

News and Intraday Jumps: Variable Selection with Regularization and Class Imbalance

Abstract

We study how information provokes intraday price jumps taking into account, besides news timing, the sentiment of news stories and other high-frequency indicators. By applying penalized logistic regression and addressing the rare nature of jumps, in addition to the previous evidence showing that causes of jumps are rate decisions and earnings announcements, we find that news impact across companies is heterogeneous, that news provoking jumps is often followed by other news about the same company, that news stories sentiment and macro-surprises sign help to predict the jump sign, and, finally, that market players sometimes anticipate company-specific news; differently from previous evidences, we find that EPS has a reduced impact compared to other news' types.

Keywords: price jumps, returns predictability, news data, financial text mining.

JEL Classification: C58, C22, G10, C55, C25, G14.

1 Introduction

Studying the occurrence of sudden and unexpected movements in stock prices, that is, price jumps, is fundamental for asset pricing and risk management. Jumps might occur as a consequence of a variety of possible events, starting from global or systemic shocks, including events related to specific companies, as well as technical market failures. The information about the events causing the jumps might be associated with news released after only a few minutes from the time of the event. Therefore, it is not surprising to observe a reaction of financial markets to news-related announcements, in particular when those announcements convey unexpected elements. Given the reactions to news announcements, the literature started to investigate the relationship between news and intraday jumps; see, among others, Bollerslev et al. (2008), Lee and Mykland (2008), Evans (2011), Lahaye et al. (2011), Rangel (2011), Lee (2012), Boudt and Petitjean (2014), and Bajgrowicz et al. (2016).

While there is an agreement on the fact that macroeconomic and selected company-specific news is indeed associated with jumps, limited interest has been given to the large number of indicators that could be derived from news textual data, in particular when going beyond the mere use of news timing and sentiment. Moreover, due to the rare nature of jumps, when analyzing their association with news, we believe that a sample imbalance correction should be included.

We first verify if the class imbalance issue does not alter the identification of the news impacting on the jumps occurrence. Furthermore, we provide two main contributions to the literature focusing on the relation between news and jumps. First, we build a large number of variables starting from a news database, including lead and lag of news timing but also distinguishing news topic, relevance, surprise, persistence, and sentiment. The introduction of several potentially relevant jump drivers requires the combination of regularization and logistic regression. The latter links the probability of the occurrence of intraday jumps to a set of indicators based on firm-specific and macroeconomic news, while the use of regularization methods allows us to avoid overfitting issues when taking into account a large

set of potential predictors and, at the same time, to perform a variable selection; such a procedure is more flexible than an a priori selection of price jump drivers (Lee , 2012). Second, we measure the impact of relevant news on the jumps occurrence going beyond the mere statistical significance and thus focusing on different quantitative tools, namely, the average partial effect (APE) and the area under the receiver operating characteristic curve (AUC) decrease after variable permutation.

Our purpose is to provide additional insights to the following open questions: What is the likelihood that a news release causes a jump? Does information conveyed by the news, beyond its mere release, have an impact on the probability of jump occurrence?

We address these questions on an empirical basis by using a unique database. It contains the unscheduled firm-specific news stories of 88 stocks included among the S&P 100 constituents. The news we consider is released by two providers, FactSet-StreetAccount and Thomson Reuters-Thomson One, both classifying news into topics depending on their content. Furthermore, we consider the companies' prescheduled earnings per share (EPS) announcements, and 23 prescheduled US macroeconomic announcements. Our data covers a relatively long time span, from January 2005 to February 2015. We integrate the news with the occurrence of price jumps at a high frequency. We identify the intraday timing of jumps relying on the method of Andersen et al. (2007b).

We perform several analyses, starting from a matching study. We observe that the news more frequently associated with jumps consists of the quarterly EPS announcements, the FOMC rate decisions, and company-specific news stories with specific topics. However, these news types are related to jumps in a very different way: EPS and some company-specific news stories present a very high conditional probability to make a jump appear after they are released, but due to their low unconditional probability to be released, only a tiny fraction of price jumps is matched with them. On the contrary, FOMC rate decisions and other company-specific news exhibit a lower conditional probability to provoke a jump but, having a higher unconditional probability to be released, are associated with a much bigger

fraction of jumps. Furthermore, we observe that the absolute size of jumps associated with EPS and firm-specific news stories is considerably higher than that of jumps due to FOMC announcements. Not all the news conveys information that could potentially lead to jumps occurrence, and the news relevance changes depending on the topic. In addition, within the news that could cause a jump we do have further heterogeneity, with some news types being more relevant than others. This is in line with previous studies; see Bollerslev et al. (2008), Lee and Mykland (2008), Lee (2012), and Boudt and Petitjean (2014). By using a more extensive dataset, we extend their results in various directions.

With the aim of reconstructing the different portions of information assimilated by heterogeneous market players, which endogenize and react to news releases at differing speeds, we build a large number of indicators. We design them by combining various time horizons, a series of concepts for news stories, surprise indicators for EPS and macro-announcements, and, finally, the sentiment we retrieve from each news story with a novel procedure based on word lists and combinations that makes use, in a different way, of the headline (a short text) and the news body (a much longer text). We then apply elastic net and adaptive lasso within the estimation of a logistic regression linking the probability of intraday jumps occurrence to the above-mentioned indicators. Additionally, we combine the previous tools with several machine-learning techniques to address the imbalanced classification problem, arising from the fact that jumps are rare events. By employing these techniques, which allow the models to reach a higher performance in terms of AUC, we are able to attribute a more correct impact to the news-based indicators associated with jumps, and to highlight a set of additional indicators that are useful to predict the jumps sign. Overall, this confirms the relevance of properly addressing the class imbalance issue.

Our innovative approach, which makes it possible to estimate the APE and the AUC decrease after variable permutation for each news based-indicator, suggests that the information most useful to explain jumps is that related to FOMC rate decisions and company-specific news stories, with the former holding a predominant role. Therefore, we robustify

the findings of Bollerslev et al. (2008), Lee and Mykland (2008), Lee (2012), and Boudt and Petitjean (2014). However, additional and novel evidences emerge from our analyses. We start by focusing on the information retrieved from our large number of news-related variables. First, we note that the news that provokes jumps is often followed, within a few minutes, by the release of other news about the same company. Second, we observe that the sentiment of the news stories and the sign of the surprise of macroeconomic news with respect to expectations are useful in the prediction of the sign of the jumps. Third, we point out that market players are sometimes able to anticipate company-specific news releases, but the same either never happens with regard to macroeconomic news or, if it happens, it does not provoke jumps. Fourth, by measuring how similar assets are in terms of their reaction to the news, we show that the main determinant of similarity of the assets within an economic sector is determined from the assets common response to macroeconomic news, and that the sectors that contain the most similar assets are Industrial Goods and Financial.

Our dataset is significantly more comprehensive with respect to those employed in the previous studies. Moreover, by distinguishing possibly delayed, instantaneous, and even anticipated market reactions to information, represented by news-based indicators that go beyond the mere release of news, our methodology substantially differs from the current literature. In fact, the latter usually defines associations between news and jumps in terms of the occurrence of jumps either within a short time after a news release or during the same day. Kanniainen and Yue (2019), similarly to us albeit with a different approach, analyze the pre- and post-news jump dynamics of stock prices, but they do not provide such a fine distinction between news types nor do they construct indicators on the basis of concepts for news stories. In addition, to our best knowledge, we are the only ones who relate such a broad variety of macroeconomic and firm-specific, both prescheduled and unscheduled, intraday news to jumps. Bajgrowicz et al. (2016) finely separate news into several topics, too. However, they report the probability of observing a jump on the same day as a news release without investigating the timing of this relationship. A further closely related study

is that of Lee (2012) that analyzes the impact of news on jumps probability. We deviate from Lee (2012) in several aspects including the equity database (number of companies, time span, jump detection approach), the news database (we include both headlines and news bodies) and, most relevant, the methodology (we filter the intraday periodic volatility, introduce a sample imbalance correction, and evaluate the impact of news with APE and AUC decrease after variable permutation). The most innovative aspects of our study relate to the methodological approaches we adopt. Accounting for sample imbalance is crucial given the nature of jumps and, to our best knowledge, we are the first to move in that direction. Moreover, although the evaluation of statistical significance of jump predictors is of interest, the evaluation of APE and AUC provides additional insights on the relevance of jump predictors, focusing on the size of their impact on the probability of jumps occurrence. The deviations we made from Lee (2012) motivate the differences between our findings and those of the mentioned study, in particular with respect to the limited role exerted by the EPS announcements as jump predictors.

The remainder of the paper is organized as follows. Section 2 describes the dataset and the procedure employed to extract the sentiment. Section 3 contains the empirical analysis, covering the matching of news to jumps, the construction of the news-based indicators, and the detection of the most important news. Section 4 presents the conclusion of the paper. A Supplementary Material document accompanies the paper; it provides a review of the literature on the relationship between news and jumps, penalized logistic regression, and the class imbalance issue; it also contains additional details on the data, the description of the jump testing procedures, the word lists and the algorithms used to extract the sentiment, summary tables for the indicators, and additional tables and figures.

2 Dataset

As we mentioned in the introduction, the dataset we use ranges from January 10, 2005 to February 25, 2015 and includes the 1-min prices of the S&P 100 stocks, their firm-specific news stories and EPS, and 23 US macroeconomic announcements. We recover the high-frequency equity prices from Kibot.com,¹ while the news stories, the EPS announcements, and the macroeconomic announcements come from two providers, StreetAccount from Factset and Thomson One from Thomson Reuters.²

In the following subsections, we describe the procedures we adopt to extract from the raw data the variables we will use in our empirical analyses, namely, the sequence of intra-daily jumps and a collection of news-related variables.

2.1 Intraday Jumps

We follow a relevant strand of the literature and estimate jumps from 5-min returns to get rid of market microstructure frictions. We work on continuously compounded 5-min returns, which are calculated as $100 \log(p_j/p_{j-1})$, where p_j denotes the price at the end of the j -th 5-min interval. From the constituents of the S&P 100 in February 2015 (end of time span), after discarding those which entered the index after January 2005 (beginning of time span) and those without continuously available news for all the time span, we remain with 88 stocks (see Table B1 in Supplementary Material B for the companies list). After filtering the periodic component of intraday volatility through the technique of Boudt et al. (2011), we rely on the method of Andersen et al. (2007b) to identify the precise intraday intervals at which jumps occur. We deviate from Andersen et al. (2007b), which adopt the realized bipower variation, as we use the corrected threshold bipower variation of Corsi et al. (2010), which was shown to be more accurate for the estimation of jumps. Supplementary Material

¹Even though this data provider is not as well-known as others, the data quality is equivalent to that of TAQ. Figure C1 in Supplementary Material C provides a comparison of the two datasets for a single stock.

²The time span covered by our dataset is dictated by the limitations we have in the access to the news providers, which prevent us from extending the analyses to the more recent years.

C reports a detailed description of the procedure we follow. This jump testing approach allows us also to recover the jump sign and to separate positive jumps from negative jumps.

Notably, our work differs from Bajgrowicz et al. (2016), who use a much finer sampling and argue that bursts of volatility are often incorrectly identified as jumps when the sampling is too sparse. Christensen et al. (2014) use ultra high-frequency tick data recorded at millisecond precision and show, indeed, that traditional measures of jump variation based on lower-frequency data tend to spuriously assign a burst of volatility to the jump component. However, they recognize that, when market participants view prices at periodic but infrequent intervals, they could be exposed to price moves that they experience as genuine jumps, in particular when a period of extreme volatility unfolds between observation points, or when their pay-off and execution costs can be impacted in a manner consistent with the presence of genuine price jumps. Therefore, our findings remain relevant for portfolio allocation and risk management strategies, as well as for monitoring the impact of news on the occurrence of price jumps.

2.2 News

As we stated, we collect firm-specific news stories and EPS announcements for 88 companies, all belonging to the S&P 100 basket. In addition, we include 23 macroeconomic announcements. All news reports the release date and time with minute precision. News stories, which are filtered for relevancy and redundancy,³ are assigned a topic by the providers, and Thomson Reuters additionally assigns a level of importance. We briefly describe the news we will use later:

- **Unscheduled News Stories.** StreetAccount news stories report a topic, while Thomson Reuters news stories report both a topic and a level of importance. The latter is assigned on the basis of the expected effect that the event will have on the company's

³The two professional news providers supply only news classified as relevant to the firm. News is also filtered for redundancy, so each news story is included only once.

Table 1: Topics list by provider.

StreetAccount	Thomson Reuters
all	all
earnings related	earnings pre-announcements
M&A	dividends
litigation	financial
regulatory	medium
newspapers	high
up/downgrades	top

all: all stories; *earnings related*: earnings related stories (e.g., guidance, quarterly results); *M&A*: mergers and acquisitions; *litigation*: citations, court disputes; *regulatory*: regulatory bodies decisions and investigations; *newspapers*: newspapers articles; *up/downgrades*: credit rating assignments by rating agencies; *earnings pre-announcements*: unscheduled earnings pre-announcements; *dividends*: unscheduled dividend pre-announcements (differently from Bajgrowicz et al. (2016), who employ prescheduled dividend announcements); *financial*: merge of the stories with the topics *equity issues*, *bond issues*, *share repurchases* and *equity investments*, all of which are related to events that have an impact on the company’s balance sheet.

operational and/or financial performance. The levels of importance are four, *low*, *medium*, *high* and *top*, and each level consists in a filter that eliminates all news with a lower level, for example, *low* gives all news and *medium* gives all news tagged with *medium*, *high* and *top*. For ease of illustration, in the following we report the levels of importance *medium*, *high* and *top* as if they were topics, and omit *low* because it coincides with *all*.⁴ We use seven topics from StreetAccount and seven topics from Thomson Reuters. We list them in Table 1.

- **Prescheduled Earnings Announcements.** StreetAccount news stories with the topic *earnings* also report the quarterly EPS announcements of each company along with the consensus forecast. The latter represents the mean of a set of surveys before the time of reporting.
- **Prescheduled Macroeconomic Announcements.** Thomson Reuters makes avail-

⁴While it is possible to filter Thomson Reuters news stories by both topic and level of importance, we apply the filter only by topic or by importance and obtain 4 topics plus 3 levels of importance. We avoid combining filters, which would yield $4 \times 4 = 16$ combinations.

Table 23 US macroeconomic announcements released during market hours. As for EPS, both released figure and consensus forecast are reported. We list them in Table 2.

Table 2: Macroeconomic announcements and release time.

Announcement	Release Time
Business Inventories	10:00
Chicago PMI	09:45/10:00
Construction Spending	10:00
Consumer Confidence	10:00
Consumer Credit	15:00
Michigan Consumer Sentiment Index	09:45/09:55/10:00
EIA Crude Oil Stocks	10:30
ECRI Weekly	10:30
IBD Economic Optimism	10:00
Employment Trends Index	10:00
Existing Home Sales	10:00
Factory Orders	10:00
Federal Budget	14:00
FOMC Rate Decisions	12:30/14:00/14:15
NAHB Housing Market	10:00/13:00
Leading Index	10:00
ISM Manufacturing Index	10:00
EIA Natural Gas Stocks	10:30
New Home Sales	10:00
New York NAPM Index	09:45
Pending Home Sales	10:00
Philadelphia Fed Business Index	10:00/12:00
Wholesale Inventories	10:00

2.3 Sentiment Detection

We introduce a novel procedure for sentiment detection. Our proposal is grounded on the method of Loughran and McDonald (2011) and is based on the combination of two distinct algorithms, specifically designed for the extraction of the sentiment from the headline (i.e., a short text) and the news body (i.e., the news story, a much longer text). Both algorithms make use of a number of word lists containing single words as well as sequences of words,

and they detect the sentiment of a text on the basis of the presence and the combination of items belonging to these lists. The word lists are set out in Table 3.

Table 3: Word lists used for sentiment detection and their rationale.

Primary	items that in a headline identify the object of the news
Secondary Positive	items that in a headline, combined with primary items, identify the positive sentiment of the news
Secondary Negative	items that in a headline, combined with primary items, identify the negative sentiment of the news
Stop Words	items often located between negations and positive or negative items
Positive	items with a positive meaning
Negative	items with a negative meaning
Negations	items that invert the sentiment of the items they precede
Neutral	items incompatible with positive and negative items (if present, negative and positive items could detect a wrong sentiment)
Primary Incompatible	items incompatible with primary items (if present, primary items could detect a wrong sentiment)
Secondary Neutral	items incompatible with secondary positive and negative items (if present, secondary items could detect a wrong sentiment)
Special Topics	items related to topics that make the sentiment detection problematic

Before applying the procedure, we pre-process the raw text. First, we strip punctuation, numbers, and symbols. Second, we remove the following words: “Limited” because it usually refers to the end of company names, “Recall” and “recall” from the news body because it is common that sentences recalling past news use this word, “COST” from the news of *CostCo Wholesale Corporation* and “LOW” from the news of *Lowe’s Companies* because these words correspond to the tickers, “Jobs” from the news of *Apple* because it often refers to “Steve Jobs”, and “Target” from the news of *Target Corporation* because it usually refers to the name of the company. Third, we convert all letters to lowercase. We do not use lemmatization nor stemming because it would cause words to change their original sentiment; as an example, consider the word “objection” which is a *negative* word, but its lemma and

stem “object” is not.

Our sentiment detection approach, which we call *H&B* (*Headline and Body of the news*), starts from the evaluation of the news headline by means of a first algorithm, we call it *H*. This procedure, suitable for short texts, is based on the identification of the main subject of the text by checking the presence of *primary* items (e.g., “dividends”, “EPS”, or “results”) and the subsequent detection of the sentiment by using *secondary* items that allow to understand if the news is good (e.g., “boost”, “grew”, or “picked up”, which indicate growth concepts) or bad (e.g., “bottoming”, “decrease”, or “lacking upside”, which indicate reduction concepts). If algorithm *H* does not retrieve a positive or negative sentiment, then we proceed to the evaluation of the news body by means of a second algorithm, we call it *B*. This procedure is suitable for long texts where the presence of *primary* items cannot be used to infer the main subject. The sentiment, the variable the algorithms provide as an output, has three possible outcomes: 0, +1 and -1 for neutral, positive and negative sentiment, respectively. Supplementary Material D contains additional details on the word lists as well as on their content, describes the two algorithms, and explains in details how they derive the sentiment from the news.⁵

3 The Relevance of News in Explaining Jumps

3.1 A First Look at the Data

Figure 1 presents the time series, over the whole sample, of the number of jumps, the median of the jumps absolute size,⁶ and the number of news stories available from each provider. All sums are taken over assets. We do not report earnings and macro-announcements, as they are released on a regular basis and their frequency is constant.

⁵We validated the procedure on a sample of a few hundred news stories, but defer the detailed results to a future study.

⁶We proxy for the jumps size with returns of the high-frequency intervals in which we detect the occurrence of a jump.

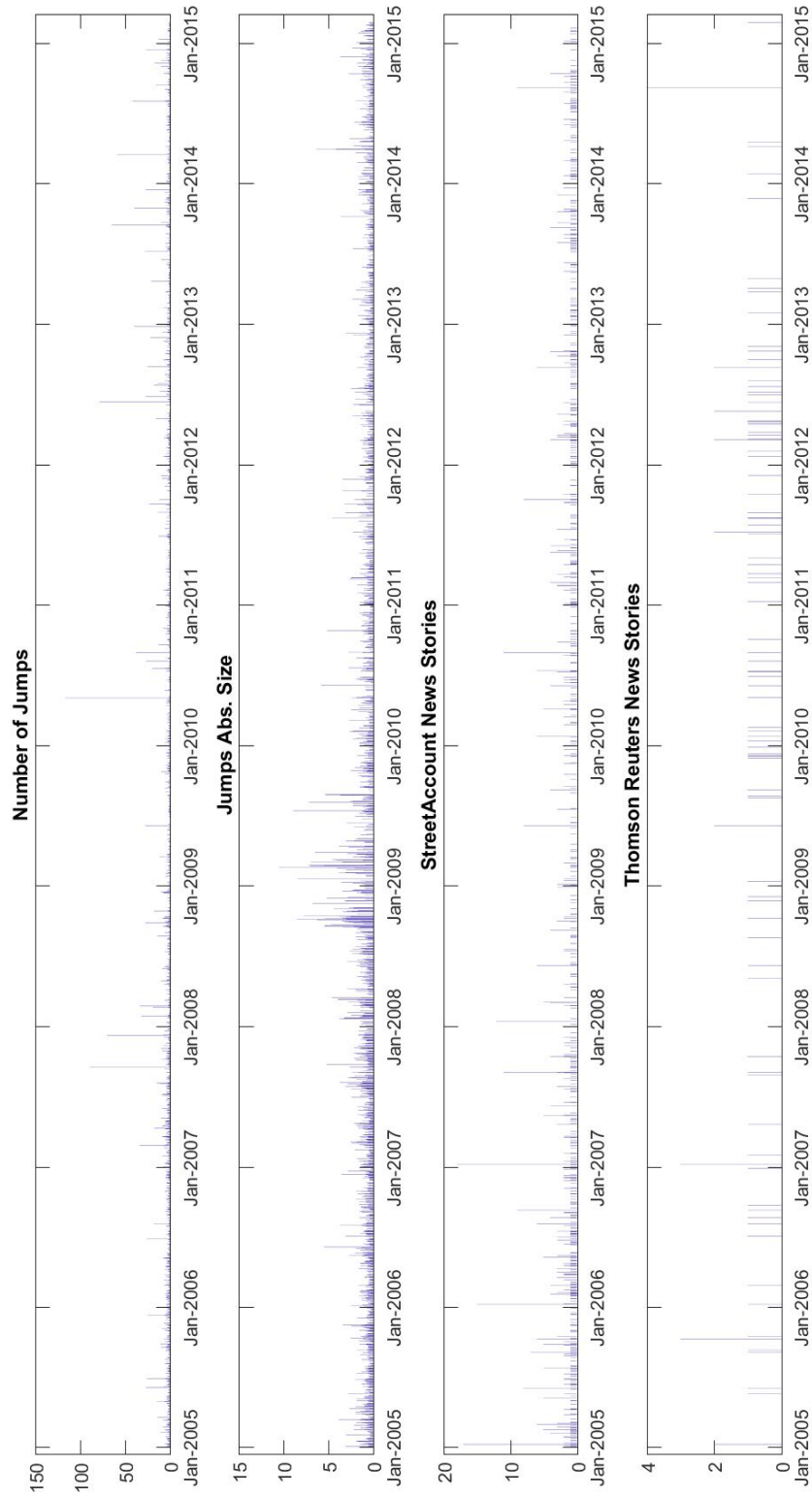
We highlight that a number of jumps higher than usual occurs from the end of 2007 to the beginning of 2009. In addition, the absolute size of jumps is higher during the same period, peaking at the beginning of 2009. This is not surprising, as we might expect intense activity during the global financial crisis, which lasted from December 2007 to June 2009.

The number of news stories released by both providers is slightly decreasing over the sample, and we interpret this as a result of an increasing selection endeavor by the providers, who are interested in the release of relevant news only. Moreover, from the end of 2008 to the beginning of 2010, that is, from the beginning of the global financial crisis until one year after its end, the number of news stories released by Thomson Reuters was higher than in the rest of the sample, possibly as a consequence of unusually high attention.

Figure 2 illustrates the distribution by intraday interval of the jumps frequency, the median of their absolute size, and the frequency of each main type of news. As all the stocks we consider trade either at NYSE or NASDAQ, regular market hours go from 9:30 to 16:00 Eastern time. We observe that the jump frequency is particularly high during the opening and closing times of the market,⁷ and that it peaks at the following times (which stand for the beginning of a 5-min intraday interval): 10:00, 14:00, 14:15, 14:20, and 15:05. The absolute jump size is characterized by a U-shaped distribution over the day and peaks at 14:50. We stress that our findings sensibly differ from those in Lee (2012), indicating that more than 80% of the jumps were occurring in the first one hour and a half after the market opening (i.e., from 9:30 to 11:00, see Lee (2012), p. 457, Table 3), as opposed to about the 20% of our analysis. This is a consequence of the pre-filtering we introduce. Moreover, as pointed out by Boudt et al. (2011), pre-filtering impacts on the price jump detection, both in the first part of the day, where volatility is high, reducing spurious jump detection, as well as during low volatility intra-day intervals, increasing the detection frequency of relatively low jumps. Consequently, a more robust jump detection procedure will lead to more robust

⁷We remind readers that we adopted the filter of Boudt et al. (2011) that accounts for the larger volatility during market opening and closing. Therefore, the jumps we observe in the first part of the day are structural jumps and are not due to the high volatility at market opening.

Figure 1: N. of jumps, median of abs. jumps size, and n. of news stories over the sample.



Notes: For each day, the figure shows the sum of the number of jumps, the median of the jumps absolute size, the sum of StreetAccount news stories, and the sum of Thomson Reuters news stories. The number of jumps and the number of news stories are summed over all the assets, while the median of the jumps absolute size is computed pooling all the jumping assets.

analyses involving the identified jumps.

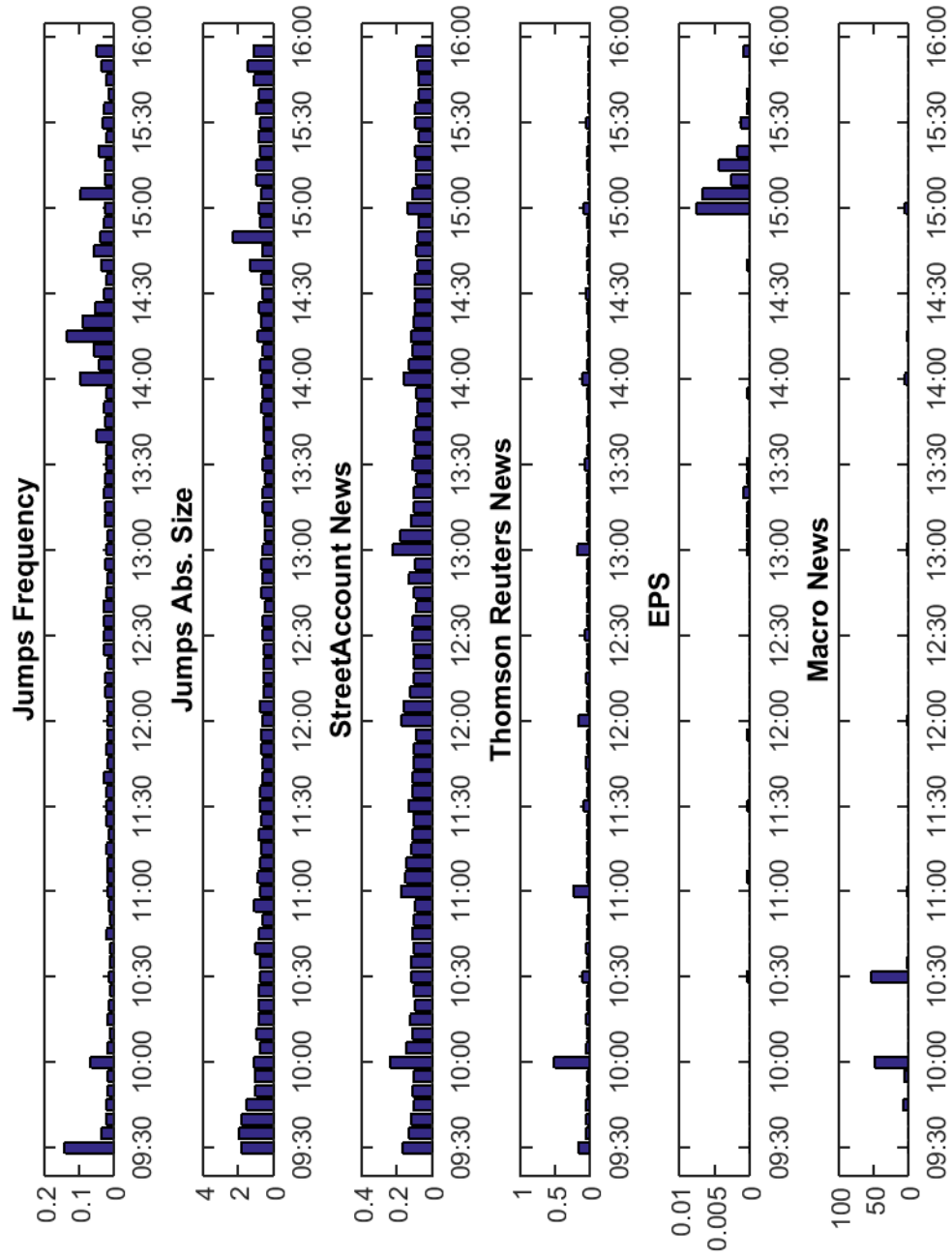
The unconditional probability of macroeconomic announcements is much higher compared to the other categories of news. Moreover, StreetAccount news stories are released more often than Thomson Reuters ones, and quarterly EPS have the lowest unconditional probability. With regard to quarterly EPS, we stress that they are released only once every three months, by definition, but it is also common, for some companies, to communicate them outside market hours. In our selected sample, 10 companies out of the 88 never release EPS during market time. Interestingly, both StreetAccount and Thomson Reuters news stories reveal a cyclic path, which is higher than usual at the beginning of the hour, especially at 10:00, and, to a lesser extent, at the half-hour. EPS are clearly primarily released at 15:00 or slightly after. Macro-news, finally, is predominantly announced at 10:00 and 10:30.

We note some potential relations between jumps and news. The frequency of jumps soars on four particular time windows: first, at 10:00, when firm-specific and macroeconomic news is released more often than usual; second, at 14:00, concomitantly with FOMC macro-announcements;⁸ third, at 14:15 and 14:20, which is concomitant with another common release time of FOMC rate decisions and during the following interval; finally, at 15:05, which is the interval following EPS and Consumer Credit announcements. Jumps absolute size is instead particularly high at 14:50, which is slightly before EPS and Consumer Credit news. Finally, we did not observe any unusual jump activity around market closing times for European markets (11:30). It seems, therefore, that all types of news are potentially related to jumps, which can happen during the same 5-min interval, with a delay or even in anticipation of news.

The previous evidences suggest investigating the relationship between news and jumps with more sophisticated techniques.

⁸See Table 2 for the release times of each macro-announcement.

Figure 2: Frequency in % of jumps, median of jumps absolute size, and frequency in % of each main type of news for each intraday interval.



3.2 Matching Analysis

We define a jump-news coincidence as the occurrence of a jump within 30 minutes after the release of a news, and use three metrics to analyze the matching of news and jumps:

- $P(J|N)$: the number of jump-news coincidences over the number of news announcements
- $\text{median}(J|N)$: the median absolute size of jumps that coincide with news
- $P(N|J)$: the number of jump-news coincidences over the number of jumps

Several works adopt similar metrics. Lahaye et al. (2011) define a jump-news match as the occurrence of jumps within 1 hour after a news announcement. Boudt and Petitjean (2014) define a match as the occurrence of jumps within 10 minutes of a news announcement. Bajgrowicz et al. (2016) report the probability of observing a jump on the same day as a news release. The choice of a time window equal to 30 minutes is a compromise between including market reactions that take place later than the first minutes after news and, at the same time, discarding jumps due to other news or sources. We repeated the analysis by defining a coincidence as the occurrence of a jump from the same interval of a news release until 30 minutes after, in order to consider faster market reactions, and we obtained similar results. News about down- and upticks of stock prices are uncommon, therefore endogeneity is not a serious issue for this alternative definition.

The metrics we adopt provide different information on the relation between jumps and news: $P(J|N)$ describes the likelihood that a news release causes a jump, while $P(N|J)$ measures what proportion of jumps are associated with a particular type of news. The median allows inferring the impact of a news announcement on the size of jumps, without being influenced by outliers.

In Table 4, we compare StreetAccount news stories (topic *all*), Thomson Reuters news stories (topic *all*), EPS, macro-announcements (aggregating all of them), and *Other Sources*,

which consists in the lack of any news of our dataset. In Tables F1–F3, in Supplementary Material F, we report detailed results for StreetAccount news stories topics, Thomson Reuters news stories topics, and single macro-announcements.

In Table 4, we see that EPS is the type of news that causes a jump with the highest probability, with a $P(J|N)$ of 5.09% (i.e., among all the EPS announcements, 5.09% are followed by a jump). In terms of $P(J|N)$, EPS are followed, in descending order, by StreetAccount news stories, Thomson Reuters news stories, macro-announcements, and Other Sources. EPS are rarely announced relative to the other news and, as a consequence, their $P(N|J)$ (the proportion of jumps associated with them) is the lowest one, equal to 0.22% (i.e., among all jumps, only 0.22% of them are preceded by an EPS announcement). EPS have the highest median($J|N$) among all types of news, equal to 1.64. StreetAccount news stories have both a higher $P(J|N)$ and a higher $P(N|J)$ with respect to Thomson Reuters news stories, therefore we read this as supporting that StreetAccount news stories are more associated with jumps than Thomson Reuters ones. The $P(N|J)$ of macro-announcements is higher than that of firm-specific news, consistent with Boudt and Petitjean (2014) and Bajgrowicz et al. (2016), due to the higher unconditional probability of macroeconomic news compared with firm-specific ones. There are no striking differences between positive and negative jumps with regard to any of the metrics, neither here nor in the tables we mention in the following, except for EPS that show a much higher $P(J|N)$ when jumps are negative.

For Other Sources, $P(N|J)$ is 84.50%, revealing that the majority of jumps are not associated with the news at our disposal. These jumps may be caused by a lack of liquidity in presence of an excessive demand for trading, or by other sources of news, for example, social media short messages such as tweets or Facebook posts.⁹ We now extend the analysis by further classifying news into news stories’ topics and single macro-announcements, results

⁹In this context see, for example, Audrino et al. (2019). They analyze the impact of sentiment and attention variables on volatility by combining RavenPack News Analytics news articles, text data from Twitter and Stock-Twits social media platforms, information consumption from Wikipedia, and search engine data from Google Trends.

are reported in Supplementary Material F.

Table 4: Main sources of news and jumps matching.

News		All	Pos	Neg
EPS	$P(J N)$	5.09	1.22	3.87
	$\text{median}(J N)$	1.64	1.78	1.74
	$P(N J)$	0.22	0.16	0.28
StreetAccount News Stories	$P(J N)$	0.20	0.11	0.09
	$\text{median}(J N)$	1.49	1.58	1.87
	$P(N J)$	4.04	3.94	4.12
Thomson Reuters News Stories	$P(J N)$	0.15	0.08	0.07
	$\text{median}(J N)$	1.34	1.29	1.63
	$P(N J)$	1.55	1.69	1.37
Macro Announcements	$P(J N)$	0.03	0.01	0.02
	$\text{median}(J N)$	0.81	0.90	0.80
	$P(N J)$	11.22	9.39	13.70
Other Sources	$P(J N)$	0.03	0.02	0.01
	$\text{median}(J N)$	0.76	0.80	0.73
	$P(N J)$	84.50	86.35	81.99

Notes: $P(J|N)$, $\text{median}(J|N)$ and $P(N|J)$ for StreetAccount news stories (all), Thomson Reuters news stories (all), macroeconomic announcements (all), EPS, and Other Sources (lack of any of the previous types of news). News is sorted in descending order by $P(J|N)$. The last three columns show the metrics by distinguishing: all jumps, positive jumps, and negative jumps. Probabilities are expressed in percentages.

When looking at StreetAccount news stories only (see Table F1), *newspapers* is the topic characterized by the highest $P(J|N)$, equal to 0.33%, and by the second highest $\text{median}(J|N)$, equal to 2.23. The topic *M&A* is the second in terms of $P(J|N)$, equal to 0.26%, and it has the highest $\text{median}(J|N)$, equal to 2.25. *Earnings related* is the third, and it is followed by *all*, therefore it seems that for StreetAccount news stories the topics *newspapers*, *M&A* and *earnings related* help to filter information potentially causing jumps. *Earnings related*, with a $P(N|J)$ equal to 1.04%, which is second only to the $P(N|J)$ of *all*, is the topic mostly associated with jumps.

With regard to Thomson Reuters news stories (see Table F2), we note a high $P(J|N)$ for the topic *top*, equal to 2.47%; this value is higher than the $P(J|N)$ of any topic of any

provider and of any macro-announcement, and only lower than the $P(J|N)$ of EPS. All other topics of Thomson Reuters news stories have a $P(J|N)$ higher than that for the topic *all*. They are, in descending order: *earnings pre-announcements*, *financial*, *dividends*, *high*, and *medium*. *Top* and *earnings pre-announcements* also have the highest median($J|N$) (1.54 and 1.52, respectively). *High* and *medium* news stories are also the ones mostly associated with jumps, with a $P(N|J)$ equal to 1.21% and 1.26%, respectively. It seems that Thomson Reuters is able to effectively classify news stories by importance and that, additionally to, as expected, *earnings pre-announcements*, news belonging to the topics *financial* and *dividends* is also relevant in causing jumps.

Finally, with regard to macro-announcements (see Table F3), it is clear that the FOMC rate decision is the macro-announcement that most likely causes jumps with a $P(J|N)$ equal to 0.72%, while the other macro-announcements have a $P(J|N)$ lower than the majority of news stories' topics. FOMC rate decision's $P(N|J)$, equal to 5.75%, is remarkably higher than for any other macro-announcement and any other type of news, indicating that it is the information associated with the highest proportion of jumps. Its median($J|N$), instead, is lower than that of EPS and firm-specific news stories.

These empirical evidences allow us to conclude that, although most of the jumps seem not to be related to the news at our disposal, several types of news represent possible causes for the occurrence of jumps. We highlight that jumps are linked to these sources in a very different way. EPS and, to a lesser extent, Thomson Reuters news with the topics *top*, *earnings pre-announcements*, *financial*, and *dividends*, present a very high conditional probability to make a jump appear after they are released, but due to their low unconditional probability to be released only a tiny fraction of jumps is matched with them. FOMC rate decisions and, to a lesser degree, StreetAccount news with topic *all*, instead, exhibit a lower conditional probability to provoke a jump but, having a higher unconditional probability to be released, are associated with a much bigger fraction of jumps. Furthermore, the absolute size of jumps associated with EPS and firm-specific news stories is considerably higher than

that of jumps due to FOMC announcements.

Our results are consistent with Boudt and Petitjean (2014), who find that macro-news is associated with jumps more often than firm-specific news but the latter, despite a lower unconditional probability, is more likely to cause jumps. Differently, Bajgrowicz et al. (2016), first claim that the other studies (which apply jump detection techniques based on sampling frequencies comparable to ours) incorrectly classify bursts of volatility as jumps. Then, they find that these bursts of volatility are generated especially by press releases following scheduled FOMC meetings and announcements related to share repurchase programs. While their findings agree with ours regarding FOMC announcements, differently from us they observe no increase in the occurrence of jumps on quarterly earnings announcements.

3.3 Penalized Logistic Regression

The information contained in the dataset may be crucial in determining the most likely sources of jumps. In the spirit of Caporin and Poli (2017), we reconstruct the different portions of information assimilated by heterogeneous market players, which incorporate and react to news disclosure at differing speeds, and build a set of news-based indicators by taking into account three lead-and-lag time horizons, a set of concepts for news stories, and the surprises from expectations of EPS and macro-announcements.

Further, in order to avoid overfitting issues and to detect the most important news-based variables, that is, those carrying relevant information that we might associate with the causation of jumps, we apply the lasso of Tibshirani (1996), the elastic net of Zou and Hastie (2005) and the adaptive lasso of Zou (2006) in the estimation of a logistic regression. The latter links the probability of the occurrence of jumps to the huge number of variables at our disposal. In addition, we address the imbalanced sample (IS) problem, which is arising from the rare nature of jumps. We adopt cost-sensitive learning, oversampling, undersampling, and synthetic sampling. To measure the importance of the variables we use the Average Partial Effect (APE) on the jump probability and the average Area Under the ROC Curve

(AUC) decrease after variable permutation.¹⁰

By distinguishing possibly delayed, instantaneous, and even anticipated market reactions to the release of information, our methodology differs substantially from Section 3.2 and the references therein, where associations between news and jumps are defined in terms of the occurrence of jumps either within a short time after a news release or during the same day. Kannianen and Yue (2019), similarly to us albeit with a different approach, examine the pre- and post-news jump dynamics of stock prices, but they do not provide such a fine distinction between news types.

Lee (2012) too focuses on the impact of announcements and company-specific variables on the probability of jump occurrence at the intra-daily level. As we anticipated in the introduction, we deviate from Lee (2012) in several aspects, including the database coverage and the time frequency. From a methodological point of view, our approach is more flexible than the one of the previously cited paper. In fact, we do not perform a priori a selection of the potentially relevant news drivers, thus allowing for relevant differences across assets in the identification of the news impacting on the jump probability; the heterogeneity among assets will be later discussed. In addition, the introduction of AUC decrease and APE is leading to more informative analyses on the impact of news on jump probability, going beyond the statistical significance of the estimated parameters, and providing insights on the impact of those variables on the jumps occurrence probability (when using APE) and on their usefulness for anticipating jumps occurrence (with AUC). Finally, our collection of company-specific news is more detailed and informative, as it is not limited to a collection of dummy variables of news releases.

¹⁰They are particularly suitable for this study because they can be easily aggregated to summarize the high number of estimated parameters; additionally, there is a lack of methods evaluating the statistical significance of parameters in this field. An exception is Audrino and Camponovo (2018) for adaptive lasso, which in our study was not employed for reasons explained later on.

3.3.1 News Indicators

For each 5-min interval from t_0 to t_1 during which a jump may occur, news-based indicators are built by looking at the information released during three lead-and-lag time horizons: the preceding 30 minutes, which allow us to infer if news Granger-causes jumps, the same interval, to detect (nearly) instantaneous reactions to news, and the following 10 minutes. The leaking of information may indeed cause capital market participants to act in advance of news releases and, consequently, it becomes necessary to also consider the effect of lead responses by the market. Table 5 illustrates the time horizons.

We point out that, looking at the relation between jumps and news released during the same interval, we may erroneously infer a fast reaction to news when in fact we are in front of a market anticipation of information. Indeed, it may happen that a jump occurs at the beginning of an interval and news is released after the jump but still inside the same interval. However, the same information is likely to be released by more than one source (e.g., other news providers, social networking services, etc.) at slightly different times and, as a consequence, even conducting this analysis at a higher frequency would not solve this issue.

Table 5: Lead-and-lag time horizons for news indicators.

Lead-and-Lag	Start	End
-1	$t_0 - 30 \text{ min}$	t_0
0	t_0	t_1
+1	t_1	$t_1 + 10 \text{ min}$

Notes: Time horizons on which information is aggregated to build news-based indicators. t_0 and t_1 are the beginning and the end of each 5-min interval during which a jump may occur. Indicators based on these time horizons cannot be related to jumps occurring during the first 30 minutes and during the last 10 minutes of the trading day.

Following Caporin and Poli (2017), we go beyond the standard techniques used to assign numbers to textual information and identify a set of concepts/events about news stories. Each concept is peculiar in the reaction it potentially causes in the market and can refer to

one or more time horizons:

- **Standard:** news occurrence, sentiment.
- **Abnormal quantity:** quantity of news over a threshold. Investors' reaction could be triggered by the release of an unusual amount of information.
- **Uncertainty:** occurrence of news with opposite sentiment within the same interval. Some market players may react to the positive news and others to the negative one, and when some players realize that they were wrong, this could lead to market frictions that ultimately end in a jump.
- **News persistence:** release of news in each of two consecutive intervals. Since providers do not supply redundant news, this event denotes persistence in the release of news stories that are related in each period to a different issue. We consider the possibility that investors base their decisions on the news of the reference interval but their attention is raised by a news released during the previous interval.

With regard to earnings announcements, from actual figures and consensus forecasts we compute the *SUE* (Standardized Unexpected Earnings) score, which measures the number of standard deviations the reported EPS differs from the mean estimates. We build the *SUE* score of company i as follows:

$$SUE_{i,t} = \frac{EPS_{i,t}^{actual} - EPS_{i,t}^{forecast}}{\hat{\sigma}_{surp, EPS_i}} \quad (1)$$

where $\hat{\sigma}_{surp, EPS_i}$ is the standard deviation of $(EPS_{i,t}^{actual} - EPS_{i,t}^{forecast})$.

With regard to macro-announcements, from reported announcements and consensus forecasts we compute the standardized surprise as well, and we call it *Std_Macro*:

$$Std_Macro_{k,t} = \frac{Macro_{k,t}^{actual} - Macro_{k,t}^{forecast}}{\hat{\sigma}_{surp, k}} \quad (2)$$

where $\widehat{\sigma}_{surp,k}$ is the standard deviation of $(Macro_{k,t}^{actual} - Macro_{k,t}^{forecast})$ and k stands for the type of macro-indicator.¹¹

Supplementary Material E illustrates the 624 variables that result from combining the lead-and-lag time horizons, the concepts for news stories, the announcement dummies, and the surprises of EPS and macro-news. We point out that our variables include both the dummy for news releases as well as a collection of derived variables that, by construction, convey more information.

3.3.2 Elastic Net, Adaptive Lasso, and Imbalanced Sample

Given the binary variable y_t that is equal to 1 for jumps occurrence and 0 otherwise and the vector of news-based indicators x_t , the objective function for the elastic net (Zou and Hastie, 2005) applied to a logistic regression is:

$$(\widehat{\beta}_0, \widehat{\beta}) = \operatorname{argmin} - \left[\frac{1}{T} \sum_{t=1}^T y_t(\beta_0 + \beta'x_t) - \log(1 + e^{\beta_0 + \beta'x_t}) \right] + \lambda [(1 - \alpha)\|\beta\|_2^2/2 + \alpha\|\beta\|_1] \quad (3)$$

where β_0 is the intercept, β is the vector of coefficients associated with the p regressors, $\lambda \geq 0$ is a complexity parameter, and $0 \leq \alpha \leq 1$ is a compromise between ridge ($\alpha = 0$) and lasso ($\alpha = 1$). We choose the λ parameter among a grid of values by selecting the one that maximizes the AUC. As no consensus exists on how to set α on the basis of the data at hand, we experiment with a set of 11 equally spaced values that range from 0 to 1.

Additionally, we estimate the two-stage adaptive lasso, which was shown to reduce the number of false positives (Zou, 2006). It has the following objective function:

$$(\widehat{\beta}_0, \widehat{\beta}) = \operatorname{argmin} - \left[\frac{1}{T} \sum_{t=1}^T y_t(\beta_0 + \beta'x_t) - \log(1 + e^{\beta_0 + \beta'x_t}) \right] + \lambda \sum_{i=1}^p \left| \frac{\beta_i}{\widehat{\beta}_{i,\text{initial}}} \right| \quad (4)$$

¹¹For the three announcements ECRI Weekly, Employment Trends and New York NAPM Index, the consensus forecast is not available. For them, we compute the standardized change with respect to the previous release.

where $\widehat{\beta}_{i,\text{initial}}$ is the first-stage estimator of β_i from equation (3) with $\alpha = 1$ and λ is set equal to the value maximizing the AUC as for elastic net.¹²

We use several techniques that address the issue of IS in the context of binary classification. Price jumps are rare events, thereby rendering our sample extremely imbalanced; among the 199,134 intraday observations, the number of jumps ranges from 22 for *Schlumberger Limited* to 112 for *American International Group*, corresponding to the minimum and maximum ratios between the minority (jumps) and majority (no jumps) classes of 0.011% and 0.056%, respectively. We experiment with the following approaches: standard (no IS technique), cost-sensitive learning, oversampling, undersampling, and synthetic sampling. Moreover, we estimate the models using block cross-validation with blocks of contiguous time to better ensure independence between cross-validation folds and achieve more reliable error estimates (Burman et al., 1994; Racine, 2000; Bergmeier and Benitez, 2012). In the spirit of the *hv*-block of Racine (2000), we omit the first observations of each test set to reduce the dependence of the training and test sets. Finally, we eliminate the possible bounce-back effects by deleting any jump that follows by a maximum of half an hour a jump with an opposite sign. We describe the details and the settings we adopt for all these procedures in Supplementary Material G.

By combining each of the five IS techniques with the 11 values of α for elastic net and adaptive lasso, we obtain 60 models. For all the assets, we estimate each model separately and find the average AUC across folds. Table 6 illustrates the mean among all the assets of the average AUC for each model. In general, oversampling and cost-sensitive techniques perform better, i.e., these techniques allow for the highest AUC to be obtained. In the descending order of performance, the remaining IS techniques are synthetic sampling, undersampling, and standard sampling. With regard to α , the smaller its value, the higher the performance of the algorithm, except for the case $\alpha = 0$. The best combination of IS technique and

¹²As the first-stage estimate comes from a lasso, we might have β_i values equal to zero. Consequently, to avoid division by zero, we drop from the second stage the coefficients that resulted equal to zero in the first-stage. We also verified that results are unchanged if we add to the null β_i a small value, i.e., 10^{-6} .

α is oversampling with $\alpha = 0.1$. For practical purposes, the cost-sensitive technique may be preferable, considering its lower computational intensity and similar performance with respect to oversampling.

We repeat the analysis separately for positive and negative jumps (see Tables H1 and H2 in Supplementary Material H). These additional empirical evidences confirm the previous findings on the technique for addressing the class imbalance, i.e., oversampling. For the case of negative jumps only, increasing α to 0.3 allows attaining the highest performance, whereas for positive jumps we have a confirmation of $\alpha = 0.1$.

3.4 Importance of Variables

We estimate the model corresponding to the best combination of the α parameter and the IS treatment approach, i.e., oversampling technique and $\alpha = 0.1$. We apply these settings across the whole sample for each asset. In order to measure the importance of the indicators, we employ two metrics: the Average Partial Effect¹³ on the jump probability and the average AUC decrease after variable permutation. While the latter approach was already used in the financial context, for example, by Wang et al. (2015), in turn inspired by Janitzka et al. (2013), as far as we know, we are the first to employ the APE to measure the relevance of a variable with regard to the causation of jumps.

¹³According to the Logit model:

$$P(y = 1|x) = G(\beta_0 + \beta x), \quad \text{with} \quad G(z) = \exp(z)/[1 + \exp(z)]$$

The APE is differently computed depending on whether the regressor is continuous or binary:

- if x_j is a continuous variable, its partial effect on $p(x) = P(y = 1|x)$ is:

$$\frac{\delta p(x)}{\delta x_j} = g(\beta_0 + \beta x)\beta_j, \quad g(z) = \frac{dG}{dz}(z) \quad (5)$$

$$APE_{x_j} = [n^{-1} \sum_{i=1}^n g(\hat{\beta}_0 + \hat{\beta}x_i)]\hat{\beta}_j \quad (6)$$

- if x_k is a binary variable, the partial effect from changing x_k from 0 to 1 is:

$$G(\beta_0 + \beta_1 x_1 + \dots + \beta_{k-1} x_{k-1} + \beta_k) - G(\beta_0 + \beta_1 x_1 + \dots + \beta_{k-1} x_{k-1}) \quad (7)$$

$$APE_{x_k} = n^{-1} \sum_{i=1}^n \left\{ G[\hat{\beta}_0 + \dots + \hat{\beta}_{k-1} x_{i,k-1} + \hat{\beta}_k] - G[\hat{\beta}_0 + \dots + \hat{\beta}_{k-1} x_{i,k-1}] \right\} \quad (8)$$

Table 6: Average AUC for each combination of IS technique and α .

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5789	0.5428	0.5428	0.5912	0.5987
0.1	0.5943*	0.6192**	0.6192**	0.5974	0.6101*
0.2	0.5916	0.6164	0.6163	0.5975*	0.6090
0.3	0.5889	0.6151	0.6152	0.5972	0.6083
0.4	0.5883	0.6139	0.6140	0.5973	0.6073
0.5	0.5890	0.6134	0.6134	0.5974	0.6071
0.6	0.5769	0.6130	0.6131	0.5970	0.6070
0.7	0.5836	0.6116	0.6112	0.5974	0.6058
0.8	0.5502	0.6079	0.6081	0.5973	0.6048
0.9	0.5621	0.6044	0.6044	0.5970	0.6040
1.0	0.5281	0.5997	0.5997	0.5964	0.6038
Ad. Lasso	0.5321	0.6110	0.6103	0.5950	0.5982
Average	0.5720	0.6057	0.6056	0.5965	0.6053

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α for each IS technique; ** indicates the global maximum across IS techniques and α .

Table 7 shows the 20 most important indicators sorted in descending order by average AUC decrease, while Tables I1 and I2 in Supplementary Material I sort the indicators by APE and absolute APE, respectively. In the following, we focus on the indicators with positive APE since we are interested in those that increase the jump probability.

From Table 7 we see that FOMC rate decisions are undoubtedly the most important type of news: announcements at lag 0 (during the same interval of jumps) and at lag -1 (in the 30 minutes before the jump interval) are the first two most important indicators. The second most important type of news is StreetAccount news stories with topic *all*, that is, independently of the topic. The persistence of a news release at lag 0 and lead +1 (during both the jump interval and the following 10 minutes) and at lags -1 and 0 (during both the half hour before and the jump interval) are the third and fourth most important indicators; StreetAccount news stories with the topic *all* appear in the table also in the form of announcement at lag 0, announcement at lead +1, sentiment not neutral at lag 0, and negative sentiment at lag 0. Another macroeconomic news is present, as well, ISM Manufacturing PMI in the form of absolute surprise at lag 0. Looking at the average AUC decrease after variable permutation, however, we can notice that FOMC announcements and the persistence of the release of StreetAccount news stories are by far more relevant than the other indicators. Table I1 (with indicators sorted by APE) confirms the previous findings, while Table I2 (with indicators sorted by absolute APE) presents only FOMC announcements and the persistence of StreetAccount news with a positive APE, and adds a series of other macroeconomic indicators with negative APE. The latter evidence supports the matching analysis of Section 3.2 where macroeconomic announcements different from FOMC decisions are not associated with jumps.

Despite the matching analysis was showing that EPS announcements are characterized by a very high $P(J|N)$ (see Table 4 of Section 3.2), they do not appear here among the most important regressors, contrary to the evidences in Lee (2012), among others. On the one side, differences might relate to methodological tools as we introduce class imbalance treatment,

we are indeed not simply detecting the statistical significance of the EPS announcements but our analyses focus on the relevance of the EPS announcement role, and we use also surprises in EPS announcements. On the other side, $P(J|N)$ focuses on the probability that a jump appears after an EPS news release, but using as denominator the total number of EPS news. Differently, the logistic regressions we run consider all the jumps in our dataset (the denominator of $P(N|J)$), where only a small fraction, 0.22% (i.e., $P(N|J)$), is matched with EPS news. Moreover, the fact that EPS announcements do not appear in our results may be driven by the informative content of our database.¹⁴

Positive jumps only and negative jumps only. We then restrict attention to positive jumps only (see Tables I3, I4 and I5 in Supplementary Material I with the indicators sorted by AUC decrease, APE and absolute APE, respectively). The most important indicators of Table I3 are almost the same of Table 7. In addition to them, we note some other variables with positive APE: the absolute surprise and the announcement of ISM Manufacturing PMI at lag 0 and the persistence of Thomson Reuters news stories with topic *dividends* and with topic *financial* at lags -1 and 0, albeit with both low AUC decrease and low APE. Table I4 (indicators sorted by APE) is in line with Table 7 and contains some additional variables; the only notable one, with both positive APE and positive AUC decrease, is the announcement of StreetAccount news stories with topic *all* and positive sentiment at lag 0. Table I5 (indicators sorted by absolute APE) is in line with Table I2 (corresponding analysis without jump sign restriction).

We also focus on negative jumps only (see Tables I6, I7 and I8, with the indicators sorted in the three ways as above). The difference with the non-restricted analysis is more striking as we can notice that the persistence indicators are no more associated with a positive APE.

¹⁴We also performed a standard logistic regression for every stock and for every indicator using each time only the indicator and the intercept (using all the indicators in a single regression cannot be applied, as explained in the paper), and reported in Tables K1, K2 and K3 (for all jumps, positive jumps and negative jumps, respectively) in Supplementary Material K the first 20 indicators sorted in descending order by number of assets with a 5% significant two-tailed test for $\beta \neq 0$. EPS indicators never appear in the tables. We conclude that in our database the EPS announcements are not statistically significant.

Table 7: Most important indicators ordered by average AUC decrease.

Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2209	0.0302
2	Macro	FOMC Rate Dec.	-1	announcement	0.1693	0.0170
3	StreetAcc.	All	0 and +1	persistence	0.0548	0.0108
4	StreetAcc.	All	-1 and 0	persistence	0.0360	0.0031
5	Macro	Nat. Gas Stocks	-1	announcement	-0.0869	0.0028
6	Macro	IBD Ec. Opt.	-1	abs(surprise)	-0.0003	0.0026
7	Macro	Nat. Gas Stocks	+1	announcement	-0.0685	0.0024
8	StreetAcc.	All	0	announcement	0.0382	0.0021
9	Macro	Factory Orders	-1	announcement	-0.0433	0.0018
10	StreetAcc.	All	-1	announcement	-0.0079	0.0016
11	Macro	ECRI	-1	announcement	-0.1018	0.0015
12	StreetAcc.	All	+1	announcement	0.0244	0.0014
13	Macro	Oil Stocks	-1	abs(min(0,surprise))	-0.0019	0.0013
14	Macro	Cons. Credit	-1	abs(surprise)	-0.0013	0.0013
15	Macro	Federal Budget	-1	announcement	-0.0324	0.0012
16	Macro	ISM Man. PMI	0	abs(surprise)	0.0007	0.0012
17	Macro	New Home Sales	-1	announcement	-0.0452	0.0010
18	StreetAcc.	Earnings	-1 and 0	persistence	-0.0030	0.0008
19	StreetAcc.	All	0	sent. \neq 0	0.0201	0.0007
20	StreetAcc.	All	0	negative sent.	0.0074	0.0007

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Furthermore, the release of StreetAccount news stories with negative sentiment at lead +1 is one of the few indicators with both positive APE and positive AUC decrease.¹⁵

Sub-sample analysis. We repeat the analysis for four sub-samples: Jan 2005 - Jun 2007 (pre-global financial crisis), Jul 2007 - Jun 2009 (global financial crisis), Jul 2009 - Dec 2012 (EU sovereign debt crisis), and Jan 2013 - Feb 2015 (post EU sovereign debt crisis), see Supplementary Material I.4 for the tables summarizing our findings. When the analysis is conducted on sub-samples, penalization and the IS treatment are even more important. Penalization allows to diminish the sample length at will, without incurring into a dimensionality issue, i.e., a number of regressors higher than the number of observations, while the IS treatment enables to use sub-samples in which the observations consisting of jumps may be even more sporadic than for the full sample.

Isolating these time periods allows to infer several facts. First, during the global financial crisis the reaction to FOMC announcements was fast and stronger than during any other period, as witnessed by the APE of the FOMC announcement at lag 0 in Table I18 (APE is equal to 0.25 and higher than the APE of any other indicator in any other table). Second, during the EU sovereign debt crisis and post EU sovereign debt crisis, the AUC decrease of the FOMC-related indicators dramatically decreases, see Tables I27–I44. Third, further restraining by jump sign the analyses by sub-sample enables to distinguish other news related to positive jumps and negative jumps: with regard to positive jumps we observe the presence of $\max(0, \text{surprise})$ of ECRI at lag -1 in the pre-global financial crisis (Table I12), $\max(0, \text{surprise})$ of ISM Manufacturing PMI at lag 0 and the announcement of StreetAccount news stories with topic *all* and positive sentiment at lag 0 during the EU sovereign debt crisis (Table I30), $\max(0, \text{surprise})$ of Employment Trends at lag -1 and $\max(0, \text{surprise})$ of Federal

¹⁵We observe that Tables I4 (sorting by APE, positive jumps) and I7 (sorting by APE, negative jumps) include both indicators with positive sentiment and indicators with negative sentiment. On the one hand, this may be due to two reasons: a few cases of concurrence of jumps with news stories that have opposite sentiment with respect to the sign of the jump, and other cases in which a human-written text of a news is too complicated for an automated procedure to correctly extract its sentiment. On the other hand, we stress that, within our sample and results, indicators with opposite sentiment with respect to the sign of the jumps are not associated with a positive AUC.

Budget at lag -1 in the post EU sovereign debt crisis (Table I39). With regard to negative jumps we have the announcement of StreetAccount news stories with topic *all* and negative sentiment at lag 0 during the global financial crisis (Table I24).¹⁶

Comparing these sub-sample results with the corresponding ones obtained without IS treatment, in addition to what already stated in the full sample analysis, we see that, when appropriate IS techniques are not employed, a set of surprise-related macroeconomic indicators are discarded; indeed, the indicators related to positive jumps and negative jumps listed above do not appear in the corresponding tables without IS treatment (Tables J10–J45 in Supplementary Material J).

Finally, after examining all the previous results (full sample, positive versus negative jumps, and sub-samples), it is possible to affirm that, differently from company-specific news, macro-indicators never appear at lead +1,¹⁷ suggesting that market players act sometimes before the release of company-specific news and either they don't anticipate macro-announcements or, if they do, they don't give rise to jumps.

Effect of IS treatment. The results obtained without class imbalance treatment (see Supplementary Material J) show that the most important variables mainly correspond to those obtained with IS treatment and that indicators related to positive (negative) sentiment again help to explain positive (negative) jumps, but some important differences emerge. First, the tables obtained without IS treatment contain only indicators with non-negative APE (in this case the tables sorted by absolute APE correspond to the tables sorted by APE). More precisely, additionally to FOMC announcements and StreetAccount news stories which present

¹⁶There are a couple of counterintuitive indicators: $|\min(0, \text{surprise})|$ of Consumer Spending at lag 0 in Table I30 (EU sovereign debt crisis) seems to cause positive jumps and $\max(0, \text{surprise})$ of Federal Budget at lag -1 in Table I42 (post EU sovereign debt crisis) seems to cause negative jumps. We consider the possibility that a negative (positive) surprise of Consumer Spending (Federal Budget) may sometimes give rise to a negative (positive) return not in the form of jumps that is followed by a jump with opposite sign. We remind readers that bounce-back jumps are deleted if they follow a jump with opposite sign but not if they follow a return with opposite sign.

¹⁷The only exception is in Table I43 which is relative to negative jumps only in the sub-sample 2013 - 2015 and sorts the indicators by APE. We can see there $|\min(0, \text{surprise})|$ of Federal Budget at lead +1 with positive APE and AUC equal to zero.

both positive APE and positive AUC decrease, they contain either company-specific indicators with positive APE and AUC decrease close to zero (among the indicators with highest APE we often find the topic *earnings*, in line with Lee (2012)), or macroeconomic indicators with positive AUC decrease and APE equal to zero. In particular, comparing the sub-sample results with IS treatment (Tables I9–I44) with the corresponding results without IS treatment (Tables J10–J45), we note that in the latter tables the indicators built as $\max(0, Std_Macro_{k,t})$ and $|\min(0, Std_Macro_{k,t})|$ (where we remind that $Std_Macro_{k,t}$ is the standardized surprise of macroeconomic announcements) never appear with both positive APE and positive AUC. Second, the APE resulting from the coefficients estimated with the IS techniques is significantly higher compared to that resulting without applying the IS treatment, usually by at least ten times. Overall, this shows how excluding class imbalance might distort the identification of jump drivers. Moreover, the differences between the inclusion of class imbalance treatment techniques and their exclusion are not limited to the selection of jump drivers, but also dramatically distort the size of the news and macroeconomic variables impact.

Assets heterogeneity. We now focus on the heterogeneity of the assets with regard to the selection of indicators that provoke jumps. Starting from the jump predictors with a positive coefficient for each pair of assets, A and B , denoting their set by $\mathcal{P}(A)$ and $\mathcal{P}(B)$, and then focusing on the fraction of common predictors with positive sign over the composite set of predictors with positive sign, we suggest computing the following news-reaction similarity index:

$$\mathcal{F}(A, B) = \frac{|\mathcal{P}(A) \cap \mathcal{P}(B)|}{|\mathcal{P}(A) \cup \mathcal{P}(B)|}. \quad (9)$$

Consequently, we evaluate $\mathcal{F}(A, B)$ for each pair of assets and use it to draw a network where the nodes are the assets and the weight of the edge between each couple of assets A and B is the similarity index in (9). We sort assets by economic sector to help interpreting

the similarity across sectors. We first draw the network using all indicators, then using only the company-specific ones, and, finally, using only the macroeconomic ones; see Figure 3. Furthermore, Table 8 reports summary statistics of the similarity index by economic sector.

From Figure 3 and Table 8 Panel (A) we observe that the sector with the highest similarity is Industrial Goods (IG) (it has the highest median of the edges' weight). The standard deviation of the similarity index is lower than that of all the assets, indicating that the similarity is not due to a few cases but is, instead, a common characteristic of IG. The sector with the less similar assets (lowest median) is instead Technology (T). Focusing on the response to company-specific news, see Panel (B), Healthcare (H) is the sector with the highest similarity, whereas for macroeconomic news, see Panel (C), we see that, as in Panel (A), IG is the sector with the highest similarity. This suggests that the similarity of IG is mainly due to the common response of the assets to macroeconomic rather than to company-specific news. Comparing Panel (B) with Panel (C), we observe that the similarity across assets is much higher (almost double, according to our metrics) when it is based on the response to macroeconomic rather than to company-specific news, and this is true both for all the assets and for the single sectors.

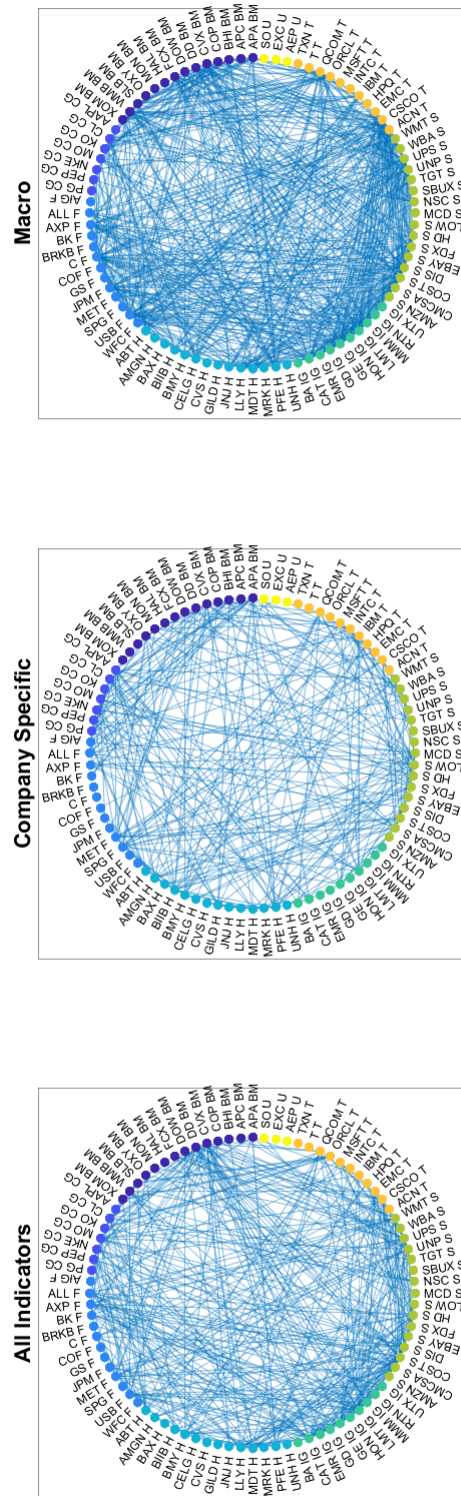
We now restrict the analysis to the response of positive jumps. Figure L1 and Table L1 Panels (A) and (C) in Supplementary Material L confirm the findings of Figure 3 and Table 8 with regard to IG. The sector with the smallest similarity (lowest median) becomes Healthcare (H). Consumer Goods (CG) is instead the most similar with regard to company-specific news, see Panel (B). With regard to the reaction of negative jumps, Figure L2 and Panels (A) and (C) of Table L2 show that the sector with the highest similarity is this time Financial (F) and that it is again mainly due to the common response of the assets to the macroeconomic indicators, while the sector with the highest heterogeneity is Basic Materials (BM). Finally, Panel (B) indicates that Services (S) is the sector whose assets react most similarly to company-specific news. Comparing the median weights of Tables L1 and L2, we see that when looking at the indicators that provoke positive jumps, the similarity is

much higher (more than doubled) compared to the negative jumps case. This might suggest that positive news has a more systematic impact while negative news tends to have a more heterogeneous impact.

Looking at the analyses by sub-sample (see Figures L3–L6 and Tables L3–L6) we notice that during the global financial crisis the similarity of the response to macroeconomic news was the highest across the four sub-samples. We attribute this result to an unusually high attention to FOMC announcements during that time, in line with the sub-sample results of the selected indicators discussed above.

Summary. Looking at the previous findings together, we infer several fundamental facts. First, FOMC announcements are the most important source of jumps. Second, company-specific news announcements provoke jumps as well and there is no dominant topic. In particular, the *persistence* indicator suggests that news associated with jumps is often followed by other news about the same company within a short time. We provide two possible interpretations: after important news is released, either news providers increase their level of attention toward the company and tend to release other news about it, or those market players able to get privileged information, they too, after increasing their level of attention toward the company, act in advance of the release to the public of other information about it. Other interpretations are possible, as well: for instance, companies can intentionally communicate information gradually over time (e.g., about management changes or new products). To the extent of our knowledge, this phenomenon we call “persistence” has not been highlighted before in the literature. Third, there is a tendency of the sign of the jumps to be related to the sentiment of the company-specific news stories and to the sign of the surprise of macroeconomic news with respect to expectations. Fourth, despite EPS announcements have a very high $P(J|N)$, they do not appear here among the most important regressors because only a small fraction of 0.22% is matched with EPS news, and in addition the market may largely anticipate their informative content. Fifth, in some cases the market

Figure 3: Assets Heterogeneity.



Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Table 8: Assets heterogeneity.

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.158	1.000	0.105
BM	14	0.000	0.111	1.000	0.130
CG	7	0.020	0.214	0.400	0.106
F	13	0.000	0.180	0.366	0.094
H	14	0.054	0.176	0.486	0.081
IG	10	0.164	0.258	0.536	0.084
S	16	0.048	0.169	0.488	0.100
T	11	0.000	0.070	0.355	0.096
(B) Company-specific					
All	88	0.000	0.102	1.000	0.116
BM	14	0.000	0.080	0.500	0.104
CG	7	0.000	0.136	0.444	0.140
F	13	0.000	0.078	0.310	0.092
H	14	0.000	0.176	0.524	0.118
IG	10	0.000	0.049	0.667	0.129
S	16	0.000	0.161	0.500	0.125
T	11	0.000	0.059	0.375	0.099
(C) Macro					
All	88	0.000	0.174	1.000	0.167
BM	14	0.000	0.111	1.000	0.250
CG	7	0.037	0.226	0.385	0.116
F	13	0.000	0.291	0.867	0.187
H	14	0.045	0.174	1.000	0.154
IG	10	0.267	0.389	0.727	0.113
S	16	0.083	0.182	0.750	0.148
T	11	0.000	0.097	0.588	0.173

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators. All stands for all the assets, BM for Basic Materials, CG for Consumer Goods, F for Financial, H for Healthcare, IG for Industrial Goods, S for services, and T for technology. We do not report the Utilities sector as it contains only three assets.

anticipates company-specific news releases, but the same either never happens with regard to macroeconomic news or, if it happens, it does not provoke jumps. Sixth, the employment of appropriate IS techniques, which allows the models to have a higher AUC, enables to attribute a higher and more correct APE to the indicators associated with jumps and to highlight a set of relevant macroeconomic indicators based on the surprise with respect

to expectations, that are useful especially to predict the sign of the jumps. Overall, this confirms the relevance of properly addressing the class imbalance issue. Seventh, EPS and Thomson Reuters news stories do not appear among the news most useful to explain jumps, supposedly because the employed regularization technique favors news occurring more frequently. Eighth, the main determinant of similarity of the assets within an economic sector consists of the assets common response to macroeconomic news, and the sectors that contain the most similar assets are Industrial Goods and Financial. Being macroeconomic news a non-diversifiable source of risk, it is therefore advisable to pay the highest attention to this driver of asset price dynamics.

With respect to Section 3.2 and the methods generally used by the literature, this methodology, which employs news-based indicators by taking into account three lead-and-lag time horizons, a set of concepts for news stories, and the surprises from expectations of EPS and macro-announcements, allows inferring in a much more detailed manner how news and jumps are related.

These findings differ from those of Boudt and Petitjean (2014), who document a marginal effect of firm-specific news always more than twice that of macro-news. Similarly to Bajgrowicz et al. (2016) with regard to what they call spurious jumps, we identify FOMC announcements as the major source of jumps, and contrast with them for what concerns firm-specific news releases, which in our study is instead a very critical source of jumps.

Given that, under this method, macro-news turns out to be a leading cause of jumps, and that macro-related jumps are likely to occur in many stocks simultaneously and propagate even to a diversified index, we might expect that macro-news related jumps are non-diversifiable, whereas jumps related to company-specific news are more likely to be diversifiable in the aggregate portfolio. This line of reasoning calls for further analyses on the importance of news in the causation of co-jumps in the cross-section of assets. In this respect, according to Merton (1976), who finds an absence of co-jumps, jump risk is diversifiable. Bollerslev et al. (2008) document that the index jumps less often than the single stocks,

therefore idiosyncratic jumps are diversifiable in the aggregate portfolio, and for Bajgrowicz et al. (2016) jumps are rare events, do not cluster in time and no co-jump affects all stocks simultaneously, therefore jump risk is diversifiable as well. Differently from the previous authors, and in line with our view, Caporin et al. (2017) associate with major financial news the simultaneous occurrence of non diversifiable co-jumps in several stocks. We believe the exploration of the impact of news and information spillovers on jumps and co-jumps deserves a more detailed and tailored analysis, which we leave to future research.¹⁸

4 Conclusions

Our work sheds additional light on the relation between the information contained in the news (going beyond the mere release time) and the occurrence of price jumps. By combining two data sources, including price jumps recorded at the 5-minute frequency and the news headlines and bodies, and using advanced econometric tools for addressing the data dimensionality and the rare nature of jumps, we provide interesting insights on the news-jump nexus.

From a matching analysis based on jump-news coincidences, it emerges that although the majority of jumps are not associated with the news at our disposal (possibly depending on a variety of reasons, such as market frictions or social media short messages), information potentially very useful to determine the causes of jumps is represented by FOMC rate decisions, quarterly earnings announcements, StreetAccount news stories, and Thomson Reuters news stories with the topics *top*, *earnings pre-announcements*, *financial*, and *dividends*. The relation of jumps with these information sources is heterogeneous, depending on both the observed frequency of news types as well as on the impact that the various news types have on the jump probability.

From a penalized logistic regression with oversampling we identify potential sources of

¹⁸About news spillovers and asset price dynamics, see, for example, Audrino and Tetereva (2019). They study the cross-industry influence of the news for a set of US and European stocks, estimating the graphical Granger causality of the news sentiments - excess return networks.

jumps and, at the same time, account for the rare nature of jumps. Addressing the sample imbalance issue allows us to attribute a higher and more appropriate APE to the indicators associated with jumps and to highlight the relevance of some macroeconomic predictors.

Our results suggest that indicators based on news with a low occurrence, such as EPS and Thomson Reuters news stories, are discarded by our regularized approach. The indicators most useful to explain jumps are those related to FOMC rate decisions and StreetAccount news stories, with the former holding a predominant role. When a company-specific news provokes a jump, it is often followed by other news about the same company. In addition, the sign of the jumps tends to be related to the sentiment of the company-specific news stories and to the sign of the surprise of macroeconomic news with respect to expectations. Market players are sometimes able to anticipate company-specific news releases, but the same either never happens with regard to macroeconomic news or, if it happens, it does not provoke jumps. Further, the main determinant of news-reaction similarity of the assets within an economic sector consists of the assets common response to macroeconomic news, and the sectors that contain the most similar assets are Industrial Goods and Financial. Being macroeconomic news a non-diversifiable source of risk, it is therefore advisable to devote it the highest attention.

Further research developments might build on our empirical evidences, in particular for further deepening our knowledge on the relation between news and systemic jumps and on the impact of information spillovers, as proxied by news, between assets, on the interactions between news, liquidity and jumps, and on the role of high-frequency traders on this mechanism.

References

- Alejo, R., García, V., Marqués, A., Sánchez, J., Antonio-Velázquez, J. Making accurate credit risk predictions with cost-sensitive mlp neural networks. In: Casillas, J. Martínez-López, F. Vicari, R. De la Prieta, F. (eds) *Management Intelligent Systems. Advances in Intelligent Systems and Computing*, **2013**, vol 220. Springer, Heidelberg, 1–8.
- Andersen, T.G., Bollerslev, T., Diebold, F.X., Vega, C. Real-time price discovery in global

- stock, bond and foreign exchange markets. *Journal of International Economics* **2007**, *73*, 251–277.
- Andersen, T.G., Bollerslev, T., Dobrev, D., No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: Theory and testable distributional implications. *Journal of Econometrics* **2007**, *138*, 125–180.
- Audrino, F., Camponovo, L. Oracle properties, bias correction, and bootstrap inference for adaptive lasso for time-series M -estimators. *Journal of Time Series Analysis*, **2018**, *39*, 111–128.
- Audrino, F., Ballinari, D., Sigrist, F. The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, **2019**.
- Audrino, F., Teterova, A. Sentiment spillover effects for US and European companies. *Journal of Banking and Finance*, **2019**, *106*, 542–567.
- Bajgrowicz, P., Scaillet, O., Treccani, A. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, **2016**, *62*, 2198–2217.
- Barndorff-Nielsen, O.E., Shephard, N. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* **2004**, *2*, 1–37.
- Barndorff-Nielsen, O.E., Shephard, N. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* **2006**, *4*, 1–30.
- Bergmeir, C., Benitez, J.M. On the use of cross-validation for time series predictor evaluation. *Information Sciences* **2012**, *191*, 192–213.
- Bollerslev, T., Law, T.H., Tauchen, G. Risk, jumps, and diversification. *Journal of Econometrics* **2008**, *144*, 234–256.
- Boudt, K., Croux, C., Laurent, S. Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, **2011**, *18*, 353–367.
- Boudt, K., Petitjean, M. Intraday liquidity dynamics and news releases around price jumps: Evidence from the DJIA stocks. *Journal of Financial Markets*, **2014**, *17*, 121–149.
- Bradley, D., Clarke, J., Lee, S., Ornathanalai, C. Are analysts recommendations informative? Intraday evidence on the impact of time stamp delays. *Journal of Finance*, **2014**, *69*(2), 645–673.
- Burman, P., Chow, E., Nolan, D. A cross-validatory method for dependent data. *Biometrika*, **1994**, *81*, 351–358.
- Caporin, M., Kolokolov, A., Renò, R. Systemic co-jumps. *Journal of Financial Economics*, **2017**, *126*(3), 563–591.
- Caporin, M., Poli, F. Building news measures from textual data and an application to volatility forecasting. *Econometrics*, **2017**, *5*, 1–46.

- Chawla, N.V., Bowyer, K., Hall, L., Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, **2002**, *16*, 321–357.
- Chawla, N.V., Cieslak, D.A., Hall, L.O., Joshi, A. Automatically countering imbalance and its empirical relationship to cost. *Data Mining and Knowledge Discovery*, **2008**, *17*, 225–252.
- Christensen, K., Oomen, R.C.A., Podolskij, M. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, **2014**, *114*(3), 576–599.
- Corsi, F., Pirino, D., Renò, R. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* **2010**, *159*, 276–288.
- Domingos, P. MetaCost: A general method for making classifiers cost-sensitive. *Proceedings of the 5th ACM SIGKDD International Conference of Knowledge Discovery and Data Mining, SanDiego, CA, USA*, **1999**, 155–164.
- Dungey, M., McKenzie, M., Smith, L.V. Empirical evidence on jumps in the term structure of the US Treasury market. *Journal of Empirical Finance*, **2009**, *16*, 430–445.
- Evans, K.P. Intraday jumps and US macroeconomic news announcements. *Journal of Banking and Finance* **2011**, *35*, 2511–2527.
- Friedman, J., Hastie, T., Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **2010**, *33*, 1–22.
- Gloß-Klußmann, A., Hautsch, N. When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, **2011**, *18*, 321–340.
- Huang, X. Macroeconomic news announcements, systemic risk, financial market volatility and jumps. *Economics Discussion Series 2015-097*. Washington: Board of Governors of the Federal Reserve System, **2015**.
- Janitza, S., Strobl, C., Boulesteix, A.L. An auc-based permutation variable importance measure for random forests. *BMC bioinformatics*, **2013**, *14*: 119.
- Japkowicz, N., Stephen, S. The class imbalance problem: A systematic study. *Intelligent Data Analysis Journal*, **2002**, *6*, 429–450.
- Jiang, G.J., Lo, I., Verdelhan, A. Information shocks, liquidity shocks, jumps, and price discovery: Evidence from the U.S. Treasury market. *Journal of Financial and Quantitative Analysis*, **2011**, *46*, 527–551.
- Kanniainen, J., Yue, Y. The Arrival of news and return jumps in stock markets: A nonparametric approach. *arXiv, Quantitative Finance*, **2019**
- Kotsiantis, S., Kanellopoulos, D., Pintelas, P. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, **2006**, *30*, 25–36.

- Kuhn, M. Caret package. *Journal of Statistical Software* **2008**, *28*(5).
- Lahaye, J., Laurent, S., Neely, C.J. Jumps, cojumps and macro announcements. *Journal of Applied Econometrics* **2011**, *26*, 893–921.
- Lee, S.S. Jumps and information flow in financial markets. *Review of Financial Studies* **2012**, *25*, 439–479.
- Lee, S.S., Mykland, P.A. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* **2008**, *21*, 2535–2563.
- López, V., Fernández, A., Moreno-Torres, J.G., Herrera, F. Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *Expert Systems with Applications*, **2012**, *7*, 6585–6608.
- Loughran, T., McDonald, B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* **2011**, *66*, 35–65.
- Maheu, J.M., McCurdy, T.H. News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance* **2004**, *59*, 755–793.
- Merton, R.C. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, **1976**, *3*(1-20), 125–144.
- Pavlou, M., Ambler, G., Seaman, S., De Iorio, M. Omar, R.Z. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine* **2016**, *35*, 1159–1177.
- Quinlan, J. C4.5: Programs for machine learning. *Morgan Kauffman*, **1993**.
- Racine, J. Consistent cross-validated model-selection for dependent data: Hv-block cross-validation. *Journal of Econometrics*, **2000**, *99*, 39–61.
- Rangel, J.G. Macroeconomic news, announcements, and stock market jump intensity dynamics. *Journal of Banking and Finance* **2011**, *35*, 1263–1276.
- Tibshirani, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society series B* **1996**, *58*, 267–288.
- Torgo, L. Data Mining with R, learning with case studies. **2010**, Publisher: Chapman and Hall/CRC.
- Wang, H., Xu, Q., Zhou, L. Large unbalanced credit scoring using Lasso-logistic regression ensemble. *PLoS ONE*, **2015**, *10*, 1–20.
- Weiss, G.M. Provost, F. Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research*, **2003**, *19*, 315–354.
- Weiss, G.M. Mining with rarity: A unifying framework. *ACM SIGKDD Explorations Newsletter*, **2004**, *6*, 7–19.

Zou, H., Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, **2005**, *67*, 301–320.

Zou, H. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, **2006**, *101*(476), 1418–1429.

Supplementary Material

News and Intraday Jumps: Variable Selection with Regularization and Class Imbalance

A Literature Review	2
A.1 News as Determinants of Jumps	2
A.2 Penalized Logistic Regression and Class Imbalance	4
B Assets List	6
C Intraday Jumps Estimation	9
D Sentiment Detection: Words Lists, Procedure Details, and Algorithms	13
E News-Based Indicators	30
F Additional Tables: Matching analysis	32
G Penalized Logistic Regression and Imbalanced Sample	36
H Additional Tables: Penalized Logistic Regression	38
I Additional Tables: Variables Importance	53
I.1 All jumps, sorting by APE and absolute APE	53
I.2 Positive Jumps	55
I.3 Negative Jumps	58
I.4 Sub-Sample Analysis	61
J Additional Tables: Variables Importance, No Imbalanced Sample Treatment	97
J.1 All Jumps	97
J.2 Positive Jumps	100
J.3 Negative Jumps	103
J.4 Sub-Sample Analysis	106
K Additional Tables: Standard Logistic Regression	142
L Additional Figures and Tables: Assets Heterogeneity	145

A Literature Review

A.1 News as Determinants of Jumps

Recent studies show that stochastic volatility models that include a discontinuous jump term in addition to a diffusive component capture in a better way the empirical characteristics of asset price returns. Consequently, the literature began to study the link of jumps with news and to monitor and measure how prices respond to public news announcements. Many studies focus on the market reaction to macroeconomic news. Andersen et al. (2007a), though not focusing on jumps, characterize the high-frequency response of US, German and British stock, bond and foreign exchange (FX) markets to real-time US macroeconomic news. Dungey et al. (2009) consider simultaneous jumps across the term structure of the US Treasury market and relate them to macroeconomic announcements. Rangel (2011) examines the effects of macroeconomic releases on stock market volatility through a Poisson-Gaussian-GARCH process developed by Maheu and McCurdy (2004). He finds evidence of the relevance of macroeconomic variables to explain jump dynamics. Lahaye et al. (2011) identify jumps and co-jumps from stock index futures, US bond futures and exchange rates, and relate them to US macroeconomic releases. Evans (2011) investigates the association of US macroeconomic announcements with intraday jumps in US equity, bond and FX markets, and finds that approximately one-third of jumps correspond to news. Furthermore, he shows that news-related jumps are larger, in average absolute terms, than jumps not related to news. Huang (2015) separates US equity and bond market responses into continuous volatility effects and jumps, and finds a larger proportion of days with jumps within macroeconomic announcement days. Caporin et al. (2017) develop a test for multivariate jumps (multi-jumps) and detect them in a panel of US stocks. They interpret multi-jumps as systemic events affecting the market as a whole and associate these rare but statistically and economically important events with relevant market-wide financial, political, and (mainly) economic news. Macroeconomic announcements are indeed the most likely sort of news to cause simultaneous jumps among many stocks.

A smaller number of studies deal with the impact of both macroeconomic and firm-specific news to price jumps in single stocks. Gloß-Klußmann and Hautsch (2011) analyze the extent to which high-frequency movements in returns, volatility and liquidity of 39 stocks traded at the London Stock Exchange can be explained by the underlying unscheduled news arrivals. They employ the trading signals from the *Reuters NewsScope Sentiment Engine*¹⁹ and find that high-frequency trading activity significantly reacts to news items that are identified as relevant. In addition, their results show that the strongest effects are in terms of volatility and cumulative trading volumes. Bollerslev et al. (2008) examine the relationship between jumps in individual stocks and jumps in an aggregate market index. They show that firm-specific news events are the dominant cause in terms of immediate price impact at the individual stock level, and find a strong tendency for the stocks to move sharply together, that is, co-jump, around 10 a.m. Eastern time, corresponding to the regularly scheduled release-time for many macroeconomic announcements. Lee and Mykland (2008) find that individual

¹⁹This service automatically classifies firm-specific news according to positive, neutral and negative author sentiments based on linguistic pattern recognition techniques and provides numeric indicators classifying the relevance of news as well as their novelty.

stock jumps are associated with prescheduled earnings announcements and other company-specific news events, while S&P 500 jumps are instead associated with general market news announcements. However, they use a small sample of three stocks over only three months. Lee (2012) relates jumps of the 30 DJIA components from 1993 to 2008 to macroeconomic and firm-specific releases, and shows that macroeconomic predictors tend to play a more important role in pinning down intraday jump dynamics than do firm-specific predictors. However, the paper does not consider information from news agencies and does not account for sample imbalance, intra-daily periodicity, news sentiment, and jumps sign. Bradley et al. (2014) examine how analyst recommendation releases are related to detected jumps in stocks. Bajgrowicz et al. (2016) relate jumps of the 30 DJIA stocks between 2006 and 2008 to macroeconomic news, prescheduled company-specific announcements and stories from news agencies that include a variety of unscheduled and uncategorized events, and consider several news types. They argue that bursts of volatility are often incorrectly identified as jumps when the sampling is too sparse, and claim that the majority of news does not cause jumps but may generate bursts of volatility. The latter are generated especially by press releases following scheduled FOMC meetings and announcements related to share repurchase programs. For them, the main reason for the small impact of news is that managers strategically shift important announcements outside market hours, and they suggest that liquidity pressures are probably an important factor of jumps. Kanniainen and Yue (2019) examine how news events such as company and macro-announcements contribute to the (pre- and post-event) jump dynamics of stock prices. They categorize firm-specific news into two groups: scheduled and non-scheduled announcements.

Previous studies support another explanation for price discontinuities, claiming that they can be the result of a local lack of liquidity on the market. Jiang et al. (2011) examine the relative importance of macroeconomic news announcements versus variation in market liquidity in explaining jumps in the US Treasury market. They find that jumps occur mostly at prescheduled macro-announcement times and announcement surprises have limited power in explaining jumps. Pre-announcement liquidity shocks have significant predictive power for jumps, even after controlling for information shocks. Boudt and Petitjean (2014) study the dynamics of liquidity and news releases around jumps for the 30 DJIA constituents between 2007 and 2009. They retrieve all macroeconomic news announcements, prescheduled or not, and all firm-specific news provided by the Dow Jones and Reuters News Service, and match one-third of the jumps with macroeconomic news, 5% with firm-specific news, and more than 50% with liquidity variations. Firm-specific news events have a higher effect on jump magnitude with respect to macroeconomic news and are identified as the dominant factor in terms of impact on the occurrence of jumps at the individual stock level. Jumps are mostly driven by variations in the demand for immediacy (trading volume), which is amplified by news and by the market’s inability to absorb them without moving the price significantly. However, they focus on the dynamics of liquidity around jumps rather than on the effect of news, as we do; in addition, they do not distinguish among news types and aggregate information into the two categories of firm-specific and macroeconomic news.²⁰

We study how both macroeconomic and company-specific news, prescheduled and un-

²⁰We leave to further research the investigation of the mechanism through which news, liquidity and their interaction may originate jumps.

scheduled, and classified into several topics, is linked to price jumps for single stocks. Differently from the literature surveyed above, we build news-based indicators that reconstruct the different portions of information assimilated by heterogeneous market players, and employ a machine-learning approach that allows us to distinguish the kind of news most likely associated with jumps and the intraday timing of this relationship in terms of possibly delayed, instantaneous, and even anticipated market reactions to information.

A.2 Penalized Logistic Regression and Class Imbalance

As mentioned in the introduction, we build several hundred predictors by considering various time horizons, a series of concepts for news stories, and the standardized surprises of EPS and macro-announcements. In order to identify within a large collection of predictors the most important ones in explaining jumps, we employ elastic net and adaptive lasso methods within a penalized logistic regression model. Friedman et al. (2010) point out that logistic regression is often plagued with degeneracies when the number of covariates is greater than the number of observations, and exhibits wild behavior even when they are close to each other. Elastic net alleviates these issues and regularizes and selects variables as well. In our empirical analyses, we will deal with a number of observations that is much higher than the number of covariates, but the part of the sample belonging to the minority class (jumps) is very small, and we consider the possibility that penalized methods could be beneficial.²¹ Summarizing, by introducing penalization methods, on the one side we reduce the risks of overfitting, while on the other side we allow for a more robust identification of the variables that might impact on the probability of price jump occurrence.

A further issue arises in our study, the so-called class imbalance problem. This is a consequence of the rare nature of jumps, an event that hinders the performance of standard classification systems, which usually assume a relatively well-balanced class distribution and equal misclassification costs (Japkowicz and Stephen, 2002). This issue occurs when one class vastly outnumbers the other one, which is usually the most important and has the highest misclassification costs (Chawla et al., 2008). Performing analyses without a proper treatment of the class imbalance might lead to an improper identification of the variables impacting on the jump probability. In turn, this negatively impacts on the economic interpretation of the empirical findings. Penalization and the imbalanced sample treatment are also fundamental to conduct the analysis separately for each stock, therefore preserving the heterogeneity of the results, and by sub-sample, where the diminished sample length is more likely to give rise to a dimensionality issue and the observations consisting of jumps can be even more sporadic than for the full sample case.

In the class imbalance framework, random over- and undersampling methodologies have

²¹Pavlou et al. (2016) review and evaluate the predictive performance of the main penalized regression methods using real and simulated data, focusing on models with low-dimensional data, binary outcomes, and few events. Their simulation study shows that maximum likelihood estimation tends to produce overfitted models with poor predictive performance in scenarios with few events, and penalized methods can offer relevant improvements. They also show that ridge regression performs well, except in scenarios with many noisy predictors, while lasso performs better than ridge in scenarios with many noisy predictors, but is worse in the presence of correlated predictors. They finally show that elastic net performs well in all scenarios. However, the analyses therein are not directly applicable to this study and we experiment a range of settings.

received particular attention (Weiss and Provost, 2003). In undersampling, instances of the minority and majority classes are selected randomly to achieve a balanced sample with equal class distributions, often using all instances of the minority class and only a subset of the majority class. Alternatively, in oversampling, the cases of the underrepresented class are replicated a number of times, so that the class distributions are more equal. Under- and oversampling generally lead to models with an enhanced discriminatory power, but both methods have their shortcomings: random undersampling can discard potentially important cases from the majority class of the sample, thus impairing an algorithm’s ability to learn the decision boundary, while random oversampling duplicates records and can lead to the overfitting of similar instances. Therefore, undersampling tends to overestimate the probability of cases belonging to the minority class, while oversampling tends to underestimate the likelihood of observations belonging to the minority class (Weiss, 2004).

Creating synthetic examples in oversampling (SMOTE, see Chawla et al., 2002) is a more sophisticated approach. The basic idea of SMOTE is that new minority class samples are synthesized through linear interpolation between two minority class samples randomly chosen from the k nearest neighbors, then they are added to the original dataset. SMOTE can generate samples the interpretation of which is difficult, and we provide more details about this in the paper.

Finally, cost-sensitive learning incorporates costs into classifiers (e.g., Domingos, 1999) and is effective in imbalanced classification tasks (López et al., 2012). In particular, it assigns a higher weight for minority class in order to make classifiers pay more attention to minority class samples. The C4.5 algorithm (Quinlan, 1993) is a typical example of this approach. However, cost-sensitive learning usually assumes that the misclassification costs of each class are known beforehand and it requires special knowledge of the classifiers themselves (Alejo et al., 2013); unfortunately, these two conditions are not easily satisfied in practice.

We will apply elastic net and adaptive lasso when estimating a logistic regression linking the probability of the occurrence of intraday jumps to the predictors, and we will address the class imbalance problem adopting the following machine-learning techniques: cost-sensitive learning, oversampling, undersampling, and synthetic sampling.²²

²²For additional details on the various techniques for the treatment of class imbalance, we refer the reader to Weiss (2004) and Kotsiantis et al. (2006).

B Assets List

Table A1: Assets list with exchange, ticker symbol, company name, and sector.

Exchange	Ticker	Name	Sector
NASDAQ	AAPL	Apple	Consumer Goods
NYSE	ABT	Abbott Laboratories	Healthcare
NYSE	ACN	Accenture plc	Technology
NYSE	AEP	American Electric Power Co., Inc.	Utilities
NYSE	AIG	American International Group, Inc.	Financial
NYSE	ALL	The Allstate Corporation	Financial
NASDAQ	AMGN	Amgen Inc.	Healthcare
NASDAQ	AMZN	Amazon.com, Inc.	Services
NYSE	APA	Apache Corp.	Basic Materials
NYSE	APC	Anadarko Petroleum Corporation	Basic Materials
NYSE	AXP	American Express Company	Financial
NYSE	BA	The Boeing Company	Industrial Goods
NYSE	BAX	Baxter International Inc.	Healthcare
NYSE	BHI	Baker Hughes Incorporated	Basic Materials
NASDAQ	BIIB	Biogen Inc.	Healthcare
NYSE	BK	The Bank of New York Mellon Corporation	Financial
NYSE	BMJ	Bristol-Myers Squibb Company	Healthcare
NYSE	BRK.B	Berkshire Hathaway Inc.	Financial
NYSE	C	Citigroup Inc.	Financial
NYSE	CAT	Caterpillar Inc.	Industrial Goods
NASDAQ	CELG	Celgene Corporation	Healthcare
NYSE	CL	Colgate-Palmolive Co.	Consumer Goods
NASDAQ	CMCSA	Comcast Corporation	Services
NYSE	COF	Capital One Financial Corporation	Financial
NYSE	COP	ConocoPhillips	Basic Materials
NASDAQ	COST	Costco Wholesale Corporation	Services
NASDAQ	CSCO	Cisco Systems, Inc.	Technology
NYSE	CVS	CVS Health Corporation	Healthcare
NYSE	CVX	Chevron Corporation	Basic Materials
NYSE	DD	E. I. du Pont de Nemours and Company	Basic Materials
NYSE	DIS	The Walt Disney Company	Services
NYSE	DOW	The Dow Chemical Company	Basic Materials
NASDAQ	EBAY	eBay Inc.	Services
NYSE	EMC	EMC Corporation	Technology
NYSE	EMR	Emerson Electric Co.	Industrial Goods
NYSE	EXC	Exelon Corporation	Utilities
NYSE	FCX	Freeport-McMoRan Inc.	Basic Materials
NYSE	FDX	FedEx Corporation	Services
NYSE	GD	General Dynamics Corporation	Industrial Goods

NYSE	GE	General Electric Company	Industrial Goods
NASDAQ	GILD	Gilead Sciences Inc.	Healthcare
NYSE	GS	The Goldman Sachs Group, Inc.	Financial
NYSE	HAL	Halliburton Company	Basic Materials
NYSE	HD	The Home Depot, Inc.	Services
NYSE	HON	Honeywell International Inc.	Industrial Goods
NYSE	HPQ	HP Inc.	Technology
NYSE	IBM	International Business Machines Corporation	Technology
NASDAQ	INTC	Intel Corporation	Technology
NYSE	JNJ	Johnson & Johnson	Healthcare
NYSE	JPM	JPMorgan Chase & Co.	Financial
NYSE	KO	The Coca-Cola Company	Consumer Goods
NYSE	LLY	Eli Lilly and Company	Healthcare
NYSE	LMT	Lockheed Martin Corporation	Industrial Goods
NYSE	LOW	Lowe's Companies, Inc.	Services
NYSE	MCD	McDonald's Corp.	Services
NYSE	MDT	Medtronic plc	Healthcare
NYSE	MET	MetLife, Inc.	Financial
NYSE	MMM	3M Company	Industrial Goods
NYSE	MO	Altria Group, Inc.	Consumer Goods
NYSE	MON	Monsanto Company	Basic Materials
NYSE	MRK	Merck & Co. Inc.	Healthcare
NASDAQ	MSFT	Microsoft Corporation	Technology
NYSE	NKE	NIKE, Inc.	Consumer Goods
NYSE	NSC	Norfolk Southern Corporation	Services
NASDAQ	ORCL	Oracle Corporation	Technology
NYSE	OXY	Occidental Petroleum Corporation	Basic Materials
NYSE	PEP	Pepsico, Inc.	Consumer Goods
NYSE	PFE	Pfizer Inc.	Healthcare
NYSE	PG	The Procter & Gamble Company	Consumer Goods
NASDAQ	QCOM	QUALCOMM Incorporated	Technology
NYSE	RTN	Raytheon Company	Industrial Goods
NASDAQ	SBUX	Starbucks Corporation	Services
NYSE	SLB	Schlumberger Limited	Basic Materials
NYSE	SO	Southern Company	Utilities
NYSE	SPG	Simon Property Group Inc.	Financial
NYSE	T	AT&T, Inc.	Technology
NYSE	TGT	Target Corp.	Services
NASDAQ	TXN	Texas Instruments Inc.	Technology
NYSE	UNH	UnitedHealth Group Incorporated	Healthcare
NYSE	UNP	Union Pacific Corporation	Services
NYSE	UPS	United Parcel Service, Inc.	Services
NYSE	USB	U.S. Bancorp	Financial
NYSE	UTX	United Technologies Corporation	Industrial Goods
NASDAQ	WBA	Walgreens Boots Alliance, Inc.	Services

NYSE	WFC	Wells Fargo & Company	Financial
NYSE	WMB	Williams Companies, Inc.	Basic Materials
NYSE	WMT	Wal-Mart Stores Inc.	Services
NYSE	XOM	Exxon Mobil Corporation	Basic Materials

C Intraday Jumps Estimation

Diffusive stochastic volatility models have problems in explaining behavior of asset prices, especially during market crashes and in general during turbulent periods, since they would require sometimes a volatility level too high for their formulation. As a solution, the total daily return variability has been decomposed into its continuous and discontinuous components based on the bipower variation measures developed by Barndorff-Nielsen and Shephard (2004, 2006). The empirical results in Andersen et al. (2007b) suggest that most of the predictable variation in the volatility stems from the strong own dynamic dependencies in the continuous price path variability, while the predictability of jumps is typically minor.

After filtering the periodic component of intraday volatility through the technique of Boudt et al. (2011), we rely on the method of Andersen et al. (2007b) to identify the precise intraday intervals at which jumps occur, and apply the following modification: when requested by the procedure, instead of the realized bipower variation we use the corrected threshold bipower variation of Corsi et al. (2010), that was shown to be more accurate for the estimation of jumps.

We assume that the scalar logarithmic asset price follows a standard jump-diffusion process:

$$dX_t = \mu_t dt + \sigma_t dW_t + dJ_t \quad (10)$$

where μ_t is predictable, σ_t is cadlag, $dJ_t = c_t dN_t$ where N_t is a non-explosive Poisson process whose intensity is an adapted stochastic process λ_t , the times of the jumps are $(\tau_j)_{j=1, \dots, N_t}$ and c_j are i.i.d. adapted random variables measuring the size, which is always positive, of the jump at time τ_j .

Quadratic variation of the process over a time window T is defined as:

$$[X]_t^{t+T} = X_{[t+T]}^2 - X_t^2 - 2 \int_t^{t+T} X_{s-} dX_s \quad (11)$$

where t indexes the day. It can be decomposed into its continuous and discontinuous component:

$$[X]_t^{t+T} = [X^c]_t^{t+T} + [X^d]_t^{t+T} \quad (12)$$

where $[X^c]_t^{t+T} = \int_t^{t+T} \sigma_s^2 ds$ and $[X^d]_t^{t+T} = \sum_{j=N_t}^{N_{t+T}} c_j^2$. To estimate these quantities, the time interval $[t, t+T]$ is divided into n subintervals of length δ and the evenly sampled returns are defined as:

$$\Delta_{j,t} X = X_{j\delta+t} - X_{(j-1)\delta+t}, \quad j = 1, \dots, n \quad (13)$$

The quadratic variation process and its separate components are, of course, not directly observable. Instead, we resort to popular model-free non-parametric consistent measures, including the familiar realized variance:

$$RV_{\Delta}(X)_t = \sum_{j=1}^n (\Delta_j X)^2 \quad (14)$$

which converges in probability to $[X]_t^{t+T}$ as $\delta \rightarrow 0$.

The theory discussed above hinges on the notion of increasingly finer sampled high-

frequency returns but, in practice, the sampling frequency is limited by the actual quotation or transaction frequency and the observed prices are contaminated by market microstructure frictions, including price discreteness and bid-ask spreads, which render the assumption of a semimartingale price process invalid at the tick-by-tick level. In response to this, we follow a relevant strand of the literature and compute our daily realized variance and jump measures from five-minute returns, using the nearest preceding or concurrent price to each five-minute mark.

In order to separately measure the jump part, we rely on the corrected threshold bipower variation (*C-TBPV*) measure, a version of the corrected threshold multipower variation (*C-TMPV*) developed by Corsi et al. (2010), which consists in turn in a modification of the realized bipower variation (*BPV*) of Barndorff-Nielsen and Shephard (2004, 2006):

$$\begin{aligned} C-TBPV_{\Delta}(X)_t &= \mu_1^{-2} C-TMPV_{\Delta}(X)_t^{1,1} \\ &= \mu_1^{-2} \sum_{j=2}^{[1/\delta]} Z_1(\Delta X_j, \vartheta_j) Z_1(\Delta X_{j-1}, \vartheta_{j-1}) \end{aligned} \quad (15)$$

where $\mu_{\alpha} = E(|Z|^{\alpha})$ for $Z \sim N(0, 1)$.

The corrected threshold multipower variation is defined as:

$$C-TMPV_{\Delta}(X)_t^{[\gamma_1, \dots, \gamma_M]} = \delta^{1 - \frac{1}{2}(\gamma_1 + \dots + \gamma_M)} \sum_{j=M}^{[1/\delta]} \prod_{k=1}^M Z_{\gamma_k}(\Delta_{j-k+1} X, \vartheta_{j-k+1}) \quad (16)$$

and the function $Z_{\gamma}(x, y)$ is:

$$Z_{\gamma}(x, y) = \begin{cases} |x|^{\gamma} & \text{if } x^2 \leq y \\ \frac{1}{2N(-c_{\vartheta})\sqrt{\pi}} \left(\frac{2}{c_{\vartheta}^2} y\right)^{\frac{\gamma}{2}} \Gamma\left(\frac{\gamma+1}{2}, \frac{c_{\vartheta}^2}{2}\right) & \text{if } x^2 > y \end{cases} \quad (17)$$

where $N(x)$ is the standard normal cumulative function, $\Gamma(\alpha, x)$ is the upper incomplete gamma function, $\vartheta = c_{\vartheta}^2 \sigma^2$ and σ^2 is the variance of $\Delta_j X$ under the assumption that $\Delta_j X \sim N(0, \sigma^2)$. Following Corsi et al. (2010), we set $c_{\vartheta} = 3$.

As $\delta \rightarrow 0$, C-TBPV converges to $\int_t^{t+T} \sigma^2(s) ds$

The difference between the realized variance and the corrected threshold bipower variation consistently estimates the part of the quadratic variation due to jumps:

$$RV_{\Delta}(X)_T - C-TBPV_{\Delta}(X)_T \xrightarrow[\delta \rightarrow 0]{P} [X^d]_t^{t+T} \quad (18)$$

As $\delta \rightarrow 0$, the test statistic

$$C-T_Z = \delta^{\frac{1}{2}} \frac{(RV_{\Delta}(X)_T - C-TBPV_{\Delta}(X)_T) \cdot RV_{\Delta}(X)_T^{-1}}{\sqrt{(\frac{\pi^2}{4} + \pi - 5) \max\left(1, \frac{C-TTriPV_{\Delta}(X)_T}{(C-TBPV_{\Delta}(X)_T)^2}\right)}} \quad (19)$$

where C-TTriPV $_{\Delta}(X)_T$ is a quarticity estimator, see Corsi et al. (2010), and is asymptotically standard normally distributed under the null hypothesis of no jumps.

Based on the above jump detection test statistic, the realized measure of the jump contribution to the quadratic variation of the logarithmic price process is then measured by:

$$\widehat{J}_t = I_{(C-T_Z > \Phi_\alpha)} \cdot (RV_t - BPV_t)^+ \quad (20)$$

where $I_{(\cdot)}$ denotes the indicator function and Φ_α refers to the appropriate critical value from the standard normal distribution.

Consequently, the realized measure for the integrated variance is:

$$\widehat{C}_t = RV_t - \widehat{J}_t \quad (21)$$

The method described above is useful to isolate days containing at least one jump. We want, instead, to identify the precise intraday intervals at which jumps occur, and rely on the procedure of Andersen et al. (2007b).

They define a randomly selected intraday return as $\Delta_\xi X = \sum_{j=1}^{1/\delta} \Delta_j X \cdot I(\xi = j)$, where ξ is an independently drawn index, uniformly distributed, from the set $\{1, 2, \dots, 1/\delta\}$. It is identified as jump if its absolute value is higher than an appropriately scaled realized bipower variation. Assuming that intraday scaled returns are distributed as $\delta^{-1/2} \cdot \Delta_\xi X \sim N(0, IV_t)$, where $IV_t = \int_t^{t+T} \sigma^2(s) ds$ is the daily integrated variance of day t , Andersen et al. (2007b) use the realized bipower variation for its empirical counterpart, such that randomly drawn intraday diffusive returns are distributed approximately as $N(0, \delta \cdot BV_t)$. Multiple intraday jumps are detected by:

$$c_j = \Delta_j X \cdot I \left[|\Delta_j X| > \Phi_{1-\beta/2} \cdot \sqrt{\delta \cdot BV_t} \right], \quad j = 1, 2, \dots, \frac{1}{\delta} \quad (22)$$

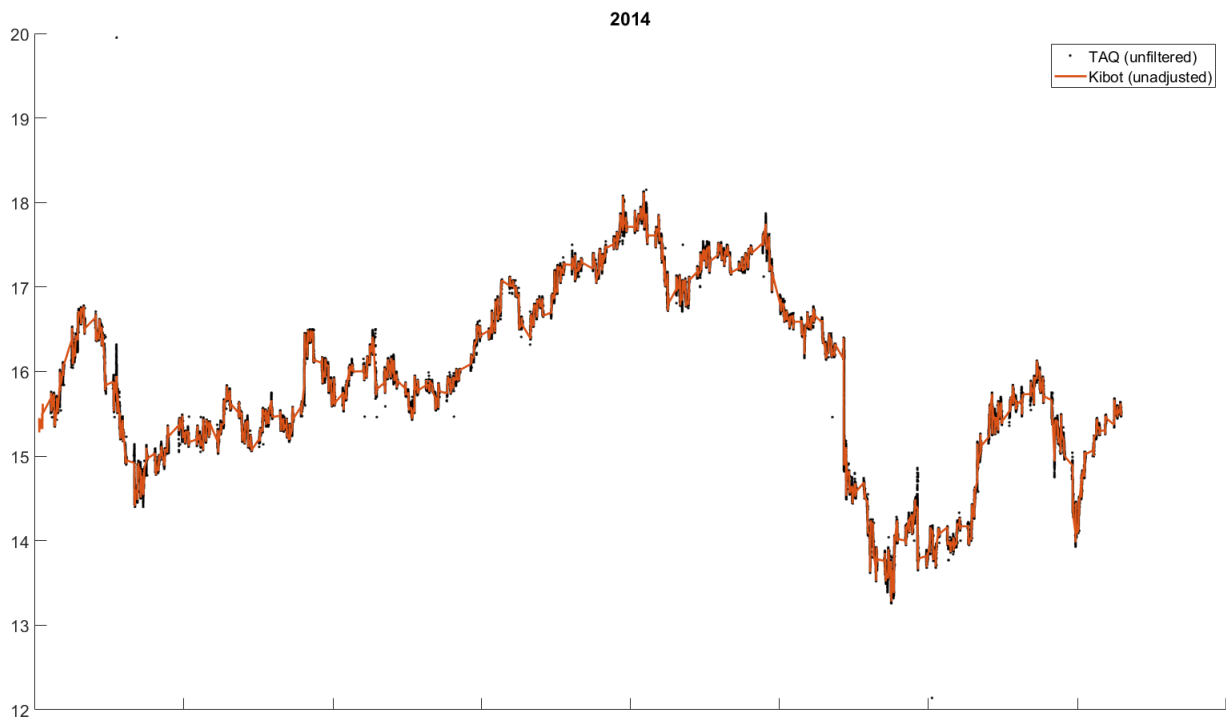
where $\Phi_{1-\beta/2}$ is the appropriate critical value from the standard normal distribution and $\delta = 1/78$, corresponding to a partition of the length of the market trading day, which is open from 9:30 to 16:00, corresponding to seventy-eight 5-min intervals. Following Andersen et al. (2007b), we choose the size of the jump test at the daily level $\alpha = 10^{-5}$ in eq. (20) and define $\beta = 1 - (1 - \alpha)^\delta = 1.28 \cdot 10^{-7}$.

Instead of the realized bipower variation of Barndorff-Nielsen and Shephard (2004, 2006) we use the corrected threshold bipower variation ($C-TBPV$) of Corsi et al. (2010), which was shown to be more accurate for the estimation of jumps. We identify, therefore, intraday jumps by:

$$c_j^* = \Delta_j X \cdot I \left[|\Delta_j X| > \Phi_{1-\beta/2} \cdot \sqrt{\delta \cdot C-TBPV_t} \right], \quad j = 1, 2, \dots, \frac{1}{\delta} \quad (23)$$

following Corsi et al. (2010), we set $c_\vartheta = 3$ when requested, see their paper for details. Finally, we remind that we pre-filter the data following the procedure of Boudt et al. (2011) before testing for the presence of jumps. This allows us to avoid over-identifying jumps occurring in the minutes after the market opening when volatility is higher. Moreover, the Andersen et al. (2007b) approach, being tailored to an intra-day return, i.e. $\Delta_j X$, preserves the signs and thus allows us to separate negative jumps, i.e. jumps associated with a negative intra-day return, from positive jumps, i.e. jumps associated with an intra-day positive return.

Figure C1: TAQ and Kibot comparison.



Notes: Comparison of TAQ and Kibot closing price minute by minute for Ford (ticker: F) in the year 2014.

D Sentiment Detection: Words Lists, Procedure Details, and Algorithms

- *Primary*: consensus, demand, div, div., dividend, dividends, earnings, ebitda, eps, estimate, estimates, f01, f02, f03, f04, forecast, fundamentals, growth, guidance, job, jobs, margin, margins, market, order, orders, outlook, output, production, productivity, profit, profits, q1, q2, q3, q4, qtr, qtr., quarter, reservations, result, results, return, returns, rev, rev., revenue, revenues, roe, sale, sales, shipment, shipments, sold, target, targets, value
- *Secondary Positive*: above, accelerate, accelerated, accelerates, accelerating, acceleration, accomplish, accomplished, accomplishes, accomplishing, accretion, accretive, achieve, achieved, achieves, achieving, active, add, added, adding, additional, adds, advantage, aggressive, ahead of, allow, allowed, allowing, allows, ample, approve, approved, approves, approving, arrive, arrived, arrives, arriving, at least, attractive, beat, beating, beats, began, begin, beginning, begins, begun, benefit, benefited, benefiting, benefits, best, better, beyond, big, bigger, biggest, blowout, boom, boomed, booming, booms, boost, boosted, boosting, boosts, brisk, broad, broaden, broadened, broadening, broadens, bull, bullish, bump, buoyant, capture, captured, captures, capturing, carried, carries, carry, carrying, chase, chased, chases, chasing, climb, climbed, climbing, climbs, comfortable, commence, commenced, commences, commencing, complete, completed, completes, completing, confidence, confident, consistent, continue, continued, continues, continuing, create, created, creates, creating, double, doubled, doubles, doubling, drive, driven, drives, driving, drove, encouraging, endorsement, enhance, enhanced, enhancement, enhances, enhancing, enter, entered, entering, enters, evolve, evolved, evolves, evolving, exceed, exceeded, exceeding, exceeds, exceptional, exciting, expand, expanded, expanding, expands, expansion, exploit, exploited, exploiting, exploits, facilitate, facilitated, facilitates, facilitating, fast, faster, favor, favorable, favored, favoring, favors, favour, favourable, favoured, favouring, favours, foster, fostered, fostering, fosters, fuel, fueled, fueling, fuelled, fuelling, fuels, fulfill, fulfilled, fulfilling, fulfills, full, gain, gained, gaining, gains, garner, garnered, garnering, garners, generate, generated, generates, generating, generator, gigantic, good, grab, grabbed, grabbing, grabs, great, greater, greatest, grew, grow, growing, grown, grows, growth, happy, healthy, help, helped, helping, helps, high, higher, highest, hit, hits, hitted, hitting, hot, huge, impressive, improve, improved, improvement, improves, improving, in line, incentive, increase, increased, increases, increasing, incredible, increment, incremental, incremented, incrementing, increments, initiate, initiated, initiates, initiating, initiation, institute, instituted, institutes, instituting, intact, killer, killing, large, larger, largest, lead, leading, leads, led, lift, lifted, lifting, lifts, maintain, maintained, maintaining, maintains, massive, match, matched, matches, matching, maximise, maximised, maximises, maximising, maximize, maximized, maximizes, maximizing, meaningful, meet, meeting, meets, met, more, new, on track, ongoing, opportunities, opportunity, optimism, optimistic, outgrown, outpace, outpaced, outpaces, outpacing, outperformance, outstrip, outstripped, outstripping, outstrips, over, overwhelming, peak, peaked, peaking, peaks, penetrate, penetrated, penetrates, penetrating, pene-

tration, pent up, permit, permits, permitted, permitting, persist, persisted, persisting, persists, pick up, picked up, picking up, picks up, pickup, please, pleased, pleases, pleasing, positive, positively, potential, power, powerful, premium, profitable, promising, promotion, push, pushed, pushes, pushing, quadruple, quadrupled, quadruples, quadrupling, quality, raise, raised, raises, raising, ramp, ramped, ramping, ramps, reaccelerate, reaccelerated, reaccelerates, reaccelerating, reach, reached, reaches, reaching, reasonable, reassure, reassured, reassures, reassuring, rebound, rebounded, rebounding, rebounds, record, recover, recovered, recovering, recovers, reinforce, reinforced, reinforces, reinforcing, reinstate, reinstated, reinstates, reinstating, reinvigorate, reinvigorated, reinvigorates, reinvigorating, renew, renewed, renewing, renews, resilience, resiliency, resilient, restart, restarted, restarting, restarts, restoration, restore, restored, restores, restoring, resume, resumed, resumes, resuming, resumption, reaffirm, reaffirmation, reaffirmed, reaffirming, reaffirms, ridden, ride, rides, riding, rise, rises, rising, robust, rode, rose, run, running, runs, satisfactory, secure, secured, secures, securing, significant, soar, soared, soars, soaring, solid, solidified, solidifies, solidify, solidifying, solidly, special, spur, spurred, spurring, spurs, stable, standout, start, start up, started, starting, starts, startup, steady, step up, stepped up, stepping up, steps up, stimulate, stimulated, stimulates, stimulating, stir up, stirred up, stirring up, stirs up, strength, strengthen, strengthened, strengthening, strengthens, strong, stronger, strongest, stunning, substantial, success, successful, supplemental, support, supported, supporting, supportive, supports, surge, surged, surges, surging, surpass, surpassed, surpasses, surpassing, surprise, surprised, surprises, surprising, sustain, sustainable, sustained, sustaining, sustains, swell, swelled, swelling, swells, tailwind, tall, top, topped, topping, tops, triple, tripled, triples, tripling, underestimate, underestimated, underestimates, underestimating, underpin, underpinned, underpinning, underpins, underway, unlimited, unprecedented, up, upbeat, upper, upside, upside, upsides, uptick, upturn, upward, well, wide, widen, widened, widening, widens, win, wins, won, wow, wows

- *Secondary Negative*: abandon, abandoned, abandoning, abandons, adverse, axe, axed, axes, axing, back end, backdrop, ban, banned, banning, bans, barred, barrier, barriers, barring, bear, bearish, below, block, blocked, blocking, blocks, bottleneck, bottlenecks, bottom, bottomed, bottoming, bottoms, break, breaking, breaks, broke, broken, bump in the road, cancel, canceled, canceling, cancelled, cancelling, cancels, cannibalisation, cannibalise, cannibalised, cannibalises, cannibalising, cannibalization, cannibalize, cannibalized, cannibalizes, cannibalizing, caution, cautious, cease, ceased, ceases, ceasing, challenge, challenged, challenges, challenging, charge, close to zero, collapse, collapsed, collapses, collapsing, compression, constrain, constrained, constraining, constrains, contracted, contraction, crush, crushed, crushes, crushing, curb, curbed, curbing, curbs, curtail, curtailed, curtailing, curtailment, curtails, cut, cuts, cutting, decelerate, decelerated, decelerates, decelerating, deceleration, decline, declined, declines, declining, decrease, decreased, decreases, decreasing, degradate, degradated, degrades, degradation, delay, delayed, delaying, delays, denied, denies, deny, denying, depress, depressed, depresses, depressing, depression, deteriorate, deteriorated, deteriorates, deteriorating, deterioration, difficult, difficulties, difficulty, dilution, dilutive, disappointing, disruption, down, downside, downsides, downtime, downturn,

downward, drag, dragged, dragging, drags, drop, dropped, dropping, drops, dwindle, dwindled, dwindles, dwindling, eliminate, eliminated, eliminates, eliminating, elimination, erase, erased, erases, erasing, erode, eroded, erodes, eroding, erosion, face, faced, faces, facing, fade, faded, fades, fading, fall, fallen, falling, falls, fell, fewer, flat, flattish, forego, foregoes, foregoing, forewent, freeze, froze, freezes, freezing, half, halt, halted, halting, halts, halve, halved, halves, halving, hamper, hampered, hampering, hampers, headwind, hiatus, hinder, hindered, hindering, hinders, hurt, hurted, hurting, hurts, inflection, interrupt, interrupted, interrupting, interruption, interrupts, lack, lack upside, lacked, lacked upside, lacking, lacking upside, lackluster, lacks, lacks upside, lag, lagged, lagging, lags, laid off, late, lay off, laying off, layoff, layoffs, lays off, light, lighter, limit, limited, limiting, limits, lose, loses, losing, loss, lost, low, lower, lowered, lowering, lowers, lowest, lull, mature, matured, matures, maturing, meager, mere, minimal, miss, missed, misses, missing, moderate, moderated, moderates, moderating, muted, narrow, narrowed, narrowing, narrows, negative, negatively, only, outage, outages, overestimate, overestimated, overestimates, overestimating, overstate, overstated, overstates, overstating, paralyse, paralysed, paralyses, paralysing, paralyze, paralyzed, paralyzes, paralyzing, pause, paused, pauses, pausing, pessimistic, poor, pressure, prevent, prevented, preventing, prevents, prohibit, prohibited, prohibiting, prohibits, reduce, reduced, reduces, reducing, reduction, risk, risks, sacrifice, sacrificed, sacrifices, sacrificing, saturated, saturation, scale back, scaled back, scales back, scaling back, seize, seized, seizes, seizing, shadow, shave, shaved, shaves, shaving, shed, shedded, shedding, sheds, short of, shortage, shortfall, shrank, shrink, shrinking, shrinks, shrunk, shrunken, shut, shut in, shut ins, shutdown, shuts, shutting, skeptical, skepticism, slacken, slackened, slackening, slackens, slash, slashed, slashes, slashing, slippage, slow, slowdown, slowed, slower, slowing, slows, sluggish, slump, slumped, slumping, slumps, small, smaller, snagged, soft, softness, sputtering, stack, stacks, staggering, stagnate, stagnated, stagnates, stagnating, stall, stalled, stalling, stalls, stop, stopped, stopping, stops, stuck, suffer, suffered, suffering, suffers, suspend, suspended, suspending, suspends, suspension, temper, tempered, tempering, tempers, threat, threaten, threatened, threatening, threatens, thwart, thwarted, thwarting, thwarts, tight, tighten, tightened, tightening, tightens, trailed off, trim, trimmed, trimming, trims, turnaround, undermine, undermined, undermines, undermining, unimpressive, unsatisfactory, wane, waned, wanes, waning, warning, weak, weaker, weakest, weakness, weaknesses, wins no fans, worse

- *Stop Words*: a, am, an, are, as, at, be, been, being, could, did, do, does, further, had, has, have, having, he, her, his, i, in, into, is, its, likely to, mean to sound, my, of, one, our, she, so, some, such, that, that's, the, their, there, there's, these, they, this, those, through, to, very, was, we, were, you, your
- *Positive*: able, abundance, abundant, acclaimed, accomplish, accomplished, accomplishes, accomplishing, accomplishment, accomplishments, achieve, achieved, achievement, achievements, achieves, achieving, adequately, advancement, advancements, advances, advancing, advantage, advantaged, advantageous, advantageously, advantages, alliance, alliances, anti trust approval, anti trust clearance, antitrust approval, antitrust

clearance, approval, approvals, approve, approved, approves, approving, receives FDA approval, assure, assured, assures, assuring, attain, attained, attaining, attainment, attainments, attains, attractive, attractiveness, beautiful, beautifully, beneficial, beneficially, benefit, benefited, benefiting, benefitted, benefitting, best, better, bolstered, bolstering, bolsters, boom, booming, boost, boosted, breakthrough, breakthroughs, brilliant, bullish, charitable, collaborate, collaborated, collaborates, collaborating, collaboration, collaborations, collaborative, collaborator, collaborators, compliment, complimentary, complimented, complimenting, compliments, conclusive, conclusively, conducive, confident, constructive, constructively, conviction buy list, cost reduction, cost reductions, courteous, creative, creatively, creativeness, creativity, delight, delighted, delightful, delightfully, delighting, delights, dependability, dependable, desirable, desired, despite, destined, diligent, diligently, distinction, distinctions, distinctive, distinctively, distinctiveness, dream, easier, easily, easy, economic expansion, effective, efficiencies, efficiency, efficient, efficiently, empower, empowered, empowering, empowers, enable, enabled, enables, enabling, encouraged, encouragement, encourages, encouraging, enhance, enhanced, enhancement, enhancements, enhances, enhancing, enjoy, enjoyable, enjoyably, enjoyed, enjoying, enjoyment, enjoys, enthusiasm, enthusiastic, enthusiastically, excellence, excellent, excelling, excels, exceptional, exceptionally, excited, excitement, exciting, exclusive, exclusively, exclusiveness, exclusives, exclusivity, exemplary, fantastic, favorable, favorably, favored, favoring, favorite, favorites, fda approval, fda approvation, fda approved, fda approves, fda approving, friendly, gain, gained, gaining, gains, good, great, greater, greatest, greatly, greatness, green light, happiest, happily, happiness, happy, high quality, high value, highest, honor, honorable, honored, honoring, honors, ideal, impress, impressed, impresses, impressing, impressive, impressively, improve, improved, improvement, improvements, improves, improving, in favor, in favour, incredible, incredibly, influential, informative, ingenuity, innovate, innovated, innovates, innovating, innovation, innovations, innovative, innovativeness, innovator, innovators, insider buying, insightful, inspiration, inspirational, integrity, international expansion, invent, invented, inventing, invention, inventions, inventive, inventiveness, inventor, inventors, leadership, leading, loyal, lucrative, meritorious, more appealing, move higher, moved higher, moves higher, moving higher, novel, on track, opportunities, opportunity, optimism, optimistic, out of trouble, outperform, outperformance, outperformed, outperforming, outperforms, pairing losses, pare losses, pared losses, pares losses, perfect, perfected, perfectly, perfects, pioneer, pioneered, pioneering, pioneers, pleasant, pleasantly, pleased, pleasure, plentiful, popular, popularity, positive, positively, powerful, preeminence, preeminent, premier, premiere, prestige, prestigious, proactive, proactively, proficiency, proficient, proficiently, profitability, profitable, profitably, progress, progressed, progresses, progressing, promising, promisingly, prospered, prospering, prosperity, prosperous, prospers, ramp up, ramped up, ramping up, ramps up, rebound, rebounded, rebounding, receptive, recover, recover, recovered, recovered, recovering, recovering, recovers, recovers, regain, regained, regaining, resolve, revolutionize, revolutionized, revolutionizes, revolutionizing, reward, rewarded, rewarding, rewards, risk less, riskless, satisfaction, satisfactorily, satisfactory, satisfied, satisfies, satisfy, satisfying, smooth, smoothing, smoothly, smooths, solves, solving, spectacular, spectacularly, stability, stabilization, stabilizations, stabilize, sta-

bilized, stabilizes, stabilizing, stable, strength, strengthen, strengthened, strengthening, strengthens, strengths, strong, stronger, strongest, succeed, succeeded, succeeding, succeeds, success, successes, successful, successfully, superior, surpass, surpassed, surpasses, surpassing, tailwind, tailwinds, temper risk, tempered risk, tempering risk, tempers risk, trade higher, traded higher, trades higher, trading higher, transparency, tremendous, tremendously, unmatched, unparalleled, unsurpassed, upgrade, upgraded, upgrades, upgrading, upside, uptick, upturn, upturns, valuable, versatile, versatility, vibrancy, vibrant, well positioned, win, winner, winners, winning, wondrous, wondrously, worthy, wow, wows

- *Negative*: abandon, abandoned, abandoning, abandonment, abandonments, abandons, abdicated, abdicates, abdicating, abdication, abdications, aberrant, aberration, aberrational, aberrations, abetting, abnormal, abnormalities, abnormality, abnormally, abolish, abolished, abolishes, abolishing, abrogate, abrogated, abrogates, abrogating, abrogation, abrogations, abrupt, abruptly, abruptness, absence, absences, absenteeism, abuse, abused, abuses, abusing, abusive, abusively, abusiveness, accident, accidental, accidentally, accidents, accusation, accusations, accuse, accused, accuses, accusing, acquiesce, acquiesced, acquiesces, acquiescing, acquit, acquits, acquittal, acquittals, acquitted, acquitting, adulterate, adulterated, adulterating, adulteration, adulterations, adversarial, adversaries, adversary, adverse, adversely, adversities, adversity, aftermath, aftermaths, against, aggravate, aggravated, aggravates, aggravating, aggravation, aggravations, alerted, alerting, alienate, alienated, alienates, alienating, alienation, alienations, allegation, allegations, allege, alleged, allegedly, alleges, alleging, annoy, annoyance, annoyances, annoyed, annoying, annoys, annul, annulled, annulling, annulment, annulments, annuls, anomalies, anomalous, anomalously, anomaly, anti competitive, anticompetitive, antitrust, antitrust fine, antitrust issue, antitrust issues, argue, argued, arguing, argument, argumentative, arguments, arrearage, arrearages, arrears, arrest, arrested, arrests, artificially, assault, assaulted, assaulting, assaults, assertions, at best, attrition, aversely, backdating, bad, bail, bailout, balk, balked, bankrupt, bankruptcies, bankruptcy, bankrupted, bankrupting, bankrupts, bans, barred, barrier, barriers, bearish, blackout, blame, blamed, blames, blaming, bottleneck, bottlenecks, boycott, boycotted, boycotting, boycotts, breach, breached, breaches, breaching, break, breakage, breakages, breakdown, breakdowns, breaking, breaks, bribe, bribed, briberies, bribery, bribes, bribing, bridge, broken, burden, burdened, burdening, burdens, burdensome, burned, calamities, calamitous, calamity, can fine, cancel, canceled, canceling, cancellation, cancellations, cancelled, cancelling, cancels, careless, carelessly, carelessness, cartel, cartels, cash outflow, cash outflows, catastrophe, catastrophes, catastrophic, catastrophically, caution, cautionary, cautioned, cautioning, cautions, cease, ceased, ceases, ceasing, censure, censured, censures, censoring, challenge, challenged, challenges, challenging, chargeoffs, circumvent, circumvented, circumventing, circumvention, circumventions, circumvents, claiming, claims, clash, clashed, clashes, clashing, class action, class action, clawback, closed, closeout, closeouts, closing, closings, closure, closures, coerce, coerced, coerces, coercing, coercion, coercive, collapse, collapsed, collapses, collapsing, collision, collisions, collude, colluded, colludes, colluding, collusion, collusions, collusive, complain, complained,

complaining, complains, complaint, complaints, compliance issue, compliance issues, complicate, complicated, complicates, complicating, complication, complications, compromised, compromises, compromising, compulsion, concealed, concealing, concede, conceded, concedes, conceding, concern, concerned, concerns, conciliating, conciliation, conciliations, condemn, condemnation, condemnations, condemned, condemning, condemns, condone, condoned, confess, confessed, confesses, confessing, confession, confine, confined, confinement, confinements, confines, confining, confiscate, confiscated, confiscates, confiscating, confiscation, confiscations, conflict, conflicted, conflicting, conflicts, confront, confrontation, confrontational, confrontations, confronted, confronting, confronts, confuse, confused, confuses, confusing, confusingly, confusion, conspiracies, conspiracy, conspirator, conspiratorial, conspirators, conspire, conspired, conspires, conspiring, contempt, contend, contended, contending, contends, contention, contentions, contentious, contentiously, contested, contesting, contraction, contractions, contradict, contradicted, contradicting, contradiction, contradictions, contradictory, contradicts, contrary, controversial, controversies, controversy, convict, convicted, convicting, conviction, convictions, corrected, correcting, correction, corrections, corrects, corrupt, corrupted, corrupting, corruption, corruptions, corruptly, corruptness, costly, counterclaim, counterclaimed, counterclaiming, counterclaims, counterfeit, counterfeited, counterfeiter, counterfeiters, counterfeiting, counterfeits, countermeasure, countermeasures, crime, crimes, criminal, criminally, criminals, crises, crisis, critical, critically, criticism, criticisms, criticize, criticized, criticizes, criticizing, crucial, crucially, culpability, culpable, culpably, cumbersome, curtail, curtailed, curtailing, curtailment, curtailments, curtails, cut, cutback, cutbacks, cuts orders, cutting orders, cyberattack, cyberattacks, cyberbullying, cybercrime, cybercrimes, cybercriminal, cybercriminals, damage, damaged, damages, damaging, dampen, dampened, danger, dangerous, dangerously, dangers, deadlock, deadlocked, deadlocking, deadlocks, deadweight, deadweights, debarment, debarments, debarred, deceased, deceit, deceitful, deceitfulness, deceive, deceived, deceives, deceiving, deception, deceptions, deceptive, deceptively, decline, declined, declines, declining, deface, defaced, defacement, defamation, defamations, defamatory, defame, defamed, defames, defaming, default, defaulted, defaulting, defaults, defeat, defeated, defeating, defeats, defect, defective, defects, defend, defendant, defendants, defended, defending, defends, defensive, defer, deficiencies, deficiency, deficient, deficit, deficits, defraud, defrauded, defrauding, defrauds, defunct, degradation, degradations, degrade, degraded, degrades, degrading, delay, delayed, delaying, delays, deleterious, deliberate, deliberated, deliberately, delinquencies, delinquency, delinquent, delinquently, delinquents, delist, delisted, delisting, delists, demise, demised, demises, demising, demolish, demolished, demolishes, demolishing, demolition, demolitions, demote, demoted, demotes, demoting, demotion, demotions, denial, denials, denied, denies, denigrate, denigrated, denigrates, denigrating, denigration, deny, denying, deplete, depleted, depletes, depleting, depletion, depletions, deprecation, depress, depressed, depresses, depressing, deprivation, deprive, deprived, deprives, depriving, derelict, dereliction, derogatory, destabilization, destabilize, destabilized, destabilizing, destroy, destroyed, destroying, destroys, destruction, destructive, detain, detained, detention, detentions, deter, deteriorate, deteriorated, deteriorates, deteriorating, deterioration, deteriorations, deterred, deterrence, deter-

rences, deterrent, deterrents, deterring, deters, detract, detracted, detracting, detriment, detrimental, detrimentally, detriments, devalue, devalued, devalues, devaluing, devastate, devastated, devastating, devastation, deviate, deviated, deviates, deviating, deviation, deviations, devolve, devolved, devolves, devolving, difficult, difficulties, difficultly, difficulty, diminish, diminished, diminishes, diminishing, diminution, disadvantage, disadvantaged, disadvantageous, disadvantages, disaffiliation, disagree, disagreeable, disagreed, disagreeing, disagreement, disagreements, disagrees, disallow, disallowance, disallowances, disallowed, disallowing, disallows, disappear, disappearance, disappearances, disappeared, disappearing, disappears, disappoint, disappointed, disappointing, disappointingly, disappointment, disappointments, disappoints, disapproval, disapprovals, disapprove, disapproved, disapproves, disapproving, disassociates, disassociating, disassociation, disassociations, disaster, disasters, disastrous, disastrously, disavow, disavowal, disavowed, disavowing, disavows, disciplinary, disclaim, disclaimed, disclaimer, disclaimers, disclaiming, disclaims, disclose, disclosed, discloses, disclosing, discontinuance, discontinuances, discontinuation, discontinuations, discontinue, discontinued, discontinues, discontinuing, discourage, discouraged, discourages, discouraging, discredit, discredited, discrediting, discredits, discrepancies, discrepancy, disfavor, disfavored, disfavoring, disfavors, disgorge, disgorged, disgorgement, disgorgements, disgorges, disgorging, disgrace, disgraceful, disgracefully, dishonest, dishonestly, dishonesty, dishonor, dishonorable, dishonorably, dishonored, dishonoring, dishonors, disincentives, disinterested, disinterestedly, disinterestedness, disloyal, disloyally, disloyalty, dismal, dismally, dismiss, dismissal, dismissals, dismissed, dismisses, dismissing, disorderly, disparage, disparaged, disparagement, disparagements, disparages, disparaging, disparagingly, disparities, disparity, displace, displaced, displacement, displacements, displaces, displacing, dispose, dispossess, dispossessed, dispossesses, dispossessing, disproportion, disproportional, disproportionate, disproportionately, dispute, disputed, disputes, disputing, disqualification, disqualifications, disqualified, disqualifies, disqualify, disqualifying, disregard, disregarded, disregarding, disregards, disreputable, disrepute, disrupt, disrupted, disrupting, disruption, disruptions, disruptive, disrupts, dissatisfaction, dissatisfied, dissent, dissented, dissenter, dissenters, dissenting, dissents, dissident, dissidents, dissolution, dissolutions, distort, distorted, distorting, distortion, distortions, distorts, distract, distracted, distracting, distraction, distractions, distracts, distress, distressed, disturb, disturbance, disturbances, disturbed, disturbing, disturbs, diversion, divert, diverted, diverting, diverts, divest, divested, divesting, divestiture, divestitures, divestment, divestments, divests, divorce, divorced, divulge, divulged, divulges, divulging, doubt, doubted, doubtful, doubts, downbeat, downgrade, downgraded, downgrades, downgrading, downside, downsize, downsized, downsizes, downsizing, downsizings, downtick, downtime, downtimes, downturn, downturns, downward, downwards, drag, drastic, drastically, drawback, drawbacks, dropped, drought, droughts, dumping, duress, dysfunction, dysfunctional, dysfunctions, easing, egregious, egregiously, embargo, embargoed, embargoes, embargoing, embarrass, embarrassed, embarrasses, embarrassing, embarrassment, embarrassments, embezzle, embezzled, embezzlement, embezzlements, embezzler, embezzles, embezzling, encroach, encroached, encroaches, encroaching, encroachment, encroachments, encumber, encumbered, encumbering, encumbers, encumbrance, encum-

brances, endanger, endangered, endangering, endangerment, endangers, enjoin, enjoined, enjoining, enjoins, erode, eroded, erodes, eroding, erosion, erratic, erratically, erred, erring, erroneous, erroneously, error, errors, errs, escalate, escalated, escalates, escalating, evade, evaded, evades, evading, evasion, evasions, evasive, evict, evicted, evicting, eviction, evictions, evicts, exacerbate, exacerbated, exacerbates, exacerbating, exacerbation, exacerbations, exaggerate, exaggerated, exaggerates, exaggerating, exaggeration, excessive, excessively, exculpate, exculpated, exculpates, exculpating, exculpation, exculpations, exculpatory, exonerate, exonerated, exonerates, exonerating, exoneration, exonerations, exploit, exploitation, exploitations, exploitative, exploited, exploiting, exploits, expose, exposed, exposes, exposing, expropriate, expropriated, expropriates, expropriating, expropriation, expropriations, expulsion, expulsions, extenuating, fail, failed, failing, failings, fails, failure, failures, fallout, false, falsely, falsification, falsifications, falsified, falsifies, falsify, falsifying, falsity, fatalities, fatality, fatally, fault, faulted, faults, faulty, fear, fears, felonies, felonious, felony, fictitious, fined, fines, fired, firing, flaw, flawed, flaws, forbid, forbidden, forbidding, forbids, force, force majeure, forced, forcing, foreclose, foreclosed, forecloses, foreclosing, foreclosure, foreclosures, forego, foregoes, foregone, forestall, forestalled, forestalling, forestalls, forfeit, forfeited, forfeiting, forfeits, forfeiture, forfeitures, forgers, forgery, fraud, frauds, fraudulence, fraudulent, fraudulently, frivolous, frivolously, frustrate, frustrated, frustrates, frustrating, frustratingly, frustration, frustrations, fudge, fudged, fudges, fudging, fugitive, fugitives, gratuitous, gratuitously, grievance, grievances, grossly, groundless, guilty, hacker, hacked, hacking, hackers, halt, halted, hamper, hampered, hampering, hampers, harass, harassed, harassing, harassment, hardship, hardships, harm, harmed, harmful, harmfully, harming, harms, harsh, harsher, harshest, harshly, harshness, hazard, hazardous, hazards, headwind, headwinds, hiccup, hiccups, hinder, hindered, hindering, hinders, hindrance, hindrances, hostile, hostility, hurt, hurting, idle, idled, idling, ignore, ignored, ignores, ignoring, ill, illegal, illegalities, illegality, illegally, illegible, illicit, illicitly, illiquid, illiquidity, imbalance, imbalances, immature, immoral, impair, impaired, impairing, impairment, impairments, impairs, impasse, impasses, impede, impeded, impedes, impediment, impediments, impeding, impending, imperative, imperfection, imperfections, imperil, impermissible, implicate, implicated, implicates, implicating, impossibility, impossible, impound, impounded, impounding, impounds, impracticable, impractical, impracticalities, impracticality, imprisonment, improper, improperly, improprieties, impropriety, imprudent, imprudently, inability, inaccessible, inaccuracies, inaccuracy, inaccurate, inaccurately, inaction, inactions, inactivate, inactivated, inactivates, inactivating, inactivation, inactivations, inactivity, inadequacies, inadequacy, inadequate, inadequately, inadvertent, inadvertently, inadvisability, inadvisable, inappropriate, inappropriately, inattention, incapable, incapacitated, incapacity, incarcerate, incarcerated, incarcerates, incarcerating, incarceration, incarcerations, incidence, incidences, incident, incidents, incompatibilities, incompatibility, incompatible, incompetence, incompetency, incompetent, incompetently, incompetents, incomplete, incompletely, incompleteness, inconclusive, inconsistencies, inconsistency, inconsistent, inconsistently, inconvenience, inconveniences, inconvenient, incorrect, incorrectly, incorrectness, indecency, indecent, infeasible, infeasibly, indict, indictable, indicted, indicting, indictment, indictments, ineffective, ineffectively,

ineffectiveness, inefficiencies, inefficiency, inefficient, inefficiently, ineligibility, ineligible, inequitable, inequitably, inequities, inequity, inevitable, inexperience, inexperienced, inferior, inferiority, inflicted, infraction, infractions, infringe, infringed, infringement, infringements, infringes, infringing, inhibited, inimical, injunction, injunctions, injure, injured, injures, injuries, injuring, injurious, injury, inordinate, inordinately, inquiry, insecure, insensitive, insider trading, insolvencies, insolvency, insolvent, instability, insubordination, insufficiency, insufficient, insufficiently, insurrection, insurrections, intentional, interfere, interfered, interference, interferences, interferes, interfering, intermittent, intermittently, interrupt, interrupted, interrupting, interruption, interruptions, interrupts, intimidation, intrusion, invalid, invalidate, invalidated, invalidates, invalidating, invalidation, invalidity, investigate, investigated, investigates, investigating, investigation, investigations, involuntarily, involuntary, irreconcilable, irreconcilably, irrecoverable, irrecoverably, irregular, irregularities, irregularity, irregularly, irreparable, irreparably, irreversible, jeopardize, jeopardized, justifiable, kickback, kickbacks, knowingly, lack, lacked, lacking, lackluster, lacks, lag, lagged, lagging, lags, laid off, lapse, lapsed, lapses, lapsing, late, laundering, lay off, laying off, layoff, layoffs, lays off, leak, leakage, leakages, leaked, leaking, leaks, lie, limitation, limitations, lingering, liquidate, liquidated, liquidates, liquidating, liquidation, liquidations, liquidator, liquidators, litigant, litigants, litigate, litigated, litigates, litigating, litigation, litigations, lockout, lockouts, lose, loses, losing, loss, losses, lost, lying, malfeasance, malfunction, malfunctioned, malfunctioning, malfunctions, malice, malicious, maliciously, malpractice, malware, manipulate, manipulated, manipulates, manipulating, manipulation, manipulations, manipulative, markdown, markdowns, mechanical issue, mechanical issues, mess, misapplication, misapplications, misapplied, misapplies, misapply, misapplying, misappropriate, misappropriated, misappropriates, misappropriating, misappropriation, misappropriations, misbranded, miscalculate, miscalculated, miscalculates, miscalculating, miscalculation, miscalculations, mischaracterization, mischief, misclassification, misclassifications, misclassified, misclassify, miscommunication, misconduct, misdated, misdemeanor, misdemeanors, misdirected, mishandle, mishandled, mishandles, mishandling, misinform, misinformation, misinformed, misinforming, misinforms, misinterpret, misinterpretation, misinterpretations, misinterpreted, misinterpreting, misinterprets, misjudge, misjudged, misjudges, misjudging, misjudgment, misjudgments, mislabel, mislabeled, mislabeling, mislabelled, mislabels, mislead, misleading, misleadingly, misleads, misled, mismanage, mismanaged, mismanagement, mismanages, mismanaging, mismatch, mismatched, mismatches, mismatching, misplaced, misprice, mispricing, mispricings, misrepresent, misrepresentation, misrepresentations, misrepresented, misrepresenting, misrepresents, miss, missed, misses, misstate, misstated, misstatement, misstatements, misstates, misstating, misstep, missteps, mistake, mistaken, mistakenly, mistakes, mistaking, mistrial, mistrials, misunderstand, misunderstanding, misunderstandings, misunderstood, misuse, misused, misuses, misusing, monopolistic, monopolists, monopolization, monopolize, monopolized, monopolizes, monopolizing, monopoly, moratoria, moratorium, moratoriums, mothballed, mothballing, move lower, moved lower, moves lower, moving lower, negative, negatively, negatives, neglect, neglected, neglectful, neglecting, neglects, negligence, negligences, negligent, negligently, no assurance, no guarantee, nonattainment, non-

competitive, noncompliance, noncompliances, noncompliant, noncomplying, nonconforming, nonconformities, nonconformity, nondisclosure, nonfunctional, nonpayment, nonpayments, nonperformance, nonperformances, nonperforming, nonproducing, nonproductive, nonrecoverable, nonrenewal, nuisance, nuisances, nullification, nullifications, nullified, nullifies, nullify, nullifying, objected, objecting, objection, objectionable, objectionably, objections, obscene, obscenity, obsolescence, obsolete, obstacle, obstacles, obstruct, obstructed, obstructing, obstruction, obstructions, offence, offences, offend, offended, offender, offenders, offending, offends, omission, omissions, omit, omits, omitted, omitting, onerous, operational issue, operational issues, opportunistic, opportunistically, oppose, opposed, opposes, opposing, opposition, oppositions, ousting, outage, outages, outdated, outmoded, overage, overages, overbill, overbilled, overbilling, overbills, overbuild, overbuilding, overbuilds, overbuilt, overburden, overburdened, overburdening, overcapacities, overcapacity, overcharge, overcharged, overcharges, overcharging, overcome, overcomes, overcoming, overdue, overestimate, overestimated, overestimates, overestimating, overestimation, overestimations, overhang, overload, overloaded, overloading, overloads, overlook, overlooked, overlooking, overlooks, overly optimistic, overpaid, overpayment, overpayments, overproduced, overproduces, overproducing, overproduction, overrun, overrunning, overruns, overshadow, overshadowed, overshadowing, overshadows, overstate, overstated, overstatement, overstatements, overstates, overstating, oversupplied, oversupplies, oversupply, oversupplying, overtly, overturn, overturned, overturning, overturns, overvalue, overvalued, overvaluing, panic, panics, penalize, penalized, penalizes, penalizing, penalties, penalty, peril, perils, perjury, perpetrate, perpetrated, perpetrates, perpetrating, perpetration, persist, persisted, persistence, persistent, persistently, persisting, persists, pervasive, pervasively, pervasiveness, pessimism, pessimistic, petty, picket, picketed, picketing, plaintiff, plaintiffs, plea, plead, pleaded, pleading, pleadings, pleads, pleas, pled, poor, poorly, poses, posing, postpone, postponed, postponement, postponements, postpones, postponing, precipitated, precipitous, precipitously, preclude, precluded, precludes, precluding, predatory, prejudice, prejudiced, prejudices, prejudicial, prejudicing, premature, prematurely, pressing, pressure on margins, pretrial, preventing, prevention, prevents, problem, problematic, problematical, problems, prolong, prolongation, prolongations, prolonged, prolonging, prolongs, prone, prosecute, prosecuted, prosecutes, prosecuting, prosecution, prosecutions, protest, protested, protester, protesters, protesting, protestor, protestors, protests, protracted, protraction, provoke, provoked, provokes, provoking, punished, punishes, punishing, punishment, punishments, punitive, purport, purported, purportedly, purporting, purports, question, questionable, questionably, questioned, questioning, questions, quit, quitting, racketeer, racketeering, raise eyebrows, raised eyebrows, raises eyebrows, raising eyebrows, rationalization, rationalizations, rationalize, rationalized, rationalizes, rationalizing, reassessment, reassessments, reassign, reassigned, reassigning, reassignment, reassignments, reassigns, recall, recalled, recalling, recalls, recession, recessionary, recessions, reckless, recklessly, recklessness, redact, redacted, redacting, redaction, redactions, redefault, redefaulted, redefaults, redress, redressed, redresses, redressing, refusal, refusals, refuse, refused, refuses, refusing, reject, rejected, rejecting, rejection, rejections, rejects, relinquish, relinquished, relinquishes, relinquishing, relinquishment, relinquish-

ments, reluctance, reluctant, renegotiate, renegotiated, renegotiates, renegotiating, renegotiation, renegotiations, renounce, renounced, renouncement, renouncements, renounces, renouncing, reparation, reparations, reported issue, reported issues, repossessed, repossesses, repossessing, repossession, repossessions, repudiate, repudiated, repudiates, repudiating, repudiation, repudiations, resign, resignation, resignations, resigned, resigning, resigns, restate, restated, restatement, restatements, restates, restating, restructure, restructured, restructures, restructuring, restructurings, retaliate, retaliated, retaliates, retaliating, retaliation, retaliations, retaliatory, retribution, retributions, revocation, revocations, revoke, revoked, revokes, revoking, ridicule, ridiculed, ridicules, ridiculing, riskier, riskiest, risky, sabotage, sacrifice, sacrificed, sacrifices, sacrificial, sacrificing, sad, scandalous, scandals, scrutinize, scrutinized, scrutinizes, scrutinizing, scrutiny, secrecy, seize, seized, seizes, seizing, sentenced, sentencing, serious, seriously, seriousness, setback, setbacks, sever, severe, severed, severely, severities, severity, sharply, shocked, shortage, shortages, shortfall, shortfalls, shrinkage, shrinkages, shut, shut in, shut ins, shutdown, shutdowns, shuts, shutting, slander, slandered, slanderous, slanders, slippage, slippages, slow, slowdown, slowdowns, slowed, slower, slowest, slowing, slowly, slowness, sluggish, sluggishly, sluggishness, sob, solvencies, solvency, spam, spammers, spamming, spill, spilled, spilling, spills, staggering, stagnant, stagnate, stagnated, stagnates, stagnating, stagnation, stall, stalled, stalling, stalls, standstill, standstills, steal, stealing, steals, stole, stolen, stoppage, stoppages, stopped, stopping, stops, strain, strained, straining, strains, stress, stressed, stresses, stressful, stressing, stringent, sub growth, subjected, subjecting, subjection, subpoena, subpoenaed, subpoenas, substandard, sue, sued, sues, suffer, suffered, suffering, suffers, suing, summoned, summoning, summons, summonses, susceptibility, susceptible, suspect, suspected, suspects, suspend, suspended, suspending, suspends, suspension, suspensions, suspicion, suspicions, suspicious, suspiciously, taint, tainted, tainting, taints, tampered, tense, terminate, terminated, terminates, terminating, termination, terminations, testify, testifying, theft, thefts, threat, threaten, threatened, threatening, threatens, threats, tightening, tolerate, tolerated, tolerates, tolerating, toleration, too optimistic, tortuous, tortuously, trade lower, traded lower, trades lower, trading lower, tragedies, tragedy, tragic, tragically, traumatic, trouble, troubled, troubles, troublesome, troubling, turbulence, turmoil, unable, unacceptable, unacceptably, unaccounted, unannounced, unanticipated, unapproved, unattractive, unauthorized, unavailability, unavailable, unavoidable, unavoidably, unaware, uncollectable, uncollected, uncollectibility, uncollectible, uncollectibles, uncompetitive, uncompleted, unconscionable, unconscionably, uncontrollable, uncontrollably, uncontrolled, uncorrected, uncover, uncovered, uncovering, uncovers, undeliverable, undelivered, under pressure, undercapitalized, undercut, undercuts, undercutting, underestimate, underestimated, underestimates, underestimating, underestimation, underfunded, underinsured, undermine, undermined, undermines, undermining, underpaid, underpayment, underpayments, underpays, underperform, underperformance, underperformed, underperforming, underperforms, underproduced, underproduction, underreporting, understate, understated, understatement, understatements, understates, understating, underutilization, underutilized, undesirable, undesired, undetected, undetermined, undisclosed, undocumented, undue, unduly, uneasy, uneconomic, uneconomical, uneconom-

ically, unemployed, unemployment, unenthusiastic, unenthusiastically, unethical, unethically, unexcused, unexpected, unexpectedly, unfair, unfairly, unfavorability, unfavorable, unfavorably, unfavourable, unfeasible, unfit, unfitness, unforeseeable, unforeseen, unforeseen, unfortunate, unfortunately, unfounded, unfriendly, unfulfilled, unfunded, unhappy, uninsured, unintended, unintentional, unintentionally, unjust, unjustifiable, unjustifiably, unjustified, unjustly, unknowing, unknowingly, unlawful, unlawfully, unlicensed, unliquidated, unmarketable, unmerchantable, unmeritorious, unnecessarily, unnecessary, unneeded, unobtainable, unoccupied, unpaid, unperformed, unplanned, unpopular, unpredictability, unpredictable, unpredictably, unpredicted, unproductive, unprofitability, unprofitable, unqualified, unrealistic, unreasonable, unreasonableness, unreasonably, unreceptive, unrecoverable, unrecovered, unreimbursed, unreliable, unremedied, unreported, unresolved, unrest, unsafe, unsalable, unsaleable, unsatisfactory, unsatisfied, unsavory, unscheduled, unsellable, unsold, unsound, unstabilized, unstable, unsubstantiated, unsuccessful, unsuccessfully, unsuitability, unsuitable, unsuitably, unsuited, unsure, unsuspected, unsuspecting, unsustainable, untenable, untimely, untrusted, untruth, untruthful, untruthfully, untruthfulness, untruths, unusable, unwanted, unwarranted, unwelcome, unwilling, unwillingness, upset, urgency, urgent, usurious, usurp, usurped, usurping, usurps, usury, vandalism, verdict, verdicts, vetoed, victims, violate, violated, violates, violating, violation, violations, violative, violator, violators, violence, violent, violently, vitiate, vitiated, vitiates, vitiating, vitiating, voided, voiding, volatile, volatility, vulnerabilities, vulnerability, vulnerable, vulnerably, walk away, walked away, walking away, walks away, warn, warned, warning, warnings, warns, wary, wasted, wasteful, wasting, weak, weaken, weakened, weakening, weakens, weaker, weakest, weakly, weakness, weaknesses, weather issue, weather issues, willfully, woe, woes, worried, worries, worry, worrying, worse, worsen, worsened, worsening, worsens, worst, worthless, wound, wounded, wounding, wounds, write down, write downs, writedown, writedown, writedowns, writedowns, writeoff, writeoffs, wrong, wrongdoing, wrongdoings, wrongful, wrongfully, wrongly

- *Negations*: absence of, any, as opposed to, barely, by no means, can't, cannot, cant, despite, didn't, didnt, difficult to, doesn't, doesnt, don't, dont, far from, few, fewer, hard to, hardly, improbable, improbably, in lieu of, in no way, in place of, in spite of, instead of, less, little, mustn't, neither, never, never to, no, no big, no longer, no longer, no longer in, no more, no sign of, no significant, no way, nobody, none, nor, not, not all, not always, not as, not at all, not big, not enough, not every, not expect, not going to, not many, not much, not significant, not so, not to, not very, nothing, nowhere, offset, offsets, offsetting, only marginally, out of, rarely, rather than, rather than, reduce, reduced, reduces, reducing, reduction, reductions, remove, removed, removes, removing, scarcely, seldom, seldomly, short of, sporadic, sporadically, too few, too little, unlikely to, without, won't, wont
- *Neutral*: a little, apologies for any confusion, apologies for any error, apologies for the confusion, apologies for the error, apologize for any confusion, apologize for any error, apologize for the confusion, apologize for the error, at closing, bear stearns, benefit, best case, boxed warning, breaking, breaks, charge offs, chargeoffs, closed, closing,

corrected comment, correction, corrections, critical requirement, critical requirements, critical to, crohn's disease, decline comment, decline comments, decline to, declined comment, declined comments, declined to, declines comment, declines comments, declines to, declining comment, declining comments, declining to, default rate, delinquencies, delinquency, delinquent, delinquents, denies talks, deny talks, destructive testing, disclose, disclosed, discloses, disclosing, disclosure, discontinued operations, divest, divested, divesting, divestiture, divestitures, divestment, divestments, divests, effective, effective date, effective tax, fall, halted, hazard ratio, heart attacks, hedge against, hedged against, hedges against, hedging against, incident management, incorrect figure, incorrect figures, incorrecly reported, incorrectly attributable, infection risk, infection risks, late apr, late april, late aug, late august, late dec, late december, late feb, late february, late jan, late january, late jul, late july, late jun, late june, late maj, late mar, late march, late nov, late november, late oct, late october, late sep, late september, late this, lease termination, lease terminations, misleading, net chargeoffs, net writeoffs, nondisclosure agreement, overestimate, overestimated, overestimates, overestimating, overestimation, overestimations, provision for losses, question, recall that, requirement critical, requirements critical, rewards program, rewards programme, right of first refusal, sharply, shortage of discussion, underestimate, underestimated, underestimates, underestimating, underestimation, undisclosed, unsure if comparable, we apologise, we apologize, we apology, weight loss, worst case

- *Primary Incompatible*: cost, costs, inventories, inventory, price, prices, taxes
- *Secondary Neutral*: bear stearns, bottom end, continuing operations, continuing ops, follow up, full year, headline only, headlines only, high end, higher end, low end, lower end, ongoing eps, only for, significant impact, too high, top end, up end, up to, upper end
- *Special Topics - Legal*: case against, court, defend, defendant, defendants, defended, defenders, defending, defends, defense, dispute, judge, judges, jury, lawsuit, lawsuits, lawyer, lawyers, litigation, no impact, plaintiff, plaintiffs, ruling, rulings, settlement, trial, verdict
- *Special Topics - Medical*: clinical, heart attacks, patient, patients, therapy, treatment, treatments
- *Special Topics - Trading Events*: bad trade, post resumption, trading miscue
- *Special Topics - Officer Changes*: leadership change, leadership changes, officer changes
- *Special Topics - Officer Changes Comb 1* ²³ : board of directors, bod, ceo, cfo, chief executive officer, chief financial officer, leadership change, leadership changes

²³*Special Topics - Officer Changes Comb 1* and *Special Topics - Officer Changes Comb 2* are used in combination: in case at least one *Special Topics - Officer Changes Comb 1* item and at least one *Special Topics - Officer Changes Comb 2* item occur within a 10-word window, the news will be considered to belong to the *Special Topics* field.

- *Special Topics - Officer Changes Comb 2*: appoint, appointed, appointing, appointment, appoints, name, named, names, naming, new

H&B sentiment detection It applies *Algorithm H* and *Algorithm B* to the news headline and the news story as detailed in the algorithm flow 1. The approach works as follows:

- apply *Algorithm H* to the headline
- if *Algorithm H* produces a positive or negative sentiment, assign that sentiment to the news and exit
- if *Algorithm H* produces a neutral sentiment due to the lack of *primary* or *secondary* items or due to the presence of *secondary positive* and *secondary negative* items in the same quantity, apply *Algorithm B* to the headline
- if *Algorithm B* applied to the headline produces a positive or negative sentiment, assign that sentiment to the news and exit
- if *Algorithm B* applied to the headline produces a neutral sentiment, apply *Algorithm B* to the news story, assign the obtained sentiment to the news and exit

Algorithm 1: *H&B Sentiment Detection Procedure*

Input : (*News Headline*, *News Story*)
Output: (*sentiment*)

```

1 apply Procedure 1 to News Headline;
2 if flag for special topics = 1 then
3   | return neutral sentiment;
4 else if sentiment ≠ 0 then
5   | return sentiment;
6 else
7   | apply Procedure 2 to News Headline;
8   | if flag for special topics = 1 then
9     | return neutral sentiment;
10  | else if sentiment ≠ 0 then
11  |   | return sentiment;
12  | else
13  |   | apply Procedure 2 to News Story;
14  |   | return sentiment;
15  | end

```

Algorithm H This procedure is suitable for short texts such as headlines and it is based on the identification of the main subject of the text by checking the presence of *primary* items and the subsequent detection of the sentiment by using *secondary* items that allow to understand if the news is good or bad. *Primary* items (e.g., “dividends”, “EPS”, or

“results”) refer to very important information about a company, *secondary positive* items (e.g., “boost”, “grew”, or “picked up”) indicate growth concepts, and *secondary negative* items (e.g., “bottoming”, “decrease”, or “lacking upside”) indicate reduction concepts. The procedure is summarized in algorithm flow 2 and is structured in this way:

- check for the presence of *primary* items
- in case there are no *primary* items, assign the text neutral sentiment and exit
- if there are *primary* items, check for the presence of *primary incompatible* items (e.g., “prices” or “taxes”) that, when combined with *secondary* items, give rise to an opposite sentiment compared to the *primary* items. In so doing, we get rid of potential wrong sentiment assignments when *secondary* items refer to *primary incompatible* items, as in the sentence “EPS results: Lower taxes contributed \$0.14 to the quarter”, where “lower” is a *secondary negative item* not related to the *primary* item “EPS” but referred to the *primary incompatible* item “taxes”
- in case there are both *primary* and *primary incompatible* items, assign the text neutral sentiment and exit
- if there are *primary* items and no *primary incompatible* items, check for the presence of *secondary neutral* items (e.g., “low end”), that are neutral expressions containing *secondary positive* and *secondary negative* items or part of them, and exclude their words from the rest of the procedure
- remove the items from the list *stop words* (e.g., “a”, “of” or “being”), which are sometimes located between a negation and a *secondary* item—such as “being” in the sentence “quarterly results instead of being positive...”, which is located between the *negation* item “instead of” and the *secondary positive* item “positive”—. In order to avoid to remove an item from the list *stop words* when it is part of a negation, the procedure removes *stop words* items only if they are not part of *negation* items. E.g., “of” is not removed if the text contains the *negation* item “instead of”
- attribute to *secondary positive* items the value +1 and to *secondary negative* items -1
- multiply by -1 the value of the *secondary* items preceded by a negation or followed by “not true”. Double negations are checked as well and the value of items preceded by double negations is not changed
- assign the text a sentiment on the basis of the sign of the average of the previous values: a positive sign corresponds to a positive sentiment, a negative sign to a negative sentiment, and an average value of zero to a neutral sentiment

Algorithm B This procedure is suitable for long texts where the presence of *primary* items cannot be used to infer the main subject and for short texts without *primary* items or containing both *primary* and *primary incompatible* items. It is detailed in algorithm flow 3 and it works as follows:

- detect all *neutral* items (e.g., “hedge against”) and exclude them from the rest of the procedure to avoid to attribute sentiment to items in the lists *positive* and *negative* when they are part of neutral expressions. E.g., the word “against” is a *negative* item but it will not be considered negative if it is part of the *neutral* item “hedge against”
- remove from the text the items from the list *stop words* which are not part of negations occurring in the text
- attribute +1 to *positive* items (e.g., “high quality”) and -1 to *negative* items (e.g., “insider trading”)
- invert the value of the *positive* and *negative* items preceded by a negation or followed by “not true”, but not of those preceded by a double negation
- assign the text a sentiment on the basis of the sign of the average of the previous values, as in *Algorithm H*

Additional remarks on sentiment detection. In case a news contains *special topics* items, *positive* and *negative* items and the combinations of *primary* and *secondary* items can be misleading. *Special topics* are the legal, medical, officer changes and trading events topics. Both *Algorithm H* and *Algorithm B* begin with the identification of these fields using the list *special topics* in the following way: if a text contains an item of this list, the news will be considered related to one of those fields. News belonging to these fields are assigned neutral sentiment because of the unreliability of the sentiment that would be produced. Some examples are the following ones: the negative word “accusations” may refer to accusations against a company but the news is about a verdict in its favor; the negative word “disease” may refer to a disease against which a company is testing a new drug; the positive word “succeeds” may refer to the appointment of a new CEO; the event of a “bad trade” is likely to have an impact on a stock price but it is unclear in which direction. The item “no impact” is included as well in the *special topics* list so this expression makes a news neutral.

Some words are part of several items, which in turn can belong to different lists. E.g., “of” is an item of the list *stop words* and part of the item “short of” in the list *secondary negative*. In order to avoid to count a word more than once or to remove words that are part of other items, items of a list are checked in descending order of length in terms of number of words in such a way that after an item is identified, its words are not used to check for the presence of other items. The same is true with regard to items of different lists: after items of a list are identified, their words are not used to check for the presence of items of other lists.

Algorithm 2: H

Input : (*Text*)
Output: (*sentiment, flag for special topics*)

- 1 **if** *special topics items found* **then**
- 2 | *flag for special topics* = 1;
- 3 | return (*neutral sentiment, flag for special topics*);
- 4 **else if** *presence of primary items and absence of primary incompatible items* **then**
- 5 | *flag for special topics* = 0;
- 6 | identify *secondary neutral* items and don't consider their words in the rest of the procedure;
- 7 | remove *stop words* items not part of *negation* items;
- 8 | attribute +1 and -1 to *secondary positive* and *secondary negative* items, respectively;
- 9 | multiply by -1 *secondary positive* and *secondary negative* values preceded by *negation* items and not preceded by a double negation;
- 10 | sentiment = sign(mean(items values));
- 11 | return (*sentiment, flag for special topics*);

Algorithm 3: B

Input : (*Text*)
Output: (*sentiment, flag for special topics*)

- 1 **if** *special topics items found* **then**
- 2 | *flag for special topics* = 1;
- 3 | return (*neutral sentiment, flag for special topics*);
- 4 **else**
- 5 | *flag for special topics* = 0;
- 6 | identify *neutral* items and don't consider their words in the rest of the procedure;
- 7 | remove *stop words* items not part of *negation* items;
- 8 | attribute +1 and -1 to *positive* and *negative* items, respectively;
- 9 | multiply by -1 *positive* and *negative* values preceded by *negation* items and not preceded by a double negation;
- 10 | sentiment = sign(mean(items values));
- 11 return (*sentiment, flag for special topics*);

E News-Based Indicators

Tables E1–E3 illustrate the set of news-based indicators that result from combining the lead-and-lag time horizons, the concepts for news stories and the surprises of EPS and macro-news of Section 3.3. We build 336 news stories-based, 12 EPS-based, and 276 macro-based indicators, in total 624 indicators for each stock.

Table E1: News stories standard indicators.

Variable	N. Transf.	N. Time Int.
STANDARD		
news occurrence flag	1 ^a	3
sentiment	3 ^b	3
ABNORMAL QUANTITY		
n. news ≥ 2	1	3
UNCERTAINTY		
pos and neg news in same interval	1	3
NEWS PERSISTENCE		
news occurrence in 2 consecutive intervals	1	2 ^c
pos. sent., persistence	1	2
neg. sent., persistence	1	2
total for each topic		24
grand total (24 x 14 topics) ^d		336

Notes: The first column shows the variables grouped by the concepts that originated them. The second column shows the number of transformations. The third column shows the number of time intervals over which the measures are computed. The total number of measures obtained is reported at the end of the third column.

^a: When the number of transformations equals 1, the measure consists of a flag (1 for the occurrence of the event, and 0 otherwise).

^b: We apply the following 3 transformations: flag for sentiment $\neq 0$, flag for positive sentiment, flag for negative sentiment.

^c: This indicator is based on the aggregation of information over consecutive intervals. From 3 lead-and-lag intervals we obtain 2 couples of consecutive intervals.

^d: There are 14 topics (7 topics for StreetAccount news stories plus 7 topics for Thomson Reuters news stories).

Table E2: EPS indicators.

Variable	N. Transf.	N. Time Int.
SUE	4 ^a	3
grand total (4 x 3)		12

^a: We apply the following 4 transformations: flag for announcement, $\text{abs}(x)$, $\text{max}(0, x)$, and $\text{abs}(\text{min}(0, x))$. x stands for standardized surprise.

Table E3: Macro-indicators.

Variable	N. Transf.	N. Time Int.
Std_Macro	4	3
total for each macro-announcement		12
grand total (12 x 23 macro-announcements)		276

F Additional Tables: Matching analysis

Table F1: StreetAccount news stories and jumps matching.

Topic		All	Pos	Neg
Newspapers	$P(J N)$	0.33	0.07	0.25
	$\text{median}(J N)$	2.23	0.98	3.09
	$P(N J)$	0.21	0.16	0.29
M&A	$P(J N)$	0.26	0.13	0.13
	$\text{median}(J N)$	2.25	3.08	2.65
	$P(N J)$	0.29	0.30	0.27
Earnings Related	$P(J N)$	0.23	0.09	0.14
	$\text{median}(J N)$	1.41	1.52	1.49
	$P(N J)$	1.04	0.74	1.45
All	$P(J N)$	0.20	0.11	0.09
	$\text{median}(J N)$	1.49	1.58	1.87
	$P(N J)$	4.04	3.94	4.12
Litigation	$P(J N)$	0.16	0.08	0.07
	$\text{median}(J N)$	2.06	1.88	1.59
	$P(N J)$	0.19	0.21	0.15
Regulatory	$P(J N)$	0.15	0.04	0.11
	$\text{median}(J N)$	1.60	1.37	1.82
	$P(N J)$	0.14	0.13	0.14
Up/Downgrades	$P(J N)$	0.09	0.03	0.06
	$\text{median}(J N)$	1.35	0.86	1.84
	$P(N J)$	0.21	0.19	0.23

Notes: $P(J|N)$, $\text{median}(J|N)$ and $P(N|J)$ for StreetAccount news stories. Topics are sorted in descending order by $P(J|N)$.

Table F2: Thomson Reuters news stories and jumps matching.

Topic		All	Pos	Neg
Top	$P(J N)$	2.47	1.35	1.11
	$\text{median}(J N)$	1.54	1.69	2.20
	$P(N J)$	0.43	0.47	0.35
Earnings Pre-Ann.	$P(J N)$	1.12	0.27	0.84
	$\text{median}(J N)$	1.52	1.79	1.62
	$P(N J)$	0.32	0.17	0.50
Financial	$P(J N)$	0.63	0.55	0.08
	$\text{median}(J N)$	0.97	0.88	1.57
	$P(N J)$	0.59	0.86	0.26
Dividends	$P(J N)$	0.52	0.38	0.14
	$\text{median}(J N)$	0.98	0.96	1.05
	$P(N J)$	0.53	0.87	0.16
High	$P(J N)$	0.31	0.15	0.16
	$\text{median}(J N)$	1.51	1.55	1.82
	$P(N J)$	1.21	1.23	1.19
Medium	$P(J N)$	0.28	0.14	0.14
	$\text{median}(J N)$	1.47	1.52	1.79
	$P(N J)$	1.26	1.28	1.23
All	$P(J N)$	0.15	0.08	0.07
	$\text{median}(J N)$	1.34	1.29	1.63
	$P(N J)$	1.55	1.69	1.37

Notes: $P(J|N)$, $\text{median}(J|N)$ and $P(N|J)$ for Thomson Reuters news stories. Topics are sorted in descending order by $P(J|N)$.

Table F3: Macro-announcements and jumps matching 1/2.

Announcement		All	Pos	Neg
FOMC	$P(J N)$	0.72	0.26	0.46
Rate Decision	median($J N$)	0.74	0.75	0.78
	$P(N J)$	5.75	3.87	8.34
Federal	$P(J N)$	0.10	0.03	0.08
Budget	median($J N$)	0.59	0.70	0.57
	$P(N J)$	1.27	0.59	2.08
NAHB	$P(J N)$	0.03	0.02	0.01
	median($J N$)	0.83	0.91	0.75
	$P(N J)$	0.43	0.44	0.40
Consumer	$P(J N)$	0.03	0.02	0.01
Credit	median($J N$)	0.69	0.74	0.66
	$P(N J)$	0.38	0.38	0.36
Natural Gas	$P(J N)$	0.02	0.01	0.01
Stocks	median($J N$)	1.25	1.39	1.13
	$P(N J)$	0.85	0.79	0.94
Michigan	$P(J N)$	0.02	0.01	0.01
Sentiment	median($J N$)	1.08	1.08	1.23
	$P(N J)$	0.54	0.64	0.44
ISM	$P(J N)$	0.02	0.01	0.01
Manufacturing	median($J N$)	1.20	1.66	0.93
	$P(N J)$	0.30	0.29	0.33
Philly Fed	$P(J N)$	0.02	0.02	0.00
Business	median($J N$)	0.84	0.91	0.37
	$P(N J)$	0.29	0.49	0.08
Chicago PMI	$P(J N)$	0.02	0.01	0.01
	median($J N$)	0.99	0.99	1.04
	$P(N J)$	0.24	0.26	0.25
Construction	$P(J N)$	0.02	0.01	0.01
Spending	median($J N$)	1.54	2.18	1.06
	$P(N J)$	0.23	0.24	0.24
Factory	$P(J N)$	0.02	0.01	0.01
Orders	median($J N$)	1.45	1.55	1.39
	$P(N J)$	0.19	0.12	0.24
New	$P(J N)$	0.02	0.01	0.01
Home Sales	median($J N$)	1.10	1.29	0.81
	$P(N J)$	0.18	0.14	0.21

Table F3: Macro-announcements and jumps matching 2/2.

Announcement		All	Pos	Neg
ECRI Weekly Index	$P(J N)$	0.01	0.01	0.01
	$\text{median}(J N)$	0.95	1.11	0.76
	$P(N J)$	0.79	0.69	0.84
Oil Stocks	$P(J N)$	0.01	0.01	0.00
	$\text{median}(J N)$	1.08	1.31	0.96
	$P(N J)$	0.50	0.48	0.54
Business Inventories	$P(J N)$	0.01	0.01	0.01
	$\text{median}(J N)$	0.92	0.72	1.12
	$P(N J)$	0.17	0.15	0.17
Consumer Confidence	$P(J N)$	0.01	0.01	0.00
	$\text{median}(J N)$	0.93	1.08	0.69
	$P(N J)$	0.15	0.18	0.12
IBD Economic Optimism	$P(J N)$	0.01	0.00	0.01
	$\text{median}(J N)$	0.75	0.93	0.62
	$P(N J)$	0.14	0.13	0.16
Existing Home Sales	$P(J N)$	0.01	0.01	0.00
	$\text{median}(J N)$	1.19	1.27	0.74
	$P(N J)$	0.13	0.17	0.05
Leading Index	$P(J N)$	0.01	0.00	0.00
	$\text{median}(J N)$	0.90	0.86	0.93
	$P(N J)$	0.12	0.10	0.14
Pending Home Sales	$P(J N)$	0.01	0.00	0.01
	$\text{median}(J N)$	0.67	0.72	0.64
	$P(N J)$	0.09	0.04	0.15
Employment Trends	$P(J N)$	0.01	0.01	0.00
	$\text{median}(J N)$	0.66	0.52	1.09
	$P(N J)$	0.07	0.10	0.03
New York NAPM Index	$P(J N)$	0.01	0.01	0.00
	$\text{median}(J N)$	0.87	0.87	-
	$P(N J)$	0.04	0.08	0.00
Wholesale Inventories	$P(J N)$	0.00	0.00	0.00
	$\text{median}(J N)$	2.66	3.02	2.31
	$P(N J)$	0.03	0.02	0.04

Notes: $P(J|N)$, $\text{median}(J|N)$ and $P(N|J)$ for macro-announcements. Announcements are sorted in descending order by $P(J|N)$. The announcements Existing Home Sales, Philly Fed Business Index, New York NAPM Index, and Wholesale Inventories are not shown because there are no jump-news coincidences for them.

G Penalized Logistic Regression and Imbalanced Sample

Given the binary variable y_t that is equal to 1 for jumps occurrence and 0 otherwise and the vector of news-based indicators x_t , the objective function for the elastic net of Zou and Hastie (2005) for a logistic regression is:

$$(\widehat{\beta}_0, \widehat{\beta}) = \operatorname{argmin} - \left[\frac{1}{T} \sum_{t=1}^T y_t(\beta_0 + \beta' x_t) - \log(1 + e^{\beta_0 + \beta' x_t}) \right] + \lambda [(1 - \alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1] \quad (24)$$

where β_0 is the intercept, β is the vector of coefficients associated with the p regressors, $\lambda \geq 0$ is a complexity parameter and $0 \leq \alpha \leq 1$ is a compromise between ridge ($\alpha = 0$) and lasso ($\alpha = 1$).

λ is chosen among a grid of 1000 values, and is set equal to the value maximizing the area under the ROC curve (AUC). We evaluate the models on a grid of 11 values for α : 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.

Additionally, we estimate the two-stage adaptive lasso of Zou (2006). It has the following objective function:

$$(\widehat{\beta}_0, \widehat{\beta}) = \operatorname{argmin} - \left[\frac{1}{T} \sum_{t=1}^T y_t(\beta_0 + \beta' x_t) - \log(1 + e^{\beta_0 + \beta' x_t}) \right] + \lambda \sum_{i=1}^p \left| \frac{\beta_i}{\widehat{\beta}_{i,\text{initial}}} \right| \quad (25)$$

where $\widehat{\beta}_{i,\text{initial}}$ is the first-stage estimator of β_i from equation (24) with $\alpha = 1$, and λ is selected as for elastic net.

We first eliminate the possible bounce-back effects by deleting any jump that follows by a maximum of half an hour a jump with an opposite sign. Then, we use several techniques to address the issue of IS in the context of binary classification:

- Standard.
- Cost-sensitive: During each training, the weight (misclassification cost) associated with each observation is equal to the inverse of the size of its class.
- Over-sampling: During each training, observations of the minority class are randomly replicated in order to obtain a balanced sample (50% minority class, 50% majority class). Test sets are left unchanged.
- Under-sampling: During each training, observations of the majority class are randomly removed in order to obtain a balanced sample. Test sets are left unchanged.
- Synthetic sampling: During each training, we employ SMOTE (Synthetic Minority Oversampling TEchnique (Chawla et al. (2002))). It artificially generates new minority examples by interpolating between two randomly selected existing ones among

k -nearest neighbors.²⁴ We set the percentage of extra cases generated from the minority class equal to 200%, the percentage of cases that are selected from the majority class for each case generated from the minority class equal to 200%, and $k = 5$. Test sets are left unchanged.

The cost of misclassifying a jump as a non-jump is much higher than the cost of the reverse error, but there is not an established misclassification cost for jumps since it varies with, e.g., the aim and the time horizon of an investor. We want to prevent the algorithm from being biased toward the correct prediction of the majority class, therefore when applying the sampling techniques we aim to achieve a balanced sample with equal class distribution, and when applying cost-sensitive learning we assign to each observation a misclassification cost equal to the inverse of the size of its class.

Each approach is applied on 5-fold block cross-validation with blocks of contiguous time. When validation data are randomly selected for cross-validation from the entire time domain, training and validation data from nearby locations will be dependent. Consequently, if the objective is to project outside the structure of the training data, error estimates from random cross-validations will be overly optimistic (overfitting). To address this, blocks of contiguous time can be designed to better ensure independence between cross-validation folds and to achieve more reliable error estimates and higher forecasting performance (Burman et al. (1994); Racine (2000); Bergmeir and Benitez (2012)). The first half of the sample is completely sacrificed for training, and coincides with the first train set. The second half is divided into 5 contiguous folds, which constitute the test sets. In the spirit of the $h\nu$ -block of Racine (2000), we remove the observations belonging to the first 5 days of each test set to reduce the dependence of the training sets and the test sets. The model is trained on the first train set and evaluated on the adjacent first test set; then, the model is trained on the second train set, obtained by rolling forward the first train set by the length of the test set (1/10 of the whole sample), and the evaluation is performed on the adjacent second test set. The procedure continues until the last training and evaluation takes place, and the mean of the AUC across the 5 evaluations is stored.

We employ the R packages *glmnet* (Friedman et al. (2010)), *caret* (Kuhn (2008)) and *DMwR* (Torgo (2010)).

²⁴Chawla et al. (2002) interpret the underlying effect in terms of decision regions in feature space and show that this approach effectively forces the decision region of the minority class to become more general. Their approach is effective because the generated synthetic examples are relatively close in feature space to existing examples from the minority class. We point out that in our study by synthesizing new jump samples through linear interpolation we are going to create examples from the minority class that can be associated to both plausible and implausible covariates in the form of news-based indicators. Examples of plausible indicators are EPS- or macro-surprises with intermediate values among the interpolated ones, while examples of implausible indicators are binary variables such as those representing the announcement of a particular news with values between zero and one. We believe that regularization methods, being able to effectively select the useful covariates and to discard the remaining ones, overcome this issue.

H Additional Tables: Penalized Logistic Regression

Table H1: Average AUC for each combination of IS technique and α . Positive jumps.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5748	0.5425	0.5425	0.5971	0.6076
0.1	0.5966*	0.6338*	0.6344**	0.6019	0.6184*
0.2	0.5912	0.6321	0.6320	0.6020	0.6182
0.3	0.5903	0.6306	0.6306	0.6019	0.6179
0.4	0.5870	0.6279	0.6279	0.6009	0.6178
0.5	0.5843	0.6261	0.6262	0.6009	0.6176
0.6	0.5708	0.6246	0.6247	0.6006	0.6176
0.7	0.5766	0.6241	0.6241	0.6005	0.6172
0.8	0.5423	0.6222	0.6222	0.6006	0.6168
0.9	0.5458	0.6152	0.6156	0.6006	0.6155
1.0	0.5179	0.6013	0.6015	0.5996	0.6161
Ad. Lasso	0.5189	0.6192	0.6203	0.6027*	0.5982
Average	0.5664	0.6166	0.6168	0.6008	0.6149

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only positive jumps.

Table H2: Average AUC for each combination of IS technique and α . Negative jumps.

Alpha	Standard	Cost Sensitive	Overampling	Undersampling	Synthetic Sampling
0.0	0.5535	0.5189	0.5189	0.5568	0.5665
0.1	0.5890*	0.5965	0.5965	0.5664	0.5815*
0.2	0.5870	0.5964	0.5964	0.5665	0.5810
0.3	0.5816	0.5968*	0.5969**	0.5665	0.5810
0.4	0.5790	0.5960	0.5960	0.5664	0.5808
0.5	0.5766	0.5958	0.5958	0.5664	0.5809
0.6	0.5693	0.5956	0.5956	0.5664	0.5808
0.7	0.5671	0.5935	0.5935	0.5664	0.5809
0.8	0.5615	0.5928	0.5928	0.5664	0.5806
0.9	0.5572	0.5913	0.5913	0.5663	0.5804
1.0	0.5530	0.5890	0.5889	0.5672*	0.5799
Ad. Lasso	0.5496	0.5790	0.5828	0.5615	0.5640
Average	0.5687	0.5868	0.5871	0.5653	0.5782

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α for each IS technique; ** indicates the global maximum across IS techniques and α . Only negative jumps.

Table H3: Average AUC for each combination of IS technique and α . All jumps. Sample: from January 2005 to June 2007.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5043	0.4854	0.4853	0.5716	0.5459
0.1	0.5835*	0.5943**	0.5943**	0.5788	0.5652*
0.2	0.5802	0.5920	0.5920	0.5787	0.5639
0.3	0.5794	0.5908	0.5907	0.5787	0.5641
0.4	0.5777	0.5900	0.5901	0.5788	0.5641
0.5	0.5748	0.5898	0.5897	0.5788	0.5638
0.6	0.5742	0.5892	0.5893	0.5788	0.5639
0.7	0.5664	0.5891	0.5891	0.5790	0.5639
0.8	0.5619	0.5877	0.5879	0.5790	0.5639
0.9	0.5564	0.5872	0.5873	0.5781	0.5640
1.0	0.5437	0.5795	0.5797	0.5782	0.5643
Ad. Lasso	0.5380	0.5773	0.5802	0.5849*	0.5631
Average	0.5617	0.5793	0.5796	0.5786	0.5625

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. All jumps. Sample: from January 2005 to June 2007.

Table H4: Average AUC for each combination of IS technique and α . Positive jumps. Sample: from January 2005 to June 2007.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4647	0.4459	0.4459	0.5170	0.5841
0.1	0.5340*	0.5462*	0.5463*	0.5231	0.5984**
0.2	0.5332	0.5455	0.5455	0.5231	0.5983
0.3	0.5310	0.5440	0.5441	0.5231	0.5984*
0.4	0.5280	0.5434	0.5434	0.5231	0.5984*
0.5	0.5219	0.5434	0.5434	0.5230	0.5981
0.6	0.5199	0.5429	0.5431	0.5229	0.5950
0.7	0.5164	0.5389	0.5390	0.5230	0.5950
0.8	0.5137	0.5385	0.5386	0.5229	0.5950
0.9	0.5104	0.5385	0.5389	0.5228	0.5951
1.0	0.4775	0.5360	0.5363	0.5232*	0.5956
Ad. Lasso	0.4755	0.5357	0.5363	0.5193	0.5593
Average	0.5105	0.5332	0.5334	0.5222	0.5926

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only positive jumps. Sample: from January 2005 to June 2007.

Table H5: Average AUC for each combination of IS technique and α . Negative jumps. Sample: from January 2005 to June 2007.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4704	0.4550	0.4550	0.5836	0.6271
0.1	0.6060*	0.6119*	0.6120*	0.5929	0.6404
0.2	0.6044	0.6109	0.6109	0.5931	0.6404
0.3	0.5961	0.6108	0.6108	0.5930	0.6406
0.4	0.5940	0.6107	0.6107	0.5930	0.6408
0.5	0.5924	0.6104	0.6106	0.5930	0.6408
0.6	0.5906	0.6104	0.6104	0.5930	0.6406
0.7	0.5823	0.6098	0.6098	0.5930	0.6405
0.8	0.5805	0.6028	0.6028	0.5931	0.6409
0.9	0.5769	0.6011	0.6011	0.5930	0.6408
1.0	0.5744	0.5977	0.5977	0.5933*	0.6415**
Ad. Lasso	0.5684	0.5936	0.5940	0.5882	0.6414
Average	0.5781	0.5938	0.5938	0.5919	0.6397

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only negative jumps. Sample: from January 2005 to June 2007.

Table H6: Average AUC for each combination of IS technique and α . All jumps. Sample: from July 2007 to June 2009.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5513	0.5301	0.5301	0.5649	0.5938
0.1	0.5998*	0.6108**	0.6108**	0.5694	0.5972*
0.2	0.5995	0.6060	0.6060	0.5694	0.5969
0.3	0.5994	0.6057	0.6057	0.5694	0.5956
0.4	0.5977	0.6055	0.6054	0.5694	0.5960
0.5	0.5922	0.6050	0.6052	0.5694	0.5961
0.6	0.5913	0.6051	0.6050	0.5694	0.5960
0.7	0.5942	0.6051	0.6049	0.5694	0.5960
0.8	0.5867	0.6037	0.6037	0.5694	0.5953
0.9	0.5685	0.6015	0.6015	0.5694	0.5894
1.0	0.5506	0.6006	0.6008	0.5697	0.5901
Ad. Lasso	0.5491	0.6001	0.6002	0.5704*	0.5884
Average	0.5817	0.5983	0.5983	0.5692	0.5942

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. All jumps. Sample: from July 2007 to June 2009.

Table H7: Average AUC for each combination of IS technique and α . Positive jumps. Sample: from July 2007 to June 2009.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5786	0.5635	0.5635	0.6454	0.5895
0.1	0.6470*	0.6674**	0.6673*	0.6477	0.5962*
0.2	0.6464	0.6666	0.6666	0.6477	0.5949
0.3	0.6458	0.6608	0.6607	0.6477	0.5949
0.4	0.6414	0.6608	0.6608	0.6479	0.5949
0.5	0.6371	0.6607	0.6608	0.6479	0.5949
0.6	0.6358	0.6607	0.6605	0.6479	0.5949
0.7	0.6353	0.6608	0.6607	0.6477	0.5949
0.8	0.6252	0.6609	0.6608	0.6479	0.5950
0.9	0.6117	0.6608	0.6607	0.6479	0.5501
1.0	0.6061	0.6508	0.6509	0.6480*	0.5510
Ad. Lasso	0.6080	0.6506	0.6506	0.6480*	0.5419
Average	0.6265	0.6520	0.6520	0.6476	0.5828

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only positive jumps. Sample: from July 2007 to June 2009.

Table H8: Average AUC for each combination of IS technique and α . Negative jumps. Sample: from July 2007 to June 2009.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5085	0.4988	0.4988	0.4916	0.5060
0.1	0.5443*	0.5583**	0.5583**	0.4960	0.5210
0.2	0.5428	0.5579	0.5579	0.4961	0.5209
0.3	0.5411	0.5571	0.5571	0.4962	0.5209
0.4	0.5336	0.5561	0.5560	0.4959	0.5210
0.5	0.5300	0.5540	0.5540	0.4959	0.5210
0.6	0.5279	0.5535	0.5535	0.4959	0.5210
0.7	0.5254	0.5523	0.5523	0.4959	0.5210
0.8	0.5137	0.5524	0.5524	0.4959	0.5210
0.9	0.5066	0.5461	0.5461	0.4963	0.5210
1.0	0.4954	0.5422	0.5422	0.4966*	0.5211*
Ad. Lasso	0.4949	0.5438	0.5438	0.4947	0.5046
Average	0.5220	0.5477	0.5477	0.4956	0.5184

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only negative jumps. Sample: from July 2007 to June 2009.

Table H9: Average AUC for each combination of IS technique and α . All jumps. Sample: from July 2009 to December 2012.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4573	0.4458	0.4458	0.4936	0.4902
0.1	0.5443*	0.5546	0.5545	0.5034	0.5152
0.2	0.5432	0.5545	0.5546	0.5034	0.5151
0.3	0.5407	0.5546	0.5546	0.5034	0.5153*
0.4	0.5375	0.5545	0.5546	0.5034	0.5153*
0.5	0.5317	0.5548*	0.5548**	0.5035	0.5153*
0.6	0.5309	0.5546	0.5546	0.5034	0.5151
0.7	0.5231	0.5546	0.5546	0.5034	0.5152
0.8	0.5162	0.5546	0.5546	0.5034	0.5151
0.9	0.5065	0.5539	0.5539	0.4996	0.5148
1.0	0.4943	0.5498	0.5498	0.5038*	0.5152
Ad. Lasso	0.4860	0.5420	0.5423	0.5009	0.4934
Average	0.5176	0.5440	0.5441	0.5021	0.5113

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. All jumps. Sample: from July 2009 to December 2012.

Table H10: Average AUC for each combination of IS technique and α . Positive jumps. Sample: from July 2009 to December 2012.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4475	0.4420	0.4420	0.4956	0.5200
0.1	0.5488*	0.5587**	0.5587**	0.4976	0.5272
0.2	0.5467	0.5575	0.5575	0.4977	0.5272
0.3	0.5418	0.5567	0.5567	0.4977	0.5271
0.4	0.5382	0.5561	0.5563	0.4977	0.5271
0.5	0.5303	0.5562	0.5561	0.4977	0.5268
0.6	0.5283	0.5562	0.5562	0.4977	0.5268
0.7	0.5162	0.5562	0.5562	0.4977	0.5269
0.8	0.5136	0.5562	0.5562	0.4977	0.5269
0.9	0.5057	0.5548	0.5548	0.4977	0.5269
1.0	0.4865	0.5526	0.5526	0.4980	0.5272*
Ad. Lasso	0.4817	0.5425	0.5426	0.5003*	0.5109
Average	0.5154	0.5455	0.5455	0.4978	0.5251

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only positive jumps. Sample: from July 2009 to December 2012.

Table H11: Average AUC for each combination of IS technique and α . Negative jumps. Sample: from July 2009 to December 2012.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4349	0.4009	0.4009	0.4634	0.4763
0.1	0.5017*	0.5142**	0.5142**	0.4662	0.4766
0.2	0.4908	0.5140	0.5140	0.4663	0.4767
0.3	0.4853	0.5129	0.5129	0.4663	0.4766
0.4	0.4799	0.5129	0.5129	0.4663	0.4766
0.5	0.4679	0.5121	0.5121	0.4662	0.4766
0.6	0.4654	0.5119	0.5119	0.4662	0.4766
0.7	0.4577	0.5116	0.5116	0.4663	0.4765
0.8	0.4551	0.5045	0.5046	0.4663	0.4768
0.9	0.4439	0.5012	0.5012	0.4663	0.4766
1.0	0.4418	0.4921	0.4921	0.4664	0.4774*
Ad. Lasso	0.4379	0.4966	0.4967	0.4690*	0.4720
Average	0.4635	0.4988	0.4988	0.4663	0.4763

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only negative jumps. Sample: from July 2009 to December 2012.

Table H12: Average AUC for each combination of IS technique and α . All jumps. Sample: from January 2013 to February 2015.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4801	0.4748	0.4748	0.4732	0.5161
0.1	0.5281*	0.5354*	0.5355**	0.4752	0.5233
0.2	0.5280	0.5349	0.5349	0.4752	0.5233
0.3	0.5272	0.5345	0.5345	0.4753	0.5233
0.4	0.5248	0.5341	0.5341	0.4753	0.5234*
0.5	0.5214	0.5339	0.5339	0.4753	0.5234*
0.6	0.5206	0.5339	0.5339	0.4753	0.5231
0.7	0.5144	0.5340	0.5340	0.4753	0.5228
0.8	0.5137	0.5340	0.5340	0.4753	0.5228
0.9	0.4992	0.5329	0.5329	0.4753	0.5228
1.0	0.4842	0.5290	0.5291	0.4754	0.5230
Ad. Lasso	0.4892	0.5268	0.5265	0.4776*	0.5144
Average	0.5109	0.5282	0.5282	0.4753	0.5218

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. All jumps. Sample: from January 2013 to February 2015.

Table H13: Average AUC for each combination of IS technique and α . Positive jumps. Sample: from January 2013 to February 2015.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.4299	0.4094	0.4094	0.4542	0.6046
0.1	0.5082*	0.5301*	0.5302*	0.4564	0.6060**
0.2	0.5075	0.5300	0.5300	0.4564	0.6060**
0.3	0.5056	0.5293	0.5294	0.4564	0.6060**
0.4	0.4993	0.5291	0.5292	0.4564	0.6060**
0.5	0.4946	0.5245	0.5245	0.4564	0.6060**
0.6	0.4932	0.5247	0.5247	0.4564	0.6057
0.7	0.4840	0.5247	0.5249	0.4564	0.6054
0.8	0.4763	0.5209	0.5210	0.4564	0.6053
0.9	0.4641	0.5206	0.5206	0.4564	0.6053
1.0	0.4514	0.5195	0.5196	0.4565*	0.6054
Ad. Lasso	0.4504	0.5190	0.5190	0.4565*	0.5986
Average	0.4804	0.5152	0.5152	0.4562	0.6050

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only positive jumps. Sample: from January 2013 to February 2015.

Table H14: Average AUC for each combination of IS technique and α . Negative jumps. Sample: from January 2013 to February 2015.

Alpha	Standard	Cost Sensitive	Oversampling	Undersampling	Synthetic Sampling
0.0	0.5119	0.5103	0.5103	0.4918	0.4474
0.1	0.5723*	0.5735*	0.5736**	0.4950	0.4468
0.2	0.5712	0.5730	0.5731	0.4950	0.4468
0.3	0.5693	0.5728	0.5728	0.4950	0.4468
0.4	0.5629	0.5726	0.5726	0.4951*	0.4468
0.5	0.5581	0.5728	0.5728	0.4951*	0.4468
0.6	0.5560	0.5727	0.5727	0.4951*	0.4468
0.7	0.5475	0.5727	0.5728	0.4951*	0.4468
0.8	0.5387	0.5728	0.5728	0.4950	0.4468
0.9	0.5341	0.5728	0.5728	0.4950	0.4468
1.0	0.5283	0.5711	0.5711	0.4950	0.4471
Ad. Lasso	0.5266	0.5640	0.5655	0.4950	0.4681*
Average	0.5481	0.5668	0.5669	0.4948	0.4486

Notes: Average AUC across stocks for each type of IS technique and for each α ; * indicates the maximum AUC across α 's for each IS technique; ** indicates the global maximum across IS techniques and α 's. Only negative jumps. Sample: from January 2013 to February 2015.

I Additional Tables: Variables Importance

For simplicity and easiness of comparison, in the following cases we use the parameters of the best model for the full sample (i.e., $\alpha = 0.1$ and IS treatment = oversampling, see Table 7) when the AUC corresponding to the best combination of α and IS treatment approach is very close to that obtained with the full sample parameters, and we deviate only in the following cases: $\alpha = 0.1$ and synthetic sampling for positive jumps and subsamples Jan 2005 - Jun 2007 and Jan 2013 - Feb 2015, and $\alpha = 1.0$ and synthetic sampling for negative jumps and subsample Jan 2005 - Jun 2007. Appendix H shows the AUC for each combination of α and IS treatment approach for each case. The results are presented in the following tables.

I.1 All jumps, sorting by APE and absolute APE

Table II: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2209	0.0302
2	Macro	FOMC Rate Dec.	-1	announcement	0.1693	0.0170
3	StreetAcc.	All	0 and +1	persistence	0.0548	0.0108
4	StreetAcc.	All	0	announcement	0.0382	0.0021
5	StreetAcc.	All	-1 and 0	persistence	0.0360	0.0031
6	StreetAcc.	All	+1	sent. $\neq 0$	0.0265	0.0005
7	StreetAcc.	All	+1	announcement	0.0244	0.0014
8	Macro	Constr. Spending	0	announcement	0.0217	-0.0003
9	StreetAcc.	All	0	sent. $\neq 0$	0.0201	0.0007
10	StreetAcc.	All	+1	negative sent.	0.0175	0.0007
11	Macro	ISM Man. PMI	0	announcement	0.0163	0.0000
12	StreetAcc.	All	-1	negative sent.	0.0122	-0.0001
13	StreetAcc.	All	-1	sent. $\neq 0$	0.0092	-0.0005
14	StreetAcc.	All	0	negative sent.	0.0074	0.0007
15	T. Reuters	Financial	-1 and 0	persistence	0.0072	0.0003
16	StreetAcc.	Regulatory	+1	announcement	0.0064	0.0000
17	StreetAcc.	M&A	-1	sent. $\neq 0$	0.0057	0.0000
18	StreetAcc.	M&A	-1	positive sent.	0.0057	0.0000
19	StreetAcc.	Newspapers	0	announcement	0.0053	0.0000
20	T. Reuters	Financial	-1	announcement	0.0049	0.0003

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table I2: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2209	0.0302
2	Macro	FOMC Rate Dec.	-1	announcement	0.1693	0.0170
3	Macro	ECRI	-1	announcement	-0.1018	0.0015
4	Macro	Oil Stocks	-1	announcement	-0.1010	-0.0011
5	Macro	Nat. Gas Stocks	-1	announcement	-0.0869	0.0028
6	Macro	Michigan Sent.	-1	announcement	-0.0702	-0.0005
7	Macro	ECRI	+1	announcement	-0.0685	0.0004
8	Macro	Nat. Gas Stocks	+1	announcement	-0.0685	0.0024
9	Macro	Oil Stocks	+1	announcement	-0.0630	0.0001
10	Macro	Wholesale Inv.	-1	announcement	-0.0576	-0.0004
11	StreetAcc.	All	0 and +1	persistence	0.0548	0.0108
12	Macro	Cons. Confidence	-1	announcement	-0.0544	-0.0004
13	Macro	Ex. Home Sales	-1	announcement	-0.0516	-0.0001
14	Macro	NAHB	-1	announcement	-0.0507	0.0002
15	Macro	IBD Ec. Opt.	-1	announcement	-0.0501	0.0007
16	Macro	Empl. Trends	-1	announcement	-0.0495	0.0004
17	Macro	Business Inv.	-1	announcement	-0.0491	-0.0002
18	Macro	Pend. Home Sales	-1	announcement	-0.0454	0.0002
19	Macro	New Home Sales	-1	announcement	-0.0452	0.0010
20	Macro	Cons. Credit	-1	announcement	-0.0443	0.0002

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

I.2 Positive Jumps

Table I3: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2394	0.0524
2	StreetAcc.	All	0 and +1	persistence	0.0387	0.0061
3	Macro	Nat. Gas Stocks	+1	announcement	-0.0773	0.0053
4	StreetAcc.	All	-1 and 0	persistence	0.0097	0.0050
5	StreetAcc.	All	-1	announcement	-0.0220	0.0042
6	Macro	Factory Orders	-1	announcement	-0.0572	0.0038
7	Macro	FOMC Rate Dec.	-1	announcement	0.1041	0.0033
8	StreetAcc.	Earnings	-1 and 0	persistence	-0.0032	0.0029
9	Macro	Nat. Gas Stocks	-1	announcement	-0.1242	0.0028
10	Macro	ECRI	-1	announcement	-0.1253	0.0028
11	T. Reuters	All	-1 and 0	persistence	-0.0330	0.0024
12	StreetAcc.	Up/Downgrades	0 and +1	persistence	-0.0002	0.0022
13	Macro	ISM Man. PMI	0	abs(surprise)	0.0010	0.0021
14	Macro	ISM Man. PMI	0	announcement	0.0223	0.0020
15	Macro	New Home Sales	-1	announcement	-0.0619	0.0019
16	Macro	Oil Stocks	-1	abs(min(0,surprise))	-0.0012	0.0017
17	StreetAcc.	All	0	announcement	0.0302	0.0015
18	Macro	NAHB	-1	announcement	-0.0681	0.0012
19	T. Reuters	Dividends	-1 and 0	persistence	0.0016	0.0011
20	T. Reuters	Financial	-1 and 0	persistence	0.0066	0.0010

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table I4: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2394	0.0524
2	Macro	FOMC Rate Dec.	-1	announcement	0.1041	0.0033
3	StreetAcc.	All	0 and +1	persistence	0.0387	0.0061
4	StreetAcc.	All	0	announcement	0.0302	0.0015
5	Macro	Constr. Spending	0	announcement	0.0275	-0.0004
6	Macro	ISM Man. PMI	0	announcement	0.0223	0.0020
7	StreetAcc.	All	0	sent. $\neq 0$	0.0137	0.0004
8	StreetAcc.	All	+1	announcement	0.0123	0.0005
9	StreetAcc.	All	-1	negative sent.	0.0112	0.0000
10	StreetAcc.	All	-1 and 0	persistence	0.0097	0.0050
11	StreetAcc.	All	+1	positive sent.	0.0078	0.0000
12	T. Reuters	Financial	-1 and 0	persistence	0.0066	0.0010
13	StreetAcc.	All	0	positive sent.	0.0057	0.0004
14	StreetAcc.	All	0	negative sent.	0.0055	0.0000
15	StreetAcc.	Up/Downgrades	0	announcement	0.0050	0.0000
16	StreetAcc.	M&A	-1	sent. $\neq 0$	0.0047	0.0000
17	StreetAcc.	M&A	-1	positive sent.	0.0047	0.0000
18	StreetAcc.	Regulatory	0 and +1	persistence	0.0045	0.0000
19	T. Reuters	Financial	-1	announcement	0.0042	0.0004
20	StreetAcc.	Regulatory	+1	announcement	0.0040	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table I5: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2394	0.0524
2	Macro	ECRI	-1	announcement	-0.1253	0.0028
3	Macro	Oil Stocks	-1	announcement	-0.1245	-0.0014
4	Macro	Nat. Gas Stocks	-1	announcement	-0.1242	0.0028
5	Macro	FOMC Rate Dec.	-1	announcement	0.1041	0.0033
6	Macro	Michigan Sent.	-1	announcement	-0.0809	0.0003
7	Macro	Oil Stocks	+1	announcement	-0.0790	-0.0006
8	Macro	ECRI	+1	announcement	-0.0774	0.0002
9	Macro	Nat. Gas Stocks	+1	announcement	-0.0773	0.0053
10	Macro	IBD Ec. Opt.	-1	announcement	-0.0720	0.0010
11	Macro	Wholesale Inv.	-1	announcement	-0.0697	0.0000
12	Macro	Cons. Confidence	-1	announcement	-0.0693	-0.0005
13	Macro	NAHB	-1	announcement	-0.0681	0.0012
14	Macro	Empl. Trends	-1	announcement	-0.0642	-0.0001
15	Macro	New Home Sales	-1	announcement	-0.0619	0.0019
16	Macro	Business Inv.	-1	announcement	-0.0617	0.0000
17	Macro	Ex. Home Sales	-1	announcement	-0.0617	-0.0004
18	Macro	Leading Index	-1	announcement	-0.0603	-0.0002
19	Macro	Federal Budget	-1	announcement	-0.0601	0.0010
20	Macro	Oil Stocks	0	announcement	-0.0574	0.0004

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

I.3 Negative Jumps

Table I6: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.2038	0.0231
2	StreetAcc.	All	0 and +1	persistence	-0.0089	0.0120
3	StreetAcc.	All	-1 and 0	persistence	-0.0301	0.0056
4	Macro	IBD Ec. Opt.	-1	abs(surprise)	-0.0013	0.0054
5	Macro	Michigan Sent.	-1	abs(surprise)	-0.0120	0.0044
6	StreetAcc.	All	+1	announcement	-0.0084	0.0043
7	StreetAcc.	All	0	announcement	0.0088	0.0038
8	Macro	Federal Budget	-1	announcement	-0.0965	0.0038
9	StreetAcc.	All	-1	announcement	-0.0539	0.0023
10	T. Reuters	All	-1 and 0	persistence	-0.0823	0.0020
11	Macro	Factory Orders	-1	abs(min(0,surprise))	0.0000	0.0018
12	Macro	Oil Stocks	+1	abs(surprise)	-0.0029	0.0016
13	Macro	Cons. Confidence	0	announcement	-0.0675	0.0015
14	Macro	Oil Stocks	+1	announcement	-0.1247	0.0014
15	Macro	ECRI	+1	abs(min(0,surprise))	-0.0055	0.0014
16	Macro	Empl. Trends	-1	announcement	-0.1181	0.0013
17	StreetAcc.	All	+1	sent. \neq 0	-0.0050	0.0013
18	StreetAcc.	All	+1	negative sent.	0.0057	0.0013
19	Macro	FOMC Rate Dec.	0	announcement	0.1300	0.0012
20	Macro	Business Inv.	-1	abs(surprise)	-0.0017	0.0012

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table I7: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.2038	0.0231
2	Macro	FOMC Rate Dec.	0	announcement	0.1300	0.0012
3	StreetAcc.	All	0	negative sent.	0.0178	0.0010
4	StreetAcc.	Earnings	-1	positive sent.	0.0142	0.0000
5	StreetAcc.	Earnings	-1	sent. $\neq 0$	0.0122	0.0000
6	StreetAcc.	Earnings	+1	sent. $\neq 0$	0.0106	0.0000
7	StreetAcc.	All	0	announcement	0.0088	0.0038
8	StreetAcc.	All	0	sent. $\neq 0$	0.0084	0.0010
9	T. Reuters	Top	-1	announcement	0.0075	0.0010
10	T. Reuters	Financial	-1	announcement	0.0065	0.0000
11	T. Reuters	Top	-1 and 0	persistence	0.0065	0.0010
12	StreetAcc.	All	+1	negative sent.	0.0057	0.0013
13	T. Reuters	Dividends	0	announcement	0.0049	0.0000
14	T. Reuters	Financial	-1 and 0	persistence	0.0047	0.0000
15	StreetAcc.	Up/Downgrades	-1	sent. $\neq 0$	0.0046	0.0000
16	T. Reuters	High	-1	announcement	0.0046	0.0010
17	T. Reuters	Dividends	0	sent. $\neq 0$	0.0041	0.0000
18	T. Reuters	Financial	0	sent. $\neq 0$	0.0038	0.0000
19	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0037	0.0000
20	T. Reuters	Dividends	0	positive sent.	0.0037	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table I8: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.2038	0.0231
2	Macro	ECRI	-1	announcement	-0.1711	0.0010
3	Macro	Oil Stocks	-1	announcement	-0.1677	-0.0024
4	Macro	Nat. Gas Stocks	-1	announcement	-0.1410	0.0004
5	Macro	FOMC Rate Dec.	0	announcement	0.1300	0.0012
6	Macro	Nat. Gas Stocks	+1	announcement	-0.1296	-0.0004
7	Macro	Oil Stocks	+1	announcement	-0.1247	0.0014
8	Macro	ECRI	+1	announcement	-0.1182	0.0004
9	Macro	Empl. Trends	-1	announcement	-0.1181	0.0013
10	Macro	Cons. Confidence	-1	announcement	-0.1177	-0.0005
11	Macro	Ex. Home Sales	-1	announcement	-0.1167	-0.0001
12	Macro	Wholesale Inv.	-1	announcement	-0.1164	-0.0004
13	Macro	New Home Sales	-1	announcement	-0.1143	-0.0005
14	Macro	Factory Orders	-1	announcement	-0.1136	-0.0002
15	Macro	Phil. Fed	-1	announcement	-0.1114	0.0004
16	Macro	Business Inv.	-1	announcement	-0.1101	-0.0002
17	Macro	Pend. Home Sales	-1	announcement	-0.1078	-0.0005
18	Macro	IBD Ec. Opt.	-1	announcement	-0.1078	-0.0001
19	Macro	Cons. Credit	+1	announcement	-0.1062	-0.0001
20	Macro	Oil Stocks	0	announcement	-0.1048	-0.0003

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

I.4 Sub-Sample Analysis

I.4.1 Jan 2005 - Jun 2007 (pre-global financial crisis)

Table I9: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.1780	0.0344
2	Macro	Nat. Gas Stocks	-1	announcement	-0.1247	0.0210
3	StreetAcc.	All	-1 and 0	persistence	-0.0328	0.0149
4	Macro	FOMC Rate Dec.	0	announcement	0.0281	0.0098
5	Macro	Cons. Credit	-1	announcement	-0.0840	0.0046
6	StreetAcc.	All	0 and +1	persistence	-0.0165	0.0042
7	Macro	Factory Orders	0	announcement	-0.0666	0.0042
8	Macro	ECRI	-1	announcement	-0.1387	0.0032
9	Macro	Constr. Spending	-1	max(0,surprise)	-0.0005	0.0028
10	Macro	Oil Stocks	-1	announcement	-0.1315	0.0027
11	StreetAcc.	All	0	sent. \neq 0	0.0132	0.0024
12	StreetAcc.	All	+1	announcement	-0.0195	0.0023
13	StreetAcc.	All	0	announcement	0.0064	0.0023
14	Macro	Oil Stocks	-1	max(0,surprise)	-0.0005	0.0022
15	Macro	ECRI	-1	abs(min(0,surprise))	-0.0149	0.0021
16	Macro	Business Inv.	0	announcement	-0.0458	0.0018
17	StreetAcc.	All	0	positive sent.	0.0131	0.0017
18	Macro	Factory Orders	-1	announcement	-0.0952	0.0016
19	StreetAcc.	Earnings	-1 and 0	persistence	-0.0060	0.0014
20	Macro	ECRI	-1	abs(surprise)	-0.0844	0.0013

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 83 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 73 assets for which, in addition to the condition above, EPS are released during market hours.

Table I10: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.1780	0.0344
2	Macro	FOMC Rate Dec.	0	announcement	0.0281	0.0098
3	StreetAcc.	All	0	sent. $\neq 0$	0.0132	0.0024
4	StreetAcc.	All	0	positive sent.	0.0131	0.0017
5	StreetAcc.	All	-1	negative sent.	0.0123	-0.0001
6	StreetAcc.	All	0	announcement	0.0064	0.0023
7	StreetAcc.	All	-1	n. news ≥ 2	0.0061	0.0011
8	StreetAcc.	All	+1	sent. $\neq 0$	0.0058	0.0000
9	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0054	0.0000
10	StreetAcc.	EPS	-1	announcement	0.0040	0.0000
11	StreetAcc.	Up/Downgrades	-1	sent. $\neq 0$	0.0038	0.0000
12	StreetAcc.	Earnings	+1	sent. $\neq 0$	0.0037	0.0000
13	StreetAcc.	Earnings	+1	negative sent.	0.0036	0.0000
14	StreetAcc.	All	+1	positive sent.	0.0035	0.0000
15	T. Reuters	Top	-1 and 0	persistence	0.0033	0.0000
16	T. Reuters	Top	-1	announcement	0.0032	0.0000
17	StreetAcc.	Up/Downgrades	-1	positive sent.	0.0030	0.0000
18	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0028	0.0000
19	T. Reuters	Financial	-1 and 0	persistence	0.0026	0.0000
20	StreetAcc.	All	0	n. news ≥ 2	0.0021	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 83 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 73 assets for which, in addition to the condition above, EPS are released during market hours.

Table I11: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.1780	0.0344
2	Macro	ECRI	-1	announcement	-0.1387	0.0032
3	Macro	Oil Stocks	-1	announcement	-0.1315	0.0027
4	Macro	Nat. Gas Stocks	-1	announcement	-0.1247	0.0210
5	Macro	Oil Stocks	+1	announcement	-0.1098	-0.0009
6	Macro	Michigan Sent.	-1	announcement	-0.1090	-0.0011
7	Macro	Nat. Gas Stocks	+1	announcement	-0.1063	-0.0007
8	Macro	ECRI	+1	announcement	-0.1026	-0.0007
9	Macro	Phil. Fed	-1	announcement	-0.0979	-0.0006
10	Macro	Cons. Confidence	-1	announcement	-0.0961	0.0002
11	Macro	Factory Orders	-1	announcement	-0.0952	0.0016
12	Macro	Wholesale Inv.	-1	announcement	-0.0930	-0.0006
13	Macro	IBD Ec. Opt.	-1	announcement	-0.0922	-0.0006
14	Macro	NAHB	-1	announcement	-0.0903	-0.0005
15	Macro	New Home Sales	-1	announcement	-0.0896	-0.0002
16	Macro	Leading Index	-1	announcement	-0.0874	-0.0007
17	Macro	Ex. Home Sales	-1	announcement	-0.0870	-0.0008
18	Macro	Oil Stocks	0	announcement	-0.0858	-0.0003
19	Macro	ECRI	-1	abs(surprise)	-0.0844	0.0013
20	Macro	Federal Budget	+1	announcement	-0.0842	-0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 83 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 73 assets for which, in addition to the condition above, EPS are released during market hours.

Table I12: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0 and +1	persistence	0.1017	0.0341
2	StreetAcc.	All	-1 and 0	persistence	0.1450	0.0101
3	Macro	Phil. Fed	-1	abs(surprise)	-0.0063	0.0083
4	Macro	Business Inv.	0	announcement	0.0147	0.0077
5	Macro	ECRI	-1	max(0,surprise)	0.0197	0.0074
6	Macro	FOMC Rate Dec.	0	announcement	0.1460	0.0061
7	Macro	Nat. Gas Stocks	-1	announcement	0.0447	0.0056
8	Macro	FOMC Rate Dec.	-1	announcement	0.1860	0.0029
9	T. Reuters	All	-1 and 0	persistence	0.0211	0.0006
10	StreetAcc.	Earnings	-1 and 0	persistence	0.0143	0.0003
11	StreetAcc.	All	-1	announcement	0.1025	0.0003
12	Macro	Oil Stocks	-1	abs(surprise)	0.0258	0.0003
13	StreetAcc.	Litigation	-1 and 0	persistence	0.0007	0.0003
14	StreetAcc.	All	0	announcement	0.1017	0.0002
15	Macro	Oil Stocks	-1	max(0,surprise)	0.0178	0.0001
16	Macro	Federal Budget	-1	abs(surprise)	-1.7814	0.0001
17	Macro	Federal Budget	-1	announcement	0.0300	0.0001
18	StreetAcc.	All	+1	announcement	0.0375	0.0001
19	T. Reuters	All	0 and +1	persistence	0.0161	0.0001
20	Macro	Oil Stocks	-1	abs(min(0,surprise))	-0.0047	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 25 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 20 assets for which, in addition to the condition above, EPS are released during market hours.

Table I13: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.1860	0.0029
2	Macro	FOMC Rate Dec.	0	announcement	0.1460	0.0061
3	StreetAcc.	All	-1 and 0	persistence	0.1450	0.0101
4	StreetAcc.	All	-1	announcement	0.1025	0.0003
5	StreetAcc.	All	0 and +1	persistence	0.1017	0.0341
6	StreetAcc.	All	0	announcement	0.1017	0.0002
7	StreetAcc.	All	-1	negative sent.	0.0682	0.0000
8	Macro	Factory Orders	-1	announcement	0.0623	0.0000
9	StreetAcc.	All	+1	positive sent.	0.0572	0.0000
10	Macro	Nat. Gas Stocks	-1	abs(min(0,surprise))	0.0565	-0.0008
11	Macro	Cons. Credit	-1	announcement	0.0552	0.0001
12	StreetAcc.	All	0	sent. \neq 0	0.0529	0.0001
13	StreetAcc.	All	-1	sent. \neq 0	0.0527	0.0000
14	Macro	Nat. Gas Stocks	-1	announcement	0.0447	0.0056
15	StreetAcc.	All	0	positive sent.	0.0413	0.0000
16	StreetAcc.	All	+1	announcement	0.0375	0.0001
17	StreetAcc.	All	+1	sent. \neq 0	0.0372	0.0000
18	StreetAcc.	Up/Downgrades	-1 and 0	persistence	0.0370	0.0000
19	Macro	Factory Orders	-1	abs(surprise)	0.0317	-0.0001
20	Macro	Federal Budget	-1	announcement	0.0300	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 25 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 20 assets for which, in addition to the condition above, EPS are released during market hours.

Table I14: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Federal Budget	-1	abs(surprise)	-1.7814	0.0001
2	Macro	Federal Budget	-1	abs(min(0,surprise))	-0.8693	0.0000
3	Macro	FOMC Rate Dec.	-1	announcement	0.1860	0.0029
4	Macro	FOMC Rate Dec.	0	announcement	0.1460	0.0061
5	StreetAcc.	All	-1 and 0	persistence	0.1450	0.0101
6	StreetAcc.	All	-1	announcement	0.1025	0.0003
7	StreetAcc.	All	0 and +1	persistence	0.1017	0.0341
8	StreetAcc.	All	0	announcement	0.1017	0.0002
9	Macro	Cons. Confidence	-1	abs(surprise)	-0.0780	0.0000
10	StreetAcc.	All	-1	negative sent.	0.0682	0.0000
11	Macro	Factory Orders	-1	announcement	0.0623	0.0000
12	Macro	Oil Stocks	-1	announcement	-0.0583	0.0000
13	StreetAcc.	All	+1	positive sent.	0.0572	0.0000
14	Macro	Nat. Gas Stocks	-1	abs(min(0,surprise))	-0.0565	-0.0008
15	Macro	Cons. Credit	-1	announcement	0.0552	0.0001
16	StreetAcc.	All	0	sent. \neq 0	0.0529	0.0001
17	StreetAcc.	All	-1	sent. \neq 0	0.0527	0.0000
18	Macro	Cons. Confidence	-1	abs(min(0,surprise))	-0.0474	0.0000
19	Macro	Nat. Gas Stocks	-1	announcement	0.0447	0.0056
20	StreetAcc.	All	0	positive sent.	0.0413	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 25 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 20 assets for which, in addition to the condition above, EPS are released during market hours.

Table I15: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 1.0$ and synthetic sampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.3318	0.0208
2	StreetAcc.	All	0 and +1	persistence	0.0291	0.0075
3	StreetAcc.	All	+1	n. news ≥ 2	0.0651	0.0045
4	Macro	ECRI	-1	announcement	0.0115	0.0042
5	Macro	FOMC Rate Dec.	0	announcement	0.0125	0.0022
6	T. Reuters	All	+1	announcement	0.0000	0.0020
7	Macro	Nat. Gas Stocks	-1	announcement	-0.0290	0.0010
8	StreetAcc.	All	-1	sent. $\neq 0$	-0.0002	0.0007
9	StreetAcc.	All	-1 and 0	persistence	0.0371	0.0007
10	StreetAcc.	All	-1	announcement	0.0187	0.0004
11	Macro	NAHB	-1	abs(surprise)	-0.0041	0.0002
12	Macro	Federal Budget	-1	announcement	-0.0497	0.0001
13	Macro	NAHB	-1	announcement	0.0125	0.0001
14	Macro	Federal Budget	-1	abs(surprise)	0.0000	0.0001
15	StreetAcc.	All	-1	n. news ≥ 2	0.0296	0.0001
16	Macro	New Home Sales	-1	abs(surprise)	0.0000	0.0001
17	Macro	NAHB	-1	abs(min(0,surprise))	0.0000	0.0001
18	Macro	ECRI	0	announcement	0.0250	0.0001
19	Macro	ECRI	0	abs(surprise)	0.0000	0.0001
20	Macro	Oil Stocks	0	announcement	-0.0103	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 1.0$ and synthetic sampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 23 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 16 assets for which, in addition to the condition above, EPS are released during market hours.

Table I16: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 1.0$ and synthetic sampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.3318	0.0208
2	StreetAcc.	All	0	announcement	0.1165	0.0000
3	StreetAcc.	All	+1	announcement	0.1136	0.0000
4	StreetAcc.	All	+1	n. news ≥ 2	0.0651	0.0045
5	StreetAcc.	All	-1 and 0	persistence	0.0371	0.0007
6	StreetAcc.	All	+1	sent. $\neq 0$	0.0332	0.0000
7	StreetAcc.	Up/Downgrades	0	announcement	0.0329	0.0000
8	StreetAcc.	Litigation	-1	n. news ≥ 2	0.0322	0.0000
9	StreetAcc.	Litigation	+1	announcement	0.0320	0.0000
10	Macro	Cons. Credit	-1	announcement	0.0314	0.0000
11	StreetAcc.	All	-1	n. news ≥ 2	0.0296	0.0001
12	StreetAcc.	All	0 and +1	persistence	0.0291	0.0075
13	Macro	ECRI	0	announcement	0.0250	0.0001
14	Macro	Leading Index	0	announcement	0.0240	0.0000
15	StreetAcc.	All	-1	positive sent.	0.0224	-0.0010
16	StreetAcc.	All	0	n. news ≥ 2	0.0218	0.0000
17	StreetAcc.	All	-1	announcement	0.0187	0.0004
18	Macro	Leading Index	-1	announcement	0.0170	0.0000
19	Macro	NAHB	-1	announcement	0.0125	0.0001
20	Macro	FOMC Rate Dec.	0	announcement	0.0125	0.0022

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 1.0$ and synthetic sampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 23 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 16 assets for which, in addition to the condition above, EPS are released during market hours.

Table I17: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 1.0$ and synthetic sampling IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.3318	0.0208
2	Macro	Oil Stocks	-1	announcement	-0.1201	-0.0002
3	StreetAcc.	All	0	announcement	0.1165	0.0000
4	StreetAcc.	All	+1	announcement	0.1136	0.0000
5	Macro	Federal Budget	-1	abs(min(0,surprise))	-0.0695	0.0000
6	StreetAcc.	All	+1	n. news ≥ 2	0.0651	0.0045
7	Macro	Cons. Confidence	-1	announcement	-0.0551	0.0000
8	Macro	Federal Budget	-1	announcement	-0.0497	0.0001
9	StreetAcc.	All	-1 and 0	persistence	0.0371	0.0007
10	Macro	Phil. Fed	-1	announcement	-0.0370	-0.0002
11	Macro	Phil. Fed	+1	announcement	-0.0341	0.0000
12	StreetAcc.	All	+1	sent. $\neq 0$	0.0332	0.0000
13	StreetAcc.	Up/Downgrades	0	announcement	0.0329	0.0000
14	StreetAcc.	Litigation	-1	n. news ≥ 2	0.0322	0.0000
15	StreetAcc.	Litigation	+1	announcement	0.0320	0.0000
16	Macro	Cons. Credit	-1	announcement	0.0314	0.0000
17	StreetAcc.	All	-1	n. news ≥ 2	0.0296	0.0001
18	StreetAcc.	All	0 and +1	persistence	0.0291	0.0075
19	Macro	Nat. Gas Stocks	-1	announcement	-0.0290	0.0010
20	Macro	ECRI	0	announcement	0.0250	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 1.0$ and synthetic sampling IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 23 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 16 assets for which, in addition to the condition above, EPS are released during market hours.

I.4.2 Jul 2007 - Jun 2009 (global financial crisis)

Table I18: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Wholesale Inv.	-1	abs(min(0,surprise))	-0.0001	0.0151
2	Macro	ECRI	-1	announcement	-0.2074	0.0136
3	Macro	Wholesale Inv.	-1	max(0,surprise)	-0.0007	0.0113
4	StreetAcc.	All	-1 and 0	persistence	-0.1213	0.0108
5	StreetAcc.	All	0 and +1	persistence	-0.1269	0.0101
6	Macro	FOMC Rate Dec.	-1	announcement	0.0493	0.0079
7	Macro	Oil Stocks	0	abs(surprise)	-0.0037	0.0069
8	Macro	FOMC Rate Dec.	0	announcement	0.2522	0.0069
9	StreetAcc.	All	-1	announcement	-0.1457	0.0067
10	Macro	Nat. Gas Stocks	0	announcement	-0.1696	0.0066
11	Macro	Chicago PMI	-1	announcement	-0.1737	0.0050
12	Macro	Nat. Gas Stocks	0	abs(surprise)	-0.0070	0.0022
13	Macro	Federal Budget	-1	announcement	-0.2039	0.0022
14	Macro	Oil Stocks	-1	announcement	-0.2322	0.0020
15	StreetAcc.	All	0	announcement	-0.0194	0.0019
16	StreetAcc.	All	0	sent. \neq 0	0.0096	0.0019
17	StreetAcc.	All	0	negative sent.	0.0015	0.0019
18	StreetAcc.	All	+1	announcement	-0.1400	0.0018
19	StreetAcc.	Up/Downgrades	-1 and 0	persistence	-0.0242	0.0018
20	Macro	Wholesale Inv.	0	abs(surprise)	-0.0014	0.0017

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 75 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 67 assets for which, in addition to the condition above, EPS are released during market hours.

Table I19: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2522	0.0069
2	Macro	FOMC Rate Dec.	-1	announcement	0.0493	0.0079
3	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0099	0.0000
4	StreetAcc.	All	0	sent. \neq 0	0.0096	0.0019
5	StreetAcc.	Newspapers	-1	n. news \geq 2	0.0081	0.0000
6	StreetAcc.	Earnings	0	sent. \neq 0	0.0064	0.0000
7	StreetAcc.	Earnings	0	positive sent.	0.0054	0.0000
8	T. Reuters	Medium	-1	announcement	0.0045	0.0000
9	T. Reuters	Top	0 and +1	persistence	0.0041	0.0000
10	T. Reuters	Top	+1	announcement	0.0040	0.0000
11	T. Reuters	Earnings	0	announcement	0.0037	0.0000
12	T. Reuters	Earnings	-1	sent. \neq 0	0.0037	0.0000
13	T. Reuters	Dividends	-1	sent. \neq 0	0.0034	0.0000
14	StreetAcc.	Newspapers	-1	negative sent.	0.0033	0.0000
15	T. Reuters	Dividends	-1	positive sent.	0.0033	0.0000
16	StreetAcc.	Newspapers	-1	sent. \neq 0	0.0031	0.0000
17	T. Reuters	Financial	-1	sent. \neq 0	0.0031	0.0000
18	T. Reuters	Financial	-1	positive sent.	0.0030	0.0000
19	T. Reuters	High	0	sent. \neq 0	0.0029	0.0000
20	T. Reuters	Financial	0	sent. \neq 0	0.0029	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 75 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 67 assets for which, in addition to the condition above, EPS are released during market hours.

Table I20: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2522	0.0069
2	Macro	Oil Stocks	-1	announcement	-0.2322	0.0020
3	Macro	Nat. Gas Stocks	-1	announcement	-0.2209	0.0013
4	Macro	NAHB	-1	announcement	-0.2112	0.0000
5	Macro	ECRI	-1	announcement	-0.2074	0.0136
6	Macro	Federal Budget	-1	announcement	-0.2039	0.0022
7	Macro	Ex. Home Sales	-1	announcement	-0.2025	-0.0010
8	Macro	Cons. Confidence	-1	announcement	-0.2016	0.0015
9	Macro	IBD Ec. Opt.	-1	announcement	-0.2002	-0.0005
10	Macro	Factory Orders	-1	announcement	-0.1976	-0.0009
11	Macro	ECRI	+1	announcement	-0.1967	-0.0017
12	Macro	Cons. Credit	-1	announcement	-0.1962	-0.0008
13	Macro	Wholesale Inv.	-1	announcement	-0.1951	0.0016
14	Macro	Nat. Gas Stocks	+1	announcement	-0.1947	0.0006
15	Macro	Business Inv.	-1	announcement	-0.1945	-0.0001
16	Macro	New Home Sales	-1	announcement	-0.1925	-0.0004
17	Macro	Phil. Fed	-1	announcement	-0.1923	-0.0003
18	Macro	Empl. Trends	-1	announcement	-0.1904	0.0002
19	Macro	Oil Stocks	+1	announcement	-0.1896	0.0003
20	Macro	ECRI	0	announcement	-0.1893	-0.0006

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 75 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 67 assets for which, in addition to the condition above, EPS are released during market hours.

Table I21: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	ECRI	-1	announcement	-0.2484	0.0370
2	Macro	Wholesale Inv.	-1	abs(min(0,surprise))	-0.0001	0.0219
3	Macro	FOMC Rate Dec.	0	announcement	0.1136	0.0152
4	Macro	Constr. Spending	-1	max(0,surprise)	0.0000	0.0109
5	StreetAcc.	Up/Downgrades	-1 and 0	persistence	-0.0273	0.0099
6	Macro	Nat. Gas Stocks	0	announcement	-0.2092	0.0090
7	StreetAcc.	All	+1	announcement	-0.1611	0.0078
8	Macro	New Home Sales	0	announcement	-0.1798	0.0069
9	Macro	Pend. Home Sales	-1	abs(surprise)	-0.0090	0.0069
10	StreetAcc.	All	-1	announcement	-0.1756	0.0066
11	Macro	Nat. Gas Stocks	-1	announcement	-0.2357	0.0057
12	Macro	Nat. Gas Stocks	-1	abs(surprise)	-0.0219	0.0056
13	Macro	Pend. Home Sales	-1	announcement	-0.2085	0.0055
14	Macro	Oil Stocks	-1	abs(surprise)	-0.0172	0.0044
15	StreetAcc.	All	-1 and 0	persistence	-0.1471	0.0040
16	Macro	Oil Stocks	-1	announcement	-0.2523	0.0031
17	Macro	Factory Orders	-1	max(0,surprise)	-0.0010	0.0028
18	Macro	Constr. Spending	-1	announcement	-0.1810	0.0025
19	StreetAcc.	Newspapers	-1	announcement	-0.0092	0.0024
20	Macro	NAHB	-1	abs(surprise)	-0.0049	0.0023

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 57 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 50 assets for which, in addition to the condition above, EPS are released during market hours.

Table I22: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.1136	0.0152
2	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0138	0.0000
3	StreetAcc.	Newspapers	0	announcement	0.0088	0.0000
4	StreetAcc.	Earnings	0	sent. \neq 0	0.0081	0.0000
5	StreetAcc.	Newspapers	0 and +1	pos. sent., pers.	0.0077	0.0000
6	StreetAcc.	Earnings	0	positive sent.	0.0077	0.0000
7	T. Reuters	Top	0	announcement	0.0071	0.0000
8	T. Reuters	Top	0 and +1	persistence	0.0064	0.0000
9	StreetAcc.	All	0	positive sent.	0.0061	0.0000
10	StreetAcc.	Newspapers	-1 and 0	persistence	0.0048	0.0001
11	T. Reuters	Earnings	0	announcement	0.0048	0.0000
12	T. Reuters	Earnings	-1 and 0	persistence	0.0045	0.0000
13	StreetAcc.	Newspapers	+1	positive sent.	0.0036	0.0000
14	T. Reuters	Earnings	0 and +1	persistence	0.0036	0.0000
15	Macro	FOMC Rate Dec.	0	abs(surprise)	0.0031	0.0000
16	T. Reuters	Top	-1 and 0	persistence	0.0024	0.0000
17	StreetAcc.	Litigation	0	announcement	0.0021	0.0000
18	T. Reuters	Financial	0	announcement	0.0018	-0.0005
19	Macro	FOMC Rate Dec.	0	abs(min(0,surprise))	0.0015	0.0000
20	StreetAcc.	Earnings	0	announcement	0.0014	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 57 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 50 assets for which, in addition to the condition above, EPS are released during market hours.

Table I23: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Oil Stocks	-1	announcement	-0.2523	0.0031
2	Macro	ECRI	-1	announcement	-0.2484	0.0370
3	Macro	Nat. Gas Stocks	-1	announcement	-0.2357	0.0057
4	Macro	NAHB	-1	announcement	-0.2321	0.0000
5	Macro	Federal Budget	-1	announcement	-0.2267	0.0016
6	Macro	Ex. Home Sales	-1	announcement	-0.2224	-0.0010
7	Macro	Cons. Confidence	-1	announcement	-0.2216	0.0013
8	Macro	IBD Ec. Opt.	-1	announcement	-0.2205	0.0009
9	Macro	Nat. Gas Stocks	+1	announcement	-0.2177	0.0019
10	Macro	Factory Orders	-1	announcement	-0.2175	-0.0008
11	Macro	Cons. Credit	-1	announcement	-0.2167	-0.0008
12	Macro	ECRI	+1	announcement	-0.2163	-0.0018
13	Macro	Wholesale Inv.	-1	announcement	-0.2143	-0.0009
14	Macro	Business Inv.	-1	announcement	-0.2134	0.0018
15	Macro	Empl. Trends	-1	announcement	-0.2133	0.0003
16	Macro	New Home Sales	-1	announcement	-0.2126	0.0004
17	Macro	Oil Stocks	0	announcement	-0.2113	-0.0006
18	Macro	Phil. Fed	-1	announcement	-0.2112	0.0005
19	Macro	Nat. Gas Stocks	0	announcement	-0.2092	0.0090
20	Macro	ECRI	0	announcement	-0.2090	-0.0006

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 57 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 50 assets for which, in addition to the condition above, EPS are released during market hours.

Table I24: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Wholesale Inv.	-1	max(0,surprise)	-0.0009	0.0184
2	StreetAcc.	All	-1 and 0	persistence	-0.1218	0.0164
3	Macro	Phil. Fed	-1	announcement	-0.2172	0.0129
4	Macro	FOMC Rate Dec.	-1	announcement	0.0455	0.0110
5	T. Reuters	High	-1	announcement	0.0046	0.0065
6	Macro	Wholesale Inv.	0	abs(surprise)	-0.0015	0.0061
7	T. Reuters	All	0	announcement	-0.0304	0.0055
8	StreetAcc.	All	0 and +1	persistence	-0.1390	0.0053
9	Macro	Federal Budget	-1	announcement	-0.2276	0.0049
10	Macro	Wholesale Inv.	-1	announcement	-0.2216	0.0040
11	Macro	Oil Stocks	-1	abs(min(0,surprise))	-0.0050	0.0036
12	StreetAcc.	All	0	negative sent.	0.0011	0.0025
13	StreetAcc.	All	0	sent. \neq 0	-0.0056	0.0025
14	StreetAcc.	All	0	announcement	-0.0366	0.0024
15	StreetAcc.	M&A	0 and +1	persistence	-0.0108	0.0022
16	Macro	ISM Man. PMI	-1	announcement	-0.2114	0.0020
17	T. Reuters	All	-1 and 0	persistence	-0.1729	0.0020
18	T. Reuters	Top	-1 and 0	persistence	0.0088	0.0013
19	StreetAcc.	All	+1	negative sent.	-0.0004	0.0009
20	T. Reuters	All	-1	sent. \neq 0	-0.0296	0.0008

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 55 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

Table I25: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.1217	0.0007
2	Macro	FOMC Rate Dec.	-1	announcement	0.0455	0.0110
3	StreetAcc.	Earnings	+1	sent. $\neq 0$	0.0182	0.0000
4	T. Reuters	Top	0 and +1	persistence	0.0118	0.0000
5	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0113	0.0000
6	StreetAcc.	Earnings	+1	positive sent.	0.0104	0.0000
7	T. Reuters	Top	0	announcement	0.0103	0.0000
8	StreetAcc.	Newspapers	-1	sent. $\neq 0$	0.0098	0.0000
9	StreetAcc.	Newspapers	-1	negative sent.	0.0096	0.0000
10	T. Reuters	Top	-1 and 0	persistence	0.0088	0.0013
11	T. Reuters	Medium	-1	announcement	0.0077	0.0001
12	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0074	0.0000
13	StreetAcc.	Earnings	-1	negative sent.	0.0070	0.0000
14	StreetAcc.	Earnings	0	negative sent.	0.0050	0.0000
15	StreetAcc.	All	-1 and 0	neg. sent., pers.	0.0047	0.0000
16	T. Reuters	High	-1	announcement	0.0046	0.0065
17	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0042	0.0000
18	StreetAcc.	Earnings	0 and +1	neg. sent., pers.	0.0041	0.0000
19	T. Reuters	Top	+1	announcement	0.0029	0.0000
20	T. Reuters	Earnings	-1	sent. $\neq 0$	0.0018	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 55 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

Table I26: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Oil Stocks	-1	announcement	-0.2566	-0.0026
2	Macro	Nat. Gas Stocks	-1	announcement	-0.2424	-0.0038
3	Macro	NAHB	-1	announcement	-0.2421	0.0000
4	Macro	ECRI	-1	announcement	-0.2404	-0.0060
5	Macro	Ex. Home Sales	-1	announcement	-0.2316	-0.0013
6	Macro	IBD Ec. Opt.	-1	announcement	-0.2289	-0.0019
7	Macro	Cons. Confidence	-1	announcement	-0.2287	-0.0010
8	Macro	Oil Stocks	+1	announcement	-0.2282	-0.0019
9	Macro	Federal Budget	-1	announcement	-0.2276	0.0049
10	Macro	Factory Orders	-1	announcement	-0.2250	-0.0014
11	Macro	Cons. Credit	-1	announcement	-0.2250	-0.0012
12	Macro	Business Inv.	-1	announcement	-0.2235	-0.0009
13	Macro	ECRI	+1	announcement	-0.2224	-0.0019
14	Macro	Empl. Trends	-1	announcement	-0.2219	0.0001
15	Macro	Wholesale Inv.	-1	announcement	-0.2216	0.0040
16	Macro	Oil Stocks	0	announcement	-0.2209	-0.0007
17	Macro	New Home Sales	-1	announcement	-0.2197	-0.0014
18	Macro	ECRI	0	announcement	-0.2189	-0.0008
19	Macro	Phil. Fed	-1	announcement	-0.2172	0.0129
20	Macro	Federal Budget	+1	announcement	-0.2150	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 55 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

I.4.3 Jul 2009 - Dec 2012 (EU sovereign debt crisis)

Table I27: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Wholesale Inv.	-1	announcement	-0.1131	0.0091
2	StreetAcc.	All	0 and +1	persistence	-0.0124	0.0056
3	StreetAcc.	All	-1 and 0	persistence	-0.0497	0.0037
4	StreetAcc.	All	-1	sent. $\neq 0$	-0.0165	0.0035
5	Macro	Empl. Trends	-1	announcement	-0.1321	0.0034
6	Macro	New Home Sales	0	announcement	-0.0694	0.0033
7	Macro	Nat. Gas Stocks	-1	announcement	-0.1624	0.0025
8	Macro	Federal Budget	-1	abs(surprise)	-0.0011	0.0025
9	StreetAcc.	All	-1	n. news ≥ 2	-0.0014	0.0019
10	Macro	Pend. Home Sales	-1	announcement	-0.1088	0.0018
11	Macro	Factory Orders	-1	announcement	-0.1192	0.0016
12	Macro	Business Inv.	-1	abs(surprise)	-0.0024	0.0015
13	StreetAcc.	All	+1	announcement	-0.0366	0.0015
14	Macro	Oil Stocks	+1	abs(surprise)	-0.0015	0.0014
15	Macro	Michigan Sent.	-1	announcement	-0.1253	0.0012
16	Macro	FOMC Rate Dec.	+1	announcement	-0.0426	0.0011
17	Macro	ISM Man. PMI	0	abs(surprise)	0.0013	0.0011
18	Macro	Constr. Spending	0	abs(surprise)	0.0017	0.0011
19	Macro	Michigan Sent.	-1	abs(surprise)	-0.0044	0.0010
20	Macro	ISM Man. PMI	0	max(0,surprise)	0.0009	0.0010

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 88 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 78 assets for which, in addition to the condition above, EPS are released during market hours.

Table I28: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0735	0.0003
2	Macro	ISM Man. PMI	0	announcement	0.0196	0.0010
3	StreetAcc.	All	0	announcement	0.0188	0.0010
4	Macro	Constr. Spending	0	announcement	0.0179	-0.0027
5	Macro	FOMC Rate Dec.	-1	announcement	0.0155	0.0003
6	StreetAcc.	All	0	sent. $\neq 0$	0.0108	0.0009
7	T. Reuters	High	-1	announcement	0.0099	0.0000
8	T. Reuters	Medium	-1	announcement	0.0092	0.0000
9	StreetAcc.	All	+1	negative sent.	0.0089	0.0000
10	T. Reuters	High	-1 and 0	persistence	0.0087	0.0009
11	T. Reuters	Medium	-1 and 0	persistence	0.0073	0.0007
12	StreetAcc.	All	+1	sent. $\neq 0$	0.0072	0.0003
13	StreetAcc.	All	+1	n. news ≥ 2	0.0068	0.0000
14	StreetAcc.	Up/Downgrades	0 and +1	persistence	0.0067	0.0010
15	T. Reuters	Financial	-1	sent. $\neq 0$	0.0064	0.0000
16	T. Reuters	Financial	-1	positive sent.	0.0059	0.0000
17	StreetAcc.	Up/Downgrades	0	sent. $\neq 0$	0.0056	0.0000
18	T. Reuters	Top	-1	announcement	0.0051	0.0000
19	T. Reuters	Dividends	-1	positive sent.	0.0046	0.0000
20	T. Reuters	Top	-1	sent. $\neq 0$	0.0046	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 88 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 78 assets for which, in addition to the condition above, EPS are released during market hours.

Table I29: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Oil Stocks	-1	announcement	-0.1753	-0.0044
2	Macro	ECRI	-1	announcement	-0.1715	-0.0016
3	Macro	Nat. Gas Stocks	-1	announcement	-0.1624	0.0025
4	Macro	Federal Budget	-1	announcement	-0.1369	-0.0002
5	Macro	Nat. Gas Stocks	+1	announcement	-0.1363	0.0004
6	Macro	Empl. Trends	-1	announcement	-0.1321	0.0034
7	Macro	ECRI	+1	announcement	-0.1259	0.0007
8	Macro	Michigan Sent.	-1	announcement	-0.1253	0.0012
9	Macro	Cons. Credit	-1	announcement	-0.1252	-0.0004
10	Macro	NAHB	-1	announcement	-0.1237	-0.0005
11	Macro	Cons. Confidence	-1	announcement	-0.1217	-0.0010
12	Macro	Oil Stocks	+1	announcement	-0.1199	-0.0015
13	Macro	Factory Orders	-1	announcement	-0.1192	0.0016
14	Macro	IBD Ec. Opt.	-1	announcement	-0.1184	-0.0007
15	Macro	Ex. Home Sales	-1	announcement	-0.1178	-0.0008
16	Macro	New Home Sales	-1	announcement	-0.1132	-0.0002
17	Macro	Wholesale Inv.	-1	announcement	-0.1131	0.0091
18	Macro	ECRI	0	announcement	-0.1106	0.0003
19	Macro	Pend. Home Sales	-1	announcement	-0.1088	0.0018
20	Macro	Oil Stocks	0	announcement	-0.1028	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 88 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 78 assets for which, in addition to the condition above, EPS are released during market hours.

Table I30: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Wholesale Inv.	-1	announcement	-0.1402	0.0085
2	Macro	Constr. Spending	0	abs(surprise)	0.0017	0.0068
3	Macro	ISM Man. PMI	0	max(0,surprise)	0.0012	0.0067
4	Macro	ISM Man. PMI	0	announcement	0.0225	0.0067
5	Macro	Constr. Spending	0	abs(min(0,surprise))	0.0013	0.0067
6	Macro	New Home Sales	0	announcement	-0.1014	0.0067
7	StreetAcc.	All	0 and +1	persistence	-0.0470	0.0042
8	Macro	Empl. Trends	-1	announcement	-0.1635	0.0042
9	StreetAcc.	All	-1	sent. \neq 0	-0.0175	0.0037
10	Macro	ISM Man. PMI	0	abs(surprise)	0.0018	0.0037
11	T. Reuters	High	-1 and 0	persistence	0.0067	0.0033
12	T. Reuters	Medium	-1 and 0	persistence	0.0031	0.0031
13	Macro	Factory Orders	-1	max(0,surprise)	-0.0010	0.0027
14	StreetAcc.	All	+1	announcement	-0.0741	0.0024
15	StreetAcc.	All	0	positive sent.	0.0012	0.0020
16	StreetAcc.	All	0	sent. \neq 0	0.0012	0.0020
17	Macro	Cons. Credit	+1	announcement	-0.1321	0.0019
18	T. Reuters	Dividends	-1 and 0	persistence	-0.0084	0.0019
19	StreetAcc.	All	0	announcement	0.0140	0.0019
20	Macro	Constr. Spending	0	announcement	0.0261	0.0018

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 87 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table I31: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0866	0.0000
2	Macro	Constr. Spending	0	announcement	0.0261	0.0018
3	Macro	ISM Man. PMI	0	announcement	0.0225	0.0067
4	StreetAcc.	All	0	announcement	0.0140	0.0019
5	StreetAcc.	All	+1	n. news ≥ 2	0.0114	0.0000
6	StreetAcc.	All	+1	positive sent.	0.0090	0.0000
7	T. Reuters	High	-1	announcement	0.0083	0.0000
8	T. Reuters	Dividends	-1	sent. $\neq 0$	0.0068	0.0000
9	T. Reuters	Dividends	-1	positive sent.	0.0067	0.0000
10	T. Reuters	Financial	-1	sent. $\neq 0$	0.0067	0.0000
11	T. Reuters	High	-1 and 0	persistence	0.0067	0.0033
12	T. Reuters	Medium	-1	announcement	0.0063	0.0000
13	T. Reuters	Financial	-1	positive sent.	0.0063	0.0000
14	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0060	0.0000
15	T. Reuters	Top	-1	sent. $\neq 0$	0.0045	0.0000
16	StreetAcc.	Earnings	-1	sent. $\neq 0$	0.0045	0.0000
17	T. Reuters	Financial	0	sent. $\neq 0$	0.0042	0.0000
18	T. Reuters	Financial	0	positive sent.	0.0039	0.0000
19	T. Reuters	Top	-1	positive sent.	0.0036	0.0000
20	StreetAcc.	Regulatory	0 and +1	persistence	0.0036	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 87 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table I32: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Oil Stocks	-1	announcement	-0.2044	-0.0049
2	Macro	ECRI	-1	announcement	-0.1955	-0.0021
3	Macro	Nat. Gas Stocks	-1	announcement	-0.1910	-0.0014
4	Macro	Federal Budget	-1	announcement	-0.1649	-0.0003
5	Macro	Empl. Trends	-1	announcement	-0.1635	0.0042
6	Macro	Nat. Gas Stocks	+1	announcement	-0.1623	-0.0007
7	Macro	ECRI	+1	announcement	-0.1533	-0.0004
8	Macro	Cons. Credit	-1	announcement	-0.1519	-0.0005
9	Macro	NAHB	-1	announcement	-0.1509	-0.0006
10	Macro	Michigan Sent.	-1	announcement	-0.1492	0.0002
11	Macro	New Home Sales	-1	announcement	-0.1483	-0.0010
12	Macro	IBD Ec. Opt.	-1	announcement	-0.1479	-0.0008
13	Macro	Factory Orders	-1	announcement	-0.1468	0.0007
14	Macro	Oil Stocks	+1	announcement	-0.1465	-0.0015
15	Macro	Ex. Home Sales	-1	announcement	-0.1463	-0.0009
16	Macro	Cons. Confidence	-1	announcement	-0.1453	-0.0012
17	Macro	ECRI	0	announcement	-0.1449	0.0007
18	Macro	Pend. Home Sales	-1	announcement	-0.1438	-0.0006
19	Macro	Wholesale Inv.	-1	announcement	-0.1402	0.0085
20	Macro	Business Inv.	-1	announcement	-0.1397	-0.0004

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 87 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table I33: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0 and +1	persistence	-0.0604	0.0198
2	Macro	Nat. Gas Stocks	-1	announcement	-0.2070	0.0089
3	Macro	Pend. Home Sales	-1	announcement	-0.1423	0.0086
4	Macro	Leading Index	-1	announcement	-0.1184	0.0073
5	Macro	Oil Stocks	-1	max(0,surprise)	-0.0009	0.0053
6	StreetAcc.	All	+1	announcement	-0.0523	0.0048
7	Macro	Michigan Sent.	-1	announcement	-0.1738	0.0047
8	StreetAcc.	Up/Downgrades	0 and +1	persistence	-0.0073	0.0044
9	T. Reuters	All	0 and +1	persistence	-0.0794	0.0041
10	StreetAcc.	All	-1 and 0	persistence	-0.1000	0.0036
11	Macro	Oil Stocks	0	announcement	-0.1363	0.0034
12	Macro	Nat. Gas Stocks	+1	announcement	-0.1564	0.0022
13	Macro	Constr. Spending	-1	announcement	-0.1213	0.0014
14	StreetAcc.	All	-1	announcement	-0.1055	0.0014
15	Macro	Constr. Spending	-1	abs(surprise)	-0.0016	0.0012
16	Macro	Oil Stocks	-1	abs(surprise)	-0.0136	0.0011
17	Macro	FOMC Rate Dec.	+1	announcement	-0.0826	0.0009
18	Macro	Leading Index	0	abs(surprise)	0.0001	0.0009
19	Macro	Leading Index	0	announcement	-0.0656	0.0008
20	T. Reuters	All	-1 and 0	persistence	-0.1032	0.0007

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 66 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 60 assets for which, in addition to the condition above, EPS are released during market hours.

Table I34: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	+1	negative sent.	0.0139	0.0000
2	StreetAcc.	All	0	negative sent.	0.0093	0.0000
3	StreetAcc.	All	0	sent. $\neq 0$	0.0083	0.0000
4	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0052	0.0000
5	StreetAcc.	Earnings	0	announcement	0.0049	0.0000
6	StreetAcc.	Earnings	0	negative sent.	0.0049	0.0000
7	T. Reuters	Earnings	-1 and 0	persistence	0.0039	0.0000
8	StreetAcc.	Earnings	-1	sent. $\neq 0$	0.0036	0.0000
9	StreetAcc.	Earnings	-1	negative sent.	0.0035	0.0000
10	StreetAcc.	All	+1	n. news ≥ 2	0.0032	0.0000
11	T. Reuters	Earnings	-1	announcement	0.0032	0.0000
12	StreetAcc.	All	+1	sent. $\neq 0$	0.0023	-0.0001
13	StreetAcc.	All	0	n. news ≥ 2	0.0021	0.0000
14	StreetAcc.	All	-1	negative sent.	0.0021	0.0000
15	StreetAcc.	Up/Downgrades	+1	sent. $\neq 0$	0.0020	0.0000
16	StreetAcc.	EPS	-1	announcement	0.0018	0.0000
17	StreetAcc.	All	0	positive sent.	0.0017	0.0000
18	StreetAcc.	Regulatory	0	sent. $\neq 0$	0.0015	0.0000
19	StreetAcc.	Regulatory	0	announcement	0.0015	0.0000
20	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0015	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 66 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 60 assets for which, in addition to the condition above, EPS are released during market hours.

Table I35: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	ECRI	-1	announcement	-0.2218	-0.0005
2	Macro	Nat. Gas Stocks	-1	announcement	-0.2070	0.0089
3	Macro	Oil Stocks	-1	announcement	-0.1959	-0.0050
4	Macro	Federal Budget	-1	announcement	-0.1758	-0.0003
5	Macro	Michigan Sent.	-1	announcement	-0.1738	0.0047
6	Macro	Empl. Trends	-1	announcement	-0.1727	-0.0003
7	Macro	Cons. Credit	-1	announcement	-0.1718	-0.0009
8	Macro	NAHB	-1	announcement	-0.1691	-0.0009
9	Macro	Cons. Confidence	-1	announcement	-0.1625	-0.0013
10	Macro	Oil Stocks	+1	announcement	-0.1591	-0.0025
11	Macro	Nat. Gas Stocks	+1	announcement	-0.1564	0.0022
12	Macro	IBD Ec. Opt.	-1	announcement	-0.1544	-0.0013
13	Macro	Ex. Home Sales	-1	announcement	-0.1529	-0.0011
14	Macro	Wholesale Inv.	-1	announcement	-0.1493	0.0007
15	Macro	ECRI	0	announcement	-0.1460	-0.0005
16	Macro	Pend. Home Sales	-1	announcement	-0.1423	0.0086
17	Macro	ECRI	+1	announcement	-0.1409	-0.0015
18	Macro	New Home Sales	-1	announcement	-0.1403	-0.0011
19	Macro	Business Inv.	-1	announcement	-0.1390	-0.0013
20	Macro	Factory Orders	-1	announcement	-0.1371	0.0004

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 66 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 60 assets for which, in addition to the condition above, EPS are released during market hours.

I.4.4 Jan 2013 - Feb 2015 (post EU sovereign debt crisis)

Table I36: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	-1	announcement	-0.1961	0.0443
2	StreetAcc.	All	0 and +1	persistence	-0.0450	0.0100
3	Macro	ISM Man. PMI	-1	announcement	-0.1553	0.0061
4	Macro	ISM Man. PMI	-1	abs(surprise)	-0.0009	0.0043
5	StreetAcc.	Earnings	-1 and 0	persistence	-0.0526	0.0043
6	StreetAcc.	Earnings	0 and +1	persistence	-0.0443	0.0039
7	StreetAcc.	All	+1	announcement	-0.0582	0.0033
8	T. Reuters	All	-1 and 0	persistence	-0.1084	0.0031
9	T. Reuters	All	-1	announcement	-0.0982	0.0027
10	Macro	NAHB	-1	announcement	-0.1753	0.0027
11	StreetAcc.	All	0	announcement	0.0076	0.0017
12	T. Reuters	Dividends	-1	announcement	-0.0158	0.0014
13	StreetAcc.	Earnings	-1	positive sent.	0.0019	0.0013
14	StreetAcc.	All	-1 and 0	persistence	-0.0705	0.0009
15	Macro	Constr. Spending	0	announcement	-0.1115	0.0006
16	Macro	Michigan Sent.	-1	abs(surprise)	-0.0073	0.0006
17	Macro	Oil Stocks	0	abs(surprise)	-0.0033	0.0006
18	Macro	Federal Budget	-1	announcement	-0.1534	0.0004
19	Macro	Federal Budget	-1	max(0,surprise)	0.0029	0.0002
20	Macro	Nat. Gas Stocks	+1	max(0,surprise)	-0.0036	0.0002

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 86 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table I37: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.2292	0.0000
2	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0093	0.0000
3	StreetAcc.	Earnings	+1	negative sent.	0.0080	0.0000
4	StreetAcc.	All	0	announcement	0.0076	0.0017
5	StreetAcc.	All	+1	sent. $\neq 0$	0.0072	0.0000
6	T. Reuters	Financial	0	announcement	0.0062	0.0000
7	StreetAcc.	M&A	0	announcement	0.0060	0.0000
8	StreetAcc.	M&A	-1	sent. $\neq 0$	0.0059	0.0000
9	Macro	Federal Budget	+1	abs(min(0,surprise))	0.0057	0.0000
10	StreetAcc.	All	-1 and 0	neg. sent., pers.	0.0055	0.0000
11	T. Reuters	High	-1	sent. $\neq 0$	0.0055	0.0000
12	T. Reuters	Financial	0 and +1	persistence	0.0054	0.0000
13	T. Reuters	Medium	-1	positive sent.	0.0053	0.0000
14	T. Reuters	High	-1	positive sent.	0.0052	0.0000
15	StreetAcc.	Earnings	0	positive sent.	0.0048	0.0000
16	T. Reuters	Financial	-1 and 0	persistence	0.0045	0.0000
17	StreetAcc.	Earnings	+1	sent. $\neq 0$	0.0039	0.0000
18	StreetAcc.	M&A	+1	sent. $\neq 0$	0.0038	0.0000
19	T. Reuters	High	0	announcement	0.0034	0.0000
20	StreetAcc.	All	-1	n. news ≥ 2	0.0034	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 86 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table I38: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Oil Stocks	-1	announcement	-0.2335	-0.0053
2	Macro	FOMC Rate Dec.	0	announcement	0.2292	0.0000
3	Macro	Nat. Gas Stocks	-1	announcement	-0.2171	-0.0028
4	Macro	ECRI	-1	announcement	-0.2156	-0.0054
5	Macro	Michigan Sent.	-1	announcement	-0.1961	0.0443
6	Macro	Oil Stocks	+1	announcement	-0.1892	-0.0004
7	Macro	Cons. Confidence	-1	announcement	-0.1855	-0.0007
8	Macro	Nat. Gas Stocks	+1	announcement	-0.1841	-0.0006
9	Macro	Cons. Credit	-1	announcement	-0.1827	-0.0010
10	Macro	Wholesale Inv.	-1	announcement	-0.1822	-0.0005
11	Macro	Empl. Trends	-1	announcement	-0.1805	-0.0007
12	Macro	Business Inv.	-1	announcement	-0.1803	-0.0010
13	Macro	IBD Ec. Opt.	-1	announcement	-0.1791	-0.0008
14	Macro	Pend. Home Sales	-1	announcement	-0.1787	-0.0009
15	Macro	NAHB	-1	announcement	-0.1753	0.0027
16	Macro	ECRI	+1	announcement	-0.1744	-0.0016
17	Macro	Factory Orders	-1	announcement	-0.1708	-0.0004
18	Macro	ECRI	0	announcement	-0.1689	-0.0006
19	Macro	Oil Stocks	0	announcement	-0.1653	-0.0006
20	Macro	Cons. Credit	+1	announcement	-0.1635	-0.0003

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 86 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table I39: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Empl. Trends	-1	max(0,surprise)	0.0002	0.0004
2	Macro	Empl. Trends	-1	announcement	0.0108	0.0003
3	Macro	Empl. Trends	-1	abs(surprise)	0.0002	0.0002
4	StreetAcc.	All	+1	announcement	0.1591	0.0001
5	Macro	New Home Sales	-1	announcement	0.0062	0.0001
6	Macro	New Home Sales	-1	abs(surprise)	0.0014	0.0001
7	StreetAcc.	All	-1 and 0	persistence	0.0915	0.0001
8	Macro	Federal Budget	-1	announcement	-0.0880	0.0001
9	Macro	Federal Budget	-1	max(0,surprise)	0.0018	0.0001
10	StreetAcc.	All	+1	sent. $\neq 0$	0.0693	0.0001
11	StreetAcc.	M&A	0 and +1	persistence	0.0114	0.0001
12	Macro	New Home Sales	-1	abs(min(0,surprise))	0.0012	0.0001
13	StreetAcc.	All	+1	negative sent.	0.0692	0.0001
14	Macro	Federal Budget	-1	abs(surprise)	-0.0034	0.0000
15	StreetAcc.	Earnings	+1	announcement	0.0493	0.0000
16	StreetAcc.	M&A	0	announcement	0.0114	0.0000
17	StreetAcc.	Earnings	0 and +1	persistence	0.1204	0.0000
18	StreetAcc.	Earnings	+1	negative sent.	0.0690	0.0000
19	StreetAcc.	All	0	sent. $\neq 0$	0.0712	0.0000
20	StreetAcc.	Earnings	+1	sent. $\neq 0$	0.0690	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 9 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 9 assets for which, in addition to the condition above, EPS are released during market hours.

Table I40: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.3212	0.0000
2	StreetAcc.	All	0 and +1	persistence	0.2400	0.0000
3	StreetAcc.	All	+1	announcement	0.1591	0.0001
4	StreetAcc.	Earnings	0 and +1	persistence	0.1204	0.0000
5	StreetAcc.	All	-1 and 0	persistence	0.0915	0.0001
6	StreetAcc.	All	0	announcement	0.0825	-0.0001
7	Macro	FOMC Rate Dec.	-1	announcement	0.0763	0.0000
8	Macro	Oil Stocks	-1	announcement	0.0731	-0.0036
9	StreetAcc.	All	0	sent. $\neq 0$	0.0712	0.0000
10	StreetAcc.	All	0	positive sent.	0.0711	0.0000
11	StreetAcc.	Earnings	0	announcement	0.0709	0.0000
12	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0708	0.0000
13	StreetAcc.	Earnings	0	positive sent.	0.0707	0.0000
14	StreetAcc.	Earnings	-1 and 0	persistence	0.0706	0.0000
15	StreetAcc.	All	+1	sent. $\neq 0$	0.0693	0.0001
16	StreetAcc.	All	+1	negative sent.	0.0692	0.0001
17	StreetAcc.	Earnings	+1	sent. $\neq 0$	0.0690	0.0000
18	StreetAcc.	Earnings	+1	negative sent.	0.0690	0.0000
19	StreetAcc.	All	-1	announcement	0.0689	-0.0003
20	StreetAcc.	Earnings	+1	announcement	0.0493	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 9 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 9 assets for which, in addition to the condition above, EPS are released during market hours.

Table I41: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.3212	0.0000
2	StreetAcc.	All	0 and +1	persistence	0.2400	0.0000
3	StreetAcc.	All	+1	announcement	0.1591	0.0001
4	StreetAcc.	Earnings	0 and +1	persistence	0.1204	0.0000
5	Macro	Ex. Home Sales	-1	abs(surprise)	-0.0985	0.0000
6	Macro	Ex. Home Sales	-1	announcement	-0.0942	0.0000
7	StreetAcc.	All	-1 and 0	persistence	0.0915	0.0001
8	Macro	Nat. Gas Stocks	-1	abs(surprise)	-0.0889	-0.0041
9	Macro	Federal Budget	-1	announcement	-0.0880	0.0001
10	Macro	Nat. Gas Stocks	-1	announcement	-0.0846	-0.0041
11	StreetAcc.	All	0	announcement	0.0825	-0.0001
12	Macro	FOMC Rate Dec.	-1	announcement	0.0763	0.0000
13	Macro	Oil Stocks	-1	announcement	0.0731	-0.0036
14	StreetAcc.	All	0	sent. $\neq 0$	0.0712	0.0000
15	StreetAcc.	All	0	positive sent.	0.0711	0.0000
16	StreetAcc.	Earnings	0	announcement	0.0709	0.0000
17	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0708	0.0000
18	StreetAcc.	Earnings	0	positive sent.	0.0707	0.0000
19	StreetAcc.	Earnings	-1 and 0	persistence	0.0706	0.0000
20	StreetAcc.	All	+1	sent. $\neq 0$	0.0693	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and synthetic sampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 9 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 9 assets for which, in addition to the condition above, EPS are released during market hours.

Table I42: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	-1	announcement	-0.2505	0.0665
2	T. Reuters	All	-1 and 0	persistence	-0.1556	0.0048
3	T. Reuters	All	-1	announcement	-0.1434	0.0037
4	Macro	Nat. Gas Stocks	+1	max(0,surprise)	-0.0046	0.0030
5	Macro	Factory Orders	-1	announcement	-0.2215	0.0024
6	StreetAcc.	Earnings	-1 and 0	persistence	-0.0438	0.0020
7	Macro	Oil Stocks	0	abs(surprise)	-0.0048	0.0019
8	Macro	Oil Stocks	-1	abs(surprise)	-0.0176	0.0009
9	Macro	Federal Budget	-1	announcement	-0.2100	0.0008
10	Macro	Michigan Sent.	-1	abs(surprise)	-0.0106	0.0007
11	Macro	Federal Budget	-1	max(0,surprise)	0.0050	0.0005
12	Macro	Federal Budget	-1	abs(surprise)	0.0066	0.0004
13	StreetAcc.	M&A	-1	positive sent.	0.0033	0.0000
14	StreetAcc.	M&A	-1	sent. \neq 0	0.0083	0.0000
15	StreetAcc.	M&A	+1	sent. \neq 0	0.0034	0.0000
16	T. Reuters	High	+1	positive sent.	0.0016	0.0000
17	T. Reuters	Medium	+1	sent. \neq 0	0.0007	0.0000
18	T. Reuters	High	0 and +1	persistence	-0.0129	0.0000
19	StreetAcc.	M&A	+1	positive sent.	0.0086	0.0000
20	T. Reuters	All	+1	positive sent.	0.0000	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 52 assets for which, in addition to the condition above, EPS are released during market hours.

Table I43: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0204	0.0000
2	Macro	Federal Budget	+1	abs(min(0,surprise))	0.0088	0.0000
3	StreetAcc.	M&A	+1	positive sent.	0.0086	0.0000
4	StreetAcc.	M&A	-1	sent. \neq 0	0.0083	0.0000
5	StreetAcc.	All	0	negative sent.	0.0081	0.0000
6	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0074	0.0000
7	StreetAcc.	Regulatory	+1	sent. \neq 0	0.0066	0.0000
8	Macro	Federal Budget	-1	abs(surprise)	0.0066	0.0004
9	StreetAcc.	Regulatory	+1	negative sent.	0.0065	0.0000
10	StreetAcc.	Regulatory	0	announcement	0.0062	0.0000
11	T. Reuters	High	-1	positive sent.	0.0055	0.0000
12	T. Reuters	Medium	-1	positive sent.	0.0055	0.0000
13	T. Reuters	Medium	-1	sent. \neq 0	0.0052	0.0000
14	T. Reuters	High	-1	sent. \neq 0	0.0051	0.0000
15	Macro	Federal Budget	-1	max(0,surprise)	0.0050	0.0005
16	StreetAcc.	All	+1	negative sent.	0.0049	0.0000
17	StreetAcc.	Earnings	+1	negative sent.	0.0041	0.0000
18	StreetAcc.	Regulatory	+1	announcement	0.0040	0.0000
19	StreetAcc.	M&A	+1	announcement	0.0039	0.0000
20	StreetAcc.	Regulatory	0	sent. \neq 0	0.0035	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 52 assets for which, in addition to the condition above, EPS are released during market hours.

Table I44: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Nat. Gas Stocks	-1	announcement	-0.2875	-0.0062
2	Macro	Oil Stocks	-1	announcement	-0.2869	-0.0057
3	Macro	ECRI	-1	announcement	-0.2747	-0.0066
4	Macro	Cons. Credit	-1	announcement	-0.2543	-0.0011
5	Macro	Michigan Sent.	-1	announcement	-0.2505	0.0665
6	Macro	Nat. Gas Stocks	+1	announcement	-0.2463	-0.0018
7	Macro	Oil Stocks	+1	announcement	-0.2455	-0.0016
8	Macro	Empl. Trends	-1	announcement	-0.2424	-0.0008
9	Macro	Cons. Confidence	-1	announcement	-0.2421	-0.0008
10	Macro	Wholesale Inv.	-1	announcement	-0.2410	-0.0008
11	Macro	Business Inv.	-1	announcement	-0.2400	-0.0011
12	Macro	Pend. Home Sales	-1	announcement	-0.2392	-0.0010
13	Macro	IBD Ec. Opt.	-1	announcement	-0.2371	-0.0009
14	Macro	ECRI	+1	announcement	-0.2353	-0.0016
15	Macro	NAHB	-1	announcement	-0.2327	-0.0013
16	Macro	ECRI	0	announcement	-0.2309	-0.0007
17	Macro	New Home Sales	-1	announcement	-0.2289	-0.0013
18	Macro	Ex. Home Sales	-1	announcement	-0.2247	-0.0011
19	Macro	Oil Stocks	0	announcement	-0.2227	-0.0006
20	Macro	Factory Orders	-1	announcement	-0.2215	0.0024

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and oversampling IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 52 assets for which, in addition to the condition above, EPS are released during market hours.

J Additional Tables: Variables Importance, No Imbalanced Sample Treatment

J.1 All Jumps

Table J1: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0345	0.0294
2	Macro	FOMC Rate Dec.	-1	announcement	0.0045	0.0164
3	Macro	Constr. Spending	0	abs(min(0,surprise))	0.0000	0.0079
4	StreetAcc.	All	0 and +1	persistence	0.0002	0.0071
5	StreetAcc.	All	0	announcement	0.0120	0.0040
6	Macro	Constr. Spending	0	abs(surprise)	0.0000	0.0034
7	StreetAcc.	All	+1	announcement	0.0031	0.0019
8	Macro	Nat. Gas Stocks	+1	announcement	0.0000	0.0019
9	Macro	Oil Stocks	-1	announcement	0.0000	0.0018
10	Macro	Factory Orders	-1	announcement	0.0000	0.0018
11	StreetAcc.	All	-1 and 0	persistence	0.0001	0.0017
12	Macro	Nat. Gas Stocks	-1	announcement	0.0000	0.0017
13	StreetAcc.	All	-1	announcement	0.0003	0.0017
14	Macro	Oil Stocks	-1	abs(min(0,surprise))	0.0000	0.0013
15	Macro	IBD Ec. Opt.	-1	abs(surprise)	0.0000	0.0012
16	StreetAcc.	All	+1	negative sent.	0.0014	0.0012
17	Macro	NAHB	0	announcement	0.0004	0.0011
18	Macro	ECRI	-1	announcement	0.0000	0.0010
19	Macro	ISM Man. PMI	0	max(0,surprise)	0.0000	0.0009
20	StreetAcc.	All	+1	sent. \neq 0	0.0000	0.0009

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table J2: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0345	0.0294
2	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0335	0.0000
3	T. Reuters	Earnings	0	announcement	0.0223	0.0000
4	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0180	0.0000
5	StreetAcc.	All	0	n. news ≥ 2	0.0145	0.0000
6	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0137	0.0000
7	StreetAcc.	Earnings	0	negative sent.	0.0132	0.0000
8	StreetAcc.	Up/Downgrades	0	negative sent.	0.0121	0.0000
9	StreetAcc.	All	0	announcement	0.0120	0.0040
10	StreetAcc.	Regulatory	0	positive sent.	0.0113	0.0000
11	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0113	0.0000
12	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0112	0.0000
13	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0101	0.0000
14	StreetAcc.	M&A	0	positive sent.	0.0101	0.0000
15	T. Reuters	Top	0	announcement	0.0072	0.0000
16	StreetAcc.	Regulatory	0	n. news ≥ 2	0.0071	0.0000
17	StreetAcc.	All	-1	n. news ≥ 2	0.0069	0.0007
18	StreetAcc.	All	+1	n. news ≥ 2	0.0060	0.0000
19	T. Reuters	Financial	0	announcement	0.0059	0.0000
20	Macro	Michigan Sent.	+1	announcement	0.0056	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table J3: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0345	0.0294
2	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0335	0.0000
3	T. Reuters	Earnings	0	announcement	0.0223	0.0000
4	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0180	0.0000
5	StreetAcc.	All	0	n. news ≥ 2	0.0145	0.0000
6	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0137	0.0000
7	StreetAcc.	Earnings	0	negative sent.	0.0132	0.0000
8	StreetAcc.	Up/Downgrades	0	negative sent.	0.0121	0.0000
9	StreetAcc.	All	0	announcement	0.0120	0.0040
10	StreetAcc.	Regulatory	0	positive sent.	0.0113	0.0000
11	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0113	0.0000
12	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0112	0.0000
13	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0101	0.0000
14	StreetAcc.	M&A	0	positive sent.	0.0101	0.0000
15	T. Reuters	Top	0	announcement	0.0072	0.0000
16	StreetAcc.	Regulatory	0	n. news ≥ 2	0.0071	0.0000
17	StreetAcc.	All	-1	n. news ≥ 2	0.0069	0.0007
18	StreetAcc.	All	+1	n. news ≥ 2	0.0060	0.0000
19	T. Reuters	Financial	0	announcement	0.0059	0.0000
20	Macro	Michigan Sent.	+1	announcement	0.0056	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

J.2 Positive Jumps

Table J4: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0242	0.0467
2	Macro	Constr. Spending	0	abs(min(0,surprise))	0.0000	0.0105
3	Macro	Constr. Spending	0	abs(surprise)	0.0000	0.0079
4	Macro	Nat. Gas Stocks	+1	announcement	0.0000	0.0054
5	StreetAcc.	All	-1 and 0	persistence	0.0001	0.0045
6	Macro	Factory Orders	-1	announcement	0.0002	0.0044
7	Macro	FOMC Rate Dec.	0	abs(surprise)	0.0000	0.0037
8	StreetAcc.	All	0	announcement	0.0029	0.0035
9	StreetAcc.	All	-1	announcement	0.0002	0.0031
10	Macro	FOMC Rate Dec.	-1	announcement	0.0017	0.0026
11	StreetAcc.	All	0 and +1	persistence	0.0001	0.0025
12	Macro	Nat. Gas Stocks	-1	announcement	0.0000	0.0024
13	Macro	ECRI	-1	announcement	0.0000	0.0024
14	Macro	NAHB	0	announcement	0.0002	0.0021
15	Macro	ISM Man. PMI	0	abs(surprise)	0.0000	0.0019
16	Macro	ISM Man. PMI	0	max(0,surprise)	0.0000	0.0019
17	StreetAcc.	All	0	positive sent.	0.0028	0.0016
18	Macro	Oil Stocks	-1	abs(min(0,surprise))	0.0000	0.0016
19	Macro	Oil Stocks	-1	max(0,surprise)	0.0000	0.0014
20	T. Reuters	All	0 and +1	persistence	0.0000	0.0013

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table J5: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0242	0.0467
2	T. Reuters	Earnings	0	announcement	0.0225	0.0000
3	StreetAcc.	All	0	n. news ≥ 2	0.0216	0.0000
4	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0202	0.0000
5	StreetAcc.	Regulatory	0	positive sent.	0.0115	0.0000
6	StreetAcc.	Newspapers	-1 and 0	pos. sent., pers.	0.0114	0.0000
7	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0114	0.0000
8	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0113	0.0000
9	StreetAcc.	Earnings	0	positive sent.	0.0091	0.0000
10	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0090	0.0000
11	T. Reuters	Top	0	announcement	0.0086	0.0000
12	StreetAcc.	Newspapers	0 and +1	persistence	0.0057	0.0000
13	T. Reuters	Financial	0	announcement	0.0057	0.0000
14	StreetAcc.	All	-1	n. news ≥ 2	0.0051	0.0000
15	StreetAcc.	Earnings	0	n. news ≥ 2	0.0051	0.0000
16	StreetAcc.	Earnings	+1	announcement	0.0050	0.0000
17	StreetAcc.	Up/Downgrades	0	announcement	0.0047	0.0009
18	StreetAcc.	Litigation	0	n. news ≥ 2	0.0047	0.0000
19	Macro	Constr. Spending	0	announcement	0.0042	0.0006
20	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0039	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table J6: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0242	0.0467
2	T. Reuters	Earnings	0	announcement	0.0225	0.0000
3	StreetAcc.	All	0	n. news ≥ 2	0.0216	0.0000
4	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0202	0.0000
5	StreetAcc.	Regulatory	0	positive sent.	0.0115	0.0000
6	StreetAcc.	Newspapers	-1 and 0	pos. sent., pers.	0.0114	0.0000
7	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0114	0.0000
8	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0113	0.0000
9	StreetAcc.	Earnings	0	positive sent.	0.0091	0.0000
10	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0090	0.0000
11	T. Reuters	Top	0	announcement	0.0086	0.0000
12	StreetAcc.	Newspapers	0 and +1	persistence	0.0057	0.0000
13	T. Reuters	Financial	0	announcement	0.0057	0.0000
14	StreetAcc.	All	-1	n. news ≥ 2	0.0051	0.0000
15	StreetAcc.	Earnings	0	n. news ≥ 2	0.0051	0.0000
16	StreetAcc.	Earnings	+1	announcement	0.0050	0.0000
17	StreetAcc.	Up/Downgrades	0	announcement	0.0047	0.0009
18	StreetAcc.	Litigation	0	n. news ≥ 2	0.0047	0.0000
19	Macro	Constr. Spending	0	announcement	0.0042	0.0006
20	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0039	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

J.3 Negative Jumps

Table J7: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.0035	0.0231
2	StreetAcc.	All	0 and +1	persistence	0.0001	0.0075
3	Macro	Federal Budget	-1	announcement	0.0000	0.0040
4	StreetAcc.	All	+1	negative sent.	0.0024	0.0040
5	Macro	Michigan Sent.	-1	abs(surprise)	0.0000	0.0039
6	Macro	IBD Ec. Opt.	-1	abs(surprise)	0.0000	0.0038
7	StreetAcc.	All	-1 and 0	persistence	0.0000	0.0025
8	StreetAcc.	All	-1	announcement	0.0000	0.0024
9	Macro	Federal Budget	-1	abs(surprise)	0.0000	0.0021
10	Macro	Oil Stocks	+1	abs(surprise)	0.0000	0.0019
11	StreetAcc.	All	+1	sent. $\neq 0$	0.0000	0.0018
12	StreetAcc.	All	+1	announcement	0.0125	0.0017
13	Macro	ECRI	-1	announcement	0.0000	0.0015
14	Macro	Business Inv.	-1	announcement	0.0001	0.0014
15	Macro	Phil. Fed	-1	announcement	0.0000	0.0014
16	Macro	Empl. Trends	-1	announcement	0.0000	0.0013
17	StreetAcc.	All	0	announcement	0.0133	0.0012
18	Macro	FOMC Rate Dec.	0	announcement	0.0110	0.0012
19	Macro	NAHB	-1	abs(surprise)	0.0000	0.0012
20	StreetAcc.	Earnings	0 and +1	persistence	0.0008	0.0011

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table J8: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	+1	announcement	0.0227	0.0000
2	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0227	0.0000
3	StreetAcc.	Earnings	0	negative sent.	0.0138	0.0000
4	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0133	0.0000
5	StreetAcc.	All	0	announcement	0.0133	0.0012
6	StreetAcc.	Up/Downgrades	0	negative sent.	0.0130	0.0000
7	StreetAcc.	All	-1 and 0	pos. sent., pers.	0.0126	0.0000
8	StreetAcc.	All	+1	announcement	0.0125	0.0017
9	StreetAcc.	Newspapers	0	n. news ≥ 2	0.0114	0.0000
10	T. Reuters	Top	0	positive sent.	0.0113	0.0000
11	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0113	0.0000
12	Macro	FOMC Rate Dec.	0	announcement	0.0110	0.0012
13	StreetAcc.	Newspapers	0	negative sent.	0.0107	0.0000
14	StreetAcc.	Regulatory	0	n. news ≥ 2	0.0107	0.0000
15	StreetAcc.	M&A	+1	n. news ≥ 2	0.0088	0.0000
16	StreetAcc.	Earnings	0	announcement	0.0083	0.0000
17	StreetAcc.	All	+1	n. news ≥ 2	0.0064	0.0000
18	StreetAcc.	M&A	0	negative sent.	0.0052	0.0000
19	StreetAcc.	All	0	negative sent.	0.0046	0.0009
20	StreetAcc.	Earnings	+1	negative sent.	0.0045	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

Table J9: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	+1	announcement	0.0227	0.0000
2	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0227	0.0000
3	StreetAcc.	Earnings	0	negative sent.	0.0138	0.0000
4	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0133	0.0000
5	StreetAcc.	All	0	announcement	0.0133	0.0012
6	StreetAcc.	Up/Downgrades	0	negative sent.	0.0130	0.0000
7	StreetAcc.	All	-1 and 0	pos. sent., pers.	0.0126	0.0000
8	StreetAcc.	All	+1	announcement	0.0125	0.0017
9	StreetAcc.	Newspapers	0	n. news ≥ 2	0.0114	0.0000
10	T. Reuters	Top	0	positive sent.	0.0113	0.0000
11	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0113	0.0000
12	Macro	FOMC Rate Dec.	0	announcement	0.0110	0.0012
13	StreetAcc.	Newspapers	0	negative sent.	0.0107	0.0000
14	StreetAcc.	Regulatory	0	n. news ≥ 2	0.0107	0.0000
15	StreetAcc.	M&A	+1	n. news ≥ 2	0.0088	0.0000
16	StreetAcc.	Earnings	0	announcement	0.0083	0.0000
17	StreetAcc.	All	+1	n. news ≥ 2	0.0064	0.0000
18	StreetAcc.	M&A	0	negative sent.	0.0052	0.0000
19	StreetAcc.	All	0	negative sent.	0.0046	0.0009
20	StreetAcc.	Earnings	+1	negative sent.	0.0045	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Feb 2015 (full sample). The average APE and the average AUC decrease of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours.

J.4 Sub-Sample Analysis

J.4.1 Jan 2005 - Jun 2007 (pre-global financial crisis)

Table J10: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.0098	0.0342
2	Macro	Nat. Gas Stocks	-1	announcement	0.0000	0.0146
3	Macro	FOMC Rate Dec.	0	announcement	0.0173	0.0102
4	StreetAcc.	All	-1 and 0	persistence	0.0000	0.0049
5	Macro	Cons. Credit	-1	announcement	0.0018	0.0044
6	Macro	ECRI	-1	abs(min(0,surprise))	0.0000	0.0036
7	Macro	Oil Stocks	-1	announcement	0.0000	0.0026
8	Macro	Business Inv.	0	announcement	0.0101	0.0024
9	StreetAcc.	All	0 and +1	persistence	0.0000	0.0021
10	Macro	NAHB	-1	abs(surprise)	0.0000	0.0015
11	StreetAcc.	All	-1	n. news ≥ 2	0.0065	0.0014
12	StreetAcc.	All	0	positive sent.	0.0017	0.0014
13	Macro	Factory Orders	-1	announcement	0.0000	0.0014
14	Macro	Cons. Confidence	-1	max(0,surprise)	0.0000	0.0013
15	Macro	Oil Stocks	-1	max(0,surprise)	0.0000	0.0012
16	Macro	New Home Sales	-1	abs(min(0,surprise))	0.0000	0.0012
17	Macro	Leading Index	-1	announcement	0.0010	0.0012
18	StreetAcc.	All	0	n. news ≥ 2	0.0007	0.0011
19	Macro	IBD Ec. Opt.	-1	abs(surprise)	0.0000	0.0009
20	Macro	Cons. Credit	-1	max(0,surprise)	0.0000	0.0009

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 83 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 73 assets for which, in addition to the condition above, EPS are released during market hours.

Table J11: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	+1	announcement	0.0241	0.0000
2	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0178	0.0000
3	Macro	FOMC Rate Dec.	0	announcement	0.0173	0.0102
4	StreetAcc.	All	0	negative sent.	0.0115	0.0000
5	T. Reuters	Top	0	announcement	0.0114	0.0000
6	StreetAcc.	Up/Downgrades	0	negative sent.	0.0114	0.0000
7	Macro	Business Inv.	0	announcement	0.0101	0.0024
8	Macro	FOMC Rate Dec.	-1	announcement	0.0098	0.0342
9	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0086	0.0000
10	StreetAcc.	Earnings	0	announcement	0.0079	0.0000
11	StreetAcc.	All	-1	n. news ≥ 2	0.0065	0.0014
12	StreetAcc.	Up/Downgrades	-1	n. news ≥ 2	0.0048	0.0000
13	StreetAcc.	Up/Downgrades	0	announcement	0.0041	0.0000
14	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0036	0.0000
15	Macro	Constr. Spending	0	announcement	0.0036	-0.0001
16	StreetAcc.	Earnings	+1	negative sent.	0.0035	0.0000
17	T. Reuters	Dividends	0	announcement	0.0034	0.0000
18	StreetAcc.	Newspapers	0 and +1	neg. sent., pers.	0.0034	0.0000
19	T. Reuters	Dividends	+1	announcement	0.0031	0.0000
20	StreetAcc.	All	+1	n. news ≥ 2	0.0030	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 83 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 73 assets for which, in addition to the condition above, EPS are released during market hours.

Table J12: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	+1	announcement	0.0241	0.0000
2	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0178	0.0000
3	Macro	FOMC Rate Dec.	0	announcement	0.0173	0.0102
4	StreetAcc.	All	0	negative sent.	0.0115	0.0000
5	T. Reuters	Top	0	announcement	0.0114	0.0000
6	StreetAcc.	Up/Downgrades	0	negative sent.	0.0114	0.0000
7	Macro	Business Inv.	0	announcement	0.0101	0.0024
8	Macro	FOMC Rate Dec.	-1	announcement	0.0098	0.0342
9	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0086	0.0000
10	StreetAcc.	Earnings	0	announcement	0.0079	0.0000
11	StreetAcc.	All	-1	n. news ≥ 2	0.0065	0.0014
12	StreetAcc.	Up/Downgrades	-1	n. news ≥ 2	0.0048	0.0000
13	StreetAcc.	Up/Downgrades	0	announcement	0.0041	0.0000
14	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0036	0.0000
15	Macro	Constr. Spending	0	announcement	0.0036	-0.0001
16	StreetAcc.	Earnings	+1	negative sent.	0.0035	0.0000
17	T. Reuters	Dividends	0	announcement	0.0034	0.0000
18	StreetAcc.	Newspapers	0 and +1	neg. sent., pers.	0.0034	0.0000
19	T. Reuters	Dividends	+1	announcement	0.0031	0.0000
20	StreetAcc.	All	+1	n. news ≥ 2	0.0030	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 83 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 73 assets for which, in addition to the condition above, EPS are released during market hours.

Table J13: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.0060	0.0302
2	Macro	Nat. Gas Stocks	-1	announcement	0.0002	0.0148
3	Macro	Cons. Credit	-1	announcement	0.0001	0.0106
4	Macro	FOMC Rate Dec.	0	announcement	0.0168	0.0091
5	Macro	ECRI	-1	abs(min(0,surprise))	0.0000	0.0091
6	StreetAcc.	All	0	announcement	0.0032	0.0078
7	Macro	Factory Orders	-1	announcement	0.0000	0.0069
8	StreetAcc.	All	-1 and 0	persistence	0.0000	0.0067
9	Macro	Oil Stocks	-1	announcement	0.0001	0.0049
10	StreetAcc.	All	0 and +1	persistence	0.0001	0.0041
11	Macro	Business Inv.	0	announcement	0.0140	0.0038
12	Macro	Business Inv.	-1	abs(surprise)	0.0000	0.0029
13	StreetAcc.	Litigation	+1	announcement	0.0002	0.0026
14	StreetAcc.	Litigation	0 and +1	persistence	0.0000	0.0025
15	StreetAcc.	All	-1	n. news ≥ 2	0.0061	0.0020
16	StreetAcc.	Earnings	-1 and 0	persistence	0.0000	0.0016
17	Macro	ECRI	-1	max(0,surprise)	0.0000	0.0010
18	Macro	ECRI	-1	announcement	0.0000	0.0010
19	Macro	Nat. Gas Stocks	0	announcement	-0.0001	0.0004
20	Macro	Nat. Gas Stocks	+1	announcement	-0.0001	0.0003

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 64 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 57 assets for which, in addition to the condition above, EPS are released during market hours.

Table J14: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0168	0.0091
2	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0165	0.0000
3	T. Reuters	Top	0	announcement	0.0156	0.0000
4	Macro	Business Inv.	0	announcement	0.0140	0.0038
5	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0090	0.0000
6	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0064	0.0000
7	StreetAcc.	Newspapers	+1	positive sent.	0.0062	0.0000
8	Macro	Constr. Spending	0	announcement	0.0061	-0.0002
9	StreetAcc.	All	-1	n. news ≥ 2	0.0061	0.0020
10	Macro	FOMC Rate Dec.	-1	announcement	0.0060	0.0302
11	T. Reuters	Financial	-1	announcement	0.0060	0.0000
12	Macro	ISM Man. PMI	-1	announcement	0.0033	-0.0011
13	Macro	New Home Sales	-1	announcement	0.0032	-0.0001
14	StreetAcc.	All	0	announcement	0.0032	0.0078
15	StreetAcc.	All	0	n. news ≥ 2	0.0032	0.0000
16	StreetAcc.	Newspapers	-1	negative sent.	0.0031	0.0000
17	T. Reuters	Dividends	0	announcement	0.0030	0.0000
18	T. Reuters	Dividends	+1	announcement	0.0028	0.0000
19	Macro	Phil. Fed	+1	announcement	0.0027	0.0000
20	T. Reuters	High	0	negative sent.	0.0023	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 64 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 57 assets for which, in addition to the condition above, EPS are released during market hours.

Table J15: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0168	0.0091
2	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0165	0.0000
3	T. Reuters	Top	0	announcement	0.0156	0.0000
4	Macro	Business Inv.	0	announcement	0.0140	0.0038
5	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0090	0.0000
6	StreetAcc.	Up/Downgrades	0 and +1	pos. sent., pers.	0.0064	0.0000
7	StreetAcc.	Newspapers	+1	positive sent.	0.0062	0.0000
8	Macro	Constr. Spending	0	announcement	0.0061	-0.0002
9	StreetAcc.	All	-1	n. news ≥ 2	0.0061	0.0020
10	Macro	FOMC Rate Dec.	-1	announcement	0.0060	0.0302
11	T. Reuters	Financial	-1	announcement	0.0060	0.0000
12	Macro	ISM Man. PMI	-1	announcement	0.0033	-0.0011
13	Macro	New Home Sales	-1	announcement	0.0032	-0.0001
14	StreetAcc.	All	0	announcement	0.0032	0.0078
15	StreetAcc.	All	0	n. news ≥ 2	0.0032	0.0000
16	StreetAcc.	Newspapers	-1	negative sent.	0.0031	0.0000
17	T. Reuters	Dividends	0	announcement	0.0030	0.0000
18	T. Reuters	Dividends	+1	announcement	0.0028	0.0000
19	Macro	Phil. Fed	+1	announcement	0.0027	0.0000
20	T. Reuters	High	0	negative sent.	0.0023	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 64 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 57 assets for which, in addition to the condition above, EPS are released during market hours.

Table J16: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 1.0$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.0045	0.0189
2	Macro	Nat. Gas Stocks	-1	announcement	0.0002	0.0085
3	Macro	ECRI	-1	announcement	0.0000	0.0055
4	Macro	Oil Stocks	-1	announcement	0.0001	0.0054
5	StreetAcc.	All	-1 and 0	persistence	0.0000	0.0035
6	Macro	ECRI	-1	abs(min(0,surprise))	0.0000	0.0033
7	Macro	Cons. Confidence	-1	announcement	0.0000	0.0018
8	Macro	Business Inv.	-1	max(0,surprise)	0.0000	0.0011
9	Macro	Business Inv.	-1	announcement	0.0000	0.0010
10	Macro	New Home Sales	-1	announcement	0.0000	0.0009
11	Macro	FOMC Rate Dec.	0	announcement	0.0051	0.0008
12	Macro	Michigan Sent.	-1	announcement	0.0000	0.0008
13	Macro	ECRI	-1	max(0,surprise)	0.0000	0.0001
14	Macro	NAHB	-1	announcement	0.0012	0.0001
15	Macro	Cons. Credit	-1	announcement	0.0009	0.0000
16	Macro	Phil. Fed	-1	max(0,surprise)	0.0000	0.0000
17	Macro	New Home Sales	-1	abs(min(0,surprise))	0.0000	0.0000
18	Macro	Michigan Sent.	-1	abs(min(0,surprise))	0.0000	0.0000
19	Macro	NAHB	+1	announcement	0.0000	0.0000
20	Macro	Leading Index	-1	abs(surprise)	0.0000	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 1.0$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

Table J17: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 1.0$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	+1	announcement	0.0169	0.0000
2	StreetAcc.	EPS	-1	announcement	0.0124	0.0000
3	StreetAcc.	Up/Downgrades	0	announcement	0.0094	0.0000
4	StreetAcc.	All	+1	n. news ≥ 2	0.0093	0.0000
5	StreetAcc.	All	+1	negative sent.	0.0072	0.0000
6	StreetAcc.	All	0	positive sent.	0.0053	0.0000
7	Macro	FOMC Rate Dec.	0	announcement	0.0051	0.0008
8	StreetAcc.	Newspapers	0	announcement	0.0050	0.0000
9	Macro	Cons. Credit	0	announcement	0.0046	0.0000
10	Macro	FOMC Rate Dec.	-1	announcement	0.0045	0.0189
11	StreetAcc.	Earnings	0	positive sent.	0.0042	0.0000
12	Macro	Cons. Credit	+1	announcement	0.0028	0.0000
13	Macro	Leading Index	-1	announcement	0.0024	-0.0001
14	StreetAcc.	All	-1	n. news ≥ 2	0.0024	0.0000
15	Macro	IBD Ec. Opt.	-1	announcement	0.0023	0.0000
16	StreetAcc.	All	+1	announcement	0.0018	0.0000
17	StreetAcc.	Up/Downgrades	-1	negative sent.	0.0012	0.0000
18	Macro	NAHB	-1	announcement	0.0012	0.0001
19	StreetAcc.	All	-1	negative sent.	0.0009	0.0000
20	Macro	Cons. Credit	-1	announcement	0.0009	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 1.0$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

Table J18: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 1.0$ and no IS technique; sample = Jan 2005 - Jun 2007 (pre-global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	+1	announcement	0.0169	0.0000
2	StreetAcc.	EPS	-1	announcement	0.0124	0.0000
3	StreetAcc.	Up/Downgrades	0	announcement	0.0094	0.0000
4	StreetAcc.	All	+1	n. news ≥ 2	0.0093	0.0000
5	StreetAcc.	All	+1	negative sent.	0.0072	0.0000
6	StreetAcc.	All	0	positive sent.	0.0053	0.0000
7	Macro	FOMC Rate Dec.	0	announcement	0.0051	0.0008
8	StreetAcc.	Newspapers	0	announcement	0.0050	0.0000
9	Macro	Cons. Credit	0	announcement	0.0046	0.0000
10	Macro	FOMC Rate Dec.	-1	announcement	0.0045	0.0189
11	StreetAcc.	Earnings	0	positive sent.	0.0042	0.0000
12	Macro	Cons. Credit	+1	announcement	0.0028	0.0000
13	Macro	Leading Index	-1	announcement	0.0024	-0.0001
14	StreetAcc.	All	-1	n. news ≥ 2	0.0024	0.0000
15	Macro	IBD Ec. Opt.	-1	announcement	0.0023	0.0000
16	StreetAcc.	All	+1	announcement	0.0018	0.0000
17	StreetAcc.	Up/Downgrades	-1	negative sent.	0.0012	0.0000
18	Macro	NAHB	-1	announcement	0.0012	0.0001
19	StreetAcc.	All	-1	negative sent.	0.0009	0.0000
20	Macro	Cons. Credit	-1	announcement	0.0009	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 1.0$ and no IS technique. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

J.4.2 Jul 2007 - Jun 2009 (global financial crisis)

Table J19: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	ECRI	-1	announcement	0.0001	0.0140
2	Macro	FOMC Rate Dec.	-1	announcement	0.0064	0.0086
3	Macro	FOMC Rate Dec.	0	announcement	0.0608	0.0071
4	Macro	Nat. Gas Stocks	0	announcement	0.0018	0.0069
5	Macro	Oil Stocks	0	abs(surprise)	0.0000	0.0068
6	Macro	New Home Sales	0	announcement	-0.0001	0.0059
7	Macro	Wholesale Inv.	-1	abs(min(0,surprise))	0.0000	0.0056
8	Macro	Wholesale Inv.	-1	max(0,surprise)	0.0000	0.0044
9	Macro	Nat. Gas Stocks	0	abs(surprise)	0.0000	0.0025
10	Macro	Federal Budget	-1	announcement	-0.0001	0.0022
11	StreetAcc.	All	-1 and 0	persistence	0.0001	0.0021
12	Macro	Wholesale Inv.	0	abs(surprise)	0.0000	0.0018
13	Macro	Cons. Confidence	-1	announcement	-0.0001	0.0017
14	StreetAcc.	All	+1	announcement	0.0001	0.0017
15	Macro	Wholesale Inv.	-1	announcement	-0.0001	0.0016
16	Macro	ISM Man. PMI	-1	announcement	-0.0001	0.0015
17	StreetAcc.	All	0 and +1	persistence	-0.0001	0.0015
18	Macro	Nat. Gas Stocks	-1	announcement	0.0006	0.0013
19	StreetAcc.	All	-1	announcement	0.0000	0.0013
20	Macro	Oil Stocks	-1	abs(surprise)	0.0000	0.0012

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 75 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 67 assets for which, in addition to the condition above, EPS are released during market hours.

Table J20: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0608	0.0071
2	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0335	0.0000
3	StreetAcc.	Newspapers	0 and +1	pos. sent., pers.	0.0133	0.0000
4	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0132	0.0000
5	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0131	0.0000
6	StreetAcc.	Earnings	0	negative sent.	0.0124	0.0000
7	T. Reuters	Earnings	0	announcement	0.0121	0.0000
8	StreetAcc.	All	+1	n. news ≥ 2	0.0118	0.0000
9	T. Reuters	Top	0	announcement	0.0111	0.0000
10	T. Reuters	Top	0 and +1	persistence	0.0107	0.0000
11	StreetAcc.	Newspapers	0	positive sent.	0.0080	0.0000
12	StreetAcc.	Up/Downgrades	0	positive sent.	0.0065	0.0000
13	Macro	FOMC Rate Dec.	-1	announcement	0.0064	0.0086
14	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0058	0.0000
15	Macro	Michigan Sent.	-1	announcement	0.0057	-0.0018
16	T. Reuters	Top	-1	announcement	0.0053	0.0000
17	StreetAcc.	Earnings	0	positive sent.	0.0047	0.0000
18	StreetAcc.	All	0	negative sent.	0.0045	0.0000
19	StreetAcc.	All	0	announcement	0.0036	0.0000
20	T. Reuters	Top	+1	announcement	0.0026	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 75 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 67 assets for which, in addition to the condition above, EPS are released during market hours.

Table J21: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0608	0.0071
2	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0335	0.0000
3	StreetAcc.	Newspapers	0 and +1	pos. sent., pers.	0.0133	0.0000
4	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0132	0.0000
5	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0131	0.0000
6	StreetAcc.	Earnings	0	negative sent.	0.0124	0.0000
7	T. Reuters	Earnings	0	announcement	0.0121	0.0000
8	StreetAcc.	All	+1	n. news ≥ 2	0.0118	0.0000
9	T. Reuters	Top	0	announcement	0.0111	0.0000
10	T. Reuters	Top	0 and +1	persistence	0.0107	0.0000
11	StreetAcc.	Newspapers	0	positive sent.	0.0080	0.0000
12	StreetAcc.	Up/Downgrades	0	positive sent.	0.0065	0.0000
13	Macro	FOMC Rate Dec.	-1	announcement	0.0064	0.0086
14	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0058	0.0000
15	Macro	Michigan Sent.	-1	announcement	0.0057	-0.0018
16	T. Reuters	Top	-1	announcement	0.0053	0.0000
17	StreetAcc.	Earnings	0	positive sent.	0.0047	0.0000
18	StreetAcc.	All	0	negative sent.	0.0045	0.0000
19	StreetAcc.	All	0	announcement	0.0036	0.0000
20	T. Reuters	Top	+1	announcement	0.0026	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). The average APE and the average AUC decrease of the indicators are computed over the 75 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 67 assets for which, in addition to the condition above, EPS are released during market hours.

Table J22: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	ECRI	-1	announcement	-0.0001	0.0403
2	Macro	Nat. Gas Stocks	0	announcement	-0.0001	0.0093
3	StreetAcc.	All	+1	announcement	0.0002	0.0078
4	Macro	New Home Sales	0	announcement	0.0000	0.0069
5	Macro	Pend. Home Sales	-1	abs(surprise)	0.0000	0.0061
6	Macro	Pend. Home Sales	-1	announcement	-0.0001	0.0056
7	Macro	FOMC Rate Dec.	-1	announcement	0.0011	0.0056
8	Macro	Nat. Gas Stocks	-1	abs(surprise)	0.0000	0.0054
9	Macro	Nat. Gas Stocks	-1	announcement	0.0001	0.0054
10	Macro	Oil Stocks	-1	abs(surprise)	0.0000	0.0045
11	Macro	Wholesale Inv.	-1	abs(min(0,surprise))	0.0000	0.0040
12	Macro	FOMC Rate Dec.	0	announcement	0.0196	0.0040
13	Macro	NAHB	-1	abs(surprise)	0.0000	0.0025
14	Macro	Nat. Gas Stocks	+1	announcement	-0.0001	0.0024
15	Macro	Constr. Spending	-1	announcement	0.0000	0.0024
16	Macro	Business Inv.	-1	announcement	0.0001	0.0019
17	Macro	Federal Budget	-1	announcement	-0.0001	0.0018
18	Macro	ECRI	-1	max(0,surprise)	0.0000	0.0016
19	Macro	Cons. Confidence	-1	announcement	-0.0001	0.0015
20	Macro	ISM Man. PMI	-1	announcement	-0.0001	0.0013

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 57 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 50 assets for which, in addition to the condition above, EPS are released during market hours.

Table J23: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0196	0.0040
2	StreetAcc.	Newspapers	0 and +1	pos. sent., pers.	0.0175	0.0000
3	StreetAcc.	Newspapers	0	positive sent.	0.0175	0.0000
4	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0175	0.0000
5	T. Reuters	Top	0	announcement	0.0174	0.0000
6	T. Reuters	Earnings	0	announcement	0.0162	0.0000
7	StreetAcc.	Up/Downgrades	0	positive sent.	0.0074	0.0000
8	T. Reuters	Top	-1	announcement	0.0058	0.0000
9	StreetAcc.	Earnings	0	positive sent.	0.0050	0.0000
10	Macro	Cons. Credit	0	announcement	0.0032	0.0000
11	StreetAcc.	Up/Downgrades	0	announcement	0.0028	0.0000
12	T. Reuters	All	0	announcement	0.0025	0.0000
13	Macro	Michigan Sent.	-1	announcement	0.0022	-0.0016
14	StreetAcc.	Earnings	0	announcement	0.0020	0.0000
15	Macro	Michigan Sent.	0	announcement	0.0016	0.0000
16	StreetAcc.	Litigation	0	announcement	0.0015	0.0000
17	StreetAcc.	Regulatory	0	positive sent.	0.0011	0.0000
18	Macro	FOMC Rate Dec.	-1	announcement	0.0011	0.0056
19	StreetAcc.	All	-1	n. news ≥ 2	0.0011	0.0000
20	StreetAcc.	All	0	announcement	0.0009	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 57 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 50 assets for which, in addition to the condition above, EPS are released during market hours.

Table J24: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0196	0.0040
2	StreetAcc.	Newspapers	0 and +1	pos. sent., pers.	0.0175	0.0000
3	StreetAcc.	Newspapers	0	positive sent.	0.0175	0.0000
4	StreetAcc.	Earnings	-1 and 0	pos. sent., pers.	0.0175	0.0000
5	T. Reuters	Top	0	announcement	0.0174	0.0000
6	T. Reuters	Earnings	0	announcement	0.0162	0.0000
7	StreetAcc.	Up/Downgrades	0	positive sent.	0.0074	0.0000
8	T. Reuters	Top	-1	announcement	0.0058	0.0000
9	StreetAcc.	Earnings	0	positive sent.	0.0050	0.0000
10	Macro	Cons. Credit	0	announcement	0.0032	0.0000
11	StreetAcc.	Up/Downgrades	0	announcement	0.0028	0.0000
12	T. Reuters	All	0	announcement	0.0025	0.0000
13	Macro	Michigan Sent.	-1	announcement	0.0022	-0.0016
14	StreetAcc.	Earnings	0	announcement	0.0020	0.0000
15	Macro	Michigan Sent.	0	announcement	0.0016	0.0000
16	StreetAcc.	Litigation	0	announcement	0.0015	0.0000
17	StreetAcc.	Regulatory	0	positive sent.	0.0011	0.0000
18	Macro	FOMC Rate Dec.	-1	announcement	0.0011	0.0056
19	StreetAcc.	All	-1	n. news ≥ 2	0.0011	0.0000
20	StreetAcc.	All	0	announcement	0.0009	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 57 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 50 assets for which, in addition to the condition above, EPS are released during market hours.

Table J25: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	-1	announcement	0.0046	0.0106
2	StreetAcc.	All	-1 and 0	persistence	0.0003	0.0069
3	T. Reuters	High	-1	announcement	0.0000	0.0068
4	T. Reuters	All	0	announcement	0.0000	0.0056
5	Macro	Federal Budget	-1	announcement	0.0000	0.0049
6	Macro	Wholesale Inv.	-1	announcement	-0.0001	0.0041
7	Macro	Oil Stocks	-1	abs(min(0,surprise))	0.0000	0.0039
8	T. Reuters	All	-1 and 0	persistence	0.0000	0.0024
9	Macro	ISM Man. PMI	-1	announcement	0.0000	0.0022
10	Macro	Wholesale Inv.	-1	max(0,surprise)	0.0000	0.0021
11	StreetAcc.	All	+1	negative sent.	0.0001	0.0009
12	Macro	FOMC Rate Dec.	0	announcement	0.0438	0.0006
13	Macro	Federal Budget	+1	announcement	0.0000	0.0002
14	Macro	Federal Budget	0	announcement	0.0000	0.0002
15	T. Reuters	Medium	-1	announcement	0.0002	0.0001
16	Macro	FOMC Rate Dec.	+1	announcement	0.0000	0.0001
17	Macro	NAHB	0	announcement	0.0000	0.0001
18	Macro	NAHB	+1	announcement	-0.0001	0.0001
19	Macro	NAHB	-1	announcement	-0.0001	0.0001
20	Macro	Empl. Trends	-1	announcement	0.0002	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 55 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

Table J26: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0438	0.0006
2	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0355	0.0000
3	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0181	0.0000
4	StreetAcc.	Earnings	0	negative sent.	0.0178	0.0000
5	StreetAcc.	All	+1	n. news ≥ 2	0.0175	0.0000
6	T. Reuters	Top	0	announcement	0.0168	0.0000
7	T. Reuters	Top	0 and +1	persistence	0.0121	0.0000
8	StreetAcc.	All	0	negative sent.	0.0114	0.0000
9	Macro	FOMC Rate Dec.	-1	announcement	0.0046	0.0106
10	Macro	Michigan Sent.	-1	announcement	0.0046	-0.0022
11	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0041	0.0000
12	StreetAcc.	All	0	announcement	0.0036	0.0000
13	StreetAcc.	Earnings	-1 and 0	neg. sent., pers.	0.0022	0.0000
14	T. Reuters	Top	+1	announcement	0.0019	0.0000
15	StreetAcc.	Earnings	0 and +1	neg. sent., pers.	0.0018	0.0000
16	Macro	Nat. Gas Stocks	0	announcement	0.0013	-0.0004
17	StreetAcc.	Earnings	+1	positive sent.	0.0008	0.0000
18	T. Reuters	All	+1	announcement	0.0005	-0.0007
19	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0005	0.0000
20	StreetAcc.	Newspapers	0	announcement	0.0005	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 55 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

Table J27: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2007 - Jun 2009 (global financial crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.0438	0.0006
2	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0355	0.0000
3	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0181	0.0000
4	StreetAcc.	Earnings	0	negative sent.	0.0178	0.0000
5	StreetAcc.	All	+1	n. news ≥ 2	0.0175	0.0000
6	T. Reuters	Top	0	announcement	0.0168	0.0000
7	T. Reuters	Top	0 and +1	persistence	0.0121	0.0000
8	StreetAcc.	All	0	negative sent.	0.0114	0.0000
9	Macro	FOMC Rate Dec.	-1	announcement	0.0046	0.0106
10	Macro	Michigan Sent.	-1	announcement	0.0046	-0.0022
11	StreetAcc.	Newspapers	-1	n. news ≥ 2	0.0041	0.0000
12	StreetAcc.	All	0	announcement	0.0036	0.0000
13	StreetAcc.	Earnings	-1 and 0	neg. sent., pers.	0.0022	0.0000
14	T. Reuters	Top	+1	announcement	0.0019	0.0000
15	StreetAcc.	Earnings	0 and +1	neg. sent., pers.	0.0018	0.0000
16	Macro	Nat. Gas Stocks	0	announcement	0.0013	-0.0004
17	StreetAcc.	Earnings	+1	positive sent.	0.0008	0.0000
18	T. Reuters	All	+1	announcement	0.0005	-0.0007
19	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0005	0.0000
20	StreetAcc.	Newspapers	0	announcement	0.0005	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2007 - Jun 2009 (global financial crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 55 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 49 assets for which, in addition to the condition above, EPS are released during market hours.

J.4.3 Jul 2009 - Dec 2012 (EU sovereign debt crisis)

Table J28: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Constr. Spending	0	abs(min(0,surprise))	0.0000	0.0195
2	Macro	Wholesale Inv.	-1	announcement	0.0000	0.0081
3	StreetAcc.	All	-1	sent. $\neq 0$	0.0000	0.0045
4	Macro	Nat. Gas Stocks	-1	announcement	0.0000	0.0031
5	StreetAcc.	All	0 and +1	persistence	0.0005	0.0030
6	Macro	Federal Budget	-1	abs(surprise)	0.0000	0.0027
7	Macro	Constr. Spending	0	abs(surprise)	0.0000	0.0025
8	Macro	ISM Man. PMI	0	max(0,surprise)	0.0000	0.0025
9	Macro	Empl. Trends	-1	announcement	0.0000	0.0023
10	Macro	Nat. Gas Stocks	+1	abs(surprise)	0.0000	0.0022
11	StreetAcc.	All	-1	n. news ≥ 2	0.0023	0.0022
12	StreetAcc.	All	0	announcement	0.0129	0.0021
13	Macro	Michigan Sent.	-1	announcement	0.0000	0.0019
14	StreetAcc.	All	0	n. news ≥ 2	0.0211	0.0019
15	StreetAcc.	All	-1 and 0	persistence	0.0002	0.0018
16	Macro	Factory Orders	-1	announcement	0.0000	0.0017
17	Macro	Leading Index	-1	announcement	0.0000	0.0016
18	StreetAcc.	All	0	positive sent.	0.0116	0.0013
19	Macro	Oil Stocks	+1	abs(surprise)	0.0000	0.0012
20	Macro	ISM Man. PMI	0	abs(surprise)	0.0000	0.0012

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 88 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 78 assets for which, in addition to the condition above, EPS are released during market hours.

Table J29: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0226	0.0000
2	StreetAcc.	All	0	n. news ≥ 2	0.0211	0.0019
3	StreetAcc.	All	+1	n. news ≥ 2	0.0167	0.0000
4	Macro	FOMC Rate Dec.	0	announcement	0.0143	0.0003
5	StreetAcc.	All	0	announcement	0.0129	0.0021
6	StreetAcc.	Up/Downgrades	0	negative sent.	0.0120	0.0000
7	StreetAcc.	All	0	positive sent.	0.0116	0.0013
8	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0114	0.0000
9	StreetAcc.	Earnings	0	negative sent.	0.0114	0.0000
10	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0113	0.0000
11	StreetAcc.	Newspapers	-1 and 0	pos. sent., pers.	0.0113	0.0000
12	StreetAcc.	M&A	0	n. news ≥ 2	0.0113	0.0000
13	StreetAcc.	Regulatory	0	positive sent.	0.0097	0.0000
14	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0079	0.0000
15	StreetAcc.	Litigation	0	announcement	0.0055	0.0000
16	T. Reuters	High	0	positive sent.	0.0053	0.0000
17	StreetAcc.	Regulatory	0	n. news ≥ 2	0.0042	0.0000
18	Macro	Leading Index	0	announcement	0.0034	0.0005
19	StreetAcc.	Regulatory	+1	announcement	0.0024	0.0000
20	StreetAcc.	All	-1	n. news ≥ 2	0.0023	0.0022

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 88 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 78 assets for which, in addition to the condition above, EPS are released during market hours.

Table J30: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0226	0.0000
2	StreetAcc.	All	0	n. news ≥ 2	0.0211	0.0019
3	StreetAcc.	All	+1	n. news ≥ 2	0.0167	0.0000
4	Macro	FOMC Rate Dec.	0	announcement	0.0143	0.0003
5	StreetAcc.	All	0	announcement	0.0129	0.0021
6	StreetAcc.	Up/Downgrades	0	negative sent.	0.0120	0.0000
7	StreetAcc.	All	0	positive sent.	0.0116	0.0013
8	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0114	0.0000
9	StreetAcc.	Earnings	0	negative sent.	0.0114	0.0000
10	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0113	0.0000
11	StreetAcc.	Newspapers	-1 and 0	pos. sent., pers.	0.0113	0.0000
12	StreetAcc.	M&A	0	n. news ≥ 2	0.0113	0.0000
13	StreetAcc.	Regulatory	0	positive sent.	0.0097	0.0000
14	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0079	0.0000
15	StreetAcc.	Litigation	0	announcement	0.0055	0.0000
16	T. Reuters	High	0	positive sent.	0.0053	0.0000
17	StreetAcc.	Regulatory	0	n. news ≥ 2	0.0042	0.0000
18	Macro	Leading Index	0	announcement	0.0034	0.0005
19	StreetAcc.	Regulatory	+1	announcement	0.0024	0.0000
20	StreetAcc.	All	-1	n. news ≥ 2	0.0023	0.0022

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 88 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 78 assets for which, in addition to the condition above, EPS are released during market hours.

Table J31: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Constr. Spending	0	abs(min(0,surprise))	0.0000	0.0247
2	StreetAcc.	All	-1	sent. $\neq 0$	0.0000	0.0084
3	Macro	Wholesale Inv.	-1	announcement	0.0001	0.0068
4	StreetAcc.	All	-1 and 0	persistence	0.0002	0.0061
5	Macro	Empl. Trends	-1	announcement	0.0000	0.0057
6	Macro	ISM Man. PMI	0	max(0,surprise)	0.0000	0.0041
7	Macro	Constr. Spending	0	abs(surprise)	0.0000	0.0025
8	StreetAcc.	All	0	positive sent.	0.0035	0.0024
9	Macro	Nat. Gas Stocks	+1	abs(surprise)	0.0000	0.0023
10	Macro	Michigan Sent.	-1	abs(surprise)	0.0000	0.0013
11	Macro	ISM Man. PMI	0	abs(surprise)	0.0000	0.0012
12	Macro	Factory Orders	-1	announcement	0.0000	0.0008
13	StreetAcc.	All	0	announcement	0.0087	0.0006
14	Macro	Ex. Home Sales	-1	abs(surprise)	0.0000	0.0006
15	Macro	NAHB	0	announcement	0.0005	0.0005
16	Macro	Michigan Sent.	-1	announcement	0.0001	0.0004
17	StreetAcc.	All	0 and +1	persistence	0.0000	0.0003
18	Macro	ECRI	0	announcement	0.0000	0.0003
19	StreetAcc.	All	+1	announcement	0.0116	0.0003
20	Macro	FOMC Rate Dec.	-1	announcement	0.0007	0.0001

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 87 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table J32: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0	n. news ≥ 2	0.0250	0.0000
2	Macro	FOMC Rate Dec.	0	announcement	0.0137	0.0000
3	StreetAcc.	All	+1	announcement	0.0116	0.0003
4	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0115	0.0000
5	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0115	0.0000
6	StreetAcc.	M&A	0	n. news ≥ 2	0.0115	0.0000
7	StreetAcc.	Regulatory	0	positive sent.	0.0114	0.0000
8	StreetAcc.	Newspapers	-1 and 0	pos. sent., pers.	0.0114	0.0000
9	StreetAcc.	All	0	announcement	0.0087	0.0006
10	StreetAcc.	Litigation	0	announcement	0.0060	0.0000
11	T. Reuters	High	0	positive sent.	0.0056	0.0000
12	StreetAcc.	Newspapers	+1	positive sent.	0.0053	0.0000
13	StreetAcc.	All	0	positive sent.	0.0035	0.0024
14	StreetAcc.	All	-1	announcement	0.0026	-0.0009
15	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0019	0.0000
16	StreetAcc.	Earnings	0	negative sent.	0.0018	0.0000
17	StreetAcc.	All	-1	n. news ≥ 2	0.0015	0.0000
18	T. Reuters	Financial	0	sent. $\neq 0$	0.0014	0.0000
19	StreetAcc.	M&A	0	announcement	0.0013	0.0000
20	T. Reuters	Financial	0	positive sent.	0.0012	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 87 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table J33: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0	n. news ≥ 2	0.0250	0.0000
2	Macro	FOMC Rate Dec.	0	announcement	0.0137	0.0000
3	StreetAcc.	All	+1	announcement	0.0116	0.0003
4	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0115	0.0000
5	StreetAcc.	Regulatory	+1	n. news ≥ 2	0.0115	0.0000
6	StreetAcc.	M&A	0	n. news ≥ 2	0.0115	0.0000
7	StreetAcc.	Regulatory	0	positive sent.	0.0114	0.0000
8	StreetAcc.	Newspapers	-1 and 0	pos. sent., pers.	0.0114	0.0000
9	StreetAcc.	All	0	announcement	0.0087	0.0006
10	StreetAcc.	Litigation	0	announcement	0.0060	0.0000
11	T. Reuters	High	0	positive sent.	0.0056	0.0000
12	StreetAcc.	Newspapers	+1	positive sent.	0.0053	0.0000
13	StreetAcc.	All	0	positive sent.	0.0035	0.0024
14	StreetAcc.	All	-1	announcement	0.0026	-0.0009
15	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0019	0.0000
16	StreetAcc.	Earnings	0	negative sent.	0.0018	0.0000
17	StreetAcc.	All	-1	n. news ≥ 2	0.0015	0.0000
18	T. Reuters	Financial	0	sent. $\neq 0$	0.0014	0.0000
19	StreetAcc.	M&A	0	announcement	0.0013	0.0000
20	T. Reuters	Financial	0	positive sent.	0.0012	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 87 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table J34: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0 and +1	persistence	0.0002	0.0161
2	Macro	Pend. Home Sales	-1	announcement	0.0002	0.0087
3	Macro	Nat. Gas Stocks	-1	announcement	0.0000	0.0084
4	StreetAcc.	All	0	announcement	0.0073	0.0075
5	Macro	Leading Index	-1	announcement	0.0000	0.0073
6	Macro	Michigan Sent.	-1	announcement	0.0000	0.0049
7	T. Reuters	All	0 and +1	persistence	0.0000	0.0044
8	Macro	Oil Stocks	0	announcement	0.0000	0.0033
9	StreetAcc.	All	+1	announcement	0.0036	0.0025
10	Macro	Nat. Gas Stocks	+1	announcement	0.0000	0.0024
11	Macro	Oil Stocks	-1	max(0,surprise)	0.0000	0.0023
12	Macro	Constr. Spending	-1	announcement	0.0000	0.0015
13	Macro	Constr. Spending	-1	abs(surprise)	0.0000	0.0013
14	StreetAcc.	Earnings	+1	announcement	0.0003	0.0013
15	Macro	Oil Stocks	-1	abs(surprise)	0.0000	0.0012
16	Macro	Leading Index	0	announcement	0.0063	0.0009
17	Macro	Wholesale Inv.	-1	announcement	0.0000	0.0008
18	StreetAcc.	All	+1	negative sent.	0.0017	0.0007
19	StreetAcc.	Newspapers	+1	announcement	0.0000	0.0007
20	T. Reuters	All	-1 and 0	persistence	0.0000	0.0005

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 66 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 60 assets for which, in addition to the condition above, EPS are released during market hours.

Table J35: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	Up/Downgrades	0	negative sent.	0.0154	0.0000
2	StreetAcc.	Earnings	0	negative sent.	0.0151	0.0000
3	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0149	0.0000
4	StreetAcc.	All	0	positive sent.	0.0122	0.0000
5	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0102	0.0000
6	StreetAcc.	All	0	announcement	0.0073	0.0075
7	Macro	Leading Index	0	announcement	0.0063	0.0009
8	StreetAcc.	All	+1	n. news ≥ 2	0.0061	0.0000
9	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0036	0.0000
10	StreetAcc.	All	+1	announcement	0.0036	0.0025
11	Macro	FOMC Rate Dec.	+1	announcement	0.0026	0.0001
12	StreetAcc.	Regulatory	+1	announcement	0.0026	0.0000
13	Macro	Phil. Fed	0	announcement	0.0017	-0.0001
14	StreetAcc.	All	+1	negative sent.	0.0017	0.0007
15	Macro	NAHB	0	announcement	0.0017	0.0000
16	Macro	Cons. Confidence	0	announcement	0.0016	0.0000
17	Macro	FOMC Rate Dec.	-1	announcement	0.0010	0.0002
18	StreetAcc.	Litigation	+1	announcement	0.0010	0.0000
19	StreetAcc.	Regulatory	0	sent. $\neq 0$	0.0010	0.0000
20	T. Reuters	Dividends	0	announcement	0.0008	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 66 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 60 assets for which, in addition to the condition above, EPS are released during market hours.

Table J36: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	Up/Downgrades	0	negative sent.	0.0154	0.0000
2	StreetAcc.	Earnings	0	negative sent.	0.0151	0.0000
3	StreetAcc.	Earnings	0 and +1	pos. sent., pers.	0.0149	0.0000
4	StreetAcc.	All	0	positive sent.	0.0122	0.0000
5	StreetAcc.	Regulatory	-1 and 0	neg. sent., pers.	0.0102	0.0000
6	StreetAcc.	All	0	announcement	0.0073	0.0075
7	Macro	Leading Index	0	announcement	0.0063	0.0009
8	StreetAcc.	All	+1	n. news ≥ 2	0.0061	0.0000
9	StreetAcc.	Earnings	-1	n. news ≥ 2	0.0036	0.0000
10	StreetAcc.	All	+1	announcement	0.0036	0.0025
11	Macro	FOMC Rate Dec.	+1	announcement	0.0026	0.0001
12	StreetAcc.	Regulatory	+1	announcement	0.0026	0.0000
13	Macro	Phil. Fed	0	announcement	0.0017	-0.0001
14	StreetAcc.	All	+1	negative sent.	0.0017	0.0007
15	Macro	NAHB	0	announcement	0.0017	0.0000
16	Macro	Cons. Confidence	0	announcement	0.0016	0.0000
17	Macro	FOMC Rate Dec.	-1	announcement	0.0010	0.0002
18	StreetAcc.	Litigation	+1	announcement	0.0010	0.0000
19	StreetAcc.	Regulatory	0	sent. $\neq 0$	0.0010	0.0000
20	T. Reuters	Dividends	0	announcement	0.0008	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 66 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 60 assets for which, in addition to the condition above, EPS are released during market hours.

J.4.4 Jan 2013 - Feb 2015 (post EU sovereign debt crisis)

Table J37: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	-1	announcement	-0.0001	0.0395
2	Macro	ISM Man. PMI	-1	announcement	0.0002	0.0057
3	T. Reuters	All	-1 and 0	persistence	0.0000	0.0038
4	Macro	ISM Man. PMI	-1	abs(surprise)	0.0000	0.0037
5	StreetAcc.	All	+1	announcement	0.0046	0.0033
6	Macro	Nat. Gas Stocks	+1	announcement	0.0000	0.0031
7	Macro	NAHB	-1	announcement	0.0000	0.0026
8	T. Reuters	All	-1	announcement	0.0000	0.0025
9	StreetAcc.	All	0 and +1	persistence	0.0000	0.0021
10	T. Reuters	Dividends	-1	announcement	0.0001	0.0014
11	StreetAcc.	Earnings	-1	positive sent.	0.0002	0.0014
12	StreetAcc.	Earnings	-1 and 0	persistence	0.0001	0.0010
13	Macro	Michigan Sent.	-1	abs(surprise)	0.0000	0.0006
14	Macro	Oil Stocks	0	abs(surprise)	0.0000	0.0006
15	Macro	Constr. Spending	0	announcement	0.0114	0.0006
16	StreetAcc.	Earnings	0 and +1	persistence	0.0002	0.0004
17	Macro	Federal Budget	-1	announcement	-0.0001	0.0004
18	Macro	Federal Budget	-1	max(0,surprise)	0.0000	0.0002
19	Macro	Federal Budget	-1	abs(surprise)	0.0000	0.0001
20	Macro	Federal Budget	+1	announcement	0.0125	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 86 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table J38: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.1145	0.0000
2	T. Reuters	Financial	0	announcement	0.0236	0.0000
3	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0215	0.0000
4	StreetAcc.	M&A	0	announcement	0.0173	0.0000
5	StreetAcc.	All	0	positive sent.	0.0138	0.0000
6	StreetAcc.	All	-1 and 0	neg. sent., pers.	0.0126	0.0000
7	StreetAcc.	Up/Downgrades	0	positive sent.	0.0126	0.0000
8	Macro	Federal Budget	+1	announcement	0.0125	0.0000
9	StreetAcc.	All	+1	n. news ≥ 2	0.0117	0.0000
10	StreetAcc.	All	-1	n. news ≥ 2	0.0116	0.0000
11	StreetAcc.	M&A	0	negative sent.	0.0116	0.0000
12	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0116	0.0000
13	T. Reuters	Earnings	0	announcement	0.0114	0.0000
14	Macro	Constr. Spending	0	announcement	0.0114	0.0006
15	StreetAcc.	All	0	negative sent.	0.0105	0.0000
16	StreetAcc.	All	0	announcement	0.0095	0.0000
17	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0091	0.0000
18	StreetAcc.	M&A	-1 and 0	persistence	0.0060	-0.0001
19	StreetAcc.	Up/Downgrades	0	announcement	0.0057	0.0000
20	StreetAcc.	All	+1	announcement	0.0046	0.0033

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 86 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table J39: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.1145	0.0000
2	T. Reuters	Financial	0	announcement	0.0236	0.0000
3	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0215	0.0000
4	StreetAcc.	M&A	0	announcement	0.0173	0.0000
5	StreetAcc.	All	0	positive sent.	0.0138	0.0000
6	StreetAcc.	All	-1 and 0	neg. sent., pers.	0.0126	0.0000
7	StreetAcc.	Up/Downgrades	0	positive sent.	0.0126	0.0000
8	Macro	Federal Budget	+1	announcement	0.0125	0.0000
9	StreetAcc.	All	+1	n. news ≥ 2	0.0117	0.0000
10	StreetAcc.	All	-1	n. news ≥ 2	0.0116	0.0000
11	StreetAcc.	M&A	0	negative sent.	0.0116	0.0000
12	StreetAcc.	All	0 and +1	pos. sent., pers.	0.0116	0.0000
13	T. Reuters	Earnings	0	announcement	0.0114	0.0000
14	Macro	Constr. Spending	0	announcement	0.0114	0.0006
15	StreetAcc.	All	0	negative sent.	0.0105	0.0000
16	StreetAcc.	All	0	announcement	0.0095	0.0000
17	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0091	0.0000
18	StreetAcc.	M&A	-1 and 0	persistence	0.0060	-0.0001
19	StreetAcc.	Up/Downgrades	0	announcement	0.0057	0.0000
20	StreetAcc.	All	+1	announcement	0.0046	0.0033

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). The average APE and the average AUC decrease of the indicators are computed over the 86 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 77 assets for which, in addition to the condition above, EPS are released during market hours.

Table J40: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	NAHB	-1	announcement	-0.0001	0.0073
2	StreetAcc.	All	+1	announcement	0.0008	0.0061
3	Macro	ISM Man. PMI	-1	announcement	0.0000	0.0055
4	StreetAcc.	All	0 and +1	persistence	0.0000	0.0036
5	Macro	ISM Man. PMI	-1	abs(surprise)	0.0000	0.0034
6	T. Reuters	Dividends	-1	announcement	0.0000	0.0029
7	Macro	Nat. Gas Stocks	-1	abs(surprise)	0.0000	0.0028
8	Macro	Nat. Gas Stocks	+1	announcement	0.0001	0.0020
9	StreetAcc.	Earnings	-1	positive sent.	0.0003	0.0020
10	StreetAcc.	All	-1 and 0	persistence	0.0000	0.0019
11	Macro	Constr. Spending	0	announcement	0.0159	0.0009
12	Macro	Constr. Spending	-1	announcement	0.0000	0.0004
13	Macro	Nat. Gas Stocks	+1	abs(surprise)	0.0000	0.0001
14	Macro	Federal Budget	+1	announcement	0.0000	0.0001
15	Macro	NAHB	-1	abs(min(0,surprise))	0.0000	0.0001
16	Macro	Phil. Fed	-1	abs(min(0,surprise))	0.0000	0.0000
17	Macro	Cons. Credit	0	announcement	0.0000	0.0000
18	Macro	Wholesale Inv.	0	announcement	0.0000	0.0000
19	StreetAcc.	Regulatory	-1	sent. \neq 0	0.0000	0.0000
20	T. Reuters	Financial	-1	announcement	0.0000	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 62 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 57 assets for which, in addition to the condition above, EPS are released during market hours.

Table J41: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.1192	0.0000
2	StreetAcc.	All	0	positive sent.	0.0186	0.0000
3	StreetAcc.	M&A	0	announcement	0.0178	0.0000
4	Macro	Constr. Spending	0	announcement	0.0159	0.0009
5	StreetAcc.	Up/Downgrades	0	positive sent.	0.0128	0.0000
6	StreetAcc.	All	-1 and 0	neg. sent., pers.	0.0114	0.0000
7	StreetAcc.	All	-1	n. news ≥ 2	0.0082	0.0000
8	StreetAcc.	Up/Downgrades	0	announcement	0.0078	0.0000
9	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0047	0.0000
10	StreetAcc.	Earnings	0	positive sent.	0.0044	0.0000
11	StreetAcc.	All	+1	positive sent.	0.0042	0.0000
12	StreetAcc.	All	0	sent. $\neq 0$	0.0040	0.0000
13	StreetAcc.	All	-1	negative sent.	0.0033	0.0000
14	StreetAcc.	M&A	0	sent. $\neq 0$	0.0029	0.0000
15	StreetAcc.	M&A	0	positive sent.	0.0024	0.0000
16	StreetAcc.	Earnings	+1	negative sent.	0.0023	0.0000
17	StreetAcc.	All	0	announcement	0.0017	0.0000
18	Macro	ECRI	0	announcement	0.0016	-0.0005
19	Macro	FOMC Rate Dec.	-1	announcement	0.0012	0.0000
20	Macro	Ex. Home Sales	-1	announcement	0.0011	-0.0010

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 62 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 57 assets for which, in addition to the condition above, EPS are released during market hours.

Table J42: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	FOMC Rate Dec.	0	announcement	0.1192	0.0000
2	StreetAcc.	All	0	positive sent.	0.0186	0.0000
3	StreetAcc.	M&A	0	announcement	0.0178	0.0000
4	Macro	Constr. Spending	0	announcement	0.0159	0.0009
5	StreetAcc.	Up/Downgrades	0	positive sent.	0.0128	0.0000
6	StreetAcc.	All	-1 and 0	neg. sent., pers.	0.0114	0.0000
7	StreetAcc.	All	-1	n. news ≥ 2	0.0082	0.0000
8	StreetAcc.	Up/Downgrades	0	announcement	0.0078	0.0000
9	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0047	0.0000
10	StreetAcc.	Earnings	0	positive sent.	0.0044	0.0000
11	StreetAcc.	All	+1	positive sent.	0.0042	0.0000
12	StreetAcc.	All	0	sent. $\neq 0$	0.0040	0.0000
13	StreetAcc.	All	-1	negative sent.	0.0033	0.0000
14	StreetAcc.	M&A	0	sent. $\neq 0$	0.0029	0.0000
15	StreetAcc.	M&A	0	positive sent.	0.0024	0.0000
16	StreetAcc.	Earnings	+1	negative sent.	0.0023	0.0000
17	StreetAcc.	All	0	announcement	0.0017	0.0000
18	Macro	ECRI	0	announcement	0.0016	-0.0005
19	Macro	FOMC Rate Dec.	-1	announcement	0.0012	0.0000
20	Macro	Ex. Home Sales	-1	announcement	0.0011	-0.0010

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only positive jumps. The average APE and the average AUC decrease of the indicators are computed over the 62 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 57 assets for which, in addition to the condition above, EPS are released during market hours.

Table J43: Most important indicators ordered by average AUC decrease. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	Macro	Michigan Sent.	-1	announcement	-0.0001	0.0636
2	T. Reuters	All	-1 and 0	persistence	0.0000	0.0055
3	T. Reuters	All	-1	announcement	0.0000	0.0037
4	Macro	Factory Orders	-1	announcement	0.0000	0.0024
5	StreetAcc.	Earnings	-1 and 0	persistence	0.0000	0.0021
6	Macro	Oil Stocks	0	abs(surprise)	0.0000	0.0020
7	StreetAcc.	Earnings	0 and +1	persistence	0.0002	0.0018
8	StreetAcc.	All	-1	announcement	0.0000	0.0009
9	Macro	Federal Budget	-1	announcement	-0.0001	0.0009
10	Macro	Michigan Sent.	-1	abs(surprise)	0.0000	0.0008
11	Macro	Federal Budget	-1	max(0,surprise)	0.0000	0.0004
12	Macro	Federal Budget	-1	abs(surprise)	0.0000	0.0004
13	Macro	Federal Budget	+1	announcement	0.0245	0.0000
14	Macro	Factory Orders	-1	abs(min(0,surprise))	0.0000	0.0000
15	Macro	NAHB	0	announcement	0.0000	0.0000
16	Macro	Nat. Gas Stocks	+1	abs(min(0,surprise))	0.0000	0.0000
17	StreetAcc.	M&A	-1	positive sent.	0.0000	0.0000
18	T. Reuters	Medium	-1	sent. \neq 0	0.0000	0.0000
19	StreetAcc.	M&A	-1	sent. \neq 0	0.0000	0.0000
20	T. Reuters	Medium	-1	announcement	0.0000	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of average AUC decrease after variable permutation. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 52 assets for which, in addition to the condition above, EPS are released during market hours.

Table J44: Most important indicators ordered by APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0341	0.0000
2	Macro	Federal Budget	+1	announcement	0.0245	0.0000
3	StreetAcc.	M&A	0	negative sent.	0.0169	0.0000
4	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0162	0.0000
5	StreetAcc.	All	0	announcement	0.0116	0.0000
6	StreetAcc.	All	+1	announcement	0.0055	-0.0006
7	Macro	Cons. Credit	0	announcement	0.0049	-0.0001
8	Macro	FOMC Rate Dec.	0	announcement	0.0040	0.0000
9	StreetAcc.	All	-1	n. news ≥ 2	0.0014	0.0000
10	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0011	0.0000
11	Macro	FOMC Rate Dec.	-1	announcement	0.0008	0.0000
12	StreetAcc.	M&A	+1	sent. $\neq 0$	0.0005	0.0000
13	StreetAcc.	All	+1	n. news ≥ 2	0.0004	0.0000
14	StreetAcc.	Earnings	+1	negative sent.	0.0004	0.0000
15	StreetAcc.	All	0	n. news ≥ 2	0.0003	0.0000
16	StreetAcc.	EPS	0	announcement	0.0003	0.0000
17	StreetAcc.	Earnings	0	n. news ≥ 2	0.0003	0.0000
18	StreetAcc.	M&A	+1	n. news ≥ 2	0.0003	0.0000
19	StreetAcc.	Earnings	+1	announcement	0.0002	0.0000
20	StreetAcc.	Litigation	+1	sent. $\neq 0$	0.0002	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 52 assets for which, in addition to the condition above, EPS are released during market hours.

Table J45: Most important indicators ordered by absolute APE. Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique; sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis); negative jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	APE	Δ AUC
1	StreetAcc.	All	0 and +1	neg. sent., pers.	0.0341	0.0000
2	Macro	Federal Budget	+1	announcement	0.0245	0.0000
3	StreetAcc.	M&A	0	negative sent.	0.0169	0.0000
4	StreetAcc.	Regulatory	0 and +1	neg. sent., pers.	0.0162	0.0000
5	StreetAcc.	All	0	announcement	0.0116	0.0000
6	StreetAcc.	All	+1	announcement	0.0055	-0.0006
7	Macro	Cons. Credit	0	announcement	0.0049	-0.0001
8	Macro	FOMC Rate Dec.	0	announcement	0.0040	0.0000
9	StreetAcc.	All	-1	n. news ≥ 2	0.0014	0.0000
10	StreetAcc.	Earnings	0	sent. $\neq 0$	0.0011	0.0000
11	Macro	FOMC Rate Dec.	-1	announcement	0.0008	0.0000
12	StreetAcc.	M&A	+1	sent. $\neq 0$	0.0005	0.0000
13	StreetAcc.	All	+1	n. news ≥ 2	0.0004	0.0000
14	StreetAcc.	Earnings	+1	negative sent.	0.0004	0.0000
15	StreetAcc.	All	0	n. news ≥ 2	0.0003	0.0000
16	StreetAcc.	EPS	0	announcement	0.0003	0.0000
17	StreetAcc.	Earnings	0	n. news ≥ 2	0.0003	0.0000
18	StreetAcc.	M&A	+1	n. news ≥ 2	0.0003	0.0000
19	StreetAcc.	Earnings	+1	announcement	0.0002	0.0000
20	StreetAcc.	Litigation	+1	sent. $\neq 0$	0.0002	0.0000

Notes: First 20 regressors sorted in descending order by cross-sectional mean of absolute APE. Estimation method: Penalized Logistic Regression with $\alpha = 0.1$ and no IS technique. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis). Only negative jumps. The average APE and the average AUC decrease of the indicators are computed over the 59 assets characterized by at least one fold with at least two jumps in the train set and at least one jump in the test set; those of the EPS-related indicators are computed over the 52 assets for which, in addition to the condition above, EPS are released during market hours.

K Additional Tables: Standard Logistic Regression

Table K1: Most important indicators ordered by significant beta. Standard Logistic Regression; sample = Jan 2005 - Feb 2015 (full sample); all jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	av. β	n. β^*	n. $\beta^* > 0$
1	Macro	FOMC Rate Dec.	0	announcement	5.11	87	87
2	Macro	FOMC Rate Dec.	0	abs(surprise)	0.04	81	81
3	StreetAcc.	All	0 and +1	persistence	1.81	77	78
4	Macro	FOMC Rate Dec.	-1	announcement	1.78	75	78
5	Macro	FOMC Rate Dec.	0	min(0,surprise)	-0.02	74	0
6	StreetAcc.	All	0	announcement	0.95	68	69
7	StreetAcc.	All	-1 and 0	persistence	1.14	67	72
8	StreetAcc.	All	+1	announcement	-0.43	60	63
9	Macro	Constr. Spending	0	announcement	-1.23	56	56
10	StreetAcc.	All	0	sent. $\neq 0$	-0.78	55	55
11	Macro	ISM Man. PMI	0	announcement	-1.37	55	55
12	Macro	ISM Man. PMI	0	abs(surprise)	-1.33	54	54
13	Macro	ISM Man. PMI	0	max(0,surprise)	-0.75	53	53
14	Macro	Constr. Spending	0	abs(surprise)	-0.67	50	50
15	Macro	Constr. Spending	0	abs(min(0,surprise))	-0.52	50	52
16	StreetAcc.	All	+1	sent. $\neq 0$	-2.22	49	50
17	StreetAcc.	All	-1	announcement	-1.57	44	47
18	StreetAcc.	All	0	positive sent.	-2.22	43	43
19	Macro	Federal Budget	-1	max(0,surprise)	-181.15	43	44
20	Macro	Federal Budget	-1	abs(surprise)	-162.94	42	44

Notes: Cross-sectional mean of beta, number of assets with a 5% significant two-tailed test for $\beta \neq 0$ and number of assets with a 5% significant one-tailed test for $\beta > 0$ for the first 20 regressors sorted in descending order by number of assets with a 5% significant two-tailed test for $\beta \neq 0$.

Estimation method: standard Logistic Regression. Sample = Jan 2005 - Feb 2015 (full sample). The average beta of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours. The number of assets with a significant beta for the EPS-related indicators is adjusted by multiplying by the number of assets (88) and dividing by the number of assets for which EPS are released during market hours (78).

Table K2: Most important indicators ordered by significant beta $\neq 0$. Standard Logistic Regression; sample = Jan 2005 - Feb 2015 (full sample); positive jumps.

Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	av. β	n. β^*	n. $\beta^* > 0$
1	Macro	FOMC Rate Dec.	0	announcement	5.33	86	86
2	Macro	FOMC Rate Dec.	0	abs(surprise)	0.04	81	81
3	Macro	FOMC Rate Dec.	0	min(0,surprise)	0.02	74	74
4	StreetAcc.	All	0 and +1	persistence	-0.44	63	63
5	StreetAcc.	All	0	announcement	-0.77	57	57
6	StreetAcc.	All	-1 and 0	persistence	-0.76	57	62
7	Macro	FOMC Rate Dec.	-1	announcement	-1.75	57	58
8	Macro	Constr. Spending	0	announcement	-0.95	55	55
9	Macro	ISM Man. PMI	0	announcement	-1.10	54	54
10	Macro	ISM Man. PMI	0	abs(surprise)	-1.33	53	53
11	Macro	ISM Man. PMI	0	max(0,surprise)	-0.73	53	53
12	Macro	Constr. Spending	0	abs(min(0,surprise))	-0.53	51	52
13	Macro	Constr. Spending	0	abs(surprise)	-0.67	50	52
14	StreetAcc.	All	0	sent. $\neq 0$	-2.42	43	43
15	StreetAcc.	All	+1	announcement	-4.02	40	41
16	StreetAcc.	All	0	positive sent.	-2.66	38	38
17	StreetAcc.	All	-1	announcement	-4.52	35	40
18	T. Reuters	All	-1 and 0	persistence	-5.19	32	33
19	T. Reuters	Medium	-1 and 0	persistence	-5.06	30	32
20	T. Reuters	High	-1 and 0	persistence	-5.02	30	31

Notes: Cross-sectional mean of beta, number of assets with a 5% significant two-tailed test for beta $\neq 0$ and number of assets with a 5% significant one-tailed test for beta > 0 for the first 20 regressors sorted in descending order by number of assets with a 5% significant two-tailed test for beta $\neq 0$.

Estimation method: standard Logistic Regression. Sample = Jan 2005 - Feb 2015 (full sample). Only positive jumps. The average beta of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours. The number of assets with a significant beta for the EPS-related indicators is adjusted by multiplying by the number of assets (88) and dividing by the number of assets for which EPS are released during market hours (78).

Table K3: Most important indicators ordered by significant beta $\neq 0$. Standard Logistic Regression; sample = Jan 2005 - Feb 2015 (full sample); negative jumps.

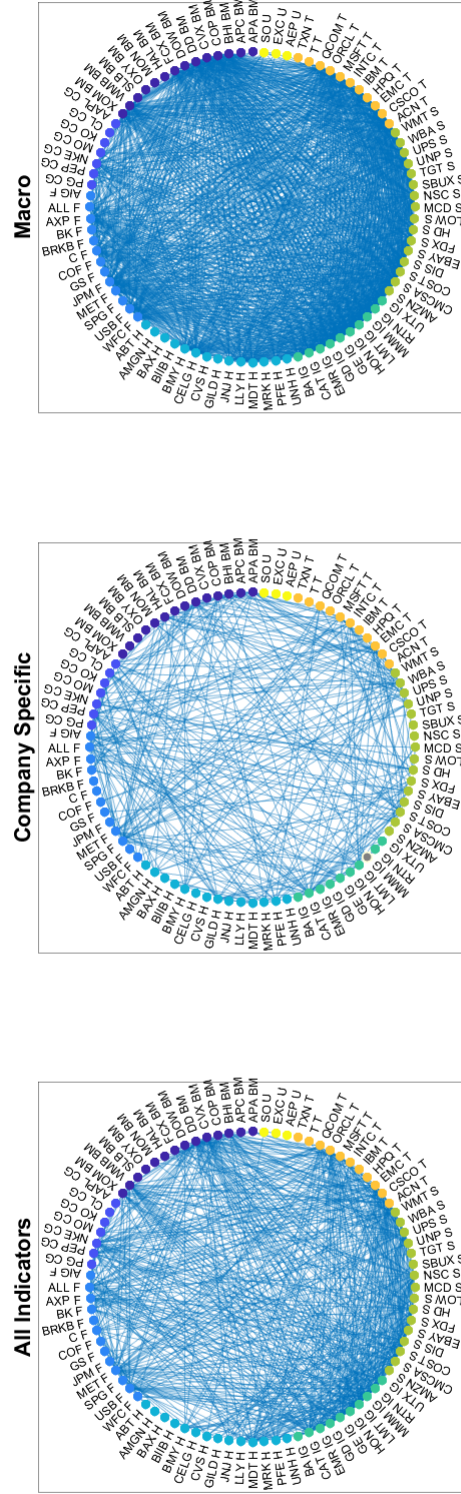
Rank	News Type	Topic/Macro Ann.	Lead-Lag	Measure	av. β	n. β^*	n. $\beta^* > 0$
1	Macro	FOMC Rate Dec.	-1	announcement	1.51	75	75
2	StreetAcc.	All	0 and +1	persistence	-0.60	59	63
3	StreetAcc.	All	-1 and 0	persistence	-1.30	51	57
4	StreetAcc.	All	+1	announcement	-2.65	49	49
5	Macro	FOMC Rate Dec.	0	announcement	-1.54	48	48
6	StreetAcc.	All	0	announcement	-2.39	47	47
7	Macro	Federal Budget	-1	abs(surprise)	-130.71	44	44
8	Macro	Federal Budget	-1	max(0,surprise)	-139.74	44	44
9	Macro	Federal Budget	-1	announcement	-4.27	37	45
10	StreetAcc.	All	-1	announcement	-3.39	36	44
11	StreetAcc.	All	-1	sent. $\neq 0$	-5.53	33	35
12	StreetAcc.	All	+1	sent. $\neq 0$	-4.77	32	32
13	StreetAcc.	All	0	sent. $\neq 0$	-4.55	30	30
14	StreetAcc.	Earnings	-1 and 0	persistence	-5.64	28	28
15	StreetAcc.	All	0	negative sent.	-4.51	25	25
16	StreetAcc.	All	+1	negative sent.	-5.11	25	25
17	StreetAcc.	Earnings	0 and +1	persistence	-5.46	24	24
18	StreetAcc.	Earnings	-1	announcement	-6.70	21	21
19	T. Reuters	All	-1 and 0	persistence	-6.96	21	22
20	T. Reuters	Medium	-1 and 0	persistence	-6.62	21	22

Notes: Cross-sectional mean of beta, number of assets with a 5% significant two-tailed test for beta $\neq 0$ and number of assets with a 5% significant one-tailed test for beta > 0 for the first 20 regressors sorted in descending order by number of assets with a 5% significant two-tailed test for beta $\neq 0$.

Estimation method: standard Logistic Regression. Sample = Jan 2005 - Feb 2015 (full sample). Only negative jumps. The average beta of the indicators are computed over all 88 assets; those of the EPS-related indicators are computed over the 78 assets for which EPS are released during market hours. The number of assets with a significant beta for the EPS-related indicators is adjusted by multiplying by the number of assets (88) and dividing by the number of assets for which EPS are released during market hours (78).

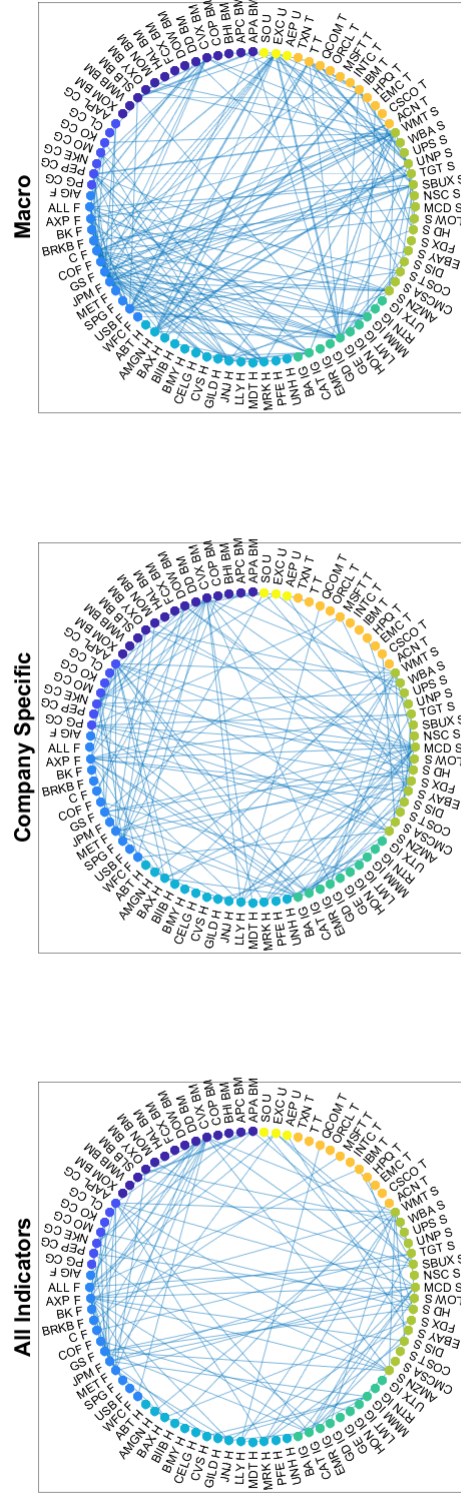
L Additional Figures and Tables: Assets Heterogeneity

Figure L1: Assets Heterogeneity. Only positive jumps. All sample.



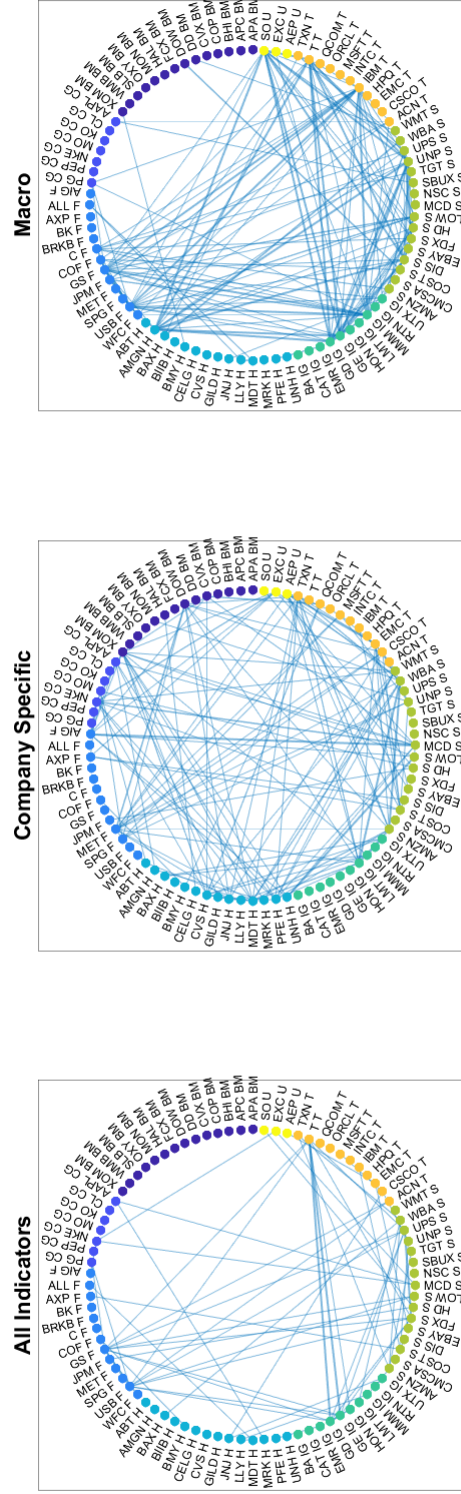
Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Figure L2: Assets Heterogeneity. Only negative jumps. All sample.



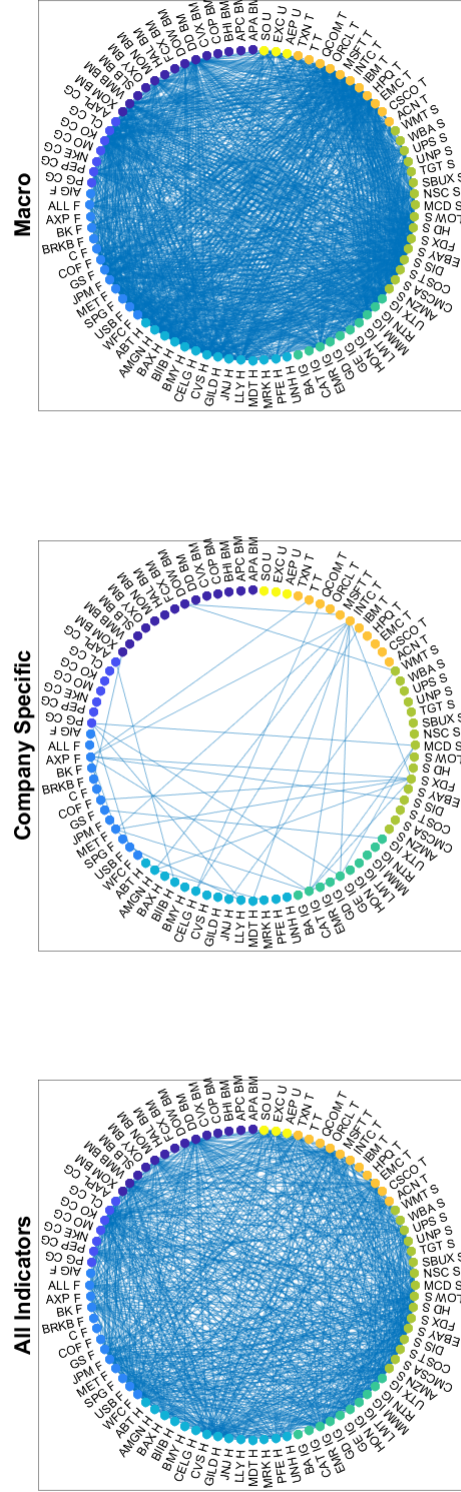
Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Figure L3: Assets Heterogeneity. All jumps. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis).



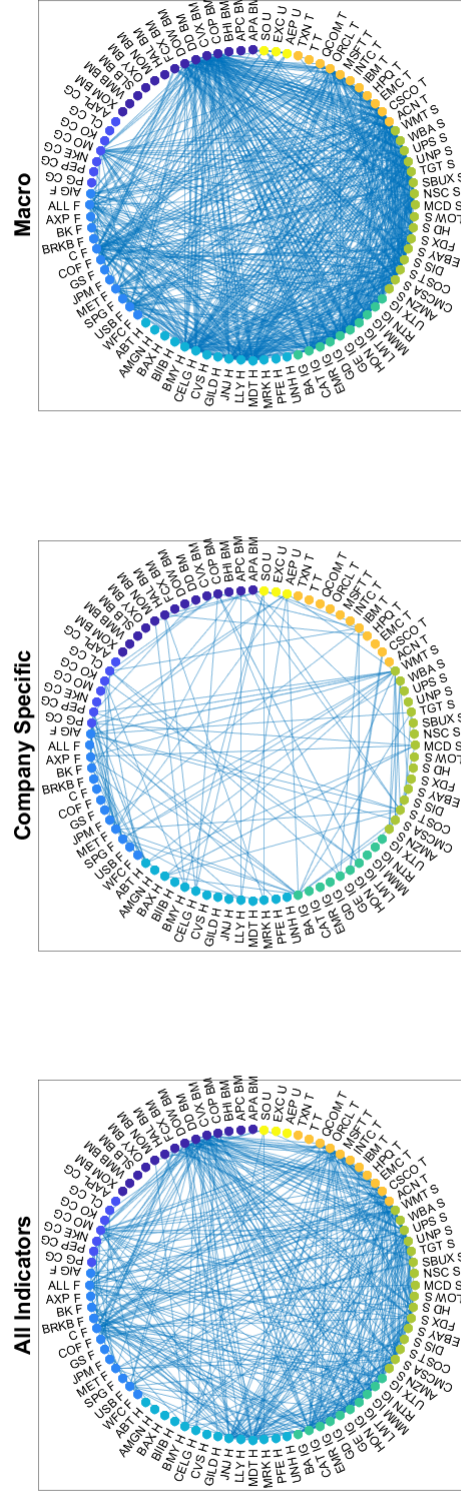
Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Figure I4: Assets Heterogeneity. All jumps. Sample = Jul 2007 - Jun 2009 (global financial crisis).



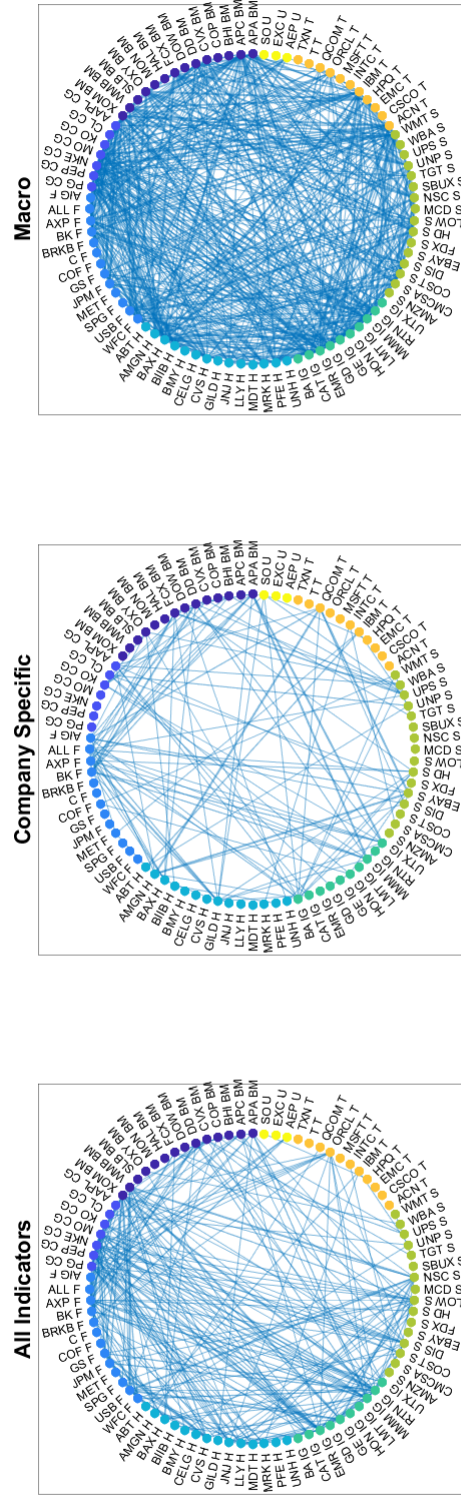
Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Figure L5: Assets Heterogeneity. All jumps. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).



Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Figure L6: Assets Heterogeneity. All jumps. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).



Notes: For each pair of assets A and B , the weight of the edge between them corresponds to $\mathcal{F}(A, B)$. Assets are ordered by sector: Basic Materials (BM), Consumer Goods (CG), Financial (F), Healthcare (H), Industrial Goods (IG), Services (S), Technology (T), Utilities (U). Left: all indicators; center: company-specific indicators; right: macro-indicators. For readability, the edges are shown only if $\mathcal{F}(A, B) > 0.30$ for all indicators and company-specific indicators, and if $\mathcal{F}(A, B) > 0.40$ for macro-indicators.

Table L1: Assets heterogeneity. Only positive jumps. All sample.

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.179	0.909	0.120
BM	14	0.000	0.156	0.818	0.151
CG	7	0.056	0.152	0.448	0.087
F	13	0.017	0.170	0.571	0.104
H	14	0.065	0.146	0.474	0.074
IG	10	0.258	0.375	0.706	0.097
S	16	0.000	0.195	0.529	0.117
T	11	0.000	0.188	0.643	0.140
(B) Company-specific					
All	88	0.000	0.059	1.000	0.122
BM	14	0.000	0.000	0.406	0.126
CG	7	0.000	0.150	0.333	0.109
F	13	0.000	0.000	0.500	0.134
H	14	0.000	0.095	0.583	0.109
IG	10	0.000	0.059	1.000	0.183
S	16	0.000	0.125	0.667	0.124
T	11	0.000	0.000	0.407	0.106
(C) Macro					
All	88	0.000	0.231	1.000	0.210
BM	14	0.000	0.100	0.900	0.275
CG	7	0.037	0.143	0.529	0.122
F	13	0.042	0.295	0.867	0.180
H	14	0.094	0.190	0.667	0.111
IG	10	0.353	0.571	0.778	0.100
S	16	0.000	0.218	0.909	0.240
T	11	0.000	0.300	0.818	0.251

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators.

Table L2: Assets heterogeneity. Only negative jumps. All sample.

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.073	1.000	0.087
BM	14	0.000	0.000	1.000	0.130
CG	7	0.031	0.125	0.625	0.124
F	13	0.025	0.160	0.571	0.097
H	14	0.000	0.083	0.419	0.083
IG	10	0.000	0.111	0.282	0.076
S	16	0.000	0.105	0.400	0.085
T	11	0.000	0.000	0.245	0.052
(B) Company-specific					
All	88	0.000	0.000	0.750	0.098
BM	14	0.000	0.000	0.500	0.091
CG	7	0.000	0.000	0.267	0.087
F	13	0.000	0.000	0.500	0.111
H	14	0.000	0.074	0.667	0.110
IG	10	0.000	0.000	0.444	0.089
S	16	0.000	0.078	0.333	0.103
T	11	0.000	0.000	0.261	0.062
(C) Macro					
All	88	0.000	0.091	1.000	0.138
BM	14	0.000	0.000	1.000	0.146
CG	7	0.053	0.200	0.625	0.127
F	13	0.071	0.262	1.000	0.178
H	14	0.000	0.067	0.500	0.112
IG	10	0.000	0.150	0.375	0.106
S	16	0.000	0.100	0.667	0.125
T	11	0.000	0.000	1.000	0.145

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators.

Table L3: Assets heterogeneity. All jumps. Sample = Jan 2005 - Jun 2007 (pre-global financial crisis).

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.034	1.000	0.087
BM	14	0.000	0.000	0.154	0.026
CG	7	0.000	0.000	0.194	0.051
F	13	0.000	0.095	1.000	0.136
H	14	0.000	0.083	0.333	0.069
IG	10	0.000	0.063	1.000	0.168
S	16	0.000	0.088	1.000	0.115
T	11	0.000	0.000	0.300	0.084
(B) Company-specific					
All	88	0.000	0.000	1.000	0.104
BM	14	0.000	0.000	0.333	0.045
CG	7	0.000	0.000	0.286	0.074
F	13	0.000	0.000	0.421	0.074
H	14	0.000	0.094	0.429	0.110
IG	10	0.000	0.000	0.500	0.076
S	16	0.000	0.000	1.000	0.129
T	11	0.000	0.000	0.364	0.105
(C) Macro					
All	88	0.000	0.000	1.000	0.136
BM	14	0.000	0.000	0.333	0.035
CG	7	0.000	0.000	0.091	0.028
F	13	0.000	0.148	1.000	0.175
H	14	0.000	0.000	0.750	0.093
IG	10	0.000	0.091	1.000	0.189
S	16	0.000	0.143	1.000	0.160
T	11	0.000	0.000	0.500	0.112

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators.

Table L4: Assets heterogeneity. All jumps. Sample = Jul 2007 - Jun 2009 (global financial crisis).

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.100	1.000	0.188
BM	14	0.000	0.000	0.750	0.132
CG	7	0.000	0.138	0.600	0.185
F	13	0.000	0.174	0.500	0.153
H	14	0.043	0.154	1.000	0.159
IG	10	0.000	0.000	0.667	0.172
S	16	0.000	0.128	1.000	0.222
T	11	0.000	0.125	1.000	0.181
(B) Company-specific					
All	88	0.000	0.000	0.667	0.056
BM	14	0.000	0.000	0.200	0.021
CG	7	0.000	0.000	0.100	0.022
F	13	0.000	0.000	0.385	0.082
H	14	0.000	0.000	0.333	0.060
IG	10	0.000	0.000	0.667	0.099
S	16	0.000	0.000	0.185	0.026
T	11	0.000	0.000	0.240	0.068
(C) Macro					
All	88	0.000	0.222	1.000	0.271
BM	14	0.000	0.000	0.750	0.148
CG	7	0.000	0.200	0.600	0.222
F	13	0.000	0.333	1.000	0.266
H	14	0.143	0.333	1.000	0.207
IG	10	0.000	0.000	0.750	0.223
S	16	0.000	0.240	1.000	0.260
T	11	0.000	0.400	1.000	0.361

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators.

Table L5: Assets heterogeneity. All jumps. Sample = Jul 2009 - Dec 2012 (EU sovereign debt crisis).

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.067	1.000	0.140
BM	14	0.000	0.000	0.538	0.095
CG	7	0.000	0.043	0.360	0.092
F	13	0.000	0.118	0.485	0.116
H	14	0.000	0.051	0.500	0.099
IG	10	0.105	0.306	0.667	0.136
S	16	0.000	0.138	0.667	0.153
T	11	0.000	0.061	0.667	0.199
(B) Company-specific					
All	88	0.000	0.000	1.000	0.090
BM	14	0.000	0.000	0.234	0.057
CG	7	0.000	0.000	0.500	0.114
F	13	0.000	0.000	0.588	0.124
H	14	0.000	0.000	0.300	0.071
IG	10	0.000	0.000	0.390	0.063
S	16	0.000	0.000	0.600	0.136
T	11	0.000	0.000	0.395	0.059
(C) Macro					
All	88	0.000	0.059	1.000	0.212
BM	14	0.000	0.000	0.778	0.158
CG	7	0.000	0.000	0.357	0.108
F	13	0.000	0.191	0.667	0.171
H	14	0.000	0.000	0.750	0.197
IG	10	0.214	0.440	0.909	0.139
S	16	0.000	0.118	1.000	0.263
T	11	0.000	0.000	0.857	0.252

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators.

Table L6: Assets heterogeneity. All jumps. Sample = Jan 2013 - Feb 2015 (post EU sovereign debt crisis).

Sector	N. Assets	Min(\mathcal{F})	Median(\mathcal{F})	Max(\mathcal{F})	Std(\mathcal{F})
(A) All Indicators					
All	88	0.000	0.056	1.000	0.131
BM	14	0.000	0.071	0.435	0.104
CG	7	0.042	0.200	1.000	0.290
F	13	0.000	0.032	1.000	0.215
H	14	0.000	0.071	1.000	0.128
IG	10	0.000	0.150	0.667	0.144
S	16	0.000	0.021	0.500	0.088
T	11	0.000	0.000	0.258	0.060
(B) Company-specific					
All	88	0.000	0.000	1.000	0.099
BM	14	0.000	0.000	0.545	0.104
CG	7	0.000	0.000	0.231	0.052
F	13	0.000	0.000	0.600	0.111
H	14	0.000	0.000	0.533	0.084
IG	10	0.000	0.000	1.000	0.168
S	16	0.000	0.000	0.500	0.091
T	11	0.000	0.000	0.259	0.051
(C) Macro					
All	88	0.000	0.071	1.000	0.233
BM	14	0.000	0.111	1.000	0.262
CG	7	0.000	0.400	1.000	0.267
F	13	0.000	0.000	1.000	0.263
H	14	0.000	0.100	1.000	0.213
IG	10	0.000	0.333	1.000	0.273
S	16	0.000	0.000	1.000	0.243
T	11	0.000	0.000	0.455	0.125

Notes: For all the assets and for each sector, the table shows the number of assets and the min, median, max, and standard deviation of \mathcal{F} . Panel (A): all indicators; Panel (B): company-specific indicators; Panel (C): macroeconomic indicators.