DISCUSSION PAPER

October 2011 ■ RFF DP 11-45

# Strategic Release of News at the EPA

**Lucija Muehlenbachs, Elisabeth Newcomb Sinha, and Nitish Ranjan Sinha**

**1616 P St. NW**
**Washington, DC 20036**
**202-328-5000   www.rff.org**

RESOURCES
FOR THE FUTURE

# Strategic Release of News at the EPA

Lucija Muehlenbachs, Elisabeth Newcomb Sinha, and Nitish Ranjan Sinha[*]

October 26, 2011

## Abstract

Using advances in text analysis, we examine the content and timing of 21,493 press releases issued by the U.S. Environmental Protection Agency (EPA) between 1994 and 2009. Press releases announcing enforcement actions or regulatory changes were issued more often on Fridays and before holidays, a time when news has the least impact on media coverage and financial markets. Changing the timing of press releases may increase deterrence through awareness of regulation and market reaction to environmental news. We find no evidence of regulatory capture. We compare text analysis techniques that allow data collection from sources previously too expensive to access.

*Jel–Classification:* Q58

*Keywords:* text analysis, computational linguistics, regulation, environment, politics, Environmental Protection Agency

On July 20, 2007, the EPA issued a press release announcing that E.I. Du Pont de Nemours & Co. would pay \$66 million to reduce its sulfur dioxide emissions after failing to comply with the Clean Air Act.[1] One of the goals of the EPA in releasing information is to provide a deterrent to other potential violators (EPA Press Director, personal communication, March 31, 2010). It would seem that this press release was an opportunity to send a message. The settlement was part of a greater campaign to reduce pollution from sulfuric acid manufacturers and air pollution generally. However, July 20, 2007, was a Friday, and Fridays are not a prime day for getting media attention. DellaVigna and Pollet (2009) document that investors pay less attention to media reports released on Fridays than to reports released on other weekdays. In addition, it is more difficult for reporters to follow up with interviews as the weekend begins, making it less likely that a story will be published at all if a press release is issued at the end of the week. Anecdotal evidence of government agencies releasing controversial announcements on Fridays abound.[2] DellaVigna and Pollet (2009) find that the president is less likely to sign a non-controversial executive order or law on a Friday. Gersen and O'Connell (2009) find that significant rules by agencies are slightly more likely to be issued on Fridays than insignificant rules, and hypothesize that agencies choose the timing to control who the audience is.

To fully exploit the effectiveness of public disclosure as a deterrent to environmental violation as well as promote awareness of environmental regulation, the EPA must manage both the content and the schedule of its public announcements. Using all the press releases (over 21,000) available from the EPA newsroom web-

---

[1]Environmental Protection Agency. "Du Pont agrees to spend \$66 million to reduce air pollution at four plants." Accessed June 20, 2011.

[2]Stephen Engelberg, "The Bad News Hour: 4 P.M. Friday," *The New York Times*, April 6, 1984: A20; Cindy Skrzycki, "New Rules Delivered Just in Time for Holidays," *Washington Post*, Jan 9, 2007: D1

site from August 1994 until October 2009, we analyze whether the press release policy of the EPA maximizes the effects of public disclosure on firms and the visibility of regulatory changes. Taking full advantage of such a strategy would imply releasing news about violations, settlements, and regulatory changes early in the week, when the public is most attentive, rather than on Friday, when there is likely less public and media scrutiny. To analyze the content of thousands of press releases, we employ multiple text analysis techniques. Machine reading allows us to examine more documents, over more dimensions, than would be possible with conventional hand-coding. We describe these text analysis techniques in some detail, as their use in the social sciences literature has been limited. We find that press releases about enforcement actions, which include descriptions of environmental violations and resulting punishments, are more often issued on Fridays and on days before holidays. Press releases mentioning environmental awards are less likely to be issued on Fridays and on days before holidays. These findings are inconsistent with the EPA trying to maximize the impact of the disclosure of enforcement actions. The EPA also frequently issues press releases about regulatory changes. Maximizing publicity for these changes could increase awareness of new regulations and advertise the EPA's activities. Consistent with the general objective of press releases—to increase awareness of regulatory change—we expect these press releases to appear early in the week. A disproportionate number of press releases mentioning regulatory changes occur on Friday, however.

Capital markets create important incentives for firms to reduce pollution. Examining the Toxics Release Inventory (TRI), one of the EPA's most significant examples of disclosure, Hamilton (1995) found that polluting firms experience significant negative returns following news about their emissions. He estimated these losses to be in the millions of dollars. Several other researchers have also

3

found a statistically negative effect on stock market performance resulting from the TRI, including Khanna, Quimio and Bojilova (1998) and Konar and Cohen (1997, 2001). Reduced stock prices can also translate into reduced pollution; firms that experience a large decrease in share price after disclosure of their pollution levels subsequently reduced their levels the most (Konar and Cohen, 1997). Foulon, Lanoie and Laplante (2002) examine the effect of the publication of lists of noncompliant companies in Canada, and find evidence that public disclosure strategies provide additional incentives beyond those provided by traditional enforcement. Though the EPA controls the format and timing of the TRI, the TRI necessarily focuses on a narrower set of environmental violations than press releases, which may address any environmental policy. Furthermore, a press release can reveal different details and nuances than TRI data.

The TRI is just one way to disclose environmental information. Researchers have also analyzed the effect of newspaper articles covering environmental news, though the results are not clear cut. Bosch, Eckard and Lee (1998) show that announcements in the *Wall Street Journal* of environmental violations result in negative returns for individual stocks. *Wall Street Journal* announcements of hazardous waste management lawsuit filings have a similar effect (Muoghalu, Robison and Glascock, 1990). Dasgupta, Laplante and Mamingi (2001) also find that stock prices react to environmental news in Argentina, Chile, Mexico, and the Philippines. Klassen and McLaughlin (1996) find a negative response to environmental incidents and a positive response to environmental recognition or performance in articles found in the NEXIS newswire database.

Negative environmental incidents reported in the *Wall Street Journal* involving oil or electricity companies had no effect on the stock prices of these companies, however (Jones and Rubin, 2001). Harper and Adams (1996) detect no stock

4

market effect for firms named potentially liable parties in a Superfund cleanup case. Karpoff, Lott and Wehrly (2005) point out that although there are statistically significant losses from environmental violations on firms' market value, these losses are of the same magnitude as the fines or penalties paid, suggesting that there is no reputational effect on stock value from environmental non-compliance.

In this paper, we take as given that disclosure of environmental violations matters for compliance, and turn our attention to the way that disclosure is managed. Though the evidence about the effects of environmental news on stock prices is mixed, there is evidence that firms benefit from strategically timing the release of bad news. Firms release bad news after securities markets have closed (Patell and Wolfson, 1982), later in the earnings quarter (Begley and Fischer, 1998), or on Fridays (DellaVigna and Pollet, 2009; Damodaran, 1989; Bagnoli, Clement and Watts, 2005). A survey by Graham, Harvey and Rajgopal (2005) indicates that managers delay bad news to Friday or after hours so that investors anticipate the news, dampening the decline in the stock price. Therefore, if the EPA were concerned about protecting firms from the stock market effect, then it would release negative news concerning the firm on a Friday. To examine disclosure, we use press releases rather than news articles, which have a few advantages. The EPA neither writes the content for news articles nor decides when they will be published, as it does with the TRI. The EPA's control of press releases allows for more direct policy implications. In addition, we can use all the available press releases, as opposed to selecting news outlets and articles for analysis.

Although we find that a disproportionate number of press releases detailing enforcement actions occurred on Friday, we find no evidence that the timing of these releases was influenced by individual firms. If the EPA was acting in the

best interest of firms (e.g., there is regulatory capture), it would have timed press releases that can harm the reputation of companies for Friday, so as to minimize scrutiny. The EPA is not more likely to release information about S&P 500 firms on Friday, nor is it more likely to release news disclosing specific dollar amounts on Friday. Firms that employed (or retained a lobbying firm that employed) past or future EPA staff are not more likely to have Friday press releases, even when those press releases are about enforcement actions. A limitation of our research is that we have not yet managed to match company names directly to press releases with great accuracy. Our estimates for S&P 500 companies estimates likely suffer from measurement error.

Our results suggest that the EPA might achieve more deterrence by tweaking its press release strategy. Disclosing information about violations earlier in the week should garner more attention from the public and from the market. We found no evidence that the EPA avoids this strategy because of lobbying, but it is possible that other pressures influence the timing of its news. The EPA often is a lightning rod for criticism,[3] so perhaps certain news items are placed on Friday to reduce negative attention. If this were the case, then we would expect the Friday effect to disappear when the political stakes are lower and there is less concern about negative attention, such as during a presidential lame duck period. At the end of an administration, appointees may be less concerned about backlash from controversial announcements. We find that press releases on a regulatory change are not more likely to come out on a Friday during a lame duck period. The Friday effect is much more pronounced during the same November-January period in years without an election.

---

[3]John M. Broder, "Bashing E.P.A. Is New Theme in G.O.P. Race," *The New York Times*, August 17, 2011

# 1 EPA Press Releases and Classification

We downloaded all press releases from the EPA newsroom website and employed two different computational linguistics methods to identify the content of press releases. Specifically, we categorize each press release as to whether it pertains to enforcement actions, awards, regulatory changes, money, or health. We describe the dictionary method and machine learning method of classification, and briefly discuss our choice of approach.

## 1.1 Dictionary Method of Classification

Though the use of text analysis techniques in the social sciences dates back several decades, reductions in the cost of computing have allowed faster growth in text analysis in recent years. Typically, a social scientist creates a "dictionary" of words or phrases that are indicative of certain information being present in text. A computer algorithm then searches the text for the words in the dictionary. The dictionary method is also known as a "bag of words" approach since it models a document as a collection of words, without regard to their ordering. Researchers have used this technique to identify the politics of parties and newspapers. Laver and Garry (2000) analyze words in party manifestos in Ireland to obtain policy positions of the parties that match the assessment of the parties' positions by political experts. By examining the choice of words and phrases such as "Death Tax" and "War on Terror," Gentzkow and Shapiro (2010) deduce a newspaper's political leaning. Others have analyzed news articles to identify their stock market effects. Tetlock (2007) analyzes the choice of words in *Wall Street Journal* articles and shows that the proportion of negative words as defined by the dictionary has the ability to predict stock market returns over the follow-

ing two or three days. Hanley and Hoberg (2010) document that an abundance of boilerplate language leads to underpricing of firms' initial public offerings.

For each category of press release (enforcement, awards, regulatory changes, money, or health), we defined a list of words, or dictionary, that was likely to signify membership in one of these categories. To classify press releases that refer to enforcement actions we created a dictionary that included phrases such as "penalize," "prosecute," or "fines," and selected any press release that contained one or more phrases from the dictionary. We created a dictionary of phrases such as "commend," "grant," or "achievement" to identify press releases that announced awards. To find content referring to regulatory changes, we created a dictionary including such phrases as "Federal Register," "finalize," and "new rule." To find press releases that mentioned money, we used a dictionary that contained only a dollar sign (\$). To locate press releases discussing health effects, we chose a longer list of words associated with health, such as "fatal," "cancer-causing," or "asthma," For each category we created dummy variables indicating that at least one of our dictionary words (or symbols) appeared in the press release. Our complete dictionaries for each topic are available in the appendix.

While the dictionary approach is simple to understand and implement, it has some important limitations. First, the choice of words for the dictionaries is rather arbitrary. Second, the method assumes that the occurrence of one word is independent of other words. There are many words with multiple definitions and connotations, of which only one meaning is of interest. One particularly troublesome word for our paper is "fine," which we would have wanted to include in an enforcement dictionary for its punitive connotation. Unfortunately, it is also a word frequently used by the EPA in the phrase "fine particulates". We therefore have to exclude it from our dictionary, and only use forms of "fine" that

are likely to have the connotations we want. Loughran and McDonald (2011) discover a similar problem in the dictionary employed by Tetlock (2007). They conclude that almost three-quarters of the negative words in the dictionary do not actually have a negative association in the finance context. Since dictionaries are so sensitive to context, and can introduce measurement error, we have explored these weaknesses and applied additional techniques.

## 1.2   Machine Learning Approach

Due to the imprecision of the dictionary approach, we employ a second, somewhat more involved method of classification to categorize enforcement, award, and regulatory change press releases. We borrow a method of text analysis from computer science: a naive Bayes model augmented with an N-gram language model (Peng and Schuurmans, 2003). Some economists have used the naive Bayes approach (Antweiler and Frank, 2004), but the combination with the N-gram model is newer to economists and we therefore describe it in some detail. Antweiler and Frank (2004) use a naive Bayesian scheme to classify postings on internet message boards to estimate their effect on stock returns. The naive Bayes classification assumes that words occur independently of each other. The scientist starts with a small sample of articles, where the category of each article is known. The classifier calculates the probability of a particular word occurring in each category from the small sample. The classifier also calculates the probability of each word occurring in the set of press releases. Armed with these two pieces of information and Bayes law, a scientist can classify any article given the words that appear in the article. Despite its simplifying and obviously incorrect assumption that the occurrence of one word is independent of the occurrence of another, this technique performs well.

Cavnar and Trenkle (1994) observe that classification based on the probability of character occurrences (such as the occurrence of letters, spaces, and their sequences) is more robust than classification based on full word occurrences. We use a character-based classifier, or N-character language model, as implemented in Alias-i [2009] to classify the press release data into different categories. The independence assumption of the naive Bayes model is relaxed by allowing local Markov chain dependence. The underlying assumption for the classifier is that the typical sequence of N characters in an enforcement article will be different from the sequence of characters in a non-enforcement article. For example, an enforcement article is more likely to contain the five-character sequence "fined" than an article that does not mention an enforcement action. Unlike the dictionary approach, we do not provide the classifier with an ex ante dictionary of words. Instead we let the classifier calculate the probability of all sequences of characters from the annotated articles.

We define N contiguous characters as N-grams, where an N-gram is a subsequence of N characters from the whole sequence of characters in the press release. For example, if we were to construct N-grams for $N \leq 2$ from a press release that only contained the word "fine", there will be seven N-grams—"f," "I," "n," "e," "fi,""in," and "ne." N-grams may identify more common elements in similar press releases than a word-based classification strategy. For example, to model the probability that the press release mentions an enforcement action, we first create a training set of hand-classified press releases. Imagine if the training set of press releases does not contain the word "fined" but only contains the word "fines"; then a simple word-based classification strategy does not associate the word "fined" with an enforcement action. However, the N-gram based strategy will associate a non-zero probability with the conjugation "fined," since it contains "f,""I," "n,""e,"

"fi,""in,""ne,""fin," "ine," and "fine."[4]

We classify a subsample of press releases by hand, so we know that a press release with the sequence of characters $d = c_1 c_2 \ldots c_K$ is classified as type $t$. By the chain rule, the probability of this sequence can be written as

$$P(c_1 c_2 \ldots c_K) = \prod_{i=1}^{K} P(c_i | c_1 \ldots c_{i-1}) \tag{1}$$

We approximate each N-gram sequence probability by assuming Markov N-gram independence. Only the N-1 preceding characters are relevant for predicting the occurrence of the subsequent character:

$$P(c_i | c_1 \ldots c_{i-1}) = P(c_i | c_{i-N+1} \ldots c_{i-1}) \tag{2}$$

For each press release of type $t$ in the training set, we calculate the probability of character sequences occurring based on frequencies of sequences in each type $t$. Let $m_{i-n+1}^{i}$ be the number of occurrences of the N-gram $c_{i-N+1} \ldots c_i$ in press release type $t$ and $M_{i-n+1}^{i-1}$ be the number of occurrences of N-gram $c_{i-N+1} \ldots c_{i-1}$ in press release type $t$. Therefore the observed probability of an N-gram sequence occurring in type $t$ is calculated as

$$\hat{P}(c_i | c_{i-N+1} \ldots c_{i-1}, t) = m_{i-n+1}^{i} / M_{i-n+1}^{i-1} \tag{3}$$

The observed probabilities of the N-gram sequence occurring in a press release of type $t$ are stored and used to classify the remaining unseen press releases as type $t$ or otherwise. That is, say we want to classify an unseen release, with the

---

[4]In computational linguistics, it is popular to algorithmically reduce words to root words, but the approach introduces error as well, e.g., gallery and gall may both be stemmed to gall (Manning and Schutze, 1999).

character sequence $u = u_1 u_2 \ldots u_J$, as a type $t \in T$. We would classify it as the type that has the largest posterior probability given the character sequence in the release:

$$t^* = \arg \max_{t \in T} P(t|u) \tag{4}$$

Applying Bayes rule, Equation (4) can be rewritten as:

$$t^* = \arg \max_{t \in T} P(t)P(u|c) \tag{5}$$

The Markov N-gram independence assumption allows us to rewrite Equation (5) as:

$$t^* = \arg \max_{t \in T} P(t) \prod_{i=1}^{J} \hat{P}(u_i|u_{i-N+1} \ldots u_{i-1}, t) \tag{6}$$

where $\hat{P}(u_i|u_{i-N+1} \ldots u_{i-1}, t)$ is the probability of each N-gram sequence appearing in the unseen press release. The probability is derived from the observed probability that the N-gram sequence would occur in the training set categorized as type $t$ by applying Equation (3). We assign the document to the more probable type.

## 2 EPA Press Release Data Description

Our textual data are derived from 21,493 press releases downloaded from the U.S. EPA newsroom website. For all electronically available press releases from August 1994 until October 2009, we captured the text of the press release and the date.[5] The number of press releases published online by the EPA has increased

---

[5]We also captured the topic categories assigned by the EPA, including agriculture, air, awards, compliance, emergency, energy, grants, international, partnership, pesticides, radiation, recovery, research, superfund, trash, underground, water, waste, and other. Inclusion of these

over time, as shown in Figure 1, from just 149 in 1995, the first full year of data, to 2,088 in 2008. For each press release we created variables for day of the week, sessions of Congress, lame duck periods, EPA administrators, presidential administrations, and federal holidays.[6]
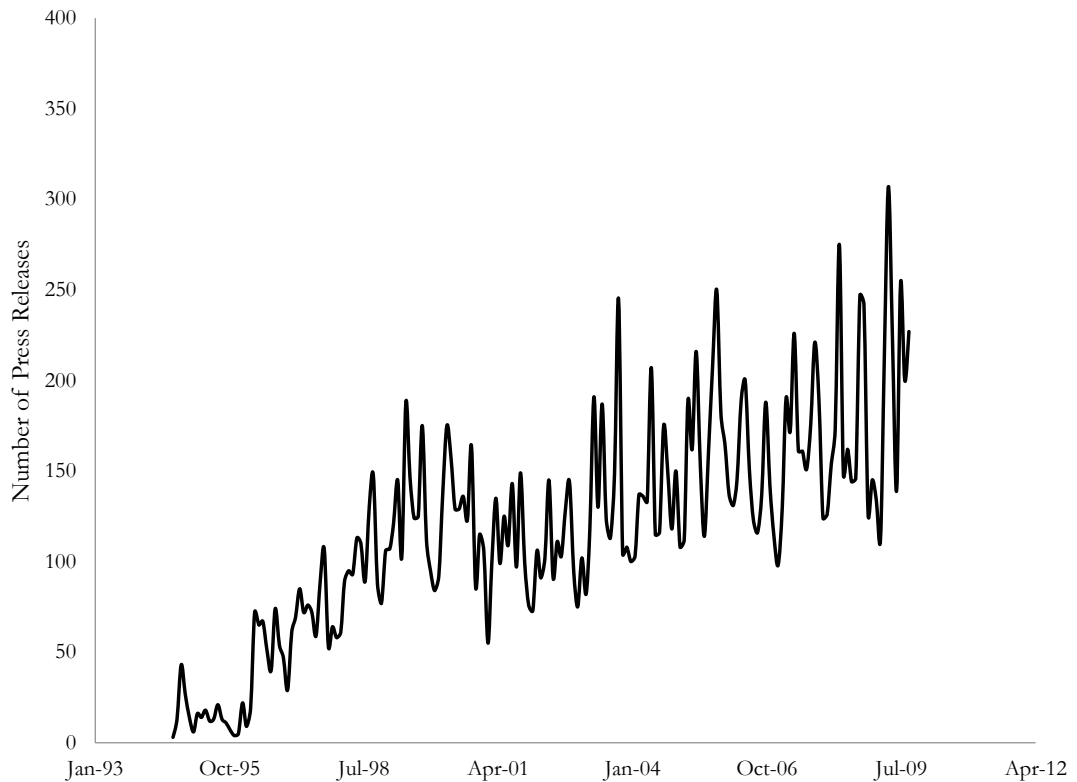


Figure 1: Time trend of number of press releases published on the EPA's website, 1994 to 2009.

To classify press releases, we randomly selected 150 press releases from our

category dummies in our regressions does not weaken our results, and yields no evidence that Friday press releases are driven by a single EPA office. We exclude these variables as they do not appear to be assigned consistently by the EPA.

[6]Specifically, the holidays examined are New Year's Day, Martin Luther King, Jr's Birthday, George Washington's Birthday, Memorial Day, Independence Day, Labor Day, Columbus Day, Veterans Day, Thanksgiving Day, and Christmas Day.

corpus of all the press releases. For each of the 150 articles, we annotated whether the article was related to an enforcement action or not; whether the article mentioned an award or not; and whether the article was about a regulatory change or not. We reserved 50 of the 150 articles as "unseen" to test the classifier. The classifier created all possible combinations of N characters in the 100 training articles, inclusive of the spaces in the text. Using the information about the category of each training article, the classifier computed the probability of a particular sequence of N characters being of the same category and not. We then created a dummy variable that takes a value of one if the likelihood of the press release being in the category exceeded the likelihood of its not being in that category. Using this method, we created variables indicating whether a press release was about enforcement actions, awards, and regulatory changes. We applied the classifier to the group of 50 "unseen" articles and observed that the accuracy for enforcement is about 87 percent, the accuracy for awards is about 94 percent, and the accuracy for regulatory change is about 58 percent for n=6 (Table 1). We find that the machine learning algorithm is more accurate in classifying the test releases than the dictionary approach. The dictionary approach (Table 2) results in more press releases being classified into a category than the machine learning method (for all categories: Reg. Change, Enforcement, and Awards). Nonetheless, our regression results for the dictionary approach are similar to those for the machine learning approach, so we present estimation results using the machine learning data only.

Table 2 also summarizes the day of the week dummies. Saturday and Sunday get three percent and two percent of press releases, respectively. Thursday gets most press releases, with 25.2 percent of press releases. Among weekdays, Monday gets least press releases, with 14.9 percent of press releases. Friday gets 20.1

14

Table 1: Comparing Algorithms: Correlation between Classification by Hand and Classification by Algorithms

|  | Algorithm | |
| --- | --- | --- |
|  | Machine Learning | Dictionary |
| Enforcement | 0.8729 | 0.6814 |
| Awards | 0.9451 | 0.7081 |
| Reg. Change | 0.5847 | 0.202 |

*Notes:* The first column shows the correlation coefficient between 50 hand-coded test press releases and the machine learning classification. The second column shows the correlation between the hand-coded press releases and the dictionary classification. The 50 test press releases are different from those used in the training dataset of the machine learning algorithm.

Table 2: Summary Statistics of Timing and Content Variables for Press Releases

|  | Mean | Std.Dev. | Min | Max |
| --- | --- | --- | --- | --- |
| Monday | 0.149 | 0.356 | 0 | 1 |
| Tuesday | 0.184 | 0.388 | 0 | 1 |
| Wednesday | 0.208 | 0.406 | 0 | 1 |
| Thursday | 0.252 | 0.434 | 0 | 1 |
| Friday | 0.201 | 0.401 | 0 | 1 |
| Saturday | 0.003 | 0.058 | 0 | 1 |
| Sunday | 0.002 | 0.045 | 0 | 1 |
| Before Holiday | 0.029 | 0.167 | 0 | 1 |
| Reg. Change (Machine Learning) | 0.095 | 0.293 | 0 | 1 |
| Reg. Change (Dictionary) | 0.151 | 0.358 | 0 | 1 |
| Enforcement (Machine Learning) | 0.297 | 0.457 | 0 | 1 |
| Enforcement (Dictionary) | 0.492 | 0.500 | 0 | 1 |
| Award (Machine Learning) | 0.198 | 0.398 | 0 | 1 |
| Award (Dictionary) | 0.356 | 0.479 | 0 | 1 |
| $\text{Firm}_{S\&P500}$ | 0.127 | 0.332 | 0 | 1 |
| $\text{Firm}_{Employed}$ | 0.006 | 0.074 | 0 | 1 |
| $\text{Firm}_{Lobby}$ | 0.078 | 0.268 | 0 | 1 |
| Lame Duck President | 0.029 | 0.166 | 0 | 1 |
| Non-lame Duck Tue. in Nov. to Jan. 20 | 0.126 | 0.332 | 0 | 1 |
| Health | 0.513 | 0.500 | 0 | 1 |
| Dollar | 0.499 | 0.500 | 0 | 1 |
| Word Count | 410.201 | 321.115 | 9 | 15,659 |
| Observations | 21,493 | | | |

*Notes:* Summary statistics for our variables based on the 21,493 press releases issued by the EPA from August 1994 to October 2009.

percent of all the press releases. Looking only at the volume of press releases, Friday does not appear to be a special day (Figure fig:bwhistogramforASSA). However, we observe that enforcement actions and regulatory changes are more likely to be announced on Fridays, while awards are more likely to be announced on Mondays (Figure 3). "Before Holiday" in Table 2 indicates the press release

was released on a weekday before a federal holiday. On days preceding federal holidays, 2.9 percent of press releases were released.
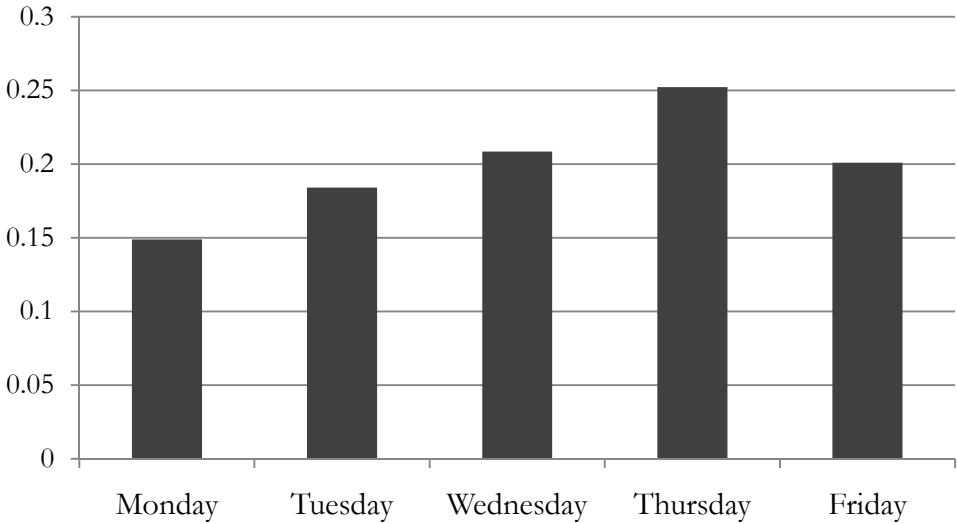


Figure 2: The distribution of press releases by the EPA across weekdays, 1994-2009.
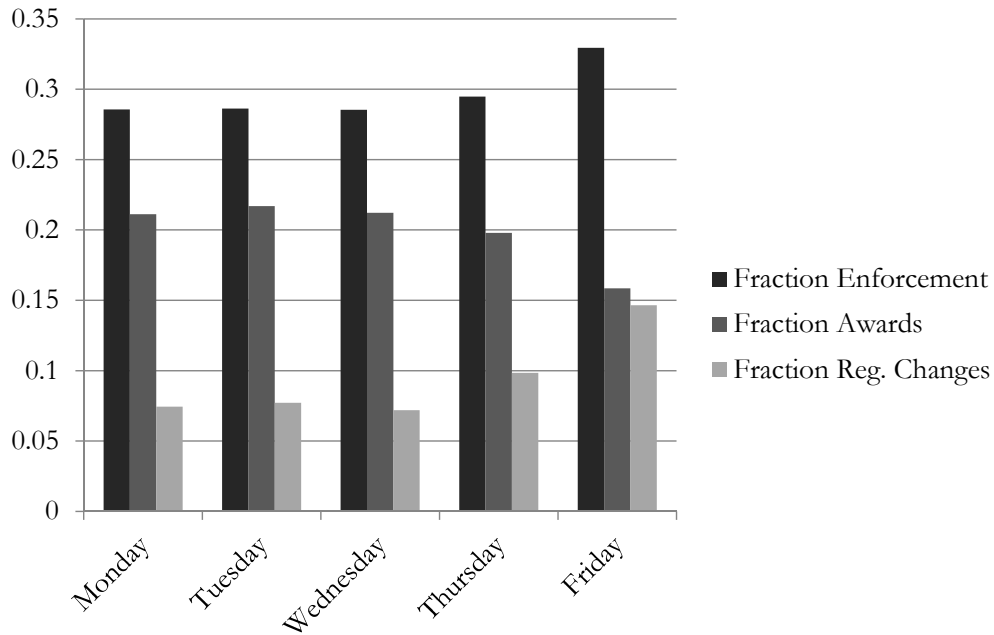
Figure 3: The fraction of press releases for three different categories (enforcement, awards, and regulatory changes) across weekdays.

"Lame duck President" indicates the press release being released during the lame duck tenure of a president. Almost three percent of press releases in our sample were released during a lame duck presidential period. "Non-lame duck Tue. in Nov. to Jan. 20" indicates the same period of time in years without presidential elections. A little over half of the press releases mention a word from our dictionary of words pertaining to health, and almost half contain the dollar symbol, $ (see "Health" and "Dollar" in Table 2). "Word Count" indicates the number of words in a press release; the average is 410 words.

## 2.1 Regulatory Capture Variables

We were also interested in whether companies (particularly those with a record of lobbying) seemed able to influence the timing of information from the EPA. We took several approaches to the problem of matching company names to individual

press releases. First, we generated a dictionary of names of companies that were part of the S&P 500 between 1994 and 2009. Due to changes in the composition of the S&P 500 and changes to company names during that time, the number of company names was close to 2000. We then simplified the names as much as possible to make matches with our data more likely. For example, "Kroger Company" became "Kroger," as this was still a unique identifier of the company, but was also more likely to be matched to a press release. In some cases, we were not able to simplify the name; we did not remove "Company" from "Southern Company," as "Southern" would have had many matches unrelated to "Southern Company." We then used this dictionary to develop a dummy variable taking a value of one if at least one S&P 500 company was mentioned in the press release. Almost 13 percent of the press releases in our sample contain an S&P 500 firm (Table 2).

Another approach was to identify companies that employed current or past EPA employees. The assumption is that a "revolving door" of employment between companies and the EPA may signal a particularly close relationship, and these companies may therefore have more influence on the timing of disclosure. We extracted names of companies that had hired current or past employees of the EPA according to the Center for Responsive Politics' website.[7] In our sample, 0.6 percent of press releases have firms that employ EPA current or past staff (Table 2). We also created a list of companies that retained lobbying firms employing past or future EPA staff. Each of these lists was used as a dictionary that generated a dummy variable equal to one if at least one company was mentioned in a press release. In our sample, 7.8 percent of press releases have firms that retained lobbyists with current or past EPA employees.

---

[7]U.S. Environmental Protection Agency Revolving Door data is available at Opensecrets.org.

# 3 Empirical Analysis

We seek to determine the effect of the content of a press release on its timing of release. To do so, we estimate probit models of the decision to release press release $i$ on a day that picks up less public attention or not. We estimate a model where the EPA chooses to release news on Fridays (where the dependent variable equals 1 if released on a Friday, 0 if not):

$$Pr(\text{Friday}_i) = Pr(\lambda_0 + \lambda_1 \text{Enforcement}_i + \lambda_2 \text{Awards}_i + \lambda_3 \text{Regulatory Change}_i$$
$$+ \mathbf{X}_i \lambda_4 + \mathbf{Z}_i \lambda_5 + \mathbf{D}_i \lambda_6 + \mathbf{W}_i \lambda_7 + \mathbf{L}_i \lambda_8 + \delta_i \geq 0)$$

Similarly, we estimate a model where the EPA may choose whether to release news on the weekday before a holiday:

$$Pr(\text{Before Holiday}_i) = Pr(\alpha_0 + \alpha_1 \text{Enforcement}_i + \alpha_2 \text{Awards}_i + \alpha_3 \text{Regulatory Change}_i$$
$$+ \mathbf{X}_i \alpha_4 + \mathbf{Z}_i \alpha_5 + \mathbf{D}_i \alpha_6 + \mathbf{W}_i \alpha_7 + \mathbf{L}_i \alpha_8 + \epsilon_i \geq 0)$$

Of interest are the variables indicating whether the press release is about enforcement actions, awards, or regulatory change. We also control for a vector, $\mathbf{X}$, which includes variables indicating if the content mentions health or money, whether an S&P 500 company is mentioned, and whether a company is mentioned that hired an EPA employee or retained a lobbying firm that did. Interaction terms are included in $\mathbf{Z}$, and to control for unobserved period-specific effects, administrator dummies (or in some specifications, year dummies) are in $\mathbf{D}$. Word count, word count squared, and word count cubed are included in $\mathbf{W}$. We also include covariates, $\mathbf{L}$, measuring whether the regulatory change occurred during a presidential lame duck period. This dummy might pick up end of year effects, and

therefore, we also include a dummy for the same time of year as lame duck periods (the Tuesday after the first Monday in November to January 20) in years that were not lame duck periods. And $\delta_i$ and $\epsilon_i$ account for unobserved idiosyncratic characteristics of the press release.

# 4    Results

To see if the content of press releases varied by day of week, we ran probit regressions with each weekday and the day before a holiday as dependent variables (Table 3). Friday is distinct from other days in several ways: the Friday probit and the before-holiday probit were the only specifications where enforcement press releases were more likely to be issued, and Friday has the only significant enforcement coefficient over all weekdays. Conversely, awards press releases have a positive coefficient on Monday through Thursday, and only on Fridays and before holidays are they significantly less likely to be issued. Regulatory changes are less likely to be announced early in the week and more likely to be announced later in the week and before holidays. Only on Fridays and before holidays are there significantly more regulatory change press releases issued.

We include word count, word count squared, and word count cubed, which are all highly significant, and suggest a non-linear relationship between length of press releases and day they are released. The length of a press release may affect the attention it generates independent of its timing or content. We speculate that shorter press releases contain less information and therefore garner less attention. Somewhat larger press releases contain more information, whereas the largest press releases may contain too much information, about disparate topics, to be easily digested by the media and the public.

Table 3: Effect of content on probability of press release by day of week

|  | (1) Monday | (2) Tuesday | (3) Wednesday | (4) Thursday | (5) Friday | (6) Before Holiday |
|---|---|---|---|---|---|---|
| Enforcement | -0.007 | -0.004 | -0.011 | -0.000 | 0.018*** | 0.003 |
|  | (0.006) | (0.006) | (0.006) | (0.007) | (0.006) | (0.002) |
| Award | 0.006 | 0.016** | 0.009 | 0.001 | -0.030*** | -0.005** |
|  | (0.007) | (0.007) | (0.007) | (0.008) | (0.007) | (0.003) |
| Reg. Change | -0.035*** | -0.034*** | -0.055*** | 0.011 | 0.105*** | 0.020*** |
|  | (0.008) | (0.009) | (0.009) | (0.010) | (0.011) | (0.005) |
| Word Count | 0.039** | 0.092*** | 0.042** | 0.030 | -0.389*** | -0.062*** |
|  | (0.015) | (0.019) | (0.017) | (0.023) | (0.029) | (0.016) |
| (Word Count)$^2$ | -0.008 | -0.026*** | -0.011* | -0.002 | 0.211*** | 0.051* |
|  | (0.006) | (0.010) | (0.007) | (0.014) | (0.023) | (0.024) |
| (Word Count)$^3$ | 0.000 | 0.002* | 0.001 | -0.000 | -0.027*** | -0.013 |
|  | (0.000) | (0.001) | (0.000) | (0.001) | (0.004) | (0.008) |
| n | 21,489 | 21,489 | 21,489 | 21,489 | 21,489 | 21,489 |
| Mean of Dep. Var. | 0.1489 | 0.1840 | 0.2085 | 0.2522 | 0.2010 | 0.0288 |
| Pseudo-R$^2$ | 0.0018 | 0.0030 | 0.0023 | 0.0003 | 0.0201 | 0.0128 |

*Notes:* Marginal effects from separate probit regressions. In turn, the dependent variable equals 1 if the news release occurred on Monday, Tuesday, Wednesday, Thursday, Friday, and the weekday before a holiday. *** Statistically significant at the 1 percent level; ** 5 percent level; * 10 percent level. Standard errors shown in parentheses.

Table 4: Effect of Content, Firms, and Politics on Probability of Friday or Before-Holiday Release

| | (1) Friday | (2) Friday | (3) Friday | (4) Friday | (5) Before Holiday |
|---|---|---|---|---|---|
| **Enforcement**: | | | | | |
| Enforcement | 0.008 (0.007) | 0.026*** (0.009) | 0.031** (0.012) | 0.021* (0.012) | 0.004 (0.004) |
| Health*Enforcement | | -0.035*** (0.011) | -0.035*** (0.011) | -0.029** (0.011) | -0.004 (0.003) |
| Dollar*Enforcement | | | 0.002 (0.014) | 0.005 (0.014) | -0.002 (0.004) |
| Firm$_{S\&P500}$*Enforcement | | | -0.009 (0.019) | -0.010 (0.019) | 0.001 (0.007) |
| Firm$_{Employed}$*Enforcement | | | -0.044 (0.074) | -0.026 (0.080) | -0.007 (0.017) |
| Firm$_{Lobby}$*Enforcement | | | -0.002 (0.025) | -0.001 (0.025) | -0.003 (0.007) |
| **Awards:** | | | | | |
| Award | -0.010 (0.008) | -0.021* (0.011) | -0.044*** (0.013) | -0.034** (0.014) | 0.000 (0.005) |
| Health*Award | | 0.021 (0.016) | 0.018 (0.016) | 0.013 (0.016) | -0.008* (0.004) |
| Dollar*Award | | | 0.048*** (0.018) | 0.049*** (0.018) | 0.001 (0.005) |
| Firm$_{S\&P500}$*Award | | | 0.014 (0.031) | 0.011 (0.030) | -0.002 (0.010) |
| Firm$_{Lobby}$*Award | | | 0.003 (0.039) | -0.002 (0.038) | -0.007 (0.010) |
| **Regulatory Changes:** | | | | | |
| Reg. Change | 0.098*** (0.011) | 0.132*** (0.016) | 0.110*** (0.017) | 0.111*** (0.017) | 0.045*** (0.010) |
| Health*Reg. Change | | -0.048*** (0.015) | -0.048*** (0.015) | -0.058*** (0.014) | -0.015*** (0.002) |
| Dollar*Reg. Change | | | 0.004 (0.024) | -0.004 (0.024) | 0.003 (0.008) |
| Lame Duck*Reg. Change | | | 0.007 (0.045) | 0.007 (0.045) | -0.009 (0.006) |
| Non-lame Duck*Reg. Change | | | 0.093*** (0.028) | 0.093*** (0.028) | -0.012*** (0.003) |
| **Other:** | | | | | |
| Health | 0.001 (0.006) | 0.014* (0.008) | 0.014* (0.008) | 0.015* (0.008) | 0.009*** (0.003) |
| Dollar | 0.001 (0.006) | 0.000 (0.006) | -.010 (0.009) | -0.013 (0.009) | 0.002 (0.003) |
| Firm$_{Lobby}$ | | | .024 (0.017) | 0.025 (0.017) | 0.005 (0.006) |
| Firm$_{Employed}$ | | | 0.029 (0.052) | 0.024 (0.052) | 0.013 (0.020) |
| Firm$_{S\&P500}$ | | | -0.014 (0.013) | -0.016 (0.013) | -0.004 (0.004) |
| Lame Duck President | | | 0.000 (0.017) | 0.058*** (0.020) | 0.083*** (0.016) |
| Non-lame Duck Tue. in Nov. to Jan. 20 | | | 0.001 (0.009) | -0.017* (0.009) | 0.058*** (0.007) |
| Word Count | -0.340*** (0.033) | -0.338*** (0.033) | -0.335*** (0.033) | -0.322*** (0.033) | -0.051*** (0.013) |
| (Word Count)$^2$ | 0.170*** (0.028) | 0.167*** (0.028) | 0.164*** (0.028) | 0.155*** (0.028) | 0.035* (0.017) |
| (Word Count)$^3$ | -0.020*** (0.005) | -0.020*** (0.005) | -0.019*** (0.005) | -0.018*** (.005) | -0.008 (0.005) |
| EPA Admin. Effects | Yes | Yes | Yes | No | No |
| Year Effects | No | No | No | Yes | Yes |
| n | 21,489 | 21,489 | 21,480 | 21,480 | 21,480 |
| Mean of Dep. Var. | 0.201 | 0.201 | 0.201 | 0.201 | 0.029 |
| Pseudo-R$^2$ | 0.047 | 0.048 | 0.050 | 0.068 | 0.084 |

*Notes:* Dependent variable in probit regressions equals 1 if the news release occurred on a Friday (columns (1)-(4)) and equals 1 if news release occured the weekday before a holiday (column (5)). Effects for EPA administrators (Jackson, Shapiro, Johnson, Leavitt, Horinko, Fisher, Whitman, McCabe, and Browner) are included in columns (1)-(3). Of the 21,493 news releases, 4,320 occurred on a Friday and 619 occurred before a holiday. Marginal effects are reported. *** Statistically significant at the 1 percent level; ** 5 percent level; * 10 percent level. Robust standard errors shown in parentheses.

In Table 4, we expand our analysis of Friday press releases. The coefficient for regulatory change is positive and significant at the one percent level after controlling for EPA administrator effects. The coefficients for enforcement and awards have the same signs as in Table 3 (without controls), but are insignificant, as are the variables for health and dollar.

When we include interactions between health and enforcement and awards and regulatory change, the coefficients on enforcement, awards, and regulatory change are significant. The probability that a press release is released on a Friday is 20 percent; this probability increases by two to three percentage points when the press release is announcing an enforcement action, increases when mentioning health, but decreases when an enforcement action is mentioned. Perhaps the EPA wishes to publicize its activities to mitigate health effects, or perhaps it couches its press releases on enforcement in the context of health effects when they are not issued on Friday.

Awards press releases are less likely to be issued on Fridays. This may be because giving awards is less likely to generate negative attention for the EPA, or because the EPA wants to draw attention to awardees, in hopes that other individuals will follow their example. Again, the subject of health seems to partly determine the timing of awards press releases. After controlling for awards releases that mention health, the coefficient on awards is significant at the 10 percent level. The coefficient for regulatory change, though already positive and significant at the one percent level in our most parsimonious specification, increases after including an interaction with health. Press releases on regulatory change are then 13 percentage points more likely to be issued on Fridays, a 68 percent increase in probability, after accounting for press releases that also mention health effects. Similar to enforcement press releases, regulatory change releases issued earlier in

the week are more likely to mention health effects. Regulatory changes also are more likely to come out the day before a holiday (whereas awards and enforcement are not statistically significant after controlling for covariates). About three percent of press releases are issued on the day before a holiday, but if the press release is about a regulatory change, this probability increases by 250 percent. The inclusion of year effects instead of EPA administrator effects increases the pseudo R-squared and also accounts for some of the variation in our variables of interest. Though the coefficients for awards and enforcement are smaller using year effects, and the significance is somewhat less, the same patterns of Friday releases persist.

Our dollar variable, which indicates that a particular press release contained at least one dollar sign ($), did not have a significant effect on the probability of Friday or before-holiday release. The only significant dollar coefficient comes when dollar is interacted with awards: press releases mentioning both dollar amounts and awards are more likely to come out on Fridays. It is possible that a measure of the actual amount of money mentioned would be informative (the range of dollar amounts is quite large, and includes such things as the price of a CFL light bulb and the total estimated expenditure of a particular congressional action). Unfortunately, we have yet to extract a usable measure of the values of money mentioned.

We find no evidence that the timing of EPA press releases varies for major public companies, companies that employed past or current EPA employees, or companies retaining lobbying firms with past or current EPA employees. S&P 500 companies have an interest in how markets perceive them, and information released early in the week may have greater influence on market perception. However, we find no evidence that press releases on enforcement actions mentioning

24

S&P 500 companies are more likely to come out on Fridays or before holidays. Companies that retained lobbying firms with past or current EPA employees, and companies that employ past or current EPA employees, have the desire and perhaps the ability to influence EPA actions, yet press releases on enforcement actions mentioning these companies are not more likely to appear on Fridays or before holidays. Firms do not seem to effectively influence the timing of EPA disclosures, though it is possible that they influence the content of press releases in unobserved ways.

We also test whether EPA press release patterns are different during presidential lame duck periods (Election Day to Inauguration Day in the years when a new president is inaugurated). The president appoints the administrator of the EPA, and therefore during this time, the outgoing administration would have less accountability to the public. If there is less fear of public reaction to press releases during lame duck periods, we would expect the Friday effect we observed on regulatory changes to be less pronounced. To assess whether the timing of regulatory press releases is different during presidential lame duck periods, we interact an indicator for lame duck periods with an indicator for regulatory change. There are typically more regulation changes during lame duck times (Cochran, 2001; Loring and Roth, 2005; Davies and de Rugy, 2008; Brito and de Rugy, 2009), so we include a dummy for whether the release was during a lame duck period. Because lame duck periods always fall during a particular time of year, we compare them to similar November-January periods when the president was not replaced. We find that regulatory change releases during lame duck times are not more likely to occur on a Friday or before a holiday, corresponding to the hypothesis that the EPA is minimizing public reaction to potentially controversial news.

# 5  Conclusion

Using machine learning techniques, we identified the content of EPA press releases along several dimensions. We analyzed which information the EPA was most likely to release on Fridays, a day of the week notorious for receiving the least attention. Non-controversial press releases, such as releases announcing awards, are timed to receive more media and investor attention than more controversial press releases such as enforcement actions (after controlling for health-related information) or press releases announcing regulatory changes.

These findings suggest that there may be scope for improving the communication of regulatory actions, increasing the reaction of capital markets to environmental violations, and increasing the visibility, and therefore the deterrence, of enforcement actions. Future work could match company names from press releases directly to their stock performance, and quantify the effect of timing for environmental violations. Another possible research avenue would be to match press releases to subsequent media coverage to determine which factors predict greater attention.

We also compared the accuracy of the dictionary approach and a naive Bayes N-gram language model in identifying the content of press releases. Text analysis techniques have broad applicability for economists interested in distilling data from the details of public policies or social interactions. As these techniques are just beginning to be used in economics, we attempted to explain clearly their advantages, limitations, and implementation.

# References

**Alias-i.** 2009. *http://alias-i.com/lingpipe.*

**Antweiler, Werner, and Murray Z. Frank.** 2004. "Is all that talk just noise? The information content of internet stock message boards." *Journal of Finance,* 59(3): 1259–1294.

**Bagnoli, Mark E., Michael B. Clement, and Susan G. Watts.** 2005. "Around-the-clock media coverage and the timing of earnings announcements." , (ACC-02-06).

**Begley, Joy, and Paul E. Fischer.** 1998. "Is there information in an earnings announcement delay?" *Review of Accounting Studies,* 3(4): 347–363.

**Bosch, J.C., E. Woodrow Eckard, and Insup Lee.** 1998. "EPA enforcement, firm response strategies, and stockholder wealth: An empirical examination." *Managerial and Decision Economics,* 19(3): 167–177.

**Brito, Jerry, and Veronique de Rugy.** 2009. "Midnight regulations and regulatory review." *Administrative Law Review,* 61: 163–195.

**Cavnar, William B., and John M. Trenkle.** 1994. "N-gram based text categorization."

**Cochran, Jay.** 2001. "The Cinderella constraint: Why regulations increase significantly during post-election quarters." *Unpublished report, George Mason University.*

**Damodaran, Aswath.** 1989. "The weekend effect in information releases: A study of earnings and dividend announcements." *Review of Financial Studies,* 2(4): 607.

**Dasgupta, Susmita, Benoit Laplante, and Nlandu Mamingi.** 2001. "Pollution and capital markets in developing countries." *Journal of Environmental Economics and Management,* 42(3): 310–335.

**Davies, Antony, and Veronique de Rugy.** 2008. "Midnight Regulations: An update." Mercatus Center Working Paper 08-06.

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

**Foulon, Jerome, Paul Lanoie, and Benoit Laplante.** 2002. "Incentives for pollution control: Regulation or information?" *Journal of Environmental Economics and Management,* 44(1): 169–187.

**Gentzkow, Matthew, and Jesse M. Shapiro.** 2010. "What drives media slant? Evidence from US daily newspapers." *Econometrica*, 78(1): 35–71.

**Gersen, Jacob E., and Anne Joseph O'Connell.** 2009. "Hiding in plain sight? Timing and transparency in the administrative state." *University of Chicago Law Review*, 76: 1157–1407.

**Graham, John R., Campbell R. Harvey, and Shiva Rajgopal.** 2005. "The economic implications of corporate financial reporting." *Journal of Accounting and Economics*, 40(1-3): 3–73.

**Hamilton, James T.** 1995. "Pollution as news: Media and stock market reactions to the Toxics Release Inventory data." *Journal of Environmental Economics and Management*, 28(1): 98–113.

**Hanley, Kathleen Weiss, and Gerard Hoberg.** 2010. "The Information Content of IPO Prospectuses." *Review of Financial Studies*, 23(7): 2821–2864.

**Harper, Richard K., and Stephen C. Adams.** 1996. "CERCLA and deep pockets: Market response to the Superfund program." *Contemporary Economic Policy*, 14(1): 107–115.

**Jones, Kari, and Paul H. Rubin.** 2001. "Effects of harmful environmental events on reputations of firms." *Advances in Financial Economics*, 6: 161–182.

**Karpoff, Jonathan M, John R. Jr. Lott, and Eric W Wehrly.** 2005. "The reputational penalties for environmental violations: Empirical evidence." *Journal of Law & Economics*, 48(2): 653–675.

**Khanna, Madhu, Wilma Rose H. Quimio, and Dora Bojilova.** 1998. "Toxics Release Information: A policy tool for environmental protection." *Journal of Environmental Economics and Management*, 36(3): 243–266.

**Klassen, Robert D., and Curtis P. McLaughlin.** 1996. "The Impact of Environmental Management on Firm Performance." *Management Science*, 42(8): 1199–1214.

**Konar, Shameek, and Mark A. Cohen.** 1997. "Information as regulation: The effect of community right to know laws on toxic emissions." *Journal of Environmental Economics and Management*, 32(1): 109–124.

**Konar, Shameek, and Mark A. Cohen.** 2001. "Does the market value environmental performance?" *Review of Economics and Statistics*, 83(2): 281–289.

**Laver, Michael, and John Garry.** 2000. "Estimating policy positions from political texts." *American Journal of Political Science*, 44(3): 619–634.

**Loring, Jason M., and Liam R. Roth.** 2005. "Empirical study: After midnight: The durability of the "midnight" regulations passed by the two previous outgoing administrations." *Wake Forest Law Review*, 40: 1441.

**Loughran, T., and B. McDonald.** 2011. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance*, 66(1): 35–65.

**Manning, Christoper D., and Hinrich Schutze.** 1999. *Foundations of statistical natural language processing.* Vol. 59, MIT Press.

**Muoghalu, Michael I., H. David Robison, and John L. Glascock.** 1990. "Hazardous waste lawsuits, stockholder returns, and deterrence." *Southern Economic Journal*, 57(2): 357–370.

**Patell, James M., and Mark A. Wolfson.** 1982. "Good news, bad news, and the intraday timing of corporate disclosures." *Accounting Review*, 57(3): 509–527.

**Peng, Fuchun, and Dale Schuurmans.** 2003. "Combining naive Bayes and n-Gram language models for text classification." In *Advances in Information Retrieval*. Vol. 2633 of *Lecture Notes in Computer Science*, , ed. Fabrizio Sebastiani, 547. Springer Berlin / Heidelberg.

**Tetlock, Paul C.** 2007. "Giving content to investor sentiment: The role of media in the stock market." *Journal of Finance*, 62(3): 1139–1168.

# A  Appendix

Table 5: Comparison of Classification by Machine Learning and Dictionary Approaches

| Date | Headline | Fines | | | Awards | | | Reg. Ch. | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | H | M | D | H | M | D | H | M | D |
| 9-May-00 | Asthma and Allergen Control Conference and Summit in Dallas | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17-Mar-05 | Cairo High School Science Club & Biology Students' Environmental Outreach Program Receives Presidential Environmental Youth Award | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 14-Jun-02 | 1. Stop Sale Order Issued to Missouri Disinfectant Company, 2. Florida Man, Virginia Real Estate Firm Plead Guilty, 3. Ship's Chief Engineer Pleads Guilty to Oil Pollution in Alaska Waters, 4. Two Men Sentenced in Las Vegas for Illegal Waste Discharge | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1-Oct-08 | EPA Picks Cleanup Plan for Two Creeks Near Nease Superfund Site | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27-Jul-01 | EPA Calls for Reports of Ground Water Contamination by Hydraulic Fracturing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 8-Mar-00 | Virginia Gets $2.3 Million EPA Grant for Safe Drinking Water | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 29-Jun-04 | U.S. EPA Proposes Nonattainment Areas under New Federal Air Quality Standard | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| 12-Aug-98 | EPA Advises Car Owners to Keep Their Air Conditioners in Good Working Order This Summer | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19-Jan-00 | Restroom Supply Company to Pay $13,750 for Not Registering Product | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18-Jun-01 | Cape Charles/Northampton County Chosen as EPA National Brownfields Showcase Community | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5-May-03 | EPA Orders 28 Facilities to Apply for Biosolids Permit | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21-Jul-04 | Iani Announces Resignation as EPA Region 10 Administrator | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 13-Apr-07 | EPA Awards $200,000 to the City of Camden to Establish a Local Brownfields Job Training Program | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 29-Nov-01 | Proposal to Ease Transition from Winter to Summer Grade Cleaner-Burning Gasoline | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 19-Mar-99 | OECA Solicits Public Comments on Enforcement | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 5-May-00 | Ten Indicted for Underground Tank Testing Fraud | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10-Dec-98 | EPA Says "Put Mother Earth on Your Holiday Gift List" | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 29-Sep-05 | EPA Preserves Program to Minimize Emissions of Nitrogen Oxides | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 9-Apr-09 | EPA Withdraws Water Permit Fee Incentive Rule | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 26-Apr-01 | Florida Woman Pleads Guilty to Illegal Storage of Chemicals | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25-Sep-06 | EPA Removes 19th Avenue Landfill from Superfund List | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1-Dec-98 | EPA Orders Companies to Begin Cleanup at Motorola Site in Phoenix | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24-Jun-97 | Updated Press Office Call List | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1-Mar-07 | EPA Actions Will Assure Air Permitting Programs Run Consistently and Smoothly | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| 27-Aug-08 | Hanover Water System Operator Receives EPA Regional Award | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

Table 5 – Continued

| Date | Headline | Fines | | | Awards | | | Reg. Ch. | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | H | M | D | H | M | D | H | M | D |
| 19-Oct-00 | Huston Township Sewer Joint Authority Receives $186,000 Grant from the EPA | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 22-Sep-08 | New England Experienced Fewer Smog Days during Recent Summer | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 23-Jun-05 | The Republic of Korea Joins Methane to Markets Partnership | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17-Oct-03 | U.S. EPA Funds Almond Pesticide Study in Three California Counties | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| 8-Jan-09 | January is National Radon Action Month | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1-Nov-05 | Inaccurate Water Pollution Reporting Results in $20,000 Fine against International Seafoods | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11-Mar-97 | U.S. EPA Creative Settlement Beefs Up Tulare Fire Department Resources | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22-Feb-01 | Houston Employer Fails to Protect Workers from Asbestos | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 24-Jan-03 | Court Finds Defendants Liable for Cleanup of Metal Bank Superfund Site in Philadelphia | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9-Jun-00 | Draft Guidance Issued to State and Local Governments to Improve Air Quality | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 13-Sep-05 | EPA Administrator to Brief Reporters on EPA Efforts in Gulf Coast | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10-Apr-08 | EPA Awards City of Asheville and Land-of-Sky Regional Council with with Grants to Revitalize Old Properties | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 2-May-08 | EPA Awards City of Spartanburg with Grants to Revitalize Old Properties | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 18-Feb-09 | United States Files Clean Air Lawsuit against Louisiana Generating | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 1-Apr-05 | Second Man Sentenced in Michigan Waste Treatment Facility Case | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27-Apr-06 | Crittenden County Children Get Cleaner Buses | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 18-May-06 | EPA Cites Reilly Industries for Clean-air Violations | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18-Nov-08 | Sierra Club Recognizes EPA Regional Administrator Alan J. Steinberg for Outstanding Environmental Leadership | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| 7-Jan-00 | EPA and University of New Hampshire Settle Claims of Hazardous Waste Violations | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 26-Jan-09 | EPA and General Electric Update Hudson Dredging Agreement | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 22-Dec-05 | EPA Selects Two New York Communities for Brownfields Grants | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 4-Dec-08 | EPA Enforcement in New Mexico Continues to Cut Pollution | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 20-May-09 | Take Control of Your Asthma, EPA Advises | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28-Sep-09 | U.S. EPA Administrator Lisa P. Jackson to Visit Bay Area | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 28-Jan-99 | Boaters No Longer Allowed to Discharge Sewage in Bay Waters near East Hampton | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |

*Notes:* H: Hand Classification; M: Machine Learning Classification; D: Dictionary Classification. Only the headline is listed here; however, classification was based on all text in the body of the press release.

Table 6: Dictionaries

| Penalty | Penalty Continued | Award | Reg. Change | Health |
|---|---|---|---|---|
| agreement | refuses | accolade | adjust | abdominal |
| alleged | refusing | achievement | adjusted | affect children |
| allegedly | reparation | achievements | adjustment | anemia |
| arbitration | resolution | award | adjustments | asthma |
| claims | ruled | awards | adjusts | blood |
| complain | rules | celebrate | amend | brain |
| complaint | ruling | celebrated | amended | breathing |
| compliance | seize | celebrates | amends | cancer-causing |
| comply | seized | celebrating | federal register | carcinogen |
| confiscate | seizes | celebration | final decision | cardiac |
| confiscating | seizing | commend | final rule | children |
| enforce | sentence | commended | finalize | chronic bronchitis |
| enforced | sentenced | commends | finalized | confusion |
| enforces | sentences | earn | finalizes | developmental impairment |
| enforcing | sentencing | earned | go into effect | disease |
| face | settle | earns | goes into effect | diseases |
| faces | settled | fellowship | into law | dizziness |
| fail | settles | grant | modification | elderly |
| failed | sued | granted | modifications | eyes |
| failing | sues | grants | modified | fatal |
| fails | suing | honor | modifies | fever |
| files | to pay | honors | modify | headaches |
| fine | trial | laud | new limit | health |
| fines | violate | praise | new regulation | heart |
| guilty | violated | praises | new regulations | high blood pressure |
| illegal | violation | prestigious | new requirement | human health |
| illegally | violations | recognition | new requirements | illness |
| indict | warn | recognize | new restriction | illnesses |
| infraction | warned | recognized | new rule | immune |
| infractions | warns | recognizes | new rules | ingestion |
| judgement | | scholarship | new standard | inhalation |
| judgment | | win | new standards | internal organs |
| lawsuit | | wins | reform | joint pain |
| lawyer | | | reformed | kidney |
| limitation | | | reforms | lethal |
| obstructed | | | regulatory changes | liver |
| ordered | | | revise | lung |
| orders | | | revised | memory |
| penalize | | | revises | memory loss |
| penalized | | | revision | muscle |
| penalizes | | | revisions | nerve disorders |
| penalties | | | revoke | nervous system |
| penalty | | | revoked | nose |
| prosecute | | | revokes | pregnancy |
| prosecuted | | | revoking | respiratory problems |
| prosecutes | | | take effect | scarring |
| prosecution | | | takes effect | severe burns |
| punished | | | | sleepiness |
| punishes | | | | throat |
| punishment | | | | unconsciousness |
| refused | | | | |