Financial Services Review 18 (2009) 87–104

FINANCIAL SERVICES REVIEW

The CAN-SPAM Act of 2003 and stock spam emails

Bill Hu[b], Thomas McInish[a,*], Li Zeng[b]

[a]*Fogelman College of Business and Economics, The University of Memphis, Memphis, TN 38152, USA*
[b]*College of Communications, Arkansas State University, Jonesboro, AR 72401, USA*

Abstract

One goal of the Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003 is to combat pump and dump stock spam email schemes aimed at individual investors. The Act specifies requirements for those who send commercial emails. We find that only 60% of the 40,000 stock spam emails analyzed follow these disclosure requirements and emails that disclose conflicts of interest have a lower market impact. After the peak spam email day, stock prices decline, indicating that individual investors lose money. A Securities and Exchange Commission crackdown in 2007 reduced the impact of stock spam emails.

JEL classification: C33; G11; G18; K22

Keywords: Stock spam; Market manipulation; Text classification

1. Introduction

Promoters of pump and dump schemes target individual investors by hyping microcap stocks through false or misleading statements (Kyle and Viswanathan, 2008; Allen and Gale, 1992). Once the price increases these promoters sell (dump) their shares, the price falls, and investors lose money.[1] In part to address concerns about these schemes, the United States enacted the Controlling the Assault of Non-Solicited Pornography and Marketing Act (CAN-SPAM) of 2003, which specifies requirements for those who send commercial emails. The Act bans false or misleading header information, prohibits deceptive subject lines, requires an opt-out option in the email, and requires that commercial email be identified as an advertisement and includes the sender's valid physical postal address. The Federal Trade

* Corresponding author. Tel.: +1-901-277-9202; fax: +1-901-678-3006.
E-mail address: tmcinish@memphis.edu (T. McInish)

1057-0810/09/$ – see front matter

Commission is authorized to enforce the Act.[2] Significant penalties and criminal sanctions can result from violations of the Act.

We investigate whether spammers comply with the disclosure requirements of the CAN-SPAM Act. Using computational linguistics methods, we classify over 40,000 stock spam emails to ascertain whether spammers meet U.S. Securities Exchange Commission (SEC) disclosure requirements. We find that 60% of the emails make the required conflict of interest disclosure concerning whether or not the email campaign results from a payment. We find that emails that do not include the disclosure have a greater market impact than those that do. We know of no previous research that analyzes the content of stock spam emails.

On March 8, 2007, the SEC launched Operation Spamalot. Citing questions about the adequacy and accuracy of information available about 35 penny stocks, the SEC suspended trading in these stocks for a minimum of 10 days and promised to hold the people behind spam emails accountable.[3] We investigate whether the SECs Operation Spamalot affected market behavior following the enforcement action. We find a significant decline in abnormal returns and trading volume for touted firms after March 2007. Hence, we provide additional support for Bhattacharya and Daouk (2002) who also show, in the context of insider trading, that law enforcement can be effective.

2. Literature review

Study of spam emails is new. Because of the availability of stock trading price data, previous researchers have largely focused on the evaluation of market reaction to a spate of stock spam emails. Several studies document significant positive one-day returns and negative longer-term returns for stocks being touted. Frieder and Zittrain (2006) show the effectiveness of stock spam emails as a manipulation tool. Through an analysis of 3,669 firm-date groups containing over 500 distinct stocks listed in the Pink Sheets between January 2004 and July 2005, these authors find that a spammer who buys on the day before touting begins and sells on the day of heaviest touting earns about 4.29% before transaction costs. The stocks in their sample usually are touted over several days. There is a significant increase in price and volume during touting, but returns in the days after touting are significantly negative. These authors indicate that although SEC disclosure requirements are weak, enforcement can deter insider trading.

Similarly, Hanke and Hauser (2008) show that stock spam emails have a significant impact on returns, intraday volatility, and trading volume, but no lasting positive effect on stock prices. Hanke and Hauser (2008) identify liquidity as a major factor in the success of spam emails and report that spamming on successive days sustains excess demand for target stocks and enlarges the time window for liquidation of spammers' positions. Hanke and Hauser (2008) formulate an economic story to explain the observed effect of spam emails. They reason that spammers build up a position in the target stock before the spam campaign, and, subsequently, inexperienced investors become victims of the pump and dump scheme.

Bohme and Holz (2006) examine 111 stocks between November 2004 and February 2006. They find that stock spam email has a significant effect on return and volume. They also find a positive dependence between turnover and the number of spam events. D'Alessio (2007) studies the effects of spam emails and finds that a model in which spammers send out a single spam message to which spammees respond within a few hours is marginally successful in explaining observed positive returns on the spam day.

There are many theoretical models related to rumors and stock market manipulation. Van Bommel (2003) demonstrates that in equilibrium rumors are informative and rumor-mongers can profit at the expense of uninformed liquidity traders. Allen and Gale (1992), Aggarwal and Wu (2002), John and Narayanan (1997), and others have modeled stock market manipulations.

In summary, prior research shows that spammers earn positive returns on the spam day and noise traders lose if they hold the spammed stock for more than one day. Most of the research focuses on the market reaction to spam events. The content of spam emails and whether these emails follow SEC regulations have not been explored.

3. Background and hypotheses development

The SEC has brought many cases against fraudsters and suspended numerous stocks from trading.[4] In April 2007, a Florida man was sentenced to five years in prison for committing securities fraud and email fraud related to stock manipulation schemes. In June 2007, Adam Vitale pleaded guilty for violating the CAN-SPAM Act and was sentenced to 30 months in prison and ordered to pay $180,000 in restitution.[5] Although the CAN-SPAM Act forces domestic email marketers to either give up the practice or risk jail, some suggest that CAN-SPAM has proven to be largely ineffective.

Spammers are becoming increasingly sophisticated. First, it is hard to analyze the reputation of the sender because spammers can use infected computers to send out spam messages without the computer owner's knowledge. Second, it is hard to analyze the content in spam messages by automatic computer programs. Spammers can misspell words and blend in common phrases. In addition, spammers use images or even dynamically change a few pixels of images during transmission.

Stock spam emails cannot be banned because the U.S. Constitution's First Amendment protects freedom of speech. Many supposedly independent and unbiased investment newsletters, research reports, radio, or television shows are often paid by either touted companies or some third parties who hold significant amount of shares in the touted stocks. Paid promoters can reach tens of thousands of people easily by sending mass emails, building websites or blogs, posting messages on online bulletin boards or chat rooms. It is very difficult to determine whether the messages are true and credible.

3.1. Trade suspensions

The SEC may suspend trading when serious questions arise about the credibility of a company's financial information or when public information about a company is not current, accurate, or adequate. The SEC maintains a list of trade suspensions since 1995.[6] Some of the suspensions are directly related to dissemination of public information such as spam emails. For example, on March 8, 2007 the SEC suspended 35 stocks from trading, each for a period of at least 10 trading days. Although the SEC does not give the exact reason for a trading suspension, analysis of the announcements indicates that these suspensions were because of involvement in spamming.

3.2. Hypotheses

Bhattacharya and Daouk (2002) show that law enforcement can effectively reduce the extent of insider trading. One reason for enforcement effectiveness may be because of fraudsters' fear of sanctions. In addition, the publicity associated with enforcement may alert investors. Therefore, after the SEC Operation Spamalot crackdown on spam emails, we expect a decline in the number of stock spam emails and a reduction in their effectiveness. This leads to our first hypothesis:

Hypothesis 1: Following SEC enforcement, the (1) number and (2) effectiveness of spam emails decreases.

Spam emails complying with SEC disclosure requirements contain conflict of interest information. Most of time, they disclose that they have received compensation from a third party who is going to sell a significant amount of shares that may have a negative impact on stock prices. Though the disclosure is in small print and at the end of the message, it nonetheless conveys some negative information. Therefore, we expect these kinds of spam emails are less effective in promoting the targeted stock than emails that do not make the required disclosure. This leads to our second hypothesis.

Hypothesis 2: Spam emails complying with SEC disclosure regulations are less effective.

4. Data and methodology

4.1. Data

We obtain 41,135 emails touting 785 firms from November 2004 to August 2007 from www.crummy.com, a Website constructed by Leonard Richardson.[7] Hanke and Hauser (2008) also use this dataset to investigate the effects of stock spam emails. Bohme and Holz (2006) compare this spam dataset to their own collection of stock spam emails and conclude that this dataset is quite reliable. Richardson informs us that there was no spam collection method change during our sample period.[8] Hence, we believe that our stock spam email dataset is representative over time. However, of course, we cannot guarantee that changes we find over time do not result from changes in the data set rather than from the CAN-SPAM Act, which is a limitation of our study.

We are able to obtain daily closing prices, high and low intraday prices, and volume data from Datastream for 395 of these firms. Following Frieder and Zittrain (2006), a Spam Campaign (SC) is defined as a period of spamming activity with no more than five consecutive days without a spam email. There are 675 stock SCs in our sample.

Spam emails sent by intermediates such as email marketers include a conflict of interest disclosure, but those sent by others do not. We use the well-established Naive Bayes algorithm from computational linguistics to classify the spam emails. The Appendix provides a detailed explanation. Additional explanation can be found in Antweiler and Frank (2004). We employ the Rainbow package developed by McCallum (1996) to accomplish the Naive Bayes text classification.[9] We classify all spam emails into those that include (Type D emails) and those that do not include (Type ND emails) a conflict of interest disclosure. Then, we classify the 395 firms in our sample as follows: Type D Firms are touted only by Type D emails, Type ND Firms are touted only by Type ND emails, and Type DND Firms are touted by both types of emails.

4.2. Methodology

We first identify SCs and study campaign level spam effects. PeakDay is defined as the day within the campaign with the maximum number of spam emails, taking the first such day if there are ties.

We apply event study methodology to determine abnormal returns, volume, volatility, and their cumulative effects surrounding the SC. Abnormal return is defined as the difference between the stock return and the Russell 2000 return, which is a proxy for market return. Brown and Warner (1985) show that a simple market model works well for most event studies. Our dependent variables are turnover, return, and volatility, in turn. Specifically, we define the following variables:

Turnover—$Turnover = \ln(1+Dolvol_t/\overline{DolVOl})$;
Return—$R_t = \ln S_{t-1}$, where S_t is the adjusted closing price of the stock on day t; and
Volatility—$RISK_t$ = intraday price range/average intraday price range where the intraday price range = ln (intraday high price – intraday low price).

We also calculate two measures of abnormal return, *AR* and *AR1*, and one measure of abnormal volume (*AVOL*):

$AR = R_t - R_{m,t}$, where $R_{m,t}$ is the Russell 2000 return on day t;
$AR1 = R_t - \bar{R}$, where $\bar{R}$ is the average stock return during the sample period; and
$AVOL_t = (DolVol_t - \overline{DolVol})/\overline{DolVol}$, where $DolVol_t$ is dollar volume at day t and $\overline{DolVol}$ is the average dollar volume during the sample period.

The return in the panel regression below is the time demeaned return. In the event study, we use *AR* or *AR1*. Buy and hold returns are the sum of log price relatives over the indicated periods.

We follow the Hanke and Hauser (2008) panel regression approach with modifications. First, we exclude the lagged dependent variables. According to Wooldridge (2002), with lagged dependent variables, the underlying strict erogeneity assumption is violated for both fixed effect and random effect models. Second, we investigate the PeakDay, the day with the most spam emails during a spam campaign.

Because we find significant unobserved fixed cross-sectional effects, we analyze the market impact of spam emails on different types of firms and the effect of the SEC crackdown using panel regressions. Analyses suggest that there are no significant unobserved time effects. However, average returns might vary systematically across stocks and there might be unobserved heterogeneity cross-sectional effects. Therefore, we use the standard fixed effects transformation (Wooldridge, 2002) methodology.

Our regression equation is as follows:

$$\begin{aligned} Y_{i,t} = \beta_0 &+ \beta_1\, Tue_{i,t} + \beta_2\, Wed_{i,t} + \beta_3\, Thu_{i,t} + \beta_4\, Fri_{i,t} \\ &+ \beta_5\, PeakDay_{i,t} + \beta_6\, Pre\text{-}SC_{i,t} + \beta_7\, Post\text{-}SC_{i,t} \\ &+ \beta_8\, Pre\text{-}SC{*}Firmtype_{i,t} + \beta_9\, PeakDay{*}Firmtype_{i,t} \\ &+ \beta_{10}\, Post\text{-}SC{*}Firmtype_{i,t} + \varepsilon_{i,t} \end{aligned}$$

where $Y_{i,t}$ is the time-demeaned data for turnover, return, and risk, in turn, β_0 is an intercept term and ε is a random error term. We adjust standard errors for intergroup correlation.

The following seven dummy variables appear in all of our regressions. *Tue, Wed, Thu, and Fri* each equal 1 if the observation is for the respective day of the week and 0 otherwise. *PeakDay* equals 1 if the observation is for the PeakDay during an SC and 0 otherwise. *Pre-SC* (*Post-SC*) equals 1 if the observation is for one of the five days before (after) the SC and 0 otherwise. Hanke and Hauser (2008) do not examine *PeakDay*.

Next, we define two sets of three dummy variables that appear in alternate regressions. The first set is: *Pre-SC*Firmtype*, *PeakDay*Firmtype*, and *Post-SC*Firmtype*, which are the product of *Pre-SC*, *PeakDay*, and *Post-SC*, in turn, multiplied by *Firmtype*. *Firmtype* takes the value 1 for Type D Firms, and 0 otherwise. In some regressions, instead of the three dummy variables just described, we substitute: *Pre-SC*SEC*, *PeakDay*SEC*, and *Post-SC*SEC*, which are the product of *Pre-SC*, *PeakDay*, and *Post-SC*, in turn, multiplied by *SEC*. *SEC* equals 1 for observations occurring after the Operation Spamalot (OS) and 0 otherwise.

5. Results

5.1. The spamming process

From our reading and categorization of spam emails, we determine that about 60% of stock spam emails go through the email marketers' route.[10] Among those messages that disclose the compensation details, some of the email messages explicitly mention that they are conducting a stock awareness campaign and directors or employees of the listing company may start to sell off their positions or that the service is paid for by a nonaffiliated third party. Some of the messages reveal the email marketer's name and physical address. We categorize the 40% of emails that do not disclose conflict of interest information as spammer-spammee direct relationship.

5.2. Descriptive statistics

From Fig. 1a, we can see that the spammers are active during weekdays. There are fewer spamming activities during weekends. Therefore, the spammers might anticipate immediate

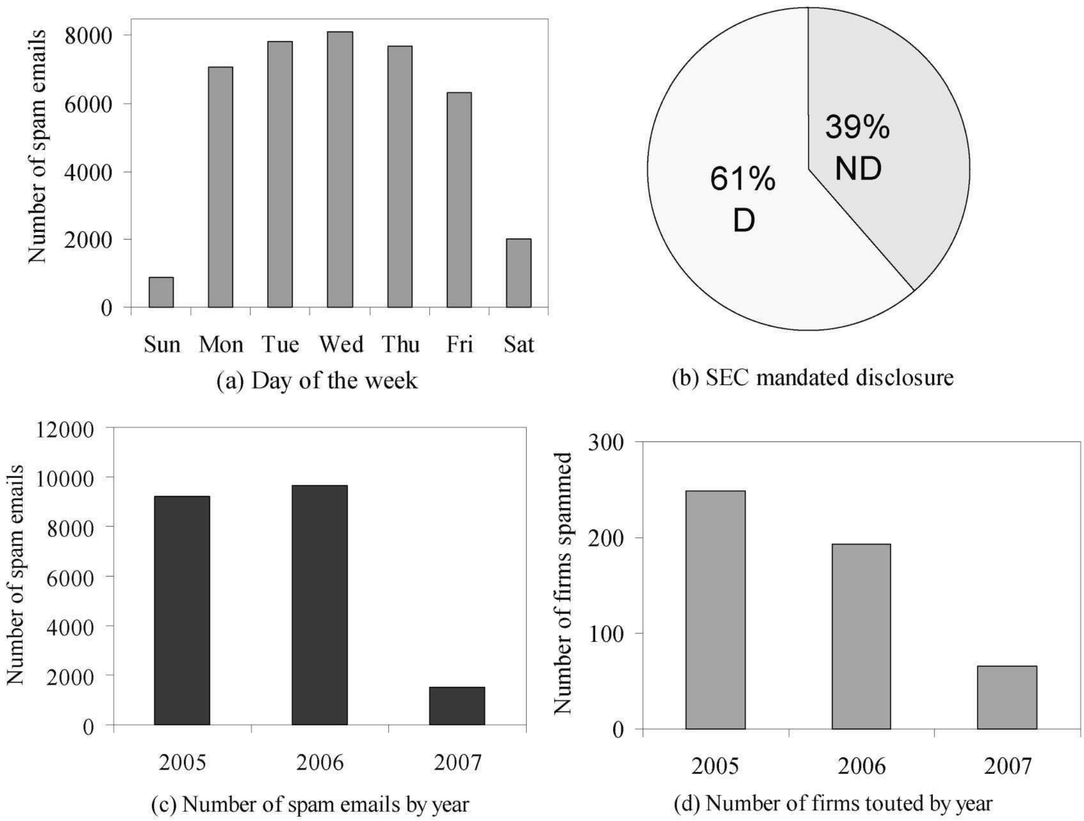

Fig. 1. Distribution of spam events. We present the number of spam events by (a) day of the week, (b) the percentage of spam for D firms and ND firms, (c) the number of spam emails by year, and (d) the number of firms touted. Some emails include a conflict of interest disclosure (Type D) and others do not (Type ND).

results from spam email campaigns. In addition, spammers are more active in the morning and the spamming activity declines toward the end of the trading day. Overall, about 60% of the spam emails meet SEC disclosure requirements as shown in Fig. 1b. Because the SEC can suspend trading of the listing company if illegal spamming activity is suspected, the spammers are less at risk from losses caused by trading suspensions if they choose to meet SEC disclosure requirements. On the other hand, the information disclosed may reduce the effectiveness of the spam emails.

5.2.1. Effect of sec crackdown

We first examine whether the number of stock spam emails decreases after the SEC crackdown in 2007. As seen from Fig. 1c, we clearly observe a downward trend of spam emails from 2006 to 2007. The number of spam emails received during January to August declines dramatically, more than 80%, from 9,186 in 2005 and 9,634 in 2006 to 1,533 in 2007. Moreover, the number of firms touted by spam emails declines about 70% from 248 firms in 2005 to 66 firms in 2007 as shown in Fig. 1d.[11] The average daily number of stock spam emails is 51.1 before the SEC crackdown and declines to 7.6 after the SEC crackdown.

The average monthly number of firms touted is 27.8 before the SEC crackdown and declines to 7.2 after the SEC crackdown. A mean difference test shows that the reductions in both the daily number of spam emails (t = 10.8) and the monthly number of firms touted (t = 4.8) are statistically significant at the 1% level. These results support Hypothesis 1a. Therefore, we conclude that the SEC crackdown on spam emails effectively reduced both the number of spam email campaigns and the number of firms touted by spam emails.

5.2.2. Firm characteristics by disclosure types

Table 1 reports descriptive statistics for our sample. On an average day during January 2004 to December 2007, the mean price is 51.7 cents, which is slightly lower than the 67 cents reported by Frieder and Zittrain (2006). The mean dollar volume is $78,039 and the mean intraday price range of 16 cents is quite high compared to the average price. The three types of firms have similar prices, dollar volume, and intraday price range. Overall, the SCs mainly target penny stocks.

Table 1, Panel B, reports a significant increase in price, dollar volume, and intraday price range on spam days. Dollar volume of $217,006 is almost triple the average daily dollar volume of $78,039. The dollar volume on spam days for Type D Firms is the least among the three firm types. Results of t-tests show that the dollar volume on spam days for Type D Firms is significantly lower than the dollar volume on spam days for Type ND Firms (t = 2.97) or Type DND Firms (t = 5.45). These results support Hypothesis 2. The price on a spam day increases modestly from 52.4 cents to 63.2 cents for Type D Firms. Type DND Firms, in contrast, have the biggest increases. The price almost doubled to 94.1 cents and dollar volume tripled to $294,842. This is not surprising because Type DND Firms are spammed by at least two types of spammers. Therefore, the pump and dump interest is the highest for Type DND Firms. Nevertheless, the intraday price range only increases slightly from 18.2 cents on an average day to 20.7 cents for Type DND Firms on a spam day.

5.3. Campaign level summary statistics

Table 2 reports campaign level summary statistics. Overall, there are 675 stock SCs in our dataset. On the PeakDay of a stock SC, we observe a significant positive abnormal return, 1.2% for market adjusted abnormal returns (significantly different from 0 at the 5% level) and 1.7% for mean adjusted abnormal returns (significantly different from 0 at the 1% level). In addition, there is a significant increase in abnormal dollar volume. Both turnover and intraday volatility increase significantly on the PeakDay of a stock SC. Our results are similar to the ones reported by Bohme and Holz (2006), Hanke and Hauser (2008), and Frieder and Zittrain (2006). In summary, SCs have a significant impact on the touted stock with respect to stock return, volume, and volatility.

The SCs are less effective after the SECs Operation Spamalot. Table 2 shows that before the SEC crack down there was a 1.3% (statistically significant at the 5% level) increase in market adjusted return. After the SEC crackdown, the SCs no longer increase the PeakDay stock returns, or even slightly decrease PeakDay stock returns by about 1.2% (statistically insignificant). In addition, the SC does not increase trading volume significantly after the SEC crackdown, which is in sharp contrast to the 260% increase in trading volume before

Table 1 Descriptive statistics

Variable	Firm type											
	All			ND			D			DND		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
Panel A: Daily mean for entire sample (*N* = number of firms)												
Price	395	0.517	0.760	102	0.524	0.642	203	0.524	0.877	90	0.494	0.584
Dollar volume	394	78,039	372,089	101	69,014	221,803	203	76,368	395,035	90	91,937	449,001
Price range (high-low)	395	0.159	0.102	102	0.129	0.103	203	0.164	0.096	90	0.182	0.108
Panel B: Spam days only (*N* = number of spam days)												
Price	1,852	0.797	1.108	265	0.717	0.843	669	0.632	0.835	918	0.941	1.314
Dollar volume	1,665	217,006	916,584	216	142,848	294,125	609	135,949	702,855	840	294,842	1,128,637
Price range (high-low)	1,665	0.227	0.439	216	0.270	0.857	609	0.241	0.442	840	0.207	0.227
Percentage of spam emails		100.0			29.8			59.3			10.9	

Some emails include a conflict of interest disclosure (Type D) and others do not (Type ND).

We classify firms into three groups according to whether they are touted by Type D emails only (D), Type ND emails only (ND), or both types of emails (DND). Panel A presents the daily mean price, dollar volume, and price range for all the days for January 2004 through December 2007. Panel B presents the same statistics only for days with spam email. In Panel A, the number of firms is N and in Panel B the number of days with spam emails is N. Dollar volume is the product of price and volume. Price range is the difference between the intraday high and low price. Price and volume data are from DataStream. Spam emails data are from www.crummy.com.

Table 2 Spam Campaign (SC) level summary statistics

Variable	All			Before			After		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
AR	675	0.012**	0.152	640	0.013**	0.154	35	−0.012	0.102
AR1	675	0.017***	0.152	640	0.018***	0.154	35	−0.008	0.100
AVOL	596	2.560***	6.845	571	2.616***	6.932	25	1.275	4.276
Turnover	596	1.032***	0.866	571	1.046***	0.864	25	0.716***	0.862
RISK	596	1.342***	2.070	571	1.364***	2.104	25	0.856***	0.912

An SC is defined as a period of spamming activity with no more than five consecutive days without a spam email. PeakDay is defined as the day within the SC with the maximum number of spam emails, taking the first such day if there are ties. N is the number of SCs. Abnormal return (*AR*) is the difference between PeakDay stock return and the market return, proxied by the Russell 2000 return. *AR1* is the difference between the PeakDay stock return and average stock return during the sample period. Abnormal volume (*AVOL*) is the difference between PeakDay stock dollar volume and average stock dollar volume during the sample period standardized by the average stock dollar volume. *Turnover* = log (1 + dollar volume/average dollar volume). Intraday price range = log(intraday high price–intraday low price). *RISK* = intraday price range/average intraday price range. On March 8, 2007, the SEC launched Operation Spamalot to crack down on stock spam emails. Before (after) refers to the period before (after) Operation Spamalot. All data are for January 2004 through December 2007.

** Significance at the 0.05 level.

*** Significant at the 0.01 level.

the SEC crackdown. Moreover, intraday volatility (*RISK*) on the PeakDay of the SC also declines to about 86% of average day volatility after the SEC crackdown. These results provide support for Hypothesis 1b. In conclusion, the SEC crackdown significantly reduces both the number of stock spam emails and the market impact of SCs.

5.4. Cumulative abnormal returns around spam campaigns

We apply event study methodology to analyze the market impact of stock SCs for the three types of spam emails. Specifically, we investigate the cumulative abnormal returns (CARs) before, during, and after the SCs and on the PeakDay

Table 3 reports CARs around SCs for three types of spam emails. Table 3, Panel A, presents unadjusted buy-and-hold returns. Table 3, Panel B, presents market adjusted CARs. Because the overall results are similar in both panels, we mainly focus on the market adjusted CARs. Overall, the market adjusted return increases 1.2% on PeakDay, which is statistically significant at the 0.05 level. However, the run up of about 1.0% before PeakDay is not statistically significant. The trend reverses after PeakDay, producing statistically significant negative abnormal returns with larger losses the longer the ex post period. From PeakDay to 1 day after the campaign, the CAR is about −4.0% and from just before the campaign to 255 days after the campaign ends the price declines about 34%.

For Type ND Firms, we do not observe a significant increase in returns during the periods immediately before or after the SC. However, for Type DND Firms and Type D Firms, we observe a run up in returns before the PeakDay and also positive returns on the PeakDay. Then the returns become negative immediately after the SC. Over the long run, we observe

Table 3 Cumulative abnormal returns around Spam Campaigns (SCs)

Variable	All (N = 675)		Firm ND (N = 119)		Firm D (N = 283)		Firm DND (N = 273)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Panel A: Unadjusted buy-and-hold returns								
PeakDay	0.013**	0.152	−0.004	0.118	0.011	0.146	0.022**	0.169
Beg-1 to PeakDay	0.011	0.205	−0.014	0.125	0.013	0.204	0.020	0.232
PeakDay to end+1	−0.040***	0.161	−0.022*	0.124	−0.044***	0.163	−0.044***	0.172
PeakDay to end+5	−0.093***	0.398	0.012	0.773	−0.115***	0.242	−0.116***	0.259
Beg-1 to end+1	−0.028***	0.253	−0.034**	0.167	−0.033**	0.228	−0.019	0.304
Beg-1 to end+5	−0.087***	0.411	0.003	0.786	−0.107***	0.266	−0.105***	0.272
Beg-1 to end+255	−0.339**	4.188	−0.240	3.069	−0.297	4.086	−0.426	4.695
Panel B: Market adjusted cumulative abnormal returns								
PeakDay	0.012**	0.152	−0.004	0.119	0.010	0.146	0.021**	0.170
Beg-1 to PeakDay	0.010	0.209	−0.027	0.186	0.013	0.191	0.024*	0.233
PeakDay to end+1	−0.050***	0.362	−0.094	0.758	−0.048***	0.198	−0.033***	0.185
PeakDay to end+5	−0.107***	0.283	−0.068***	0.200	−0.120***	0.300	−0.110***	0.294
Beg-1 to end+1	−0.033**	0.398	−0.105	0.761	−0.033**	0.249	−0.002	0.272
Beg-1 to end+5	−0.090***	0.312	−0.079***	0.233	−0.105***	0.329	−0.079***	0.324

An SC is defined as a period of spamming activity with no more than five consecutive days without a spam email. PeakDay is defined as the day within the SC with the maximum number of spam emails, taking the first such day in case of ties. Some emails include a conflict of interest disclosure (type D) and others do not (type ND). We classify firms into three groups according to whether they are touted by type D emails only (D), type ND emails only (ND), or both types of emails (DND). "Beg" ("End") is the beginning (end) day of the SC. N is the number of SCs. We use the Russell 2000 as our proxy for the market return. Abnormal return is the difference between the stock return and the Russell 2000 return. CARs are the cumulative abnormal returns. Buy and hold return is the sum of the log price relatives over the indicated period. All data are for January 2004 through December 2007.

* Significance at the 0.10 level.

** Significance at the 0.05 level.

*** Significant at the 0.01 level.

a statistically insignificant decline in buy-and-hold returns over a one year holding period for each firm type, but the decline in buy-and-hold returns is statistically significant for the aggregated sample.

Fig. 2 graphs CARs around SCs for Type D Firms, Type ND Firms, and Type DND Firms. We can clearly see the run up and subsequent decline in CARs for Type D Firms and Type DND Firms. Recall from Table 3 that the run up in stock price is statistically insignificant. The return build-up is about 1% to 2% for these firms and then declines more than 5% from the start of the SC. However, the spam emails not disclosing conflicts of interest do not drive up stock prices before or on the PeakDay of SC. The CARs decline immediately after the PeakDay, but recover quickly for Type ND Firms. All three types of firms suffer a large drop in CARs of about 5% after the stock SC. In summary, for all of the sample firms, there is a positive but statistically insignificant run up in stock prices in the short term and a big decline in the long term after SCs.

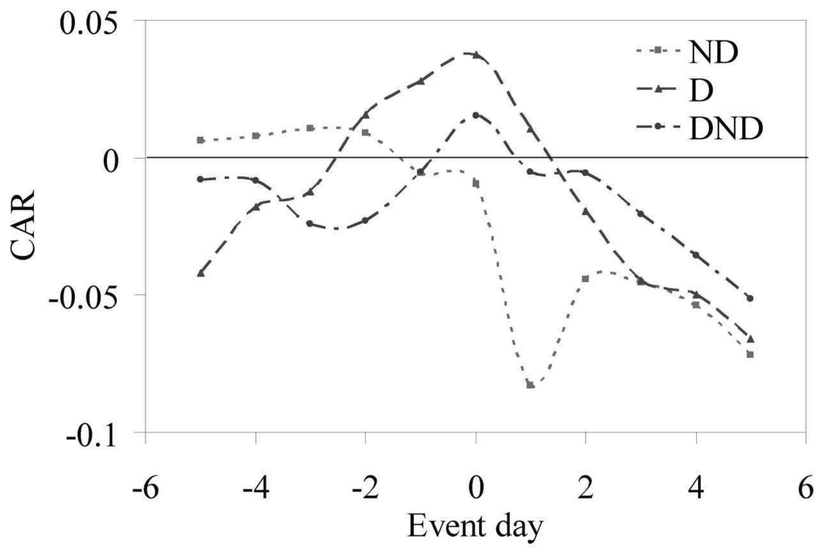

Fig. 2. Cumulative abnormal returns (CAR) around Spam Campaigns (SCs). An SC is defined as a period of spamming activity with no more than five consecutive days without a spam email. PeakDay is the day within the SC with the maximum number of spam emails, taking the first such day for ties. Event days are relative to the PeakDay of the SC. Abnormal return (*AR*) is the difference between PeakDay stock return and the Russell 2000 return, which is a proxy for market return. CAR is the cumulative abnormal return from Day -5 through the event day. Some emails include a conflict of interest disclosure (Type D) and others do not (Type ND). We classify firms into three groups according to whether they are touted by Type D emails only (D), Type ND emails only (ND), or both types of emails (DND).

5.5. Panel analysis of firm type effects

We continue our investigation of whether the spam emails disclosing conflicts of interest are less effective than those that do not using panel regressions. Table 4 reports results with dependent variables for turnover, return, and *RISK*, in turn. We include interactive dummies for firm type and days surrounding spam events. Our panel regressions control for unobserved heterogeneities across sample firms.

5.5.1. Liquidity

Table 4, Columns 2 and 3, report panel regression results with turnover as a dependent variable. We use turnover as a proxy for liquidity. The coefficients for weekday dummies are positive and statistically significant for Tuesday through Thursday, indicating that traders are more willing to supply liquidity on these days relative to Monday. For all three dummies related to the days around SCs, the coefficients are positive and highly significant, especially given the *PeakDay* coefficient of 0.3501, which signifies a 35% increase in turnover.

We also include the interaction of *PeakDay* and *Firmtype*. The coefficient for *PeakDay***Firmtype* is −0.1109, which is statistically significant at the 5% level. For the spam emails disclosing conflicts of interest, the increase in liquidity is 11.09% less compared

Table 4 Panel regression results for the effect of firm types

	Turnover		Return		RISK	
Variable	Coefficient	*p*-value	Coefficient	*p*-value	Coefficient	*p*-value
Intercept	0.4692	0.0000	−0.0079	0.0000	1.0150	0.0000
Tue	0.0131	0.0000	0.0015	0.2800	−0.0138	0.1680
Wed	0.0132	0.0000	0.0034	0.0120	−0.0198	0.0490
Thu	0.0071	0.0550	0.0055	0.0000	−0.0237	0.0190
Fri	−0.0024	0.5140	0.0117	0.0000	−0.0227	0.0240
Pre-SC	0.1023	0.0000	0.0046	0.4240	0.0083	0.8190
PeakDay	0.3985	0.0000	0.0189	0.1350	0.2909	0.0000
Post-SC	0.1430	0.0000	−0.0119	0.0370	−0.0820	0.0220
Pre-SC*Firmtype	−0.0285	0.1640	0.0066	0.4500	0.0962	0.0850
PeakDay*Firmtype	−0.1109	0.0120	−0.0034	0.8610	0.1192	0.3190
Post-SC*Firmtype	−0.0616	0.0020	−0.0030	0.7290	0.1438	0.0090
N	212,197		344,057		211,922	
R^2	0.0023		0.0003		0.0002	

We report results for fixed effects panel regressions with three dependent variables—turnover, return, and risk, in turn. *Turnover* = ln((1 + dollar volume)/average dollar volume). *Return* = $R_t = ln\ S_t - ln\ S_{t-1}$ where S_t is the adjusted closing price of the stock on day t. *RISK* = (intraday price range)/(average intraday price range) where Intraday price range = log(intraday high price–intraday low price). *Tue, Wed, Thu,* and *Fri* are dummy variables that equal 1 if the observation is for the respective day of the week and 0 otherwise. *Pre-SC* (*Post-SC*) is a dummy variable that equals 1 if the observation is for one of the five days before (after) the SC and 0 otherwise. *PeakDay* is a dummy variable that equals 1 if the observation is for the PeakDay and 0 otherwise. We define a dummy variable, *Firmtype*, that equals 1 for Type D Firms and 0 otherwise. We create three interactive dummy variables, *Pre-SC*Firmtype*, *PeakDay*Firmtype*, and *Post-SC*Firmtype*, that are the product of *Firmtype* and *Pre-SC, PeakDay,* and *Post-SC,* in turn. We also adjust standard errors for intra-group correlation. All data are for January 2004 through December 2007.

to the overall increase in liquidity for all spam emails. Therefore, the spam emails disclosing conflicts of interest are less effective. This result supports Hypothesis 2.

We further include the interactions of *Pre-SC*, and *Post-SC* and *Firmtype*. We find that Type D Firms, which is spammed only by emails disclosing conflicts of interest, has lower turnover on the PeakDay and during the five days after the SC. The increase in liquidity is about 6.16% less during the five days after an SC for Type D Firms. These results further support Hypothesis 2.

5.5.2. Return

Turning to returns, the results reported in Table 4 confirm our earlier findings obtained using the CAR approach. *PeakDay* return is about 1.89%. The average return during the five days after an SC is −1.19%, or −5.95% cumulatively for the five days. The decline in stock price after an SC is statistically significant at the 5% level. These results are similar to Hanke and Hauser (2008) who report 1.9% for the spam day and −1.3% for the days after a spam day.

We also include the interaction of *PeakDay* and *Firmtype*. The coefficient for *PeakDay*Firmtype* is −0.0034. Although PeakDay return is less for the spam emails disclosing conflicts of interest, the reduction is neither economically nor statistically significant. We further include the interactions of *Pre-SC*, and *Post-SC* and *Firmtype*. Again, we find that Type D Firms are not significantly different for the PeakDay return compared to the return for overall sample firms.

5.5.3. RISK

For the regression reported in Table 4, we use standardized intraday return volatility (*RISK*) as a proxy for risk. The coefficient for *PeakDay* is 0.2909, which is statistically significant at the 1% level. Spammers increase return volatility on the PeakDay of an SC. The coefficient for *Post-SC* is -0.0820, which is statistically significant at the 5% level.

Next, we examine the interaction of *PeakDay* and *Firmtype*. The coefficient for *PeakDay*Firmtype* is 0.1192, which is not statistically significant. We further include the interactions of *Pre-SC*, and *Post-SC* and *Firmtype*. If the Type D emails are in fact less effective, we expect that the reduction in volatility after the SC for Type D Firms will be smaller compared to the reduction by other firms. We find that the coefficient of *Post-SC*Firmtype* is 0.1438, which is statistically significant at the 1% level. These results support Hypothesis 2.

Overall, the results in Table 4 show that the spam emails disclosing conflicts of interest are less effective. Our results support the liquidity hypothesis proposed by Hanke and Hauser (2008). However, the SEC mandatory disclosure requirement tends to reduce the liquidity benefit for spammers, making Type D spam emails less detrimental.

5.6. Panel analysis of SEC crackdown effects

Table 5 presents panel analysis of the effect of the SEC crackdown. From the turnover-based regressions, the coefficients for the interactions of *SEC* and days surrounding SCs are all negative and statistically significant. For example, the PeakDay turnover is reduced by 23.39% after the SEC crackdown. For return based regressions, the coefficients are all negative, but statistically insignificant. After the SEC crackdown, the spammers earn less return on their spam efforts. For the *RISK* based regressions, the coefficient for *PeakDay*SEC* is -0.5132, which is statistically significant at the 10% level. The SEC crackdown reduces return volatility on the PeakDay of a SC. These results provide further support for Hypothesis 1. Overall, we find that SCs do not increase liquidity as much after the SEC crackdown as they did earlier.

6. Conclusions

Fraudsters frequently use spam email to make false or misleading statements to induce individual investors to purchase stocks. When the stock price increases because of this artificial demand, the promoters of these pump and dump schemes sell their shares and investors lose their money when the stock price subsequently declines. To combat these spam email campaigns, the United States adopted the CAN-SPAM of 2003. We investigate the efficacy of the CAN-SPAM Act in combating these stock scams.

Using statistical text classification techniques, we examine and classify over 40,000 stock spam emails. We find that about 60% of spam emails follow the SEC conflict of interest

Table 5 Panel analysis of SEC Operation Spamalot

	Turnover		Return		RISK	
Variable	Coefficient	*p*-value	Coefficient	*p*-value	Coefficient	*p*-value
Intercept	0.4693	0.0000	−0.0079	0.0000	1.0149	0.0000
Tue	0.0131	0.0000	0.0015	0.2800	−0.0138	0.1690
Wed	0.0132	0.0000	0.0034	0.0120	−0.0197	0.0490
Thu	0.0071	0.0540	0.0055	0.0000	−0.0237	0.0180
Fri	−0.0024	0.5170	0.0117	0.0000	−0.0227	0.0250
Pre-SC	0.0952	0.0000	0.0080	0.0740	0.0524	0.0630
PeakDay	0.3601	0.0000	0.0189	0.0560	0.3650	0.0000
Post-SC	0.1226	0.0000	−0.0131	0.0030	−0.0145	0.5990
Pre-SC*SEC	−0.1283	0.0130	−0.0116	0.5560	−0.0973	0.4880
PeakDay*SEC	−0.2339	0.0310	−0.0269	0.5360	−0.5132	0.0820
Post-SC*SEC	−0.1470	0.0040	−0.0042	0.8290	−0.1682	0.2220
N	212,197		344,057		211,922	
R^2	0.0023		0.0003		0.0002	

We report results for fixed effects panel regressions with three dependent variables—turnover, return, and risk, in turn. *Turnover* = ln((1 + dollar volume)/average dollar volume). *Return* = $R_t = ln\ S_t - \ln S_{t-1}$ where S_t is the adjusted closing price of the stock on day t. *RISK* = (intraday price range)/(average intraday price range) where Intraday price range = log(intraday high price–intraday low price). *Tue, Wed, Thu,* and *Fri* are dummy variables that equal 1 if the observation is for the respective day of the week and 0 otherwise. *Pre-SC* (*Post-SC*) is a dummy variable that equals 1 if the observation is for one of the five days before (after) the SC and 0 otherwise. *PeakDay* is a dummy variable that equals 1 if the observation is for the PeakDay and 0 otherwise. We define *SEC* as a dummy variable the equals 1 for observations occurring before Operation Spamalot and 0 otherwise. We define three dummy variables that are the product of *SEC* and *Pre-SC, PeakDay*, and *Post-SC*, in turn, to produce *Pre-SC*SEC*, *PeakDay*SEC*, and *Post-SC*SEC*. We adjust standard errors for intra-group correlation. All data are for January 2004 through December 2007.

disclosure requirements. Overall, the SEC crackdown on spam emails on March 8, 2007 reduced the number of spam emails and diminished their effectiveness.

We find evidence that emails that disclose conflicts of interest have a lower market impact than those that do not. However, all email campaigns show a statistically significant decline in stock price from the peak spam day to a subsequent day, indicating losses for individual investors.

Notes

1 http://www.sec.gov/answers/pumpdump.htm.
2 http://www.ftc.gov/spam/.
3 For example, see http://www.sec.gov/news/press/2007/2007–34.htm, or http://news.bbc.co.uk/1/hi/business/6433721.stm.
4 For example, see http://stockpromoters.com/Stockpromoters_In_News.asp and http://www.reuters.com/article/domesticNews/idUSN1120537620070611.
5 http://news.cnet.com/8301–1001_3–9992381–92.html.
6 http://www.sec.gov/litigation/suspensions.shtml.
7 We thank Leonard Richardson from www.crummy.com for supplying the spam email dataset.

8 We received a private email communication.
9 The Rainbow software package can be obtained from http://www.cs.cmu.edu/~mc-callum/bow/. Bin Zhang provides several patches. Details about our implementation of Rainbow algorithm are available upon request.
10 The percentage of spam emails classified as Type D is 60.3% in the initial sample and 57.1% in the final sample.
11 The firms may appear in all three years. Therefore, the number of firms across all three years exceeds 395.

Acknowledgment

We would like to thank Stuart Michelson (the editor) for valuable comments and suggestions.

Appendix

Spam email classification using computational linguistic methods

We use the Naive Bayes Classification to assign the most probable target value (1 for disclosing conflicts of interest and 0 for not disclosing) to a spam email, given the attribute values of the message. The Naive Bayes classifier requires 2,000 email messages of training data to estimate the means and variances necessary for classification. Mathematically, let P(h) be the probability that a hypothesis h holds. For example, P(the new spam message discloses conflicts of interest). Let P(T) be the probability that training data T will be observed. Let P(T h) probability of observing data T given that hypothesis h holds and P(h T) probability that h holds given training data T. Based on Bayes Theorem,

$$P(h|T) = \frac{P(T|h)P(h)}{P(T)}$$

Therefore, Bayes Theorem provides a way to calculate P(h T) from P(h), together with P(T) and P(T h). In our case, parameter estimation for naive Bayes models uses the method of

Table A1 Naive Bayes classification accuracy within sample and overall classification distribution

Class	By hand	By algorithm		Accuracy
		D	ND	
D	1,560	1,551	9	99.42%
ND	388	0	388	100.00%

The first number column shows the number of messages in each class hand-coded by the authors. The D-ND matrix entries show the in-sample prediction accuracy in terms of number of messages using the classification algorithm with respect to the learned samples. The last column shows the classification accuracy.

maximum likelihood. Given the attributes in the spam message, the most probable target value can be assigned by maximizing

$$h = \max_{h_j} P(h_j \mid a_1, \ldots, a_n)$$

where $a_1, \ldots, a_n$ are the attributes appeared in the spam email to be classified. Simplify based on Bayes Theorem

$$h = \max_{h_j} \frac{P(a_1, \ldots, a_n \mid h_j) P(h_j)}{P(a_1, \ldots, a_n)}$$

The denominator is a constant; therefore, we can just maximize the following equation:

$$h = \max_{h_j} P(a_1, \ldots, a_n \mid h_j) P(h_j)$$

Naive Bayes assumes that the attribute values are conditionally independent given the target value h. The above equation can be simplified further.

$$h = \max_{h_j} P(h_j) \prod_i P(a_i \mid h_j)$$

Because all the terms in the above equation can be calculated, we can assign the spam email a target value and hence its class.

To implement the Rainbow algorithm, we first train the software using a set of pre-classified emails and construct two vectors, one vector for each class (Type D or ND). This serves as learned model. We then use the model to classify all the spam emails and assign them to the class with the most similar vector.

References

Aggarwal, R. K., & Wu, G. (2006). Stock market manipulations. *Journal of Business*, *79*, 1915–1953.

Allen, F., & Gale, D. (1992). Stock-price manipulation. *Review of Financial Studies*, *5*, 503–529.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, *59*, 1259–1294.

Bhattacharya, U., & Daouk, H. (2002). The world price of insider trading. *Journal of Finance*, *57*, 75–108.

Bohme, R., & Holz, T. (2006). *The Effect of Stock Spam on Financial Markets*. Working paper, SSRN Working Paper No. 897431.

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, *14*, 3–31.

D'Alessio, D. (2007). A preliminary evaluation of the impact of unsolicited commercial email promoting stocks on the price of the stock. *New Media & Society*, *9*, 1011–1027.

Frieder, L., & Zittrain, J. (2006). Spam works: Evidence from stock touts and corresponding market activity. *Berkman Center Research Publication.* No. 2006–2011.

Hanke, M., & Hauser, F. (2008). On the effects of stock spam e-mails. *Journal of Financial Marketso*, *11*, 57–83.

John, K., & Narayanan, R. (1997). Market manipulation and the role of insider trading regulations. *Journal of Business*, *70*, 217–247.

Kyle, A. S., & Viswanathan, S. (2008). How to define illegal price manipulation. *American Economic Review*, *98*, 274–279.

McCallum, A. (1996). *Bow: A Toolkit for Statistical Language Modeling, Text Retrieval, Classification and Clustering*. Working paper, School of Computer Science, Carnegie-Mellon University.

Van Bommel, J. (2003). Rumors. *Journal of Finance*, *58*, 1499–1519.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.