) + \text{Abs}(\text{Return } t+6, t+120)}$$
 over the 120 seconds around the news event.<sup>12</sup> This variable

measures the amount of the two-minute price change that takes place in the first five seconds after the news release. It is in the spirit of DellaVigna and Pollet (2008). It captures the degree of under-reaction by decomposing the market reaction into its short- and long-term components. The higher

---

<sup>12</sup> We use  $\text{Abs}(\text{Return } t-1, t+5) + \text{Abs}(\text{Return } t+6, t+120)$  rather than  $\text{Abs}(\text{Return } t-1, t+120)$  in the denominator to constrain the variable between 0 and 1 rather than to allow it to approach infinity in cases where  $\text{Abs}(\text{Return } t-1, t+120)$  is close to zero.



the value of *Speed of Stock Price Response*, the more the reaction to the news event concentrates in the first few seconds after the event – i.e., the less under-reaction. We run the following article-level regression including only HRH and HRL articles:

$$\text{Speed of Stock Price Response}_a = \alpha_t + \alpha_f + \beta_1 * D(HRH)_a + \gamma * \text{Controls} + \varepsilon_a$$

where  $D(HRH)$  is a dummy variable that equals one if the article was released to the market as highly relevant (HRH) and zero if it was (incorrectly) released as having low relevance (HRL). The regression is estimated at the article level, thus allowing for both HRH and HRL articles that were released for the same firm or on the same day. This allows us to control in all regressions for unobserved heterogeneity with firm fixed effects and daily fixed effects ( $\alpha_t$  and  $\alpha_f$ ). In addition, we include the same article-level fixed effects and control variables as before.

We report the results in Table 4. In regressions 1 to 3, we estimate our main specification during the time in which RavenPack was live (Apr 1, 2009 – Sept 10, 2012). We observe a positive and significant relation between *Speed of Stock Price Response* and  $D(HRH)$ , indicating that the stock price response is much quicker for an HRH article than for an HRL article. This result holds across all the different specifications and samples. It is not only statistically significant, but also economically relevant. HRH articles exhibit a *Speed of Stock Price Response* that is 1.3 percentage points higher (10% relative to the mean). We find similar results if we compute *Speed of Stock Price Response* using market-adjusted and industry-adjusted returns (reported in Internet Appendix 3). This finding suggests that news analytics increase market efficiency by increasing the speed of reaction to news.

Although we showed in section 2.3 that HRH and HRL articles are similar along most dimensions, one potential concern in this set-up is that the results are driven by the two article categories having different informational content. To address this issue, we use the fact that

RavenPack has back-filled the data to February 2004. If our results are driven by general differences in the two categories, then there should be a difference in stock price reaction before RavenPack went live. In regressions 4 to 6, we report a placebo test for the time period where RavenPack was not yet released to investors (February 1, 2004 – March 31, 2009). This placebo test does not show a statistically significant relation between  $D(HRH)$  and the *Speed of Stock Price Response*, thereby confirming that our main test is appropriate.

Another potential concern is that there might be a trend in the difference of informational content between HRH and HRL articles, which is driving our results. To address this concern, we examine the relation between *Speed of Stock Price Response* and  $D(HRH)$  for different years before and after RavenPack went live. We follow Gormley and Matsa (2011) and plot in Figure 3 the point estimates of a modified version of regression 3 in Table 4, in which we interact  $D(HRH)$  with yearly fixed effects<sup>13</sup>. We plot the coefficients of the interaction of  $D(HRH)$  with one-year dummy variables in Panel A. In Panel B, we report the same regression but interacting  $D(HRH)$  with two-year dummy variables. We report 95% confidence intervals for the coefficients in both panels. In Panel C, we report the simple difference between *Speed of Stock Price Response* for HRH and HRL articles without any controls over different years.

The results confirm that RavenPack has a causal effect on the *Speed of Stock Price Response*. Before the introduction of RavenPack, the difference between HRH and HRL hovers around zero and there is no obvious time trend. After the introduction of RavenPack, the difference is much larger. This suggests the delivery of news analytics by RavenPack has an impact on the market

---

<sup>13</sup> Because RavenPack went “live” in the second quarter of 2009, we assign the first quarter of every year to the prior year. This way, years 2004 to 2008 were entirely before RavenPack went live, while years 2009 to 2011 were completely after RavenPack went live.

that is distinct from the underlying informational content of the news and that our results are not driven by a spurious trend.

### 3.2.2 Regression analysis – directional stock price response

We now ask whether there is a relation between the stock price response and the sentiment direction of the news. That is, does the magnitude of the RavenPack-related stock price response (via correctly-labelled HRH articles) depend on whether the news is positive, negative, or neutral? For this purpose, we ask whether the sentiment indicator in RavenPack better predicts the short run stock price reaction if an article is correctly classified as relevant (HRH) in RavenPack. We use the following article-level regression specification:

$$\text{Return}(t-1, t+5)_a = \alpha_t + \alpha_f + \beta_1 * D(HRH)_a * \text{Sentiment Direction}_a + \beta_2 * \text{Sentiment Direction}_a + \beta_3 * D(HRH)_a + \gamma * \text{Controls} + \varepsilon_a$$

We use the same fixed effects as before, but exclude any sentiment-related control variables as the effect of sentiment will be captured by *Sentiment Direction*.

We report the results in Table 5. We observe a significant positive relation between returns and the interaction of *D(HRH)* and *Sentiment Direction* at the time where RavenPack was live (regressions 1 to 3). That is, the RavenPack-induced stock price reaction is significantly more positive for positive news stories than for negative news stories. This result holds across all the different specifications. Similar results for market- and industry-adjusted returns are reported in Internet Appendix 3. As before, the placebo test in Regressions 4 to 6 shows no effect on returns before RavenPack was live. These results confirm that news analytics have a directional impact on stock prices over and above the one of the underlying news.

While RavenPack improves market efficiency, the information contained in RavenPack does not fully get incorporated into prices instantaneously. In the Internet Appendix in IA-Table 2, we

show that RavenPack sentiment predicts returns from t+5 to t+120 seconds after the article by about 1.35 basis points. Comparing this result with the causal effect of RavenPack on t-1 to t+5 second returns of 0.45 basis points in Table 5, we estimate that about 25% of the information of RavenPack gets incorporated into prices in the first five seconds after the article.<sup>14</sup>

### 3.3 News analytics and trade volume response

We have showed above that news analytics increase the speed of price adjustment after news is publicly released via the DJ Newswire. While the DJ Newswire constitutes a public signal, RavenPack enables faster reaction to such signals for its subscribers. Such a reaction speed advantage is modelled by Foucault, Hombert, and Rosu (2016) who predict that investors trade very aggressively when they receive a signal earlier than other market participants.

Therefore, we investigate whether the faster stock price response to an HRH article is accompanied by a faster trade volume response as well. We define *Speed of Trade Volume Response* as:  $\frac{\text{Turnover}_{t-1,t+5}}{\text{Turnover}_{t-1,t+120}}$ . The variable is defined using the same intervals as *Speed of Stock Price Response*. It captures the amount of trade volume that is concentrated in the first 5 seconds after the news event relative to the trading volume in the two minutes following the news event. We regress *Speed of Trade Volume Response* on  $D(HRH)$  using the same fixed effects and control variables defined above.

We report the results in Panel A of Table 6. We find a strong positive and significant relation between *Speed of Trade Volume Response* and  $D(HRH)$ . This result holds across all specifications.

---

<sup>14</sup> This calculation assumes that the information of the article is fully incorporated into prices within 2 minutes. If the market was fully efficient, RavenPack sentiment information would not predict returns from t+5 to t+120 seconds after the article (because all RavenPack sentiment information would be instantaneously incorporated into prices). Instead, we observe that only 25% ( $0.45/(0.45+1.35)$ ) of this information is incorporated into prices. This finding also shows that it remains profitable to trade on RavenPack sentiment information.

*Speed of Trade Volume Response* is 0.5 percentage points larger for HRH articles than for HRL articles, or 9% relative to the mean. The placebo test shows no significant difference in the speed of trade volume response between HRH and HRL articles before RavenPack went live.

We expect traders using RavenPack to trade in the direction of the article sentiment, so we now examine directional trading volume. Using the methodology of Lee and Ready (1991), we first determine whether a trade is initiated in the direction of the article, i.e. buyer initiated for positive articles and seller initiated for negative articles.<sup>15</sup> We then define *Speed of Directional Trade Volume Response* as  $\frac{\text{Turnover in Direction of Article } t-1,t+5}{\text{Turnover } t-1,t+120}$ , and use it as the dependent variable using the same regression specification (Panel B, Table 6). *Speed of Directional Trade Volume Response* is 0.4% larger for HRH than for HRL articles. Comparing it to the 0.5 percentage point increase in *Speed of Trade Volume Response* suggests that close to 80 percent of the increase in trading volume in the 5 seconds after the article is due to trading in the direction of the article. This finding suggests that RavenPack triggers fast and informed trading.

Taken together, the results in this section show that stock prices react faster and traders trade more aggressively after the release of articles covered in RavenPack. Combined, these results confirm that news analytics have a measurable impact on stock prices in addition to the information content of the news itself and improve price efficiency.

---

<sup>15</sup> Different from Lee and Ready (1991), we use the quote at the end of the previous second as the prevailing quote rather than the quote 5 seconds ago to account for the faster trading processes today.

## 4. News analytics and article type

### 4.1 Types of News: Press releases vs other media-reported news

Our analyses above, suggest that RavenPack has a causal effect on stock prices and turnover and thereby makes markets more efficient. In this section, we dig a little deeper and examine for which types of articles the effect of RavenPack is most relevant. The main distinction between different articles on the Dow Jones News Wire is whether they are press releases that are directly released by companies or articles that are written by Dow Jones' journalists. Press releases make up 23.7% of articles in our sample.

Ex ante, it is not obvious whether we would expect RavenPack to have a stronger or weaker effect for press releases. It might have a stronger effect if press releases are timelier or include more important information. On the other hand, it might have less of an effect if press releases contain less relevant information or the positive spin given by the company makes it difficult for RavenPack to determine the true sentiment of the article.

We examine this question by re-estimating specifications from Tables 4 to 6 but interacting  $D(HRH)$  with  $D(Press Release)$ , which is a dummy variable equal to one if the article is a press release (and zero otherwise). We report the results in Table 7. In regressions 1 and 2, the dependent variable is *Speed of Stock Price Response*. The coefficient of the interaction between  $D(HRH)$  and  $D(Press Release)$  is economically large at about 1.8% and statistically significant at the 5% significance level. This suggests that RavenPack has a significantly stronger effect on the *Speed of Stock Price Response* for press releases than for other articles. In fact,  $D(HRH)$  is only marginally significant, suggesting that our findings are almost exclusively due to press releases and are insignificant for other articles.

In regressions 3 and 4, we find similar results using *Speed of Trade Volume Response* as the dependent variable. Here, the difference between press releases and non-press releases is even larger: Being correctly covered in RavenPack increases the *Speed of Trade Volume Response* by about 1.4% for press releases but only by an insignificant 0.2% for other articles.

Finally, we study the directional effect of RavenPack in regressions 5 and 6 of Table 7. Studying the directional effect is interesting because it requires RavenPack to be able to identify the sentiment of the article correctly, which may be more difficult for press releases that usually entail a positive spin. The dependent variable is the signed return from  $t-1$  to  $t+5$  seconds around the article. The coefficient of the interaction between  $D(HRH)$  and  $D(Press Release)$  is once again positive and statistically significant, suggesting that also the directional effect of RavenPack is mainly driven by press releases. Taken together, our results suggest that RavenPack mainly affects the market reaction to press releases while having limited effects on other articles.

In Panel B of Table 7, we repeat the same analysis using the time period before RavenPack was live. In this robustness check, we do not find significant results and the coefficient of the interaction between  $D(HRH)$  and  $D(Press Release)$  is generally small and insignificant. This finding suggests that our results are not driven by a mechanical difference between press releases and other articles.

#### *4.2 Why are press releases special?*

In this section we ask why RavenPack has a larger effect for press releases than for other articles. We look at several characteristics by comparing their mean for press releases and non-press releases in Table 8. We start by comparing whether there are more incorrectly classified articles for press releases than for other articles. We find that this is not the case: for both press releases and other articles 97% of articles are correctly released as relevant in both versions of

RavenPack, i.e. they are HRH articles. This suggests that the results are not driven by a different fraction of HRH articles among press releases.

Next, we look at a variety of market reaction measures. There is no clear sign that there is a stronger market reaction to press releases. While the return reaction in the 2 minutes after the article is 2 basis points higher, the turnover reaction over the same time period is actually lower. Similarly, the return reaction over the whole trading day is also lower for press releases. This suggests that the results are not driven by the fact that press releases are generally more important.

We confirm that press releases have a more positive sentiment on average, which makes sense considering that companies give stories a positive spin and might not release negative news during trading hours.

Finally, we show that press releases are less stale: They are 8% more likely to be the first article for the company on that day and 17% more likely to be a new news story rather than a reprint as identified by RavenPack. This may be one of the explanations why RavenPack has a larger effect for press releases. The other is that RavenPack is more accurate for press releases. For press releases, the RavenPack sentiment predicts the market reaction in the 2 minutes around the article correctly 56% of the time, while it is only 53% for other articles. This difference is statistically significant. Taken together, our results suggest that RavenPack has a larger market impact for press releases because they are timelier and RavenPack is better in evaluating their sentiment.

#### *4.3 Learning about precision in news analytics*

We find that RavenPack has a stronger effect for press releases in part because it is more accurate for press releases. This suggests that algorithmic traders have learned that RavenPack is more



accurate for press releases and thus use RavenPack more for trading on press releases. This finding raises the more general question whether RavenPack users also learn dynamically about the changing signal precision of RavenPack over time. Such learning could be programmed into their algorithms (machine learning) or come from manual updates to the algorithms. If algorithmic traders learn about the precision of RavenPack, we expect them to rely more on RavenPack’s sentiment indicators if these indicators were more informative in the past. In that case, there should be a stronger price reaction to news analytics at times where news analytics have been more informative in the past.

This raises two issues: how to best measure the informativeness of RavenPack and whether this informativeness is persistent and thus predictable. We choose to measure informativeness by studying how well the *Sentiment Direction* of RavenPack predicts stock returns in the two minutes following an article.<sup>16</sup> In IA-Table 3 in the Internet Appendix, we show that informativeness is persistent. In particular, *Sentiment Direction* predicts two-minute post-article returns better if it was better in predicting these returns in the prior 3-6 months. Thus, we define *Past Informativeness* as the average *signed* two-minute post-article return for all articles during the previous six months.

We study whether RavenPack subscribers trade more on articles with higher *Past Informativeness* by estimating the following article-level regression specification:

$$\text{Signed Ret}(t-1, t+5)_a = \alpha_t + \alpha_f + \beta_1 * D(HRH)_a * \text{Past Informativeness}_a + \beta_2 * D(HRH)_a + \gamma * \text{Controls} + \varepsilon_a$$

The explanatory variable of interest is the interaction between  $D(HRH)$  and *Past Informativeness*. Its coefficient tests whether RavenPack’s causal effect on 5-second signed returns following the article is stronger if RavenPack has been more informative in the past.

---

<sup>16</sup> It may seem more intuitive to use the number of “mistakes” RavenPack makes in assigning the relevance score. However this information is not available to the market because it uses the New RavenPack dataset, which was only released years later. Therefore, traders could not have conditioned their trading on the number of “mistakes”.

The results are presented in Table 9. We find a significant increase in the causal effect of RavenPack on 5 second signed stock returns if *Past Informativeness* is high. In regressions 4 to 6, we show in a placebo test that this effect does not occur before Ravenpack went “live”. In IA-Table 4 in the Internet Appendix, we report robustness checks using different definitions of *Past Informativeness*. In particular, we use 3 months instead of 6 months and we use *Stock-Level Past Informativeness*, where we only use articles for that specific stock in the past 6 month when measuring past informativeness. Furthermore, we use informativeness based on whether the *direction* of the news and the two-minute stock price reaction agree. This last robustness check shows that our results are not driven by any persistence in volatility. For all three alternative measures of informativeness, we obtain similar results as those presented in Table 9.

In total, these results suggest that algorithmic traders learn dynamically about the precision of RavenPack and base their trades more on RavenPack’s sentiment scores if these scores have been informative in the past. This finding has interesting implications for how the market reacts to inaccurate news analytics. Because inaccurate news analytics are uninformative, subscribers will base their future trades less on RavenPack’s sentiment score. Thus, inaccurate news analytics can potentially reduce the market’s responsiveness to news analytics for several months following the inaccurate news analytic.

## **5. High frequency traders (HFTs) and news analytics**

Our results above suggest that algorithmic traders use news analytics to trade on news releases. So far, we have identified algorithmic traders mainly through the fact that they are able to react to the article within 5 seconds, which is too fast for a human trader. In this section, we focus on a subset of algorithmic traders, namely high frequency traders (HFTs). HFTs are traders that use special technology such as co-location, microwave towers between exchanges, and specialized algorithms

to trade extremely fast. Common trading strategies associated with HFTs include market making and cross-venue arbitrage (Boehmer, Li, and Saar (2017), Zhang (2017)). These activities require submitting limit orders which are at risk of being picked off when corporate news hits the market. Therefore, they are amongst the traders most likely to use news analytics to avoid adverse selection. Whether they use news analytics to avoid adverse selection or to make directional bets or whether they do not use them at all remains an empirical question that we address in this part of the paper.

### *5.1 HFT response to news analytics – total trading*

To analyze HFT trading, we use NASDAQ high frequency trading data first used in Brogaard, Hendershott and Riordan (2014). These data contain all trades on NASDAQ for a sample of 120 stocks for the years 2008 and 2009. For each trade, the data identifies which side of the trade demanded liquidity and whether that trader is an HFT or a non-HFT (which can include algorithmic traders that are not HFTs). HFTs are identified by NASDAQ based on the firms' trading styles and the description on their websites.

We merge the NASDAQ data with our RavenPack data. Unfortunately, due to the fact that the NASDAQ data contain only 2 years and only 120 stocks, our sample is significantly reduced. For example, the merged data include only 112 HRL articles in the period after RavenPack was “live”. Therefore, we are not able to use the regression set-up estimated in the previous sections. Instead, we use a simpler set up in which we compare the reaction to all articles with high relevance (HRL *and* HRH) before and after RavenPack was released on April 1, 2009. By using both HRL and HRH articles, our sample size is significantly increased. We ask how HFTs changed their trading behavior following company specific articles after the release of RavenPack.

We focus on *HFT Trading Fraction*, which we define as shares traded by HFTs (double counting a trade if HFTs are on both sides of the trade) divided by total shares traded on NASDAQ. In line with our previous analysis, we focus on the *HFT Trading Fraction Response*, which we define as  $\frac{\text{HFT Trading Fraction}_{t-1,t+5}}{\text{HFT Trading Fraction}_{t-1,t+5} + \text{HFT Trading Fraction}_{t+6,t+120}}$ . Thus it measures the *HFT Trading Fraction* in the 5 seconds after an article standardized by the *HFT Trading Fraction* in the 120 seconds after the article. Standardizing in this way controls for the overall level of HFT trading in that stock during that time of the day and focuses on how HFT trading might be different immediately after the article.

We regress *HFT Trading Fraction Response* on  $D(\text{RavenPack Release})$ , which is a dummy variable equal to 1 after RavenPack went “live” on April 1, 2009 and 0 before that. We also include our usual article and firm specific controls and two-way cluster standard errors at the firm and date level. We report the results in Table 10. The coefficient of  $D(\text{RavenPack Release})$  is about 1.8 and statistically significant at the 1% level. This suggests that the *HFT Trading Fraction Response* increases by 1.8 percentage points after RavenPack is released to the market, which is 3.6% relative to the unconditional median. This suggests that HFTs make up a larger fraction of trading in the 5 seconds after an article once RavenPack is live. This finding is consistent with HFTs using RavenPack to trade on company specific news releases rather than to reduce their trading to avoid adverse selection.

Next, in regressions 3 to 6, we split the sample into press releases and other articles. Similar to our prior results, we find a much stronger reaction in press releases, where *HFT Trading Fraction Response* increases by 3.8 percentage points (compared to a 1.3 percentage point increase for other articles).

## 5.2 HFT response to news analytics – Liquidity demanding and directional trades

If HFTs conducted directional trades based on RavenPack information, we would expect them to conduct more liquidity-demanding trades after an article. In contrast, if their limit orders were being picked off by other traders using RavenPack, we would expect them to have more liquidity-supplying trades. We address this question in Panel A of Table 11. We compute two dependent variables: *Active HFT Trading Fraction Response* and *Passive HFT Trading Fraction Response*. *Active (Passive) HFT Trading Fraction Response* is constructed in the same way as *HFT Trading Fraction Response* but uses *Active (Passive) HFT Trading Fraction*, which is defined as shares traded using liquidity-demanding (liquidity-supplying) trades by HFTs divided by all shares traded on NASDAQ. Using the same specification as above, we regress these two variables on  $D(\text{RavenPack Release})$ . After the introduction of RavenPack, we observe an increase in *Active HFT Trading Fraction Response* of 2.3% that is significant at the 5% level. In contrast, the increase in *Passive HFT Trading Fraction Response* is only 0.9% and not statistically significant. This finding suggests that HFTs mainly increase their liquidity-demanding trades following news articles after the RavenPack release, which is consistent with them exploiting the information in RavenPack to conduct directional trades rather than to be picked off by other traders using RavenPack.

Next, we study whether HFTs indeed trade in the direction of the RavenPack sentiment. Analogously to the variables above, we compute *With News HFT Trading Fraction Response* and *Against News HFT Trading Fraction Response* by counting only those trades that HFTs do in the direction of the article sentiment as identified by RavenPack (we exclude articles with neutral sentiment). The results are presented in Panel B of Table 11. After the introduction of RavenPack, we observe an increase in *With News HFT Trading Fraction Response* of 1.9% that is significant

at the 10% level. In contrast, the increase in *Against News HFT Trading Fraction Response* is only 0.6% and not statistically significant. This finding is consistent with HFTs trading more in the direction of the RavenPack sentiment.

To conclude, we do not find any evidence that HFTs use RavenPack to avoid being picked off following an article. Rather, at least some NASDAQ HFT traders seem to use RavenPack for speculative bets on company-specific news releases: After RavenPack goes “live” there is more HFT trading in the 5 seconds after an article, and this trading consists mainly of liquidity demanding trades and trades in the direction of the article sentiment. As a caveat, given that the results are based on a purely time-series comparison, they might be driven by other factors that change in the news trading environment at the same time RavenPack gets released.

### *5.3 News analytics and market liquidity*

The results above suggest that RavenPack causes HFTs to conduct more liquidity demanding trades after news releases. In this section, we ask whether such trades lead to a worsening of liquidity after an article is released. Such a drop in liquidity may be caused by users of RavenPack picking off liquidity-supplying orders in the order book. In addition, liquidity providers (that are not using RavenPack) may notice such trading and reduce liquidity provision to avoid being picked off. Both of these effects would lead to a decrease in liquidity after the release of the article. We study whether RavenPack contributes to such a drop in liquidity.

Because we do not have to rely on the HFT data for this question, we can employ our more robust regression set-up using the comparison between HRH and HRL articles. In particular, we regress changes in liquidity on our  $D(HRH)$  dummy and a set of control variables defined as in the previous specifications. We use two standard proxies for liquidity – the effective spread is the most widely-used and reliable measure of the bid-ask spread when using transaction-level data like the

TAQ data used here, and the Amihud (2002) illiquidity measure is widely-considered the most reliable measure of price impact (see Goyenko, Holden and Trzcinka (2009)). The Amihud illiquidity measure is defined as:

$$\text{Amihud Illiquidity}_{ij} = \frac{1}{N_{ij}} \sum_t^{N_{ij}} \frac{|r_{it}|}{\text{dolv}_{it}},$$

where  $r_{it}$  is the return for stock  $i$  during second  $t$ ;  $\text{dolv}_{it}$  is the dollar volume for stock  $i$  during second  $t$ ; and  $N_{ij}$  is the number of seconds in which stock  $i$  traded during interval  $j$ . Effective spread is defined as:

$$\text{Effective Spread}_{ij} = \frac{1}{N_{ij}} \sum_t^{N_{ij}} \text{sign}(\text{buys}_{it} - \text{sells}_{it}) * \frac{\text{price}_{it} - \text{midquote}_{it-1}}{\text{midquote}_{it-1}},$$

where  $\text{buys}_{it}$  ( $\text{sells}_{it}$ ) is the number of stocks bought (sold) for stock  $i$  during second  $t$ ;  $\text{price}_{it}$  is the last execution price for stock  $i$  during second  $t$ ;  $\text{midquote}_{it}$  is the last bid-ask midpoint for stock  $i$  during second  $t$  and  $N_{ij}$  is the number of seconds in which stock  $i$  traded during interval  $j$ .

Because these liquidity measures are positively autocorrelated, we standardize them with respect to their average, computed in the t-5 to t-2 minutes before the article is released. Specifically, we compute:

$$\text{Change in Amihud Illiquidity} = \frac{\text{Amihud Illiquidity}_{t-1,t+5}}{\text{Amihud Illiquidity}_{t-1,t+5} + \text{Amihud Illiquidity}_{t-300,t-120}} \quad ; \text{ and}$$

$$\text{Change in Effective Spread} = \frac{\text{Effective Spread}_{t-1,t+5}}{\text{Effective Spread}_{t-1,t+5} + \text{Effective Spread}_{t-300,t-120}} .$$

The regression set-up is the same as in Tables 4 and 6 and we report the results in Table 12. During the time period where RavenPack was live (Panel A), we observe an increase in both Amihud illiquidity and effective spread if an article is correctly released as relevant (HRH), while there is no significant effect in the placebo sample (Panel B). These results show that illiquidity

increases (liquidity decreases) more after a news release delivered via RavenPack. This confirms that the directional trading triggered by RavenPack reduces liquidity after an article is released.

## 6. Robustness checks

### 6.1 Difference in difference specification

Until now we have mainly focused on the significant effect of RavenPack on the stock market during the period when it was live and showed in placebo tests that there is no effect for the pre-RavenPack period. However, it is possible that the placebo tests might not find significant results because of weak power. Therefore, we estimate a difference-in-difference specification for our entire sample period (February 1, 2004 – September 10, 2012) to study whether the difference between the pre- and post-RavenPack periods is statistically significant.

We report the results in Table 13. In Regressions 1-2 and 3-4, the dependent variables are *Speed of Stock Price Response* and *Speed of Trade Volume Response*, respectively. In Regressions 1 and 4, we estimate the following difference-in-difference specification:

$$DepVar_a = \alpha_t + \alpha_f + \beta_1 * D(HRH)_a * RavenPack Release_t + \beta_2 * D(HRH)_a + \gamma * Controls + \varepsilon_a$$

where  $\alpha_t$  and  $\alpha_f$  are firm and date fixed effects; *RavenPack Release* is a dummy variable equal to 1 after the release of RavenPack on April 1, 2009, and zero otherwise; and  $D(HRH)$  is a dummy equal to 1 for HRH articles and 0 for HRL articles. Endogenous control variables should not be included in difference-in-difference specifications (see for example Gormley and Matsa (2014)). To account for the fact that our control variables may be endogenous, we report specifications without any control variables in regressions 1, 3 and 5. The usual specification with firm controls is reported in Internet Appendix 3. The difference between the two specifications is minimal.



For specifications 1-4, the explanatory variable of interest is the interaction between  $D(HRH)$  and *RavenPack Release*. Its coefficient is significantly positive, implying the effect of an HRH article on the speed of reaction increased significantly after the introduction of RavenPack. This is in line with our previous findings.

In regression 5 and 6, the dependent variable is the return from 1 second before to 5 seconds after the article. Because our baseline analysis used an interaction, the difference-in-difference specification uses a triple interaction:

$$\begin{aligned}
 Ret(t-1, t+5)_a = & \alpha_t + \alpha_f + \beta_1 * Sentiment Direction_a * D(HRH)_a * RavenPack Release_t + \beta_2 * D(HRH)_a * RavenPack Release_t \\
 & + \beta_3 * Sentiment Direction_a * RavenPack Release_t + \beta_4 * D(HRH)_a * Sentiment Direction_t + \beta_5 * D(HRH)_a \\
 & + \beta_6 * Sentiment Direction_t + \gamma * Controls + \varepsilon_a
 \end{aligned}$$

The explanatory variable of interest is this triple interaction between  $D(HRH)$ , *RavenPack Release* and *Sentiment Direction*. Its coefficient is statistically significantly positive, suggesting an increase in the effect of *Sentiment Direction* on returns for articles classified as HRH after RavenPack went live and confirming our previous finding.

## 6.2 RavenPack overfitting its algorithm

One concern is that our results may be related to RavenPack overfitting the algorithm used to identify the relevance score by training it on market reactions to articles. We think this issue is unlikely to be relevant for two reasons. First, RavenPack’s relevance score, which we use for identification, is solely based on where and how often the company name appears in an article. It is not based or trained on market reactions to specific articles. Second, if RavenPack were to train its relevance scoring algorithm on past returns, we would expect a change in the 2-minute market reaction to LRH and HRL articles after RavenPack went live. The intuition is that Old RavenPack could not have been trained on articles released after RavenPack went “live” while New

RavenPack could have been. Therefore, HRL articles that New RavenPack classifies as more relevant, should have a larger 2-minute stock market reaction (leading New RavenPack to upgrade them to highly relevant), while LRH articles should have a smaller 2-minute stock market reaction (leading New RavenPack to downgrade them to low relevance). However, as we show in the Internet Appendix in IA-Table 1, neither HRL nor LRH articles experience a change in 2-minute stock price or trading volume response after RavenPack went live. This finding suggests that RavenPack did not overfit its relevance-score algorithm to market responses.

### *6.3 Alternative placebo tests*

Our base sample for the placebo test is Feb 2004 – Apr 2009. This time period includes the financial crisis and the introduction of Regulation National Market System (Reg NMS) which brought several changes to market structure that increased high frequency trading (Hasbrouck and Saar (2013)) and fragmentation of U.S. markets (O’Hara and Ye (2011)). To address the possible effect of these events on our analysis, we conduct two additional placebo tests: one that covers the period Feb 1, 2004 to Dec 31, 2007, thereby excluding the financial crisis; and one that covers the period July 9, 2007 to April 1, 2009, thereby excluding the time before the introduction of Reg NMS. We report the results in Panel A and B of IA-Table 5 in the Internet Appendix. For both alternative placebo tests and for all specifications, we find no significant effects and the coefficients of interest are small. This suggests that the absence of significant results in our placebo sample is not driven by the inclusion of the financial crisis or the pre-Regulation NMS time period.

### *6.4 “Old RavenPack” definition: RavenPack 1.0 versus RavenPack 2.0*

In our main specification, Old RavenPack includes both RavenPack 1.0 and RavenPack 2.0. A concern is that the difference in reaction before and after the release of New RavenPack might be related to the transition from v.1.0 to v.2.0 in July 2011. Therefore, our next robustness check

focuses only on RavenPack 1.0. We re-estimate the same specifications as before, but include only the period when RavenPack 1.0 was live, i.e. April 1, 2009 to July 6, 2011. We report results in Panel C of IA-Table 5 in the Internet Appendix. All specifications confirm previous results and are similar in terms of economic magnitude.

#### *6.5 Adding fixed effects for number of firms mentioned in the article*

As we have shown in Section 2.3, HRL articles mention somewhat more firms than HRH articles. Thus it is a concern that this difference in the number of firms drives the difference in market reaction to these two article types. In our main tests, we address this issue by including a control variable for the number of firms and running placebo tests. In this robustness check, we control even more carefully for this issue using fixed effects. In particular, we exclude articles that mention more than 3 firms and use fixed effects for whether the article mentions 1, 2, or 3 firms. The results are presented in Panel D of IA-Table 5 (Internet Appendix). All specifications confirm previous results and are similar in terms of economic and statistic magnitude.

#### *6.6 Controlling for the information environment*

Another concern is that HRL and HRH articles have different transmission rates among non-algorithmic traders. This is unlikely given that we do not find any difference between HRL and HRH in our placebo tests before RavenPack went live, but, nonetheless, we conduct an additional robustness check to address this issue. Following Manela (2014), we use media coverage as a measure for the rate of transmission among non-algorithmic traders. Accordingly, we control for different transmission rates by including the following controls: three dummy variables equal to one if there was an article for the same firm in (1) the last hour, (2) earlier in the trading day or (3) in the past trading day. In addition, we add  $\log(1 + \text{number of articles in last week})$  to control for the number of articles in the prior week. The results are presented in Panel E of IA-Table 5 in the

internet appendix. The additional controls do not meaningfully change our results, suggesting that they are not driven by differences in the transmission speeds of non-algorithmic traders.

### *6.7 Alternative length of event window*

In our analyses we compare the stock price reaction in the short run, during which only algorithmic traders can react to an article, to the stock price reaction in the long run during which human traders will have read and traded on the article. In all of our prior analyses, we used 5 seconds as the short-run window and 120 seconds as the long-run window. Here we consider other window lengths.

In IA-Table 6 in the Internet Appendix, we show robustness to choosing longer windows lengths. In particular, we use 10 seconds as the alternative short window and 300 seconds as the alternative long window. The results vary slightly in magnitude, but remain statistically significant.

## **7. Conclusion**

News analytics have revolutionized the way in which financial markets process news. In this paper, we study how news analytics affect market efficiency, for which articles news analytics are most relevant, and which traders use news analytics.

To establish a causal effect of news analytics on the market, we exploit an identification strategy based on inaccuracies in news analytics that were released to the market by RavenPack, a major provider of news analytics for algorithmic traders. Comparing the market reaction to similar news items depending on whether the news has been correctly released to customers or not, we are able to determine the causal effect of news analytics on stock prices, irrespective of the informational content of the news.

We show that the market temporarily responds to errors in news analytics but the reaction is small and reverts after 30 seconds, suggesting that the market is quite resilient. More generally,

news analytics increase market efficiency: The speed of adjustment of both stock prices and trade volume in response to a highly-relevant article is faster if the article was originally released by RavenPack as being relevant than if it was incorrectly released as not relevant. Interestingly, this finding is almost completely concentrated in press releases, likely because they are timelier and news analytics tend to be more accurate for press releases.

This finding suggests that users of news analytics understand for which articles news analytics are most accurate. Furthermore, we show that they understand *when* news analytics are most accurate: RavenPack has a stronger market impact at times after it had been more informative in the past. This finding suggests that algorithmic traders learn about the informativeness of news analytics dynamically. Thus, inaccuracies in news analytics can reduce the market's sensitivity to news analytics for a prolonged period of time.

Finally, we show that high frequency traders trade more following company-specific articles after the release of RavenPack. In particular, they engage in more liquidity demanding trades in the direction of the article sentiment. This finding suggests that HFTs, which are usually seen as trading exclusively on price signals, might also trade on corporate-specific news using news analytics. We show that such trading leads to a decrease in liquidity following an article.

Our findings have normative implications in terms of the recent regulatory debate on high-speed information and the effects of algorithmic and high-frequency trading. We show that news analytics improve market efficiency by speeding up the market reaction to news, in particular for press releases. In contrast, temporary price distortions due to erroneous news analytics are corrected quickly. News analytics provide an additional source of information advantage to high frequency traders, who use it for directional speculative bets. However, these directional bets absorb liquidity and contribute to a decrease in liquidity following the articles.

## References

- Amihud, Y., 2002. Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets* 5, 31–56.
- Baron, Matthew, Jonathan Brogaard, Björn Hagströmer, and Andrei Kirilenko, 2017, The Risk and Return in High Frequency Trading, *Journal of Financial and Quantitative Analysis*, forthcoming.
- Benos, Evangelos and Satchit Sagade, 2016, Price discovery and the cross-section of high-frequency trading, *Journal of Financial Markets* 30, 54-77.
- Biais, Bruno, Thierry Foucault and Sophie Moinas, 2015, Equilibrium Fast Trading, *Journal of Financial Economics*.
- Boehmer, Ekkehart, Kingsley Fong, and Julie Wu, 2015, International evidence on algorithmic trading, *Working Paper*.
- Boehmer, Ekkehart, Dan Li, and Gideon Saar, 2018, The Competitive Landscape of High-Frequency Trading Firms, *Review of Financial Studies*, forthcoming.
- Boudoukh, Jacob, Ronen Feldman, Shimon Kogan and Matthew Richardson, 2016, Information, Trading, and Volatility: Evidence from Firm-Specific News, *Working Paper*.
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High frequency trading and price Discovery, *Review of Financial Studies* 27, 2267-2306. .
- Brogaard, Jonathan, Al Carrion, Thibaut Moyaert, Ryan Riordan, Andriy Shkilko, Konstantin Sokolov, 2015, High-Frequency Trading and Extreme Price Movements, *Working Paper*.
- Chaboud, Alain, Ben Chiquoine, Erik Hjalmarsson, and Clara Vega, 2014, Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, *Journal of Finance* 69, 2045-2084.
- Chan, W. S., 2003, Stock price reaction to news and no-news Drift and reversal after headlines, *Journal of Financial Economics* 70, 223-260.
- Chordia, Tarun, T. Clifton Green, and Badrinath Kottimukkalur, 2017, Rent Seeking by Low Latency Traders: Evidence from Trading on Macroeconomic Announcements, *Working Paper*.
- Clark-Joseph, Adam D., 2013, Exploratory Trading, *Working Paper*.
- Das, Sanjiv R., and Mike Y. Chen, Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, *Management Science* 53, 1375-1388.
- DellaVigna, S., and Pollet, J., 2009, Investor Inattention, Firm Reaction, and Friday Earnings Announcements, *Journal of Finance* 64, 709-749.
- Dougal, Casey, Joseph Engelberg, Diego Garcia and Christopher Parsons, 2012, Journalists and the Stock Market, *Review of Financial Studies* 25, 639-679.
- Dugast, Jerome and Thierry Foucault, 2017, Data Abundance and Asset Price Informativeness, *Working Paper*.
- Dzielinski, Michal, and Henrik Hasseltoft, 2017, News Tone Dispersion and Investor Disagreement, *Working Paper*.

- Engelberg, Joseph and Parsons, Christopher A., 2011, The Causal Impact of Media in Financial Markets, *Journal of Finance* 66, 67-97.
- Fang, Lily H. and Peress, Joel, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance* 64, 2023-2052.
- Ferguson, Nicky J., Dennis Philip, Herbert Y.T. Lam and Jie Michael Guo, 2015, Media Content and Stock Returns: The Predictive Power of Press, *Multinational Finance Journal* 19, 1-31.
- Foucault, Thierry, Johan Hombert and Ioanid Rosu, 2016, News Trading and Speed, *Journal of Finance* 71, 335-382.
- Gai, Jiading, Chen Yao and Mao Ye, 2013, The Externalities of High Frequency Trading, *Working Paper*.
- Garcia, Diego, Sentiment during Recessions, *Journal of Finance* 68, 1267–1300.
- Gerig, Ausin, 2015, High-Frequency Trading Synchronizes Prices in Financial Markets, *Working Paper*.
- Golub, Anton, John Keane and Ser-Huang Poon, 2012, High Frequency Trading and Mini Flash Crashes, *Working Paper*.
- Gormley, Todd A., and David A. Matsa, 2011, Growing Out of Trouble? Corporate Responses to Liability Risk, *Review of Financial Studies* 24, 2781-2821.
- Griffin, John M., Nicholas H. Hirschey, and Patrick J. Kelly, How Important Is the Financial Media in Global Markets?, *Review of Financial Studies* 24, 3941-3992.
- Groß-Klußmann, Axel and Nikolaus Hautsch, 2011, When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions, *Journal of Empirical Finance*, 18, 321-340.
- Hasbrouck, Joel and Gideon Saar, 2013, Low-Latency Trading, *Journal of Financial Markets* 16, 646–679.
- Hendershott, Terrence and Ryan Riordan, 2013, Algorithmic Trading and the Market for Liquidity, *Journal of Financial and Quantitative Analysis* 48, 1001-1024.
- Hendershott, Terrence, Charles M. Jones and Albert J. Menkveld, 2011, Does Algorithmic Trading Improve Liquidity?, *Journal of Finance*, 66, 1-33.
- Heston, Steven L. and Nitish R. Sinha, 2016, News versus Sentiment: Predicting Stock Returns from News Stories, *Working Paper*.
- Hirschey, Nicholas, 2018, Do High-Frequency Traders Anticipate Buying and Selling Pressure?, *Working Paper*.
- Hu, Grace X., Jun Pan and Jiang Wang, 2017, Early Peek Advantage? Efficient price discovery with tiered information disclosure, *Journal of Financial Economics* 126, 399-421.
- Jegadeesh, Narasimhan, and Di Wu, 2013, Word power: A new approach for content analysis, *Journal of Financial Economics* 110, 712-729.
- Jones, Charles, 2013, What do we know about high-frequency trading?, *Working Paper*.

- Jovanovic, Boyan and Albert J. Menkveld, 2011, Middlemen in Limit-Order Markets, *Working Paper*.
- Kyle, Albert, 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.
- Lee, Charles M. and Mark J. Ready, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46, 733-746.
- Loughran, Tim and Bill McDonald, 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66, 35-65.
- Manela, Asaf, 2014, The value of diffusing information, *Journal of Financial Economics* 111, 181-199.
- Martinez, Victor H., and Ioanid Rosu 2013, High Frequency Traders, News and Volatility, *Working Paper*.
- Menkveld, Albert, 2013, High frequency trading and the new market makers, *Journal of Financial Markets*, 16, 712-740.
- O'Hara, Maureen and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics*, 100, 459-474.
- Peress, Joel, 2014, The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes, *Journal of Finance*, 69, 2007-2043.
- Riordan, Ryan, Andreas Storkenmaier, Martin Wagener and S. Sarah Zhang, 2013, Public information arrival: Price discovery and liquidity in electronic limit order markets, *Journal of Banking and Finance*, 37, 1148-1159.
- Riordan, Ryan and Andreas Storkenmaier, 2012, Latency, liquidity and price discovery, *Journal of Financial Markets*, 15, 416-437.
- Sinha, Nitish Rajan, 2016, Underreaction to News in the US Stock Market, *Quarterly Journal of Finance* 6.
- Tetlock, Paul, 2007, Giving content to investor sentiment The role of media in the stock market, *Journal of Finance*, 62, 1139-1168.
- Tetlock, Paul, 2011, All the News That's Fit to Reprint: Do Investors React to Stale Information?, *Review of Financial Studies* 24, 1481-1512.
- Weller, Brian, 2015, Efficient Prices at Any Cost: Does Algorithmic Trading Deter Information Acquisition?, *Working Paper*.
- Zhang, Sarah S., 2017, Need for Speed: Hard information processing in a high-frequency world, *The Journal of Futures Markets* 38, 3-21.



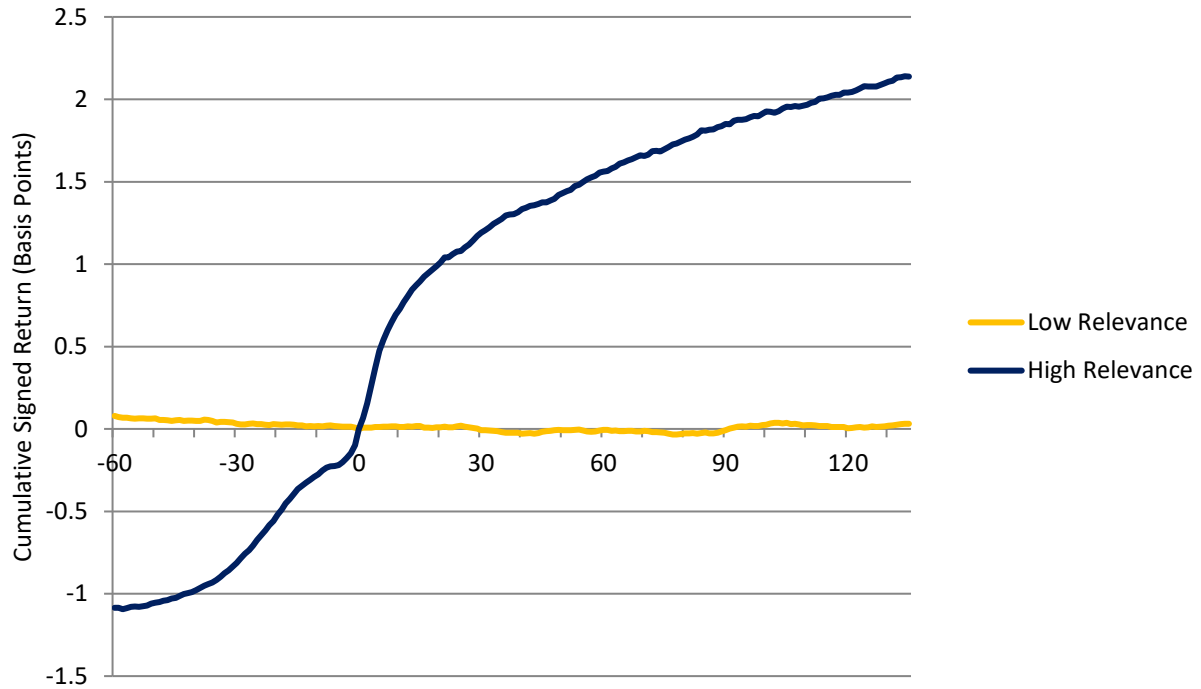
## Figure 1: Market Reaction by Relevance Score

This figure displays cumulative signed return (relative to the time of the article) from 60 seconds before to 120 seconds after the article for news events from April 1, 2009 to September 10, 2012 (the time when RavenPack was live). Signed returns are returns multiplied with the sentiment direction of the article. We exclude articles with neutral sentiment. *Low Relevance* refers to articles with a Relevance Score below 90 in both RavenPack versions, while *High Relevance* refers to articles that have a Relevance Score greater or equal than 90 in both RavenPack versions.

---

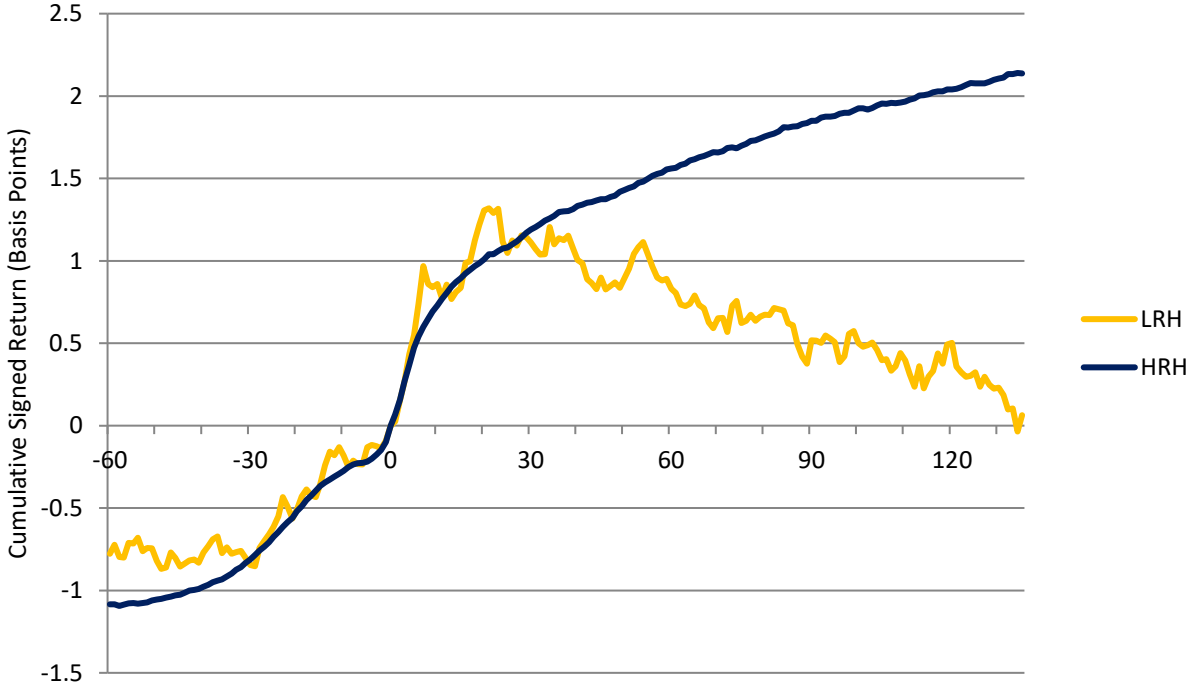
*Cumulative signed return*

---



## Figure 2: Difference in Stock Price Response between HRH and LRH Articles

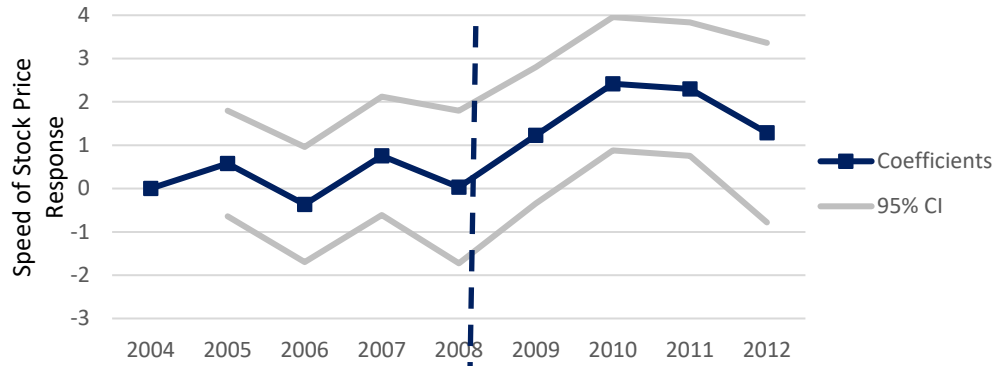
This figure displays cumulative signed return (relative to the time of the article) from 60 seconds before to 120 seconds after the article for news events from April 1, 2009 to September 10, 2012 (the time when RavenPack was live). Returns are multiplied with the sentiment direction of the article. We exclude articles with neutral sentiment. HRH refers to articles that have a relevance scores greater or equal 90 in both RavenPack versions, while LRH refers to articles that had a relevance score greater or equal 90 in the old RavenPack version while having Relevance below 90 in the new RavenPack version.



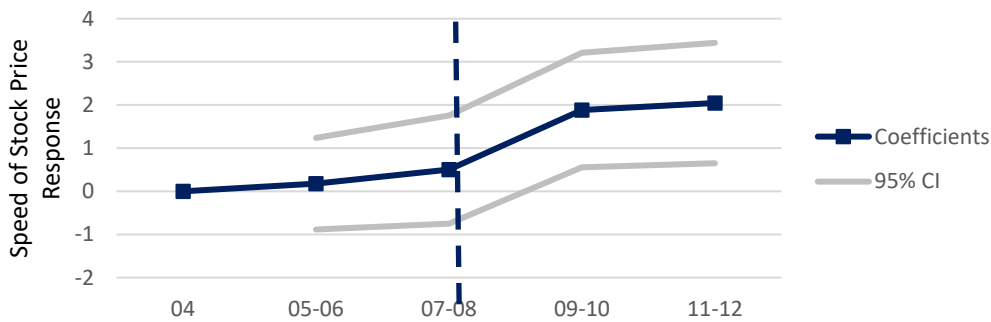
### Figure 3: Difference in Speed of Stock Price Response Over Time

The figure in Panel A reports the point estimates from an OLS regression of *Speed of Stock Price Response* ( $\frac{Abs(Return_{t-1,t+5})}{Abs(Return_{t-1,t+5})+Abs(Return_{t+6,t+120})}$ ) expressed in percent on  $D(HRH)$  interacted with yearly dummy variables from 2004 to 10 September 2012. We assign the first quarter of a year to the prior year, i.e. the 2009 dummy covers a time period from 1 April 2009 to 1 April 2010. Controls and fixed effects are the same as in table 3 regression 3. The vertical line indicates the introduction of RavenPack on 1 April 2009. In Panel B, we report the same regression but interacting the HRH dummy variable with two-year dummy variables (with the first quarter shifted backwards). In Panel C, we report the difference between Speed of Stock Price Response for HRH and HRL articles over different years (with the first quarter shifted backwards).

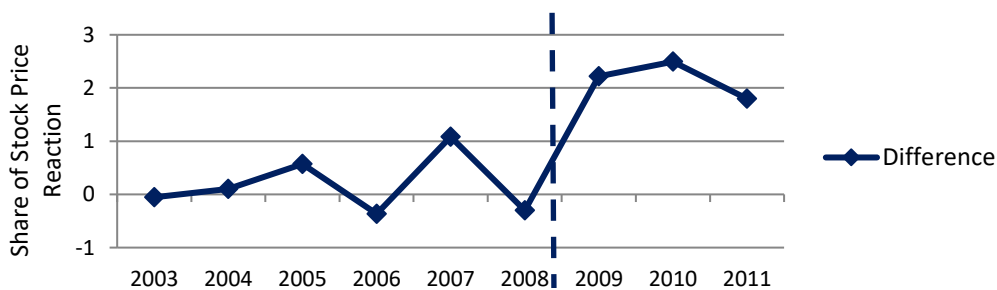
Panel A: Estimate of coefficient on  $D(HRH)$  interacted with yearly dummies



Panel B: Estimate of coefficient on  $D(HRH)$  interacted with two-year dummies



Panel C: Comparing the difference in mean



**Table 1: Overview of Four Article Types**

In Panel A, we present our predictions for the market reaction to different articles. In Panel B, we present the results of article-level regressions that examine the market reaction to different types of articles in the time period where RavenPack was not yet sold to investors. The dependent variables are the absolute returns and turnover in the two minutes after the article. Returns are based on mid-quotes and expressed in basis points. The explanatory variable of interest are D(HRH), D(HRL) and D(LRH), which are dummy variables for these article categories (LRL is the omitted category). At the bottom of the table we display the t-statistic for the difference between HRH and HRL articles and for the difference between RLH and LRL articles. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

*Panel A: Predictions for the market reaction of the different article types*

		New RavenPack	
		High Relevance Article	Low Relevance Article
Old RavenPack	High Relevance Article	HRH: Fast and persistent market reaction	LRH: Fast market reaction that mean-reverts (overreaction).
	Low Relevance Article	HRL: Slow market reaction (underreaction)	LRL: No stock price reaction

*Panel B: Stock price reaction to the different article types BEFORE RavenPack went “live”*

Dependent Variable:	Absolute Return t-1,t+120 (in bp)	Turnover t-1,t+120 (in bp)
D (HRH)	3.235*** (33.02)	0.335*** (30.30)
D (HRL)	3.118*** (11.66)	0.325*** (8.90)
D (LRH)	2.494*** (9.07)	0.311*** (9.27)
Number of Observations	2214677	2214677
R <sup>2</sup>	0.151	0.184
Hour Fixed Effects	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes
Difference between HRH and HRL (t-stat)	0.117 (0.43)	0.011 (0.29)
Difference between LRH and LRL (t-stat)	2.494*** (9.07)	0.311*** (9.27)

**Table 2: Summary Statistics – Relevant articles, Apr 2009 to Sept 2012**

This table displays summary statistics for the 321,912 article-company combinations after RavenPack went “live” (April 1, 2009 to September 10, 2012). These article-company observations are classified as relevant in the new RavenPack (i.e. they are HRH or HRL). *Market capitalization* is the number of shares outstanding multiplied by the prior day closing price. *Average volatility prior month* is the mean of daily squared return in the 20 trading days before the article. *Average turnover prior month* is the mean of daily trading volume divided by shares outstanding in the 20 trading days before the article. *Absolute return t-1, t+5* is the absolute stock return from 1 second before to 5 seconds after the article. *Speed of Stock Price Response* is defined as  $\frac{\text{Abs}(\text{Return } t-1, t+5)}{\text{Abs}(\text{Return } t-1, t+5) + \text{Abs}(\text{Return } t+6, t+120)}$ . *Turnover t-1, t+5* is trading volume divided by shares outstanding from 1 second before to 5 seconds after the article. *Speed of Trade Volume Response* is defined as  $\frac{\text{Turnover } t-1, t+5}{\text{Turnover } t-1, t+120}$ . *Return on trading day* is the stock return over the entire trading day that the article was released. *Absolute return on trading day* is its absolute value. *Time since last company article* is the time since the company was last mentioned in an article. *Number of firms in article* defines the number of companies mentioned in the article. *Composite Sentiment Score* is a sentiment score that is provided by RavenPack and takes a value from 100 (positive) to 0 (negative). *Absolute Composite Sentiment Score* is defined as  $\text{Abs}(\text{Composite Sentiment Score} - 50)$ . *Neutral Composite Sentiment Score* is a dummy variable equal to 1 if the *Composite Sentiment Score* equals 50. *Article Category Identified* is a dummy variable equal to 1 if the article category (e.g. merger and acquisitions) is identified by RavenPack. *Event Sentiment Score* is a sentiment score that is provided by RavenPack and takes a value from 100 (positive) to 0 (negative); this is available only for articles for which the category is identified. *Absolute Event Sentiment Score* is defined as  $\text{Abs}(\text{Event Sentiment Score} - 50)$ . *Neutral Event Sentiment Score* is a dummy variable equal to 1 if the *Event Sentiment Score* equals 50. In Panel A, we report descriptive statistics. In Panel B we report the difference between articles that were consistently released as relevant in both RavenPack versions (HRH) and those that were released as having low relevance (HRL). The difference is defined as the regression coefficient of D(HRH) in a regression of the respective variable on D(HRH) and Relevance, Category, Hour and Date Fixed Effects. D(HRH) is a dummy equal to 1 if the article is HRH. We also report t-statistics for the coefficient two-way clustered by stock and date. \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

*Panel A: Descriptive Statistics*

	Mean	25 <sup>th</sup> Percentile	Median	75 <sup>th</sup> Percentile	Standard Deviation
Market capitalization (\$ million)	13185.0	157.4	1782.9	30027.4	37016.1
Average return prior month (%)	0.12	-0.57	0.10	0.79	0.65
Average volatility prior month (%)	9.69	1.19	4.79	20.4	17.7
Average turnover prior month (%)	1.17	0.27	0.83	2.29	1.23
Absolute Return t-1,t+5 (basis points)	1.95	0	0	4.43	9.46
Absolute Return t-1,t+120 (basis points)	11.4	0	5.00	27.4	21.7
Speed of Stock Price Response (%)	13.2	0	0	50	24.7
Signed Return t-1,t+5 (basis points)	0.60	-1.38	0	1.97	10.2
Signed Return t-1,t+120 (basis points)	1.89	-15.1	0	18.5	25.6
Turnover t-1,t+5 (basis points)	0.041	0	0	0.084	0.14
Turnover t-1,t+120 (basis points)	0.86	0	0.24	1.81	2.22
Speed of Trade Volume Response (%)	5.80	0	0	16.8	13.9
Return on trading day (%)	0.23	-3.29	0.056	3.92	4.02
Absolute return on trading day (%)	2.48	0.22	1.45	5.65	3.18
Time since last company article (hours)	32.2	0.49	6.42	103.1	57.6
Number of companies in article	2.14	1	1	3	4.30
Composite Sentiment Score	50.0	47	50	52	4.19
Absolute Composite Sentiment Score	2.07	0	2	5	3.65
Neutral Composite Sentiment Score	0.47	0	0	1	0.50
Article category identified	0.35	0	0	1	0.48
Event Sentiment Score	51.8	37	50	67	12.9
Absolute Event Sentiment Score	3.83	0	0	13	6.71
Neutral Event Sentiment Score	0.69	0	1	1	0.46
Past Informativeness 6 month 12FF	1.67	0.58	1.39	3.46	1.15
Past Informativeness 3 month 12FF	1.58	0.41	1.22	3.29	1.23
Past Informativeness 6 month 30FF	1.69	0.39	1.34	3.66	1.46
Number of Observations	321,912				

*Panel B: Comparison between Accurately Classified as Relevant (HRH) vs. Misclassified (HRL)*

	Standard Deviation	Difference between HRH and HRL after fixed effects	T- Statistic	Difference in terms of Standard Deviations
Market capitalization (\$ million)	37016.1	-921.79	-0.25	-0.0249
Average return prior month (%)	0.65	-0.0319***	-3.01	-0.04908
Average volatility prior month (%)	17.7	-1.525*	-1.70	-0.08621
Average turnover prior month (%)	1.23	-0.0773	-0.93	-0.06285
Average illiquidity prior month (percentile)	26.4	-2.0886	-0.84	-0.07907
Absolute Return t-1,t+120 (basis points)	21.7	-0.3433	-0.77	-0.01582
Turnover t-1,t+120 (basis points)	2.22	0.0731	1.17	0.032928
Return on trading day (%)	4.02	-0.0953	-1.54	-0.02371
Absolute return on trading day (%)	3.18	-0.1242	-0.92	-0.03906
Time since last company article (hours)	57.6	3.32	1.24	0.057639
Number of companies in article	4.30	-0.95***	-3.13	-0.22093
Composite Sentiment Score	4.19	-0.0679	-0.63	-0.01621
Absolute Composite Sentiment Score	3.65	0.0105	0.06	0.002877
Neutral Composite Sentiment Score	0.50	-0.0353	-0.67	-0.0706
Event Sentiment Score	12.9	-0.7440	-1.10	-0.05767
Absolute Event Sentiment Score	6.71	-0.0986†	-1.75	-0.01469
Neutral Event Sentiment Score	0.46	-0.0020	-0.98	-0.00435

**Table 3: Overreaction to News Analytics (LRH articles)**

This table contains the results of article-level regressions that examine how well the sentiment direction of LRH articles predicts stock returns before and after the release of RavenPack. In regressions 1 to 3, the dependent variable is the return from 1 second before to 5 seconds after the article (measured in basis points). In regressions 4 to 6, we study the return from 6 to 120 seconds after the article to determine a potential reversal of the short run reaction. Returns are based on mid-quotes. The explanatory variable of interest is an interaction between *RavenPack Release* and *Sentiment Direction*. *RavenPack Release* is a dummy variable equal to 1 during the time in which RavenPack was “live” (April 1, 2009 – September 10, 2012) and equal to 0 before RavenPack was “live” (February 1, 2004 – March 31, 2009). *Sentiment Direction* is a variable indicating the sentiment of the article derived from RavenPack sentiment indices; it takes the value +1 for positive sentiment, 0 for neutral sentiment and -1 for negative sentiment. In all regressions we include the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 3, 5 and 6 we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In regressions 3 and 6, we include absolute return, turnover, and volatility each for industry and market from t-1 to t+5 seconds around the article. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable:	Return t-1, t+5 (in bp)			Return t+6, t+120 (in bp)		
	(1)	(2)	(3)	(4)	(5)	(6)
RavenPack Release * Sentiment Direction	0.465** (2.29)	0.533** (2.46)	0.563*** (2.60)	-0.602 (-1.00)	-0.666 (-1.09)	-0.660 (-1.10)
RavenPack Release	-0.252 (-1.48)	-0.258 (-1.43)	-0.309 (-1.60)	-1.739*** (-3.88)	-1.659*** (-3.67)	-1.643*** (-3.36)
Sentiment Direction	0.258** (2.31)	0.098 (0.82)	0.099 (0.83)	1.532*** (4.96)	1.517*** (4.41)	1.538*** (4.49)
Article category identified		-1.177*** (-3.14)	-1.217*** (-3.05)		0.043 (0.04)	-0.316 (-0.26)
Time since last article		0.112** (2.41)	0.107** (2.31)		0.373*** (2.97)	0.357*** (2.83)
Number of firms in article		-0.179** (-2.15)	-0.171** (-2.08)		-0.397* (-1.68)	-0.415* (-1.78)
Number of Observations	20588	20586	20586	20588	20586	20586
R <sup>2</sup>	0.003	0.008	0.014	0.007	0.013	0.018
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

### Table 4: Speed of Stock Price Response to News Articles

This table contains the results of article-level regressions that examine the effect of an article covered in RavenPack on stock price, measured by absolute returns. The dependent variable is *Speed of Stock Price Response* (in percent) defined as  $\frac{\text{Abs}(\text{Return } t-1, t+5)}{\text{Abs}(\text{Return } t-1, t+5) + \text{Abs}(\text{Return } t+6, t+120)}$  and measured in seconds around an article. Returns are based on mid-quotes. The explanatory variable of interest is D(HRH), a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). In regressions 1 to 3, we estimate the various specification during the time in which RavenPack was “live” (April 1, 2009 – September 10, 2012). In regressions 4 to 6, we run a placebo test for the time period where RavenPack was not yet sold to investors (February 1, 2004 – March 31, 2009). In all regressions we include firm and date fixed effects and the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 3, 5 and 6 we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In regressions 3 and 6, we include additional controls: the absolute return, turnover, and volatility each for industry and market and for the two horizons from  $t-1$  to  $t+5$  and  $t-1$  to  $t+120$  seconds around the article. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable:	Speed of Stock Price Response (in %)					
	Main Test - RavenPack is “live”			Placebo Test - Before RavenPack is “live”		
	(1)	(2)	(3)	(4)	(5)	(6)
D(HRH)	1.469*** (3.39)	1.333*** (3.06)	1.321*** (3.17)	-0.048 (-0.17)	-0.058 (-0.20)	-0.012 (-0.04)
Absolute Composite Sentiment Score		-0.004 (-0.23)	-0.002 (-0.12)		-0.032** (-2.35)	-0.032** (-2.38)
Neutral Composite Sentiment Score		-0.107 (-0.91)	-0.115 (-1.02)		-0.338*** (-3.44)	-0.357*** (-3.67)
Article category identified		0.522 (0.10)	1.394 (0.28)		-3.575 (-0.57)	-4.142 (-0.71)
Absolute Event Sentiment Score		0.098*** (5.20)	0.089*** (4.86)		0.021* (1.73)	0.022* (1.88)
Neutral Event Sentiment Score		-0.818 (-1.35)	-0.958* (-1.65)		-1.202*** (-2.99)	-1.150*** (-2.89)
Time since last article		0.099*** (2.78)	0.086** (2.46)		0.069*** (2.67)	0.062** (2.40)
Number of firms in article		-0.061 (-0.67)	-0.058 (-0.67)		-0.149*** (-2.59)	-0.170*** (-2.99)
Number of Observations	248849	248849	248849	400158	400158	400158
R <sup>2</sup>	0.034	0.038	0.084	0.032	0.033	0.048
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes



**Table 5: Directional Stock Price Response to Article Sentiment**

This table contains the results of article-level regressions that examine how well the sentiment direction of an article predicts the 5-second return reaction to an article depending on whether the article is covered in RavenPack. The dependent variable is the return from 1 second before to 5 seconds after the article (measured in basis points). Returns are based on mid-quotes. The explanatory variable of interest is an interaction between D(HRH) and *Sentiment Direction*. D(HRH) is a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). *Sentiment Direction* is a variable indicating the sentiment of the article derived from RavenPack sentiment indices; it takes the value +1 for positive sentiment, 0 for neutral sentiment and -1 for negative sentiment. In regressions 1 to 3, we estimate the various specification during the time in which RavenPack was “live” (April 1, 2009 – September 10, 2012). In regressions 4 to 6, we run a placebo test for the time period where RavenPack was not yet sold to investors (February 1, 2004 – March 31, 2009). In all regressions we include firm and date fixed effects and the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 3, 5 and 6 we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In regressions 3 and 6, we include absolute return, turnover, and volatility each for industry and market from t-1 to t+5 seconds around the article. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable:	Return t-1, t+5 (in bp)					
	Main Test - RavenPack is “live”			Placebo Test - Before RavenPack is “live”		
	(1)	(2)	(3)	(4)	(5)	(6)
D(HRH) * Sentiment Direction	0.407*** (3.06)	0.452*** (3.40)	0.452*** (3.41)	0.081 (0.79)	0.116 (1.10)	0.114 (1.09)
D(HRH)	0.187* (1.86)	0.125 (1.19)	0.125 (1.19)	0.137 (1.29)	0.103 (0.95)	0.102 (0.94)
Sentiment Direction	0.118 (0.91)	-0.010 (-0.08)	-0.009 (-0.07)	0.421*** (4.25)	0.184* (1.82)	0.187* (1.85)
Article category identified		1.152** (2.06)	0.933* (1.96)		0.304 (0.81)	0.370 (0.97)
Time since last article		0.043*** (3.44)	0.043*** (3.47)		0.237*** (13.19)	0.239*** (13.26)
Number of firms in article		-0.150*** (-4.96)	-0.145*** (-4.80)		-0.221*** (-9.24)	-0.220*** (-9.18)
Number of Observations	321762	321762	321762	481852	481852	481852
R <sup>2</sup>	0.060	0.063	0.067	0.056	0.062	0.063
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

**Table 6: Speed of Trade Volume Response to News Articles**

This table contains the results of article-level regressions that examine the effect of an article covered in RavenPack on the market for a stock, measured by turnover. In Panel A, the dependent variable is *Speed of Trade Volume Response* (in percent), defined as the turnover from 1 second before the article to 5 second after the article divided by the turnover from 1 second before the article to 120 seconds after the article. The explanatory variable of interest is D(HRH), a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). In regressions 1 to 3, we estimate the various specification during the time in which RavenPack was “live” (April 1, 2009 – September 10, 2012). In regressions 4 to 6, we run a placebo test for the time period where RavenPack was not yet sold to investors (February 1, 2004 – March 31, 2009). In all regressions we include firm and date fixed effects and the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 3, 5 and 6 we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In regressions 3 and 6, we include additional controls: the absolute return, turnover, and volatility each for industry and market and for the two horizons from t-1 to t+5 and t-1 to t+120 seconds around the article. In panel B, we split the trading volume by who initiated the trade using the algorithm of Lee and Ready (1991). In Panel B, the dependent variable is *Speed of Directional Trade Volume Response*, defined as the turnover initiated in direction of the article from 1 second before the article to 5 second after the article divided by the turnover from 1 second before the article to 120 seconds after the article. The control variables in Panel B are the same as in Panel A, but not reported for brevity. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

*Panel A: Trade Volume*

Dependent Variable:	Speed of Trade Volume Response (in %)					
	Main Test - RavenPack is “live”			Placebo Test - Before RavenPack is “live”		
	(1)	(2)	(3)	(4)	(5)	(6)
D(HRH)	0.656*** (3.06)	0.516** (2.36)	0.533** (2.49)	0.033 (0.24)	-0.002 (-0.02)	0.022 (0.16)
Absolute Composite Sentiment Score		-0.008 (-0.97)	-0.006 (-0.77)		-0.018*** (-2.66)	-0.017** (-2.54)
Neutral Composite Sentiment Score		-0.171** (-2.57)	-0.164** (-2.52)		-0.085 (-1.62)	-0.089* (-1.69)
Article category identified		-4.034*** (-3.02)	-3.973*** (-3.47)		-0.716 (-0.47)	-0.691 (-0.45)
Absolute Event Sentiment Score		0.059*** (5.82)	0.056*** (5.50)		0.023*** (3.55)	0.024*** (3.68)
Neutral Event Sentiment Score		-0.973*** (-2.82)	-1.018*** (-3.02)		-0.394* (-1.86)	-0.368* (-1.74)
Time since last article		0.109*** (5.58)	0.101*** (5.21)		0.091*** (5.98)	0.087*** (5.82)
Number of firms in article		-0.107** (-2.55)	-0.112*** (-2.70)		-0.168*** (-6.04)	-0.173*** (-6.25)
Number of Observations	272019	272019	272019	418029	418029	418029
R <sup>2</sup>	0.028	0.031	0.058	0.025	0.026	0.037
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

*Panel B: Trade volume in direction of article*

Dependent Variable:	Speed of Directional Trade Volume Response (in %)					
	Main Test - RavenPack is “live”			Placebo Test - Before RavenPack is “live”		
	(1)	(2)	(3)	(4)	(5)	(6)
D(HRH)	0.472*** (2.67)	0.376** (2.12)	0.385** (2.20)	0.080 (0.60)	0.085 (0.64)	0.082 (0.61)
Number of Observations	168278	168278	168278	272477	272477	272477
R <sup>2</sup>	0.046	0.051	0.063	0.039	0.040	0.045
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

**Table 7: Effect of News Analytics: Press Releases vs. Non-Press Releases**

This table contains the results of article-level regressions examining if our results are stronger for press releases. The dependent variables are *Speed of Stock Price Response* (in percent), *Speed of Trade Volume Response* (in percent), and the signed return from 1 second before to 5 seconds after the article (measured in basis points). The explanatory variable of interest is the interaction between *D(HRH)* and *D(Press Release)*. *D(Press Release)* is a dummy variable taking the value of 1 for press releases. In all regressions we include firm and date fixed effects. In regressions 2, 4, and 6, we add the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In these regressions, we also add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released as well as additional controls: the absolute return, turnover, and volatility each for industry and market from t-1 to t+5 seconds around the article. In regression 2 and 4, we also include those values for t-1 to t+120 seconds around the article. All variables are defined in Appendix 1. In Panel A, we run the various specifications during the time in which RavenPack was “live” (April 1, 2009 – September 10, 2012). In Panel B, we run a placebo test for the time period where RavenPack was not yet sold to investors (February 1, 2004 – March 31, 2009). In Panel B, we use the same control variables as in Panel A. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

*Panel A: Main Test*

Dependent Variable:	Speed of Stock Price Response (in %)		Speed of Trade Volume Response (in %)		Signed Return t-1, t+5 (in bp)	
	(1)	(2)	(3)	(4)	(5)	(6)
D(Press Release) * D(HRH)	1.686** (2.05)	1.813** (2.25)	1.044*** (2.85)	1.195*** (3.32)	1.038*** (3.15)	0.992*** (2.97)
D(HRH)	1.003* (1.96)	0.791 (1.63)	0.348 (1.50)	0.181 (0.78)	0.184 (1.07)	0.053 (0.30)
D(Press Release)	-0.885 (-1.09)	-1.487* (-1.87)	-0.760** (-2.11)	-1.019*** (-2.90)	-0.160 (-0.49)	-0.484 (-1.41)
Sentiment Direction					0.276*** (8.93)	0.231*** (7.10)
Absolute Composite Sentiment Score		-0.001 (-0.06)		-0.006 (-0.72)		
Neutral Composite Sentiment Score		-0.116 (-1.03)		-0.165** (-2.54)		
Absolute Event Sentiment Score		0.090*** (4.90)		0.056*** (5.53)		
Neutral Event Sentiment Score		-1.108* (-1.90)		-1.095*** (-3.21)		
Article category identified		1.167 (0.23)		-4.090*** (-3.55)		0.932 (1.61)
Time since last article		0.079** (2.25)		0.097*** (4.95)		0.041** (2.51)
Number of firms in article		-0.061 (-0.70)		-0.113*** (-2.73)		-0.222*** (-5.29)
Number of Observations	248888	248849	272058	272019	198263	198225
R <sup>2</sup>	0.034	0.084	0.028	0.058	0.086	0.095
Relevance, Category and Hour Fixed Effects	No	Yes	No	Yes	No	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	Yes	No	Yes	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

*Panel B: Placebo Test*

Dependent Variable:	Speed of Stock Price Response (in %)		Speed of Trade Volume Response (in %)		Signed Return t-1, t+5 (in bp)	
	(1)	(2)	(3)	(4)	(5)	(6)
D(Press Release) * D(HRH)	0.265 (0.48)	0.344 (0.63)	0.038 (0.13)	0.059 (0.20)	0.396 (1.07)	0.363 (0.98)
D(HRH)	-0.140 (-0.43)	-0.095 (-0.30)	0.034 (0.22)	0.008 (0.06)	0.177 (1.06)	0.107 (0.65)
D(Press Release)	0.112 (0.20)	0.076 (0.14)	0.222 (0.76)	0.089 (0.30)	1.283*** (3.59)	0.938*** (2.68)
Number of Observations	400304	400158	418187	418029	316945	316783
R <sup>2</sup>	0.031	0.048	0.025	0.037	0.076	0.081
Relevance, Category and Hour Fixed Effects	No	Yes	No	Yes	No	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	Yes	No	Yes	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

**Table 8: Comparison Press Releases vs. Non-Press Releases**

This table compares the characteristics of press releases and non-press releases. We report means for both groups and their difference as well as significance levels based on two-way clustered standard errors by stock and date.  $D(HRH)$  is a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). *Absolute return  $t-1, t+120$*  and *Turnover  $t-1, t+120$*  are absolute return and turnover from 1 second before to 120 seconds after the article. *Absolute return on trading day* is the absolute stock return over the entire trading day on which the article was released. *Sentiment Direction* is a variable indicating the sentiment of the article derived from RavenPack sentiment indices; it takes the value +1 for positive sentiment, 0 for neutral sentiment and -1 for negative sentiment.  $D(\text{First company article of the day})$  is a dummy variable equal to one if it is the first article for that company on that day.  $D(\text{New story})$  is a dummy variable equal to 1 if it is a new news story (RavenPack ENS=100) and 0 if it is a reprint. This variable is missing if the category of the article cannot be identified.  $D(\text{Sentiment predicts return})$  is a dummy variable equal to 1 if the sign of the return from 1 second before to 120 seconds after the article matches the sign of the article sentiment (it is set to missing for articles with neutral sentiment). We use the full sample from February 1, 2004 to September 10, 2012. \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

	Press Releases (Mean)	Non-Press Releases (Mean)	Difference
$D(HRH)$	0.97	0.97	-0.00
Absolute Return $t-1, t+120$ (basis points)	14.6	12.5	2.17***
Turnover $t-1, t+120$ (basis points)	0.92	1.05	-0.12***
Absolute return on trading day (%)	2.09	3.18	-1.01***
Sentiment Direction	0.41	0.07	0.34***
$D(\text{First company article of the day})$	0.85	0.77	0.08***
$D(\text{New story})$	0.95	0.78	0.17***
$D(\text{Sentiment predicts return})$	0.56	0.53	0.03***

**Table 9: Directional Stock Price Response conditional on Past Informativeness of RavenPack**

This table contains the results of article-level regressions that examine how well the past performance of RavenPack affects the stock price impact of RavenPack. The dependent variable is the signed return from 1 second before to 5 seconds after the article (measured in basis points), i.e. returns multiplied with the sentiment direction of the article. Returns are based on mid-quotes. The explanatory variable of interest is the interaction between *D(HRH)* and *Past Informativeness*. *Past Informativeness* is the average signed return (in basis points) from t-1 to t+120 seconds around articles over the previous 6 month. *D(HRH)* is a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). In regressions 1 to 3, we estimate the various specification during the time in which RavenPack was “live” (April 1, 2009 – September 10, 2012). In regressions 4 to 6, we run a placebo test for the time period where RavenPack was not yet sold to investors (February 1, 2004 – March 31, 2009). In IA-Table 4 in the Internet Appendix, we report a robustness check using *Stock-Level Past Informativeness* measured over the previous three months (instead of a six), using *Stock-Level Past Informativeness* and using *Direction-Based Past Informativeness*. In all regressions we include firm and date fixed effects and the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 3, 5 and 6, we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In regressions 3 and 6, we add additional controls: the absolute return, turnover, and volatility each for industry and market and for the two horizons from t-1 to t+5 seconds around the article. Control variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable:	Signed Return t-1, t+5 (in bp)					
	Main Test - RavenPack is “live”			Placebo Test - Before RavenPack is “live”		
	(1)	(2)	(3)	(4)	(5)	(6)
Past Informativeness * D(HRH)	0.342*** (3.03)	0.276** (2.45)	0.295** (2.58)	-0.191 (-0.63)	-0.189 (-0.63)	-0.181 (-0.60)
D(HRH)	-0.253 (-1.18)	-0.173 (-0.80)	-0.208 (-0.95)	0.551 (0.80)	0.635 (0.92)	0.608 (0.89)
Article category identified		1.040* (1.93)	0.732 (1.31)		-0.037 (-0.09)	0.034 (0.08)
Time since last article		0.056*** (3.43)	0.055*** (3.41)		0.251*** (12.40)	0.252*** (12.46)
Number of firms in article		-0.207*** (-4.98)	-0.197*** (-4.75)		-0.170*** (-5.24)	-0.167*** (-5.13)
Number of Observations	198225	198225	198225	310657	310657	310657
R <sup>2</sup>	0.085	0.090	0.094	0.073	0.079	0.080
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

**Table 10: HFT Trading Response to News Articles**

This table contains the results of article-level regressions that examine the effect of an article covered in RavenPack on the fraction of HFT trading on NASDAQ. The dependent variable is *HFT Trading Fraction Response* (in percent) defined as  $\frac{\text{HFT Trading Fraction}_{t-1,t+5}}{\text{HFT Trading Fraction}_{t-1,t+5} + \text{HFT Trading Fraction}_{t+6,t+120}}$  and measured in seconds around an article. HFT Trading Fraction is defined as  $\frac{\text{Shares Traded by HFT on NASDAQ}_{t,s}}{\text{Total Shares Traded on NASDAQ}_{t,s}}$ . The explanatory variable of interest is RavenPack Release, a dummy variable equal to 1 during the time in which RavenPack was “live” (April 1, 2009 – December 31, 2009) and equal to 0 during the time where RavenPack was not yet sold to investors (January 1, 2008 – March 31, 2009). In all regressions we include the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 4, and 6 we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable:	HFT Trading Fraction Response (in %)					
	All Articles		Press Releases		Non-Press Releases	
	(1)	(2)	(3)	(4)	(5)	(6)
RavenPack Release	1.735*** (3.10)	1.784*** (3.20)	3.538** (2.54)	3.880** (2.50)	1.258* (1.91)	1.343* (1.92)
Absolute Composite Sentiment Score		-0.031 (-0.39)		-0.248 (-1.40)		0.007 (0.09)
Neutral Composite Sentiment Score		-0.070 (-0.12)		-2.758* (-1.88)		0.437 (0.65)
Article category identified		-8.47e+05 (-0.00)		-8.84e+04 (-0.00)		0.000 (0.00)
Absolute Event Sentiment Score		0.148** (2.06)		0.397** (2.40)		0.099 (1.12)
Neutral Event Sentiment Score		1.064 (0.35)		1.755 (0.27)		-0.431 (-0.10)
Time since last article		-0.126 (-0.42)		0.687 (0.97)		-0.422 (-1.26)
Number of firms in article		-0.681* (-1.78)		1.323 (1.13)		-0.706* (-1.82)
Number of Observations	5163	5162	1014	1012	4149	4147
R <sup>2</sup>	0.010	0.019	0.016	0.044	0.009	0.022
Relevance, Category and Hour Fixed Effects	No	Yes	No	Yes	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

**Table 11: HFT Trading Response split by active/passive and direction**

This table contains the results of article-level regressions that examine the effect of an article covered in RavenPack on HFT trading on NASDAQ. In Panel A, the dependent variable *Active HFT Trading Fraction Response* (in percent) is defined as  $\frac{\text{Active HFT Trading Fraction}_{t-1,t+5} + \text{Active HFT Trading Fraction}_{t+6,t+120}}{\text{Total Shares Traded on NASDAQ}_{t,s}}$  measured in seconds around an article. Active HFT Trading Fraction is defined as Shares Traded by HFT using liquidity demanding trades on NASDAQ  $t,s$ . The dependent variable *Passive HFT Trading Fraction Response* is defined analogously using the stocks traded in liquidity supplying trades. The explanatory variable of interest is RavenPack Release, a dummy variable equal to 1 during the time in which RavenPack was “live” (April 1, 2009 – December 31, 2009) and equal to 0 during the time where RavenPack was not yet sold to investors (January 1, 2008 – March 31, 2009). In all regressions we include the following firm specific control variables: *Company size, Return prior month, Volatility prior month, Turnover prior month, Illiquidity prior month*. In regressions 2 and 4 we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In Panel B, we use the same set-up but the dependent variables are *With News HFT Trading Fraction Response* and *Against News HFT Trading Fraction Response*, which are defined as above but using the fraction of trading that HFTs do in direction of the news sentiment or against the news sentiment (articles with neutral sentiment are excluded). All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

*Panel A: Split by liquidity demanding (active) and liquidity supplying (passive) trades*

Dependent Variable:	Active HFT Trading Fraction Response (in %)		Passive HFT Trading Fraction Response (in %)	
	(1)	(2)	(3)	(4)
RavenPack Release	2.394** (2.54)	2.345** (2.60)	1.054 (1.21)	0.935 (1.06)
Absolute Composite Sentiment Score		0.136 (1.53)		-0.070 (-0.67)
Neutral Composite Sentiment Score		1.090 (1.34)		0.245 (0.28)
Article category identified		-4.15e+06 (-0.00)		25834.284 (0.00)
Absolute Event Sentiment Score		0.297*** (2.78)		0.094 (0.94)
Neutral Event Sentiment Score		4.500 (0.96)		0.694 (0.18)
Time since last article		0.085 (0.22)		-0.597* (-1.93)
Number of firms in article		-0.361 (-0.66)		-0.721 (-1.35)
Number of Observations	5116	5115	5088	5087
R <sup>2</sup>	0.009	0.023	0.015	0.025
Relevance, Category and Hour Fixed Effects	No	Yes	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes

*Panel B: Split by whether HFTs trade with the news sentiment or against the news sentiment*

Dependent Variable:	With News HFT Trading Fraction Response (in %)		Against News HFT Trading Fraction Response (in %)	
	(1)	(2)	(3)	(4)
RavenPack Release	2.084* (1.96)	1.881* (1.84)	0.496 (0.44)	0.601 (0.56)
Number of Observations	5116	5115	5088	5087
R <sup>2</sup>	0.009	0.023	0.015	0.025
Relevance, Category and Hour Fixed Effects	No	Yes	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes

**Table 12: Liquidity Response to News Articles**

This table contains the results of article-level regressions that examine the effect of an article being covered in RavenPack on illiquidity. In regressions 1 to 3, the dependent variable is *Change in Amihud Illiquidity* (in percent) defined as  $\frac{\text{Amihud Illiquidity}_{t-1,t+5}}{\text{Amihud Illiquidity}_{t-1,t+5} + \text{Amihud Illiquidity}_{t-300,t-120}}$  in seconds around the article. In regressions 4 to 6, the dependent variable is *Change in Effective Spread* (in percent) defined as  $\frac{\text{Effective Spread}_{t-1,t+5}}{\text{Effective Spread}_{t-1,t+5} + \text{Effective Spread}_{t-300,t-120}}$  in seconds around the article. The explanatory variable of interest is D(HRH), a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). In Panel A, we estimate our main specification during the time in which RavenPack was live (April 1, 2009 – September 10, 2012). In Panel B, we run a placebo test in the time period where RavenPack was not yet being sold to investors (February 1, 2004 – March 31, 2009). In all regressions we include firm and date fixed effects and the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In regressions 2, 3, 5 and 6, we add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released. In regressions 3 and 6, we include additional controls: the absolute return, turnover, and volatility each for industry and market from t-1 to t+5 seconds around the article. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

*Panel A: Main Specification – RavenPack is “live”*

Dependent Variable:	Change in Effective Spread (in %)			Change in Amihud Illiquidity (in %)		
	(1)	(2)	(3)	(4)	(5)	(6)
D(HRH)	0.883*	1.268***	1.281***	1.633**	1.408**	1.357**
	(1.96)	(2.84)	(2.85)	(2.35)	(2.03)	(1.99)
Absolute Composite Sentiment Score		0.042**	0.046**		-0.025	-0.011
		(2.16)	(2.35)		(-0.88)	(-0.38)
Neutral Composite Sentiment Score		-0.398***	-0.403***		0.052	0.080
		(-2.76)	(-2.80)		(0.22)	(0.36)
Article category identified		-12.021	-12.719		-4.158	-3.938
		(-1.47)	(-1.50)		(-0.50)	(-0.45)
Absolute Event Sentiment Score		0.073***	0.072***		0.067**	0.065**
		(3.36)	(3.32)		(2.31)	(2.27)
Neutral Event Sentiment Score		-1.658**	-1.768**		-2.197**	-2.341**
		(-2.21)	(-2.36)		(-2.05)	(-2.20)
Time since last article		-0.436***	-0.459***		0.350***	0.253***
		(-10.70)	(-11.35)		(5.15)	(3.75)
Number of firms in article		1.427***	1.433***		-0.637***	-0.542***
		(14.32)	(14.74)		(-4.65)	(-4.02)
Number of Observations	252077	252077	252077	115630	115630	115630
R <sup>2</sup>	0.162	0.165	0.170	0.091	0.092	0.122
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes

*Panel B: Placebo Test - Before RavenPack is “live”*

Dependent Variable:	Change in Effective Spread (in %)			Change in Amihud Illiquidity (in %)		
	(1)	(2)	(3)	(4)	(5)	(6)
-0.032	0.480	0.445	0.861	0.687	0.631	-0.032
(-0.11)	(1.60)	(1.49)	(1.43)	(1.14)	(1.05)	(-0.11)
Number of Observations	411734	411734	411734	178102	178102	178102
R <sup>2</sup>	0.156	0.160	0.167	0.067	0.069	0.077
Relevance, Category and Hour Fixed Effects	No	Yes	Yes	No	Yes	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	No	Yes	No	No	Yes
Firm specific control variables	Yes	Yes	Yes	Yes	Yes	Yes



**Table 13: Difference in Difference Analysis**

This table contains the results of article-level regressions implementing a difference in difference set-up for our whole sample from February 1, 2004 to September 10, 2012 as a robustness check to tables 4 to 6. In regressions 1 and 2, the dependent variable is *Speed of Stock Price Response* (in percent), defined as  $\frac{\text{Abs}(\text{Return } t-1, t+5)}{\text{Abs}(\text{Return } t-1, t+5) + \text{Abs}(\text{Return } t+6, t+120)}$  and measured in seconds around an article. In regressions 3 and 4, the dependent variable is *Speed of Trade Volume Response* (in percent), defined as the turnover from 1 second before the article to 5 second after the article divided by the turnover from 1 second before the article to 120 seconds after the article. In regressions 1 to 4, the explanatory variable of interest is the interaction between *D(HRH)* and *RavenPack Release*. *D(HRH)* is a dummy variable equal to 1 if an article was consistently released as highly relevant in both RavenPack versions and 0 if it was originally released as having low relevance (HRL). *RavenPack Release* is a dummy variable taking the value of 1 for articles after RavenPack went “live” on April 1, 2009, and zero otherwise. In regressions 5 and 6, the dependent variable is the return (in percent) measured from 1 second before to 5 seconds after the article. The explanatory variable of interest is a triple interaction between *HRH*, *RavenPack Release* and *Sentiment Direction*, where *Sentiment Direction* is a variable indicating the sentiment of the article derived from RavenPack sentiment indices. It takes the value +1 for positive sentiment, 0 for neutral sentiment and -1 for negative sentiment. In all regressions we include firm and date fixed effects. In regressions 2, 4, and 6, we add the following firm specific control variables: *Company size*, *Return prior month*, *Volatility prior month*, *Turnover prior month*, *Illiquidity prior month*. In these regressions, we also add fixed effects for the article category (e.g. mergers and acquisitions), the relevance score (from 90 to 100) and the hour during the day in which the article was released as well as additional controls: the absolute return, turnover, and volatility each for industry and market from  $t-1$  to  $t+5$  seconds around the article. In regression 2 and 4, we also include those values for  $t-1$  to  $t+120$  seconds around the article. All variables are defined in Appendix 1. All standard errors are two-way clustered by stock and date. T-statistics are below the parameter estimates in parenthesis; \*\*\*, \*\*, \* indicate significance at the 1%, 5%, and 10% level, respectively.

Dependent Variable:	Speed of Stock Price Response (in %)		Speed of Trade Volume Response (in %)		Return t-1, t+5 (in bp)	
	(1)	(2)	(3)	(4)	(5)	(6)
RavenPack Release * D(HRH) * Sentiment Direction					0.334** (2.10)	0.330** (2.08)
RavenPack Release * D(HRH)	1.744*** (3.63)	1.709*** (3.88)	0.506* (1.79)	0.488* (1.71)	0.121 (0.90)	0.109 (0.77)
RavenPack Release * Sentiment Direction					-0.307** (-2.00)	-0.294* (-1.91)
D(HRH) * Sentiment					0.079 (0.77)	0.114 (1.09)
D(HRH)	-0.081 (-0.28)	-0.057 (-0.21)	0.084 (0.59)	0.039 (0.27)	0.087 (0.89)	0.025 (0.24)
Sentiment Direction					0.431*** (4.35)	0.222** (2.21)
Absolute Composite Sentiment Score		-0.027*** (-2.62)		-0.013*** (-2.66)		
Neutral Composite Sentiment Score		-0.280*** (-3.72)		-0.111*** (-2.70)		
Absolute Event Sentiment Score		0.044*** (4.46)		0.035*** (6.28)		
Neutral Event Sentiment Score		-1.012*** (-3.10)		-0.529*** (-3.01)		
Article category identified		-1.672 (-0.39)		-2.300** (-2.12)		0.580** (2.38)
Time since last article		0.054*** (2.63)		0.085*** (7.15)		0.165*** (13.08)
Number of firms in article		-0.109** (-2.24)		-0.132*** (-5.63)		-0.201*** (-10.77)
Number of Observations	649435	649247	690462	690268	804002	803725
R <sup>2</sup>	0.025	0.054	0.018	0.037	0.046	0.052
Relevance, Category and Hour Fixed Effects	No	Yes	No	Yes	No	Yes
Date and Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Market control variables	No	Yes	No	Yes	No	Yes
Firm specific control variables	No	Yes	No	Yes	No	Yes

## Appendix 1: Variable Definitions

This table displays the variable definitions for all variables used in the regressions. Article variables (sentiment scores, relevant scores, etc.) are based on RavenPack 3. When we winsorize, we set outliers to the allowed extreme value; e.g., “smaller 10” means that any value below 10 is set to 10. For all variables, winsorizing affects less than 1% of observations on either side.

Variable Name	Definition	Winsorizing
HRH	High relevance article Released as High relevance article. Article with relevance score of 90 or higher in both RavenPack versions.	None
HRL	High relevance article Released as Low relevance article. Article with a relevance score of 90 or higher in the new RavenPack version, but was not assigned to the company or had a relevance score below 90 in the old RavenPack version.	None
LRH	Low relevance article Released as High relevance article. Article with a relevance score below 90 or not assigned to the company in the new RavenPack version, but had a Relevance Score greater or equal than 90 in the old RavenPack version.	None
<i>RavenPack Release</i>	Dummy variable equal to 1 after RavenPack is “live” (April 1, 2009) and equal to 0 before.	None
<i>Company size</i>	Log(prior day closing price * shares outstanding)	Smaller 10
<i>Volatility prior month</i>	Average of daily squared returns of the stock in the prior 20 trading days	Larger 2%
<i>Turnover prior month</i>	Average of daily volume divided by shares outstanding in the prior 20 trading days	Larger 10%
<i>Return prior month</i>	Average return in the prior 20 trading days	Larger 3% & Smaller -3%
<i>Illiquidity prior month</i>	Percentile rank of all article-firm combinations of a day according to Amihud Illiquidity = $\frac{ ret_{daily} }{dollar\ volume_{daily}}$ mean over past 20 trading days. The most illiquid firms are assigned 100.	None
<i>Relevance</i>	Score provided by RavenPack that indicates the relevance of an article to a company and takes values from 0 (least relevant) to 100 (most relevant).	
<i>Event Sentiment Score</i>	Sentiment score that is provided by RavenPack; takes a value from 100 (positive) to 0 (negative). It is available only for articles for which the category is identified.	None
<i>Absolute Event Sentiment Score</i>	Abs ( <i>Event Sentiment Score</i> – 50)	None
<i>Neutral Event Sentiment Score</i>	Dummy variable equal to 1 if <i>Event Sentiment Score</i> equals 50 or if it is missing.	None
<i>Composite Sentiment Score</i>	Sentiment score that is provided by RavenPack; takes a value from 100 (positive) to 0 (negative). It is available for each article.	None
<i>Absolute Composite Sentiment Score</i>	Abs ( <i>Composite Sentiment Score</i> – 50)	None
<i>Neutral Composite Sentiment Score</i>	Dummy variable equal to 1 if <i>Composite Sentiment Score</i> equals 50.	None
<i>Sentiment Direction</i>	Takes the values 1 (positive sentiment), 0 (neutral sentiment) and –1 (negative sentiment). It is first based on <i>Event Sentiment Score (ESS)</i> . If <i>ESS</i> is larger 50, this variable is 1, if <i>ESS</i> is smaller than 50, it is –1. If <i>ESS</i> is missing or 50, we consult <i>Composite Sentiment Score (CSS)</i> . If <i>CSS</i> is greater than 50 we set this variable to 1, if <i>CSS</i> is smaller than 50 we set it to –1, if <i>CSS</i> equals 50 we set it to zero.	None
<i>Article category identified</i>	Dummy variable equal to 1 if the category (e.g. “merger”) of the article is identified	None
<i>Number of firms in article</i>	Log ( Number of firms in article)	None
<i>Time since last article</i>	Log (Time since last article in seconds)	None
<i>Return t-1, t+5</i>	Stock return from 1 second before to 5 seconds after the article. Returns are computed from mid-quotes.	Larger 2%, smaller -2%
<i>Return t+6, t+120</i>	Stock return from 6 seconds after to 120 seconds after the article. Returns are computed from mid-quotes.	Larger 2%, smaller -2%
<i>Speed of Stock Price Response</i>	$\frac{Abs(Return\ t - 1, t + 5)}{Abs(Return\ t - 1, t + 5) + Abs(Return\ t + 6, t + 120)}$	None
<i>Speed of Stock Price Response – Market Adj.</i>	$\frac{Abs(Market\ Adjusted\ Return\ t - 1, t + 5)}{Abs(Market\ Adjusted\ Return\ t - 1, t + 5) + Abs(Market\ Adjusted\ Return\ t + 6, t + 120)}$ Set to missing if: $Abs(Return\ t - 1, t + 5) + Abs(Return\ t + 6, t + 120) = 0$ .	None
<i>Speed of Stock Price Response – Industry Adjusted</i>	$\frac{Abs(Industry\ Adjusted\ Return\ t - 1, t + 5)}{Abs(Industry\ Adjusted\ Return\ t - 1, t + 5) + Abs(Industry\ Adjusted\ Return\ t + 6, t + 120)}$ Set to missing if: $Abs(Return\ t - 1, t + 5) + Abs(Return\ t + 6, t + 120) = 0$ .	None

<i>Speed of Trade Volume Response</i>	$\frac{\text{Turnover } t-1, t+5}{\text{Turnover } t-1, t+120}$	<i>None</i>
<i>Speed of Directional Trade Volume Response</i>	$\frac{\text{Turnover in Direction of the Article } t-1, t+5}{\text{Turnover } t-1, t+120}$ , where <i>Turnover in Direction of the Article</i> is buyer initiated turnover for articles with positive <i>Sentiment Direction</i> and seller initiated turnover for articles with negative <i>Sentiment Direction</i> . The direction of a trade is determined using the Lee and Ready (1991) methodology, but using the quote at the end of the previous second as the prevailing quote rather than the quote 5 seconds ago.	<i>None</i>
<i>D(Press Release)</i>	Dummy variable equal to 1 if the article is a press release.	<i>None</i>
<i>D(First company article of the day)</i>	Dummy variable equal to 1 if the article is the first article of the day within our sample period.	<i>None</i>
<i>D(New Story)</i>	Dummy variable equal to 1 if the article has Event Novelty Score (ENS) of 100, which means it is a new news story.	<i>None</i>
<i>D(Sentiment predicts return)</i>	Dummy variable equal to 1 if the sign of the return from 1 second before to 120 seconds after the article matches the sign of the article sentiment (it is set to missing for articles with neutral sentiment)	
<i>Signed Return t-1, t+120</i>	$\text{Return}_{t-1, t+120} * \text{Sentiment Direction}$ This variable is set to missing if Sentiment Direction is equal to zero.	<i>Larger 2%, smaller -2%</i>
<i>Past Informativeness</i>	$\text{Mean}(\text{Signed Return}_{t-1, t+120})$ , the mean is taken over the prior six calendar months.	<i>None</i>
<i>HFT Trading Fraction Response</i>	$\frac{\text{HFT Trading Fraction}_{t-1, t+5}}{\text{HFT Trading Fraction}_{t-1, t+5} + \text{HFT Trading Fraction}_{t+6, t+120}}$ where $\text{HFT Trading fraction}_{t,s} = \frac{\text{Shares Traded by HFT on NASDAQ } t_s}{\text{Total Shares Traded on NASDAQ } t_s}$	<i>None</i>
<i>HFT Active Trading Fraction Response</i>	$\frac{\text{Active HFT Trading Fraction}_{t-1, t+5}}{\text{Active HFT Trading Fraction}_{t-1, t+5} + \text{Active HFT Trading Fraction}_{t+6, t+120}}$ where $\text{Active HFT Trading Fraction}_{t,s} = \frac{\text{Shares Traded by HFT using liquidity demanding trades on NASDAQ } t_s}{\text{Total Shares Traded on NASDAQ } t_s}$	<i>None</i>
<i>Passive HFT Trading Fraction</i>	Defined analogous to HFT Active Trading Fraction Response counting only trades using liquidity supplying trades.	
<i>With News Trading Fraction</i>	Defined analogous to HFT Active Trading Fraction Response counting only trades in direction of news sentiment (missing for neutral sentiment).	
<i>Against News Trading Fraction</i>	Defined analogous to HFT Active Trading Fraction Response counting only trades against direction of news sentiment (missing for neutral sentiment).	
<i>Amihud Illiquidity<sub>ij</sub></i>	$\frac{1}{N_{ij}} \sum_t^{N_{ij}} \frac{ r_{it} }{\text{dolvol}_{it}}$ , where $r_{it}$ is the return for stock $i$ during second $t$ ; $\text{dolvol}_{it}$ is the dollar volume for stock $i$ during second $t$ ; and $N_{ij}$ is the number of seconds in which stock $i$ traded during interval $j$ .	<i>Larger 2</i>
<i>Effective Spread<sub>ij</sub></i>	$\frac{1}{N_{ij}} \sum_t^{N_{ij}} \text{sign}(\text{buys}_{it} - \text{sells}_{it}) * \frac{\text{price}_{it} - \text{midquote}_{it-1}}{\text{midquote}_{it-1}}$ , where $\text{buys}_{it}$ ( $\text{sells}_{it}$ ) is the number of shares traded and initiated by the buyer (seller) for stock $i$ during second $t$ ; $\text{price}_{it}$ is the last execution price for stock $i$ during second $t$ ; $\text{midquote}_{it}$ is the last bid-ask midpoint for stock $i$ during second $t$ and $N_{ij}$ is the number of seconds in which stock $i$ traded during interval $j$ . The direction of a trade is determined using the Lee and Ready (1991) methodology, but using the quote at the end of the previous second as the prevailing quote rather than the quote 5 seconds ago.	<i>Larger 3</i>
<i>Change in Amihud Illiquidity</i>	$\frac{\text{Amihud Illiquidity } t-1, t+5}{\text{Amihud Illiquidity } t-1, t+5 + \text{Amihud Illiquidity } t-300, t-120}$ in seconds around the article.	<i>None</i>
<i>Change in Effective Spread</i>	$\frac{\text{Effective Spread } t-1, t+5}{\text{Effective Spread } t-1, t+5 + \text{Effective Spread } t-300, t-120}$ in seconds around the article.	<i>None</i>
<i>Market return t-1, t+5</i>	Value-weighted return of all common stocks in TAQ (which are also in CRSP) from 1 second before to 5 seconds after the article. Returns are computed from mid-quotes.	<i>None</i>
<i>Market turnover t-1, t+5</i>	Total dollar trading volume of all common stocks in TAQ (which are also in CRSP) from 1 second before to 5 seconds after the article divided by total market capitalization at $t-2$ .	<i>None</i>
<i>Market volatility t-1, t+5</i>	Value weighted average squared second return of all common stocks in TAQ (which are also in CRSP) averaged from 1 second before to 5 seconds after the article.	<i>Larger 20 bp</i>
<i>Market adjusted return t-1, t+5</i>	$\text{Return}(t-1, t+5) - \text{Market Return}(t-1, t+5)$	<i>Larger 2%, smaller -2%</i>
<i>Industry variables</i>	Follow the same definition as market variables but include only stocks within the firm's 12 Fama French industry	<i>Same as market variables</i>