ELSEVIER

Journal of Financial Markets 11 (2008) 57–83

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/finmar

On the effects of stock spam e-mails[☆]

Michael Hanke, Florian Hauser*

Innsbruck University School of Management, Department of Banking and Finance, Universitaetsstrasse 15, 6020 Innsbruck, Austria

Available online 6 October 2007

Abstract

Over the past few years we have seen a rise in the number of unsolicited e-mails that recommend buying certain stocks. The senders of these mails claim to have private information indicating that strong increases in the prices of these stocks are imminent. We first describe the common characteristics of stocks being peddled by such e-mails. Next, we investigate the effects of stock spam e-mails on excess returns, turnover, and intra-day price range, and we find that spam e-mails have a significant impact on all of these variables. Unlike the information spread on discussion forums, the positive news contained in spam e-mails has no lasting positive effect on stock prices. We analyze whether spam effects depend on stock characteristics such as price level and average turnover, and we find that liquidity is a major factor in the success of spamming. Moreover, we show that repeated spamming on successive days sustains excess demand for target stocks, enlarging the time window for liquidation of the spammers' positions.

JEL classification: G14; C23; G24

Keywords: Stock spam; Market manipulation

1. Introduction

Many internet users regularly receive e-mails recommending that they buy certain stocks. One motive for such activity could be market manipulation by first taking a

☆We would like to thank Matthias Bank, Alois Geyer, Rudolf Kerschbamer, Michael Pfaffermayer, Sava Savov, Leopold Soegner, Gerhard Winkler, participants of the 2006 conference of the German Finance Association, and Avanidhar Subrahmanyam (the editor) for valuable comments and suggestions.

*Corresponding author.

E-mail addresses: Michael.Hanke@uibk.ac.at (M. Hanke), Florian.Hauser@uibk.ac.at (F. Hauser).

1386-4181/$ - see front matter
doi:10.1016/j.finmar.2007.10.001

position in the stock and then artificially increasing demand, which in turn may lead to favorable price increases for offloading positions. Activity of this kind has received increased media attention in recent years (Wall Street Journal, 2000, 2005a,b).

The manipulation of stock markets is a phenomenon that has been around for a long time, possibly for as long as stock markets have existed. There is quite a wide range of literature on this topic (e.g., Allen and Gale, 1992; Allen and Gorton, 1992; van Bommel, 2003; for a survey of the early literature, see Cherian and Jarrow, 1995). Due to a lack of empirical data, however, most of this literature focuses on theoretical models, whereas there are only a few studies of real-world manipulation incidents (Mei, Wu, and Zhou, 2004; Jiang, Mohaney, and Mei, 2005; Khwaja and Mihan, 2005; Aggarwal and Wu, 2006). Following the classification by Allen and Gale (1992), we categorize manipulation via stock spam e-mails as information-based. In the past, information-based manipulation usually occurred more subtly, sometimes using "cooked" analyst reports. The success of such strategies depended strongly on the analyst's reputation, which was put on the line by repeatedly issuing manipulated reports. Precise timing was much more difficult, if not impossible, with traditional media. Newspapers, for example, are read by some people in the morning, by others in the evening, and by a few "insiders" the day before they go to press.

For would-be manipulators, the internet offers an ideal framework because it allows them to contact a large number of potential investors almost simultaneously. In addition to e-mail, electronic newsletters and internet discussion forums are popular tools that can be used for information-based manipulation. Whereas electronic newsletters are little more than old print newsletters transferred to a new medium, discussion forums on the web are a new framework for exchanging information on stocks. There are already several studies analyzing the explanatory power of information spread in such forums for stock price changes (see, e.g., Tumarkin and Whitelaw, 2001; Dewally, 2003; Antweiler and Frank, 2004). A major finding of these studies is that information spread in discussion forums has a lasting impact on returns: prices rise when positive news becomes available. Although there is some counter-reaction on the following days, the ensuing fall in prices is less than the initial rise, and the result is a net positive effect. This indicates that the information circulated in discussion forums is (at least partly) relevant in the sense that it has a longer-term effect on stock prices. In contrast to this, we find that the "information" contained in spam e-mails is useless: positive excess returns on spam days are completely wiped out by negative excess returns in the week after a spam event.[1]

When we started this research project, no empirical study had yet been done on stock spam e-mails. Only recently, we have become aware of two competing papers (Boehme and Holz, 2006; Frieder and Zittrain, 2006). Boehme and Holz (2006) use a data set of 111 stocks from the same source as ours, and a time frame of about 15 months (Nov. 2004–Feb. 2006). They employ a multiplicative multivariate regression model and classical event study methodology, focusing on effects of spam e-mails on the return and volume of the target stocks. For both variables, they find significant effects. Exploring the direction of the relationship between spam e-mails and volume, they find that spam results in high volume, but high volume does not cause increased spam activity. Stock prices react significantly and positively to spam e-mails, with a positive relationship between the

[1] Actually, cumulated negative returns after a spam event are on average larger than the positive return on the spam day itself. We will discuss possible reasons for this finding in Section 5.1.1.

intensity of spam mails (number of spam e-mails per day about the same stock) and the size of returns. The sample used by Frieder and Zittrain (2006) partly overlaps with ours with regard to both ticker symbols and time frame. They confine themselves to stocks listed on Pink Sheets (see Section 3.2). While we employ panel regression allowing for fixed cross-section effects, they use pooled OLS with clustered standard errors. Our analysis focuses solely on a sample of spammed stocks, while they concentrate primarily on a comparison of spammed stocks with a control group of low-priced Pink Sheets stocks that were not spammed during their sample period. Generally, they find a positive impact of spam e-mails on both the return and volume of the target stock. Further results from Boehme and Holz (2006) and Frieder and Zittrain (2006) will be discussed and compared to our findings in Section 5.

Our goal in this paper is to explore the effects of stock spam e-mails that can be found in market data. We use panel regression, allowing for fixed cross-section effects and avoiding potential biases resulting from ordinary (pooled) least squares. Our results indicate that spam e-mails have marked effects on returns, turnover, and intra-day volatility. On the day before a spam e-mail is sent out, turnover and intra-day volatility are already higher than usual. On the spam day itself, we see significantly positive excess returns, high intra-day volatility and increased turnover. For several days after a spam event, excess returns are negative at high turnover, while intra-day volatility quickly returns to normal levels. A separate investigation of the effects of ‘‘successive spamming’’ (our term for repeated spamming of the same stock on successive days) shows that this strategy sustains the effects of spam e-mails. Analyzing dependencies between spam effects and stock characteristics, we find significantly different reactions depending on price and turnover levels of target stocks. This suggests that liquidity is an important factor in stock spamming. Some of the limitations of our study are due to the unavailability of intra-day data and bid-ask quotes. These would have allowed us to assess potential profits from trading strategies.

The paper is organized as follows: Section 2 provides additional information about stock spam e-mails and discusses possible motivation and trading strategies of spammers. We formulate an economic story together with a number of interesting questions to be investigated in our empirical analysis. Section 3 presents our data set in detail. In Section 4, we describe the design of our study, providing all the regression equations estimated. Section 5 presents and discusses the results, and Section 6 concludes.

2. Stock spamming

We define stock spam e-mails as unsolicited e-mails that explicitly recommend buying a certain stock (the *target*) and/or contain ‘‘information’’ about this stock that investors will most certainly interpret as a buy recommendation. Stock spam e-mails are one of the contemporary variants of ‘‘pump-and-dump’’ or ‘‘scalping’’ strategies. The goal of such strategies is to artificially increase the demand for a certain stock by spreading false positive news in the hope of liquidating one’s own existing position (which may have been built up exclusively for this reason) at an inflated price. Before the advent of internet technology and its vast possibilities, these strategies were difficult to execute. Ideally, all recipients should get the forged information at the same time, but until e-mail there was no efficient way to disseminate it quickly to a large audience. Moreover, e-mails allow manipulators to remain anonymous. This is important for them because spamming is illegal in many jurisdictions.

The stocks that are commonly targeted by spammers share several typical characteristics (details will be provided in Section 3). Here, it suffices to note that popular target stocks are (see, e.g., Luft and Levine, 2004, p. 2)

- not traded on organized exchanges, but quoted on platforms like OTCBB or Pink Sheets[2] and traded OTC via market makers,
- very cheap, often trading markedly below one dollar a share (so-called penny stocks),
- micro-caps (i.e., have a small total market value),
- thinly traded, with average daily turnover at five-to-six-digit figures or lower,
- traded quite irregularly, with frequent gaps in daily price series, and
- very difficult or impossible to sell short.

These features, together with the observation that stock spam e-mails always indicate an imminent rise in the target's stock price, suggest two alternative possible motivations for spamming a stock: either the spammer intends to sell a certain stock he already owns and wants to boost the selling price, or the spammer buys the target stock, initiates the spam soon afterwards, and sells the stock again once the price has risen. The large number of spam e-mails suggests that there are professional spammers around who initiate stock spam e-mails on a regular basis. The negative average return of target stocks during our sample period (see Section 3.1) provides a strong disincentive against holding such stocks for longer periods of time. Thus, we assume that professional spammers build up the positions in their target stocks as shortly as possible before initiating the spam mails.

In the absence of published work on the exact way spammers proceed, we formulate a plausible economic story by putting ourselves in the shoes of the spammer. Based on descriptive statistics of our sample and straightforward economic reasoning, we believe that the following events surround a typical spam incident (we will confirm whether the data support this story in Section 5): First, the spammer selects a target. Once he thinks market conditions are favorable for his activity (e.g., because the stock price is currently low compared to the recent past), he builds up a position in the target stock. Shortly afterwards, he sends out spam e-mails which lead to a rising demand for the target stock. Market makers react to this order flow imbalance by raising quotes and/or increasing bid-ask spreads, leading to higher returns on spam days. The spammer liquidates his position and skims the profit from his activity. Once spam victims realize that the "information" spread by the spammer is false, they sell their stocks, leading to negative returns after the spam day. If markets for the stocks typically targeted were liquid and efficient, the magnitude of the cumulated negative returns after the spam event should approximately equal that of the positive return on the spam day.

Given that many spam events have occurred in recent years, investors should already know better than to act on their advice. The fact that we continue to see abnormal returns on spam days over the entire sample period is hard to explain using standard neoclassical finance theory. However, we should not forget that the stocks typically targeted by spammers are very illiquid: a handful of investors following the spammers' advice is

[2]These platforms will be described in more detail in Section 3.2.

sufficient to move the prices of such stocks significantly. We believe that spam victims are predominantly inexperienced investors hoping to make a fast buck. Even if victims have some doubts about the credibility of the information in spam e-mails, they may be willing to buy if they underestimate (or are unaware of) the liquidity risk associated with typical target stocks. Spammers only need a few such investors in order to be successful, so they may simply rely on finding new victims each time they initiate a new spam (cf. the proverb "There's a fool born every minute.").

From a spammer's point of view, many of the characteristics of popular target stocks described above make good sense. It is much easier to manipulate the price of a low-turnover stock. Small movements in absolute terms translate into large percentage changes for penny stocks, which makes cheap stocks attractive targets. Given that stocks with these characteristics are often very difficult to sell short, at least at reasonable transactions costs, we should indeed expect manipulation attempts via stock spam e-mails to contain only information which is positive for the target's stock price.

At the same time, this raises a number of questions:

- Quite frequently, we observe stock spam e-mails targeting the same stocks repeatedly on successive days. Does this strategy succeed in keeping up demand for the target stock? Is sequential spamming more effective than one-shot spams? Do the other effects we observe differ between one-off spams and sequential spamming?
- The effectiveness of spam e-mails may vary depending on whether the information spread by spammers actually reaches their target clientele. The habits of potential investors with regard to reading e-mails and tending to their finances differ: some leave financial matters for the weekend because they are too busy during the workweek, while others reserve weekends for their families. Are there any day-of-the-week effects in the sense that spam e-mails are more effective, say, on Mondays?
- Even without any spamming, stock prices can be manipulated by simply buying illiquid stocks,[3] and hoping that other market participants interpret increasing turnover and prices as a revelation of private information (see, e.g., Allen and Gale, 1992). For the spammer, however, illiquidity is a two-edged sword: a spammer's own orders might have too great a price impact on stocks that are too thinly traded, and bid-ask spreads of illiquid stocks are generally higher. Are high- or low-turnover stocks more desirable targets from the spammers' point of view?
- The spammers' success depends crucially on the willingness of other investors to invest in the targets selected by the spammers. Given that stocks trading at very low prices usually carry a negative image, is the spammers' rule really "the cheaper, the better"?

Answers to these questions will be provided in Section 5.

3. Data description

We define a *spam event* for a certain stock as a day on which at least one incident of a recorded stock spam e-mail occurs before the close of business.[4] Spam events are retrieved

[3] Allen and Gale (1992) call this *manipulation by trading*. Stocks with short-selling restrictions are particularly prone to this type of manipulation.

[4] We do not account for the number of spam e-mails received for the same ticker symbol on an event day.

from a publicly available database.[5] The database allows us to precisely match spam events to trading days. Price and volume data are gathered from Datastream. Our sample covers the year 2005.

Due to the unusual nature of our data, we describe the data collection process underlying the construction of the Crummy database in detail. The author of the database maintains a large number of so-called trap accounts—e-mail accounts set up solely for the reception of spam e-mails. Every evening, after the close of trading in the U.S., scripts automatically evaluate the e-mails received for all trap accounts, classify the subset of stock spam e-mails according to the ticker symbol, and time-stamp them. We checked a large sample of e-mails from the database for possible errors in the automated classification scripts and excluded about 20 ticker symbols that turned out to be incorrectly classified in the database from our sample. We should point out that there is absolutely no way to ensure that the database is exhaustive. There may well be stock spam e-mails from 2005 that are not recorded in the database. We strongly believe, however, that this in no way casts doubt on the validity of our results. On the contrary, a higher number of spam events in our sample should rather improve the statistical significance of our results.

Our price series consists of daily closing (transaction) prices together with intra-day highs and lows. For the stocks in the database, bid-ask quotes are not publicly available.[6] These would have been valuable for two reasons: first, price data are prone to bias from bid-ask bounce (see, e.g., Roll, 1984; Lease, Mahulis, and Page, 1991; Jegadeesh, 1990). Such bias could have been completely avoided by using mid-prices (average of bid and ask quotes) instead of closing transaction prices. Second, the size of the bid-ask spread could have been used to construct a proxy for liquidity.

To be included in our sample, stocks in the database had to fulfill the following criteria:

- Availability of price data from datastream.
- At least one spam event between January 1 and December 31, 2005.
- Trading activity resulting in at least 50 recorded daily closing prices in 2005 (eliminating extremely illiquid stocks).
- Trading at a U.S. marketplace (to ensure adequate matching of spam events to trading days).

3.1. Descriptive statistics

Of the 360 stocks in the database, 235 correspond to our criteria. For these 235 stocks included in the sample, we observe 1,241 spam events in 2005. Fig. 1 shows that the effects of spam e-mails on the stocks are quite different: while 188 stocks were subject to no more than five spam events, the most popular target was spammed on almost 60 days in 2005.

Considering the large sample used in our study, an analysis of the number of target stocks spammed on certain trading days provides a good indicator for the distribution of spam activity over time. The left panel of Fig. 2 depicts the number of spammed stocks per

[5]The Crummy database, http://www.crummy.com/features/StockSpam/reports/.

[6]Frieder and Zittrain (2006) have obtained bid-ask spread data for their dataset (which consists only of stocks listed on Pink Sheets) directly from Pink Sheets. Nevertheless, they use transaction prices to calculate returns.

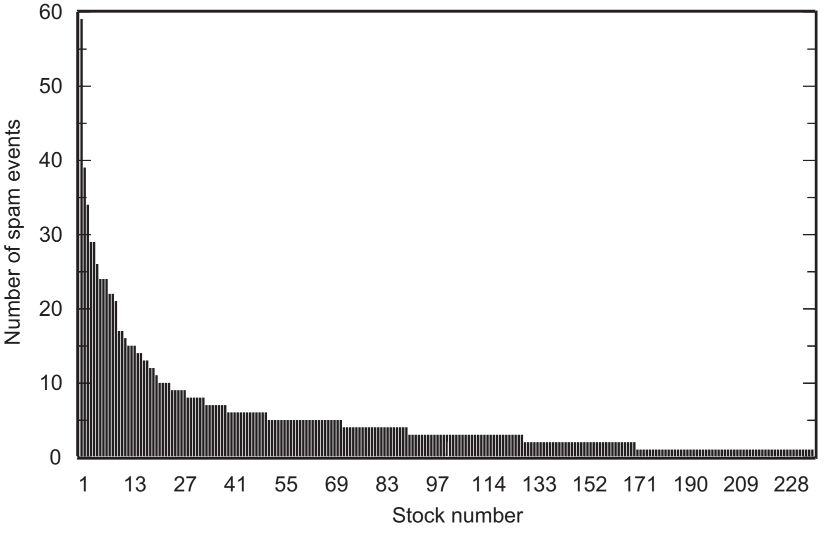

Fig. 1. Number of spam events per stock (stocks sorted by spam event frequency).

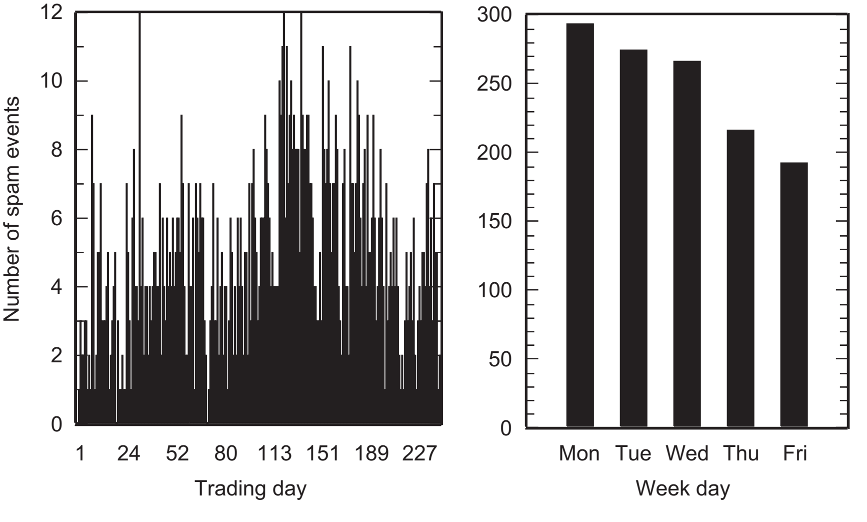

Fig. 2. Number of spam events by trading day (left panel)/by weekday (right panel).

trading day in 2005, showing a maximum of 12 spammed stocks for a single trading day. The right panel of Fig. 2 shows that there are differences between days of the week, with Monday showing the largest number of spam events and Friday the lowest. The larger amount of spam on Mondays may be explained by the fact that this number includes spam e-mails received during the weekend. In Section 5.4 we show that Friday spam e-mails do not have the intended effect, which may explain why we observe less spam activity on Fridays.

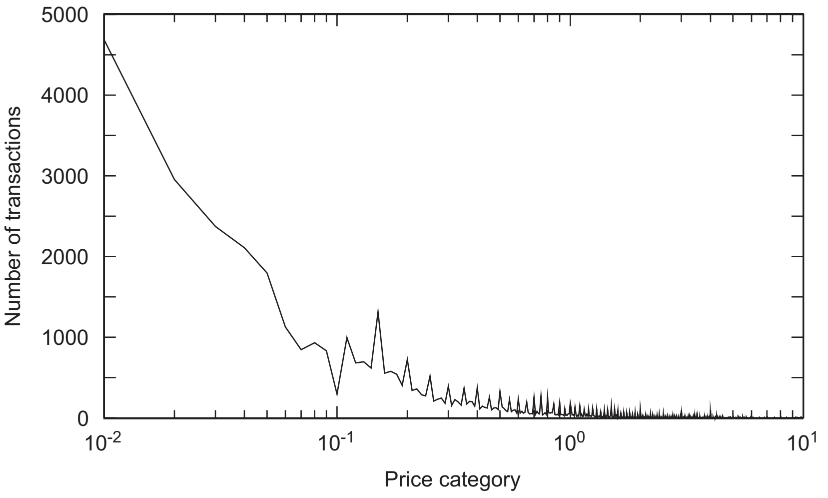

Fig. 3. Number of trades by price category (intervals of 1 cent, observe the log scale on the horizontal axis!).

A preliminary analysis of our data confirms the characterization of target stocks provided above. Trading in the target stocks is quite thin: for 59,220 stock trading days in total (252 trading days times 235 stocks), we record only 41,764 prices. Penny stocks are the predominant targets: 36,309 of the 41,764 unadjusted prices are below one dollar. Fig. 3 provides more information about the distribution of prices in our sample.

After several warnings issued by the SEC,[7] investors should already be aware that investing in penny stocks, in particular stocks that are targeted by spammers, is quite a risky business. This is confirmed by the average returns of the stocks in our sample: even diversifying across all 235 stocks targeted by spam mails in 2005 does not provide protection from huge losses. Fig. 4 shows the value of a hypothetical portfolio, starting at 100 on January 3 and growing at compounded average daily returns of the stocks in our sample. This can roughly be seen as the performance of a naïve diversification strategy. The portfolio loses almost 90% over one year.[8] During the same period, the Russell Microcap Index gained 2.57%.

Fig. 5 shows that daily returns on the stocks in our sample are leptokurtic. The solid line shows the frequency for values from a normal distribution with mean and standard deviation equal to those of the returns in our sample. The fat tails are more pronounced for

[7] http://www.sec.gov/investor/pubs/microcapstock.htm, http://www.sec.gov/news/press/2007/2007-34.htm. On March 8, 2007, the SEC suspended trading in 35 stocks listed on the Pink Sheets which had been targeted by spammers, see http://www.sec.gov/litigation/suspensions/2007/34-55420.pdf.

[8] The strongly negative performance of the stocks in our sample is puzzling. We checked for a possible selection bias (i.e., whether spammers prefer stocks with a strongly negative past performance) by calculating average returns using only prices observed *after* the first spam event for each stock. The numbers did not change significantly. An alternative explanation could be that target stocks underperform *because* they are spammed, e.g., because market participants feel that a stock's reputation is hampered when it is spammed. This is more difficult to check because we do not know anything about spam events that occurred prior to our sample period. As a rough check, we tested the hypothesis that the average return of target stocks depends on the number of spam events observed during the sample period. Again, we did not find any evidence to support this.

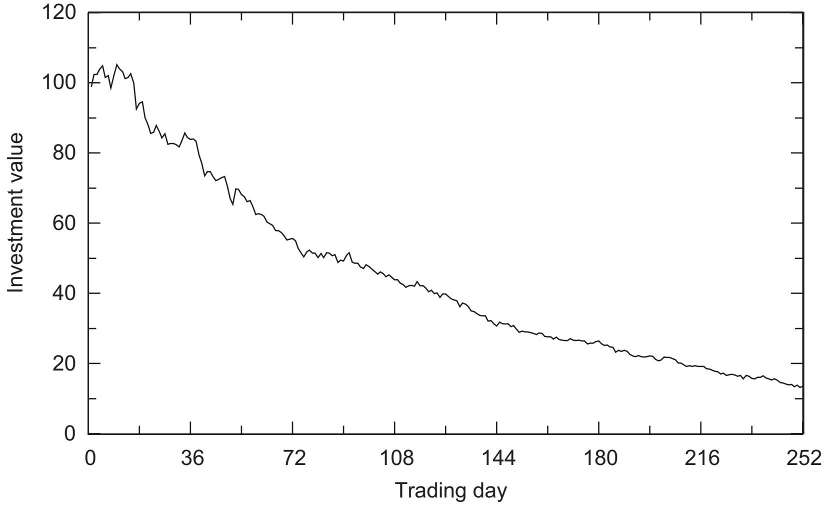

Fig. 4. Value of a hypothetical portfolio growing at compounded average daily returns of all stocks in our sample.

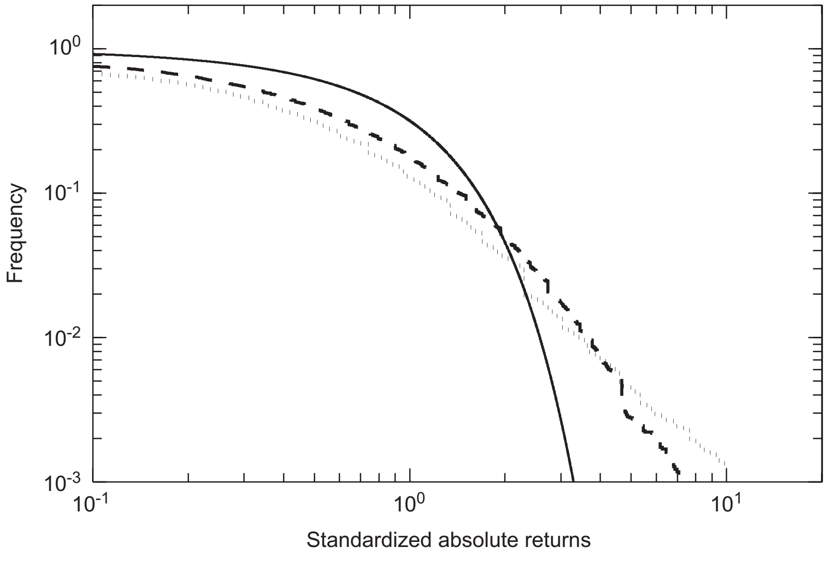

Fig. 5. Cumulated frequency of standardized absolute returns for the stocks in our sample. Dotted line is returns on spam days; dashed line is returns on unaffected days; solid line is normal distribution.

returns on spam days (dotted) than for those on unaffected days (dashed—for the exact definition of *unaffected days* see Section 4).

3.2. Electronic quotation platforms for OTC-traded stocks

The vast majority of stocks in our sample are not traded on organized exchanges, but quoted on platforms like Pink Sheets and Over-the-Counter Bulletin Board (OTCBB). In

Table 1
This table shows a comparison of OTCBB and Pink Sheets for January 2006 (data obtained from http://www.otcqx.com/docs/OTCQXBrochure.pdf)

	OTCBB	Pink sheets
Minimum financial requirements	No	No
Issuer application required	No	No
Periodic financial reporting required	Yes	No
Eligible market makers	225	189
Dollar turnover (bill.)	4.2	8.7
Exclusive securities	3269	4788

contrast to many stock exchanges, OTC markets are dealer markets where market makers quote bid and ask prices. The platforms process these quotes and provide investors with additional information. Companies quoted on these platforms usually do not meet the requirements of reputable exchanges. The listing requirements on OTC platforms are very lenient: there are no minimum financial standards, and in some cases the platforms include companies that never expressed any interest in being listed. A minor difference in regulation is that companies quoted on OTCBB are required to publish periodic financial reports (this is not necessary for stocks quoted on Pink Sheets). Table 1 compares the two most important platforms, OTCBB and Pink Sheets. More information on the market structure of OTCBB as well as the impact of liquidity and size on returns, volatility, and bid-ask spreads can be found in Larson, Luft, and Levine (2001) and Luft and Levine (2004).

4. Method

4.1. Basic setup

Our first step is the selection of a suitable method to provide answers to the questions discussed in Section 2. At first glance, our problem seems to lend itself well to the standard event study methodology, but there is one major difference: in classical event studies, the exact timing of the event is usually unknown. With our data, this is not a problem: as described in Section 3, each spam mail is time-stamped and easily associated with the corresponding trading day.

For many of the stocks in our sample, we observe only a small number of spam days. This problem can be mitigated by using panel regression with dummy variables, taking advantage of our large cross-section. In addition to a large cross-section, we also have a large number of observations in time, so many of the econometric problems related to a small number of observations in time that are discussed in the panel regression literature are not a concern for us. Panel regression is preferred over pooled OLS because we find significant unobserved fixed cross-section effects, which makes pooled OLS estimates biased.

Unobserved effects may be present both in the time and in the cross-section domain. Our data constitute an unbalanced panel, which restricts us to allowing for unobserved effects in only one of the domains. Preliminary analyses showed that there are no significant unobserved time effects. There is, however, good reason to assume a fixed unobserved cross-sectional effect, e.g., average returns vary systematically across stocks. We account for this by using the standard fixed effects transformation, also known as the *within transformation* (see, e.g., Wooldridge, 2006, p. 461). To remove any correlation from the residuals, we include lagged dependent variables (AR terms). Coefficients of AR terms up to order four are both economically and statistically significant. Higher-order AR coefficients are sometimes statistically significant (due to the size of our sample) but very small (close to zero).

To account for possible day-of-the-week effects, we use four dummies capturing the difference of Tuesdays, Wednesdays, Thursdays and Fridays from Mondays ($D_{i,\text{tue},t}$ through $D_{i,\text{fri},t}$, $i = 1, \ldots, 235$: index for the stocks in our sample). These weekday dummies are included because the descriptive statistics in Section 3.1 show less spam activity in the second half of the week. One possible reason for this could be the presence of day-of-the-week effects *even in the absence of spams* (e.g., a generally lower volume on Fridays which affects the popularity of Fridays for the spammers' activities).

In the basic model, we include different dummies for the days immediately preceding a spam day ($D_{i,p,t}$), spam days ($D_{i,s,t}$), and the five days following a spam day ($D_{i,f,t}$).[9] Days that would qualify as both following days (of an earlier spam event) and preceding days (of a later spam event) are not encoded by a dummy. In total, 2212 days are encoded as following days. Remembering that there are 1241 spam days in our sample, this gives us a ratio of 1.78 following days per spam day (this number will be useful in the interpretation of our results in Section 5.1.1.). All days that are neither spam days, preceding days or following days are called *unaffected days* (returns on such days are referred to as *unaffected returns*.).

Thus, the basic regression equation reads as follows:

$$\ddot{y}_{i,t} = \alpha_i + \sum_{j=1}^{4} \beta_{\text{AR}_\text{j}} \ddot{y}_{i,t-j} + \sum_{k \in \{\text{tue,wed,thu,fri}\}} \beta_k \ddot{D}_{i,k,t} + \sum_{l \in \{p,s,f\}} \beta_l \ddot{D}_{i,l,t} + \ddot{u}_{i,t}, \tag{1}$$

where $\ddot{y}_{i,t} = y_{i,t} - \overline{y}_i$, the time-demeaned data on y (usually called the *fixed-effects transformation*, analogously for D and u). The α reported in our results tables is the average of the α_i. We account for possible (time-) heteroskedasticity as well as autocorrelation within cross-sections by using a period SUR (PCSE) method to compute robust covariances (Beck and Katz, 1995).

4.2. Dependent variables

The following variables are used as dependent variables: RET, representing returns, TO, measuring daily turnover, and IR, a measure of intra-day volatility. Various standardization procedures are applied to the raw data to facilitate a meaningful aggregation across stocks and to linearize the dependence structure.

RET, our return measure, is based on daily raw returns $R_{i,t}$ and calculated from

$$R_{i,t} = \ln S_{i,t} - \ln S_{i,t-1}, \tag{2}$$

[9] The choice of the length of the time window is based on our economic story and preliminary analyses. Including an additional day before or after the interval chosen does not lead to significant dummy coefficients in our return regressions.

where $S_{i,t}$ denotes the closing price of stock i on day t. In many studies, the returns are also corrected for systematic effects, e.g., by calculating excess returns relative to a stock index. For the stocks in our sample, however, idiosyncratic risk clearly dominates systematic risk. In preliminary regressions, we included returns from various indices as explanatory variables (S&P 500, Russell MicroCap). None of their coefficients were significant. For this reason, $\mathrm{RET}_{i,t}$ is calculated from $R_{i,t}$ using a constant-mean-return correction as proposed in Brown and Warner (1985). This correction is handled implicitly within our model via the fixed-effects transformation (as described in Section 4.1).

As briefly mentioned in Section 3.1, our results for RET potentially suffer from bias due to bid-ask bounce. At the most elementary level, this means that a positive excess return on spam days may have different causes: the price level (measured as the mean of bid and ask quotes) might actually have increased. Alternatively, a higher number of buy orders leads to a higher probability of trades being executed at the ask. The order-flow imbalance may also induce market makers to increase bid-ask spreads, which would amplify the bid-ask bias. In practice, a combination of all three effects will be observed. Since our main goal is to detect effects of spam e-mails on market data, the exact mechanism behind the excess return is less important. Regardless which of the three effects dominates, a positive excess return on spam days provides clear evidence that spam e-mails have an impact on stock prices. Measuring the contribution of each component separately would require bid-ask quotes instead of price data. This would then allow an assessment of the economic profit from spamming. Possible bias from bid-ask bounce has further effects on our regressions using RET as the dependent variable. In particular, the coefficients of the AR terms could be biased, inducing biases in other coefficients as well. For this reason, we will perform a number of robustness checks for our return results. These checks will be described in more detail in Section 4.4.

Our second dependent variable, TO, is calculated from the daily turnover by standardizing over the turnover on unaffected days and then applying a log transformation:

$$\mathrm{TO}_{i,t} = \ln\left[1 + \frac{\mathrm{to}_{i,t}}{\overline{\mathrm{to}_i}}\right], \tag{3}$$

where $\mathrm{to}_{i,t}$ denotes the turnover of share i on trading day t, and $\overline{\mathrm{to}_i}$ is the average daily turnover of share i on unaffected days.

To construct a proxy for volatility, we follow Alizadeh, Brandt, and Diebold (2002) who show that range-based volatility proxies are less prone to noise from microstructure effects such as bid-ask bounce when compared to squared returns or absolute returns. We construct a measure of (relative) intra-day price range by starting from the daily price highs and lows in Datastream:

$$\varsigma_{i,t} = \ln[\mathrm{high}_t(S_{i,t})] - \ln[\mathrm{low}_t(S_{i,t})], \tag{4}$$

where $\mathrm{high}_t(S_{i,t})$ and $\mathrm{low}_t(S_{i,t})$ denote the intra-day high and low of stock i on day t. IR is then calculated from

$$\mathrm{IR}_{i,t} = \frac{\varsigma_{i,t}}{\overline{\varsigma_i}}, \tag{5}$$

where $\overline{\varsigma_i}$ denotes the arithmetic average of $\varsigma_{i,t}$ for unaffected days.

4.3. Explanatory variables

In the base case, we use $D_{i,k,t}$, $k \in \{\text{tue, wed, thu, fri}\}$, and $D_{i,l,t}$, $l \in \{p, s, f\}$ together with AR_j-terms ($j = 1, \ldots, 4$) as independent variables. In preliminary runs, we also included a coefficient for linear time dependence, which turned out to be zero in all regressions and has therefore been dropped. This is in contrast to the results of Boehme and Holz (2006) who find a positive linear trend in their (much smaller) sample.

In the base case we assume that following days are homogeneous. To find out whether this assumption is justified, we test for possible differences depending on the distance of following days to the spam day by including additional dummies for the first to fourth following day (D_{f1}–D_{f4}). That is, dummy D_{f1} captures any idiosyncratic deviations of the effect on the first following day compared to the behavior common to the other four. The resulting regression equation is

$$\ddot{y}_{i,t} = \alpha_i + \sum_{j=1}^{4} \beta_{\text{AR}_j} \ddot{y}_{i,t-j} + \sum_{k \in \{\text{tue,wed,thu,fri}\}} \beta_k \ddot{D}_{i,k,t} + \sum_{l \in \{p,s,f\}} \beta_l \ddot{D}_{i,l,t} + \sum_{m \in \{f1,f2,f3,f4\}} \beta_m \ddot{D}_{i,m,t} + \ddot{u}_{i,t}. \tag{6}$$

One of the questions raised in Section 2 is whether repeated spamming on successive days shows greater or smaller effects compared to one-off spams. As a simple extension to the base model, we investigate this question by defining separate spam day dummies for the first to fifth spam day in a row. These dummies are denoted by $D_{i,s,t,1}$, $D_{i,s,t,2}, \ldots, D_{i,s,t,5}$. The regression equation then reads

$$\ddot{y}_{i,t} = \alpha_i + \sum_{j=1}^{4} \beta_{\text{AR}_j} \ddot{y}_{i,t-j} + \sum_{k \in \{\text{tue,wed,thu,fri}\}} \beta_k \ddot{D}_{i,k,t} + \sum_{l \in \{p,f\}} \beta_l \ddot{D}_{i,l,t} + \sum_{m \in \{s1,s2,s3,s4,s5\}} \beta_m \ddot{D}_{i,m,t} + \ddot{u}_{i,t}. \tag{7}$$

Another issue raised in Section 2 is whether the effects of spam e-mails depend on the day of the week. We investigate this question by interacting weekday dummies $D_{i,k,t}$, $k \in \{\text{tue, wed, thu, fri}\}$ with the spam-day dummy, $D_{i,s,t}$. The resulting regression equation is given by

$$\ddot{y}_{i,t} = \alpha_i + \sum_{j=1}^{4} \beta_{\text{AR}_j} \ddot{y}_{i,t-j} + \sum_{k \in \{\text{tue,wed,thu,fri}\}} \beta_k \ddot{D}_{i,k,t} + \sum_{l \in \{p,s,f\}} \beta_l \ddot{D}_{i,l,t} + \sum_{k \in \{\text{tue,wed,thu,fri}\}} \beta_{k,s} \ddot{\Delta}_{i,k,s,t} + \ddot{u}_{i,t}, \tag{8}$$

where $\Delta_{i,k,s,t} = D_{i,k,t} \cdot D_{i,s,t}$.

Analyzing the remaining questions discussed in Section 2 using the same method would require including additional dummies, e.g., for "high", "medium" and "low-turnover" stocks. However, this approach is not applicable together with the fixed effects specification because for many cross-sections, these additional dummies would be constant in time. As an alternative approach, we split our sample into subsamples and repeat the regressions on each of the subsamples.

To investigate a possible dependence on turnover, we split our sample into three categories: The first contains all the stocks with a "high" average daily turnover (more than \$45,000), the second all the stocks with a "medium" turnover (between \$20,000 and \$45,000), and the third all the stocks with a "low" turnover (less than \$20,000). The boundaries are chosen such that the three subsamples for the return regression are of approximately equal size.

In order to find out whether the effects of spam e-mails differ depending on the price level of the target stock, we subdivide our sample into two parts, using a closing price of ten cents as the boundary. This boundary reflects the fact that stocks trading below ten cents look very cheap to investors, even in a sample where most stocks trade below one dollar.[10]

4.4. *Bias from bid-ask bounce—robustness checks*

As described in Section 4.2, our return measure RET is calculated from daily closing transactions prices. All our results using RET as the dependent variable are potentially affected in various ways by bias from bid-ask bounce.

First, selling pressure could increase the probability that the closing price on the first following day will occur at the bid. This could lead us to overestimate the magnitude of the return on that day. We address this concern by investigating any possible differential effects on the first following day compared to the other four. This is included in our analysis of possible idiosyncratic effects on particular following days (see Section 4.3).

Moreover, the bid-ask bounce is known to induce negative first-order autocorrelation. Hence, we should expect our AR coefficients to reflect this bias (in particular for AR_1, see Roll, 1984; Jegadeesh, 1990), which in turn affects our spam dummies. This effect occurs only if returns are computed over immediately adjacent time intervals. It is more pronounced for short time intervals such as trading days and therefore relevant to our analysis. To overcome this problem, Jegadeesh (1990) suggests artificially introducing gaps into the price series before calculating returns. Working with monthly data, he simply omits the last trading day in each month when calculating monthly returns. Directly transferring this approach to our daily intervals would require calculating returns from each day's closing price to the following day's second-to-last transaction price. Unfortunately, since the latter is unavailable, this option is infeasible. The closest we can get in this spirit is to omit every other day in the sample, arriving at two subsamples.

Omitting every other day effectively destroys any first-order autocorrelation that may be present in the data. Therefore, as an alternative robustness check, we simply omit the first-order AR term and repeat the analysis on the entire sample. Thus, we effectively assume that the negative first-order AR coefficient is solely due to bid-ask bounce bias.

5. Results

In the following subsections, we present and discuss the results for the regressions described in Section 4. For the spam dummy coefficients, we use one-sided tests according

[10]Other subsamples that could have been defined on the basis of such "psychological" boundaries (e.g., one dollar or one cent) contain very few observations. We therefore use only two subsamples for this analysis.

Table 2
This table shows the results for the base case, with the dependent variable RET. p-values are given for two-sided tests with period SUR corrections (one-sided for β_l, $l \in \{p, s, f\}$). The F-value is from a comparison of R^2 with/without spam dummies, ** indicates statistical significance at the 1% level

Coeff.	Est.	p-val.
α	−0.023	0.000
β_{AR_1}	−0.260	0.000
β_{AR_2}	−0.081	0.000
β_{AR_3}	−0.051	0.000
β_{AR_4}	−0.034	0.000
β_{tue}	−0.001	0.741
β_{wed}	0.009	0.010
β_{thu}	0.009	0.004
β_{fri}	0.024	0.000
β_p	0.006	0.261
β_s	0.019	0.000
β_f	−0.013	0.001
R^2	0.439	
n	55028	
F	347.15**	

to our economic story presented in Section 2. For all the other coefficients, we use two-sided tests.

5.1. Base case

5.1.1. Return

The base case panel regression is given by Eq. (1). Results for this regression using RET as the dependent variable are shown in Table 2. The spam day dummy coefficient β_s is highly significant and positive: on spam days, returns are indeed markedly higher than on other days. In contrast, there is no significant return effect on preceding days. Returns on the five following days are negative and statistically significant. The cumulated counter-reaction on following days is larger than the positive excess return on spam days. We will discuss this effect together with possible explanations below. We test whether the inclusion of our spam dummies leads to a significant increase in R^2 compared to the same regression without spam dummies. The resulting F-value is highly significant (bottom line of Table 2).

Day-of-the-week dummies are positive and statistically significant for Wednesday, Thursday and Friday. While only the coefficient for Friday can be regarded as economically significant, this effect is large enough for us to report it, even though we cannot provide an economic explanation at this stage.

We find significantly negative AR coefficients up to lag four, with a large coefficient at lag one. As discussed in Section 4.4, our best explanation for this is bias from bid-ask bounce, resulting from measuring returns over contiguous time intervals. This effect is particularly pronounced for short intervals such as the daily returns used here. Low liquidity is known to amplify the effect. This is supported by the results presented in Section 5.5 below: the lower the turnover (which can be interpreted as a rough proxy for liquidity), the higher the magnitude of the AR_1 coefficient (see Table 9).

Table 3
This table reports results of the robustness checks as discussed in Section 4.4, with the dependent variable RET. In each half of the sample, every other day is omitted. Columns 5 and 6 show results from Eq. (1) (cf. Table 2) when omitting AR_1. p-values are given for two-sided tests with period SUR corrections (one-sided for β_l, $l \in \{p, s, f\}$)

Coeff.	1st half		2nd half		All	
	Est.	p-val.	Est.	p-val.	Est.	p-val.
α	−0.023	0.000	−0.022	0.000	−0.023	0.000
β_{AR_2}	−0.020	0.018	−0.025	0.005	−0.018	0.004
β_{AR_3}					−0.035	0.000
β_{AR_4}	−0.022	0.005	−0.027	0.001	−0.021	0.000
β_{tue}	0.004	0.345	0.004	0.348	0.005	0.219
β_{wed}	0.012	0.010	0.017	0.000	0.015	0.000
β_{thu}	0.011	0.015	0.015	0.001	0.013	0.000
β_{fri}	0.028	0.000	0.029	0.000	0.028	0.000
β_p	0.006	0.352	0.005	0.352	0.003	0.381
β_s	0.015	0.026	0.013	0.047	0.015	0.001
β_f	−0.011	0.035	−0.012	0.016	−0.012	0.000
R^2	0.023		0.028		0.013	
n	14238		14231		28469	

As a first robustness check, we run the regression described in Eq. (6) to see whether the effect on the first following day differs from the other following days. The coefficient for dummy D_{f1} is insignificant (detailed results are given in Table 6 below). We conclude that as far as excess returns are concerned, the first day following a spam event does not differ from the other four. Negative excess returns on following days are not driven by a possible bid-ask bounce between the spam day and the first following day.

Additional robustness checks discussed in Section 4.4 include splitting our sample into two parts, omitting every other day in each subsample. Alternatively, we simply omit the AR_1 term, assuming that the large negative first-order autocorrelation shown in Table 2 is solely due to bias from bid-ask bounce. The results for these robustness checks are shown in Table 3. For both types of robustness checks, β_{AR_2} to β_{AR_4} are decreased markedly: As soon as the negative first-order autocorrelation is destroyed, AR coefficients of higher order become much less important. We interpret this as a clear indication that bias from bid-ask bounce is the main driver behind the large negative AR_1 coefficient in Table 2.

At the same time, the magnitude of the coefficients for spam and following days is almost unchanged. Significance levels for β_s and β_f are still very high. Thus our robustness checks show that while bias from bid-ask bounce is present in our return regressions, it does not (qualitatively) affect our findings regarding the impact of stock spam on returns. For all results to be presented later using RET as the dependent variable, we will show both the base case version and the second robustness check (dropping the AR_1 term).

Boehme and Holz (2006) also find significantly positive abnormal returns on spam days. However, they find significantly negative abnormal returns on the first day following a spam day and significantly positive abnormal returns for the next two days. They do not provide a convincing explanation for these findings. Frieder and Zittrain (2006) report negative cumulated abnormal returns for the five days following a spam event.

According to our economic story, we would have expected the positive excess return on spam days to equal the cumulated negative excess returns on the following days. Instead, we observe some over-reaction: In Section 4.1 we found that there are on average 1.78 following days per spam day. The third pair of columns in Table 3 shows that the magnitude of the positive effect on spam days is 0.013, while the counter-reaction is on average 1.78 times 0.012, or 0.021. At first glance, this puzzle bears some similarities to the well-known underperformance phenomenon in IPO markets, albeit on a much shorter time scale. In IPO markets, we also observe abnormally high short-term returns, followed by medium- to long-run underperformance (for a discussion and possible explanations see, e.g., Ritter and Welch, 2002). As with the IPO puzzle, part of the over-reaction could be caused by bias from bid-ask bounce: selling pressure from spam victims may increase the probability of trades on all following days being executed at the bid.

Apart from bid-ask bounce, there are several possible explanations for the market's over-reaction following a spam event. A simple information-based explanation for the results in Tables 2 and 3 could be that spammers are insiders who know that negative information they already possess will soon become public.[11] In this case, the apparent over-reaction would reduce to a simple omitted variables effect. Given the regularity of this type of over-reaction combined with the high overall number of spam events, we do not find this explanation is appealing.

A more plausible line of argument can be found by asking who buys (sells) on spam (following) days. On the spam day, the spammer is on the sell side, possibly alongside some investors who are long in the target stock and take advantage of rising prices to sell out, while spam victims buy. On the following days, spam victims—in our opinion, predominantly inexperienced investors—realize that the large returns promised in the spam e-mails are not forthcoming. When they subsequently try to sell their positions, they are hampered by a lack of liquidity: in the absence of enough demand in the market, market makers will lower prices to avoid being swamped with the target stock (or to make sure that they are at least adequately compensated for the risk associated with holding a large position in the target).[12] Spam victims see the value of their investment decline, and some of them may be prepared to sell anyway: they just want to get out of their investment quickly before it tanks, even if it means selling at a loss. If this explanation holds, we should observe a negative relation between the magnitude of over-reaction and liquidity. Support for this line of argument is provided by the results discussed in Sections 5.5 and 5.6.

5.1.2. Turnover

Results for the base regression using TO as the dependent variable are given in Table 4. The turnover of spammed stocks shows a significant increase on spam days, with a much smaller positive effect on following days. The fact that turnover is already up on the preceding day could be attributed to the spammer himself building up his position shortly before spreading his e-mails. This would support our economic story. Alternatively, it

[11] Note that in our analysis, we use only the spam e-mails we observe. We do not include any information on the target dissipated on following days through other media.

[12] See Stoll (1978) for a basic analysis of the supply of market making services. In a paper focusing on manipulation by trading, Khwaja and Mihan (2005) show how market makers in an illiquid emerging market without efficient regulation exploit investors by colluding.

Table 4
This table provides results for the base case, with the dependent variable TO. p-values are given for two-sided tests with period SUR corrections (one-sided for β_l, $l \in \{p, s, f\}$). The F-value is from a comparison of R^2 with/without spam dummies, ** indicates statistical significance at the 1% level

Coeff.	Est.	p-val.
α	0.112	0.000
β_{AR_1}	0.402	0.000
β_{AR_2}	0.150	0.000
β_{AR_3}	0.094	0.000
β_{AR_4}	0.088	0.000
β_{tue}	0.011	0.134
β_{wed}	−0.003	0.709
β_{thu}	−0.006	0.458
β_{fri}	−0.015	0.035
β_p	0.253	0.000
β_s	0.446	0.000
β_f	0.056	0.000
R^2	0.078	
n	28469	
F	8.84**	

could indicate that spammers choose high-turnover days to build up their positions because high turnover makes it easier for them to hide their trades. Boehme and Holz (2006) find that this alternative explanation does not hold.[13] Increased volume on following days combined with negative excess returns (cf. Table 2) suits our interpretation that spam victims try to get rid of spammed stocks shortly after a spam event.

All AR coefficients up to lag four are highly significant. Strong positive autocorrelation indicates a clustering of high- and low-turnover days. Fridays show statistically significantly lower turnover than other days. However, the magnitude of the coefficient is quite small, indicating limited economic significance. We test whether the inclusion of our spam dummies leads to a significant increase in R^2 compared to the same regression without spam dummies. The resulting F-value is highly significant (bottom line of Table 4).

Boehme and Holz (2006) also find a positive dependence between turnover and spam events, but they do not consider effects on preceding or following days. Frieder and Zittrain (2006) report a significantly positive impact of spam e-mails on volume on spam and following days.

5.1.3. Intraday range

Table 5 shows the results for the dependent variable IR. β_p and β_s are highly significant with positive signs, indicating larger spreads on spam days and days preceding spam days than on unaffected days. Interestingly, β_p is much larger than β_s. This may be viewed as supporting our story of the spammer building up his position shortly before he initiates the spam. Combined with the insignificant β_p of the return regression (see Table 2), it could also hint at a possible timing strategy of spammers: if professional spammers buy

[13]The dependence is still significant when all spam mails received during business hours are omitted, leaving only those sent before the spammer could have observed this day's volume.

Table 5
This table provides results for the base case, with the dependent variable IR. p-values are given for two-sided tests with period SUR corrections (one-sided for β_l, $l \in \{p, s, f\}$). The F-value is from a comparison of R^2 with/without spam dummies, ** indicates statistical significance at the 1% level

Coeff.	Est.	p-val.
α	0.622	0.000
β_{AR_1}	0.212	0.000
β_{AR_2}	0.079	0.000
β_{AR_3}	0.057	0.000
β_{AR_4}	0.048	0.000
β_{tue}	0.025	0.173
β_{wed}	−0.008	0.630
β_{thu}	−0.018	0.294
β_{fri}	−0.034	0.051
β_p	0.354	0.000
β_s	0.158	0.000
β_f	−0.007	0.387
R^2	0.0983	
n	29782	
F	19.84**	

immediately after a significant decrease in the target's price (when the stocks are very cheap compared to recent price levels), this will lead to large spreads on preceding days but will leave excess returns on these days unaffected.

The dummy for following days is insignificant: increased (intra-day) volatility reverts back to normal fast, showing that the increased uncertainty caused by the spam event vanishes quickly. The F statistic for testing the joint significance of spam dummies is again highly significant, indicating that the inclusion of the spam dummies indeed adds explanatory power.

All AR coefficients are positive and highly significant up to lag 4, with the first lag clearly dominating. Days with large intra-day price changes tend to cluster. The only day-of-the-week effect we find is a lower intra-day range on Fridays (p-value 5.1%).

5.2. Homogeneity of following days

In the base regression, we (implicitly) assume that the five days following a spam event are homogeneous. We check whether this is really the case by running the regression described in Eq. (6). The results are shown in Table 6.

Apart from the reduction in the magnitude of β_{AR_2} to β_{AR_4} discussed in Section 5.1.1, there are no important differences between the return regressions with and without a robustness check (omission of AR_1). In both cases, our initial assumption is confirmed: excess returns on the five days following a spam event are negative and do not differ significantly from each other. In particular, the excess return on the first following day is no different from the others (compare the robustness check discussion in Section 4.4).

Turnover remains at increased levels over the week following a spam event. Here, we note an additional positive effect on the first following day, when the coefficient is about

Table 6
This table provides results for the homogeneity of following days, using the regression from Eq. (6). The top line shows the dependent variables; "robust" refers to the robustness check where AR_1 is omitted, **/* indicate statistical significance at the 1%/5% level (two-sided tests with period SUR corrections, one-sided for β_l, $l \in \{p, s, f\}$)

	RET	RET (robust)	TO	IR
α	−0.023**	−0.023**	0.112**	0.622**
β_{AR_1}	−0.260**		0.401**	0.212**
β_{AR_2}	−0.081**	−0.018**	0.151**	0.079**
β_{AR_3}	−0.051**	−0.035**	0.095**	0.057**
β_{AR_4}	−0.034**	−0.021**	0.088**	0.048**
β_{tue}	−0.001	0.005	0.011	0.024
β_{wed}	0.008*	0.015**	−0.003	−0.008
β_{thu}	0.009**	0.013**	−0.006	−0.018
β_{fri}	0.024**	0.028**	−0.015*	−0.034
β_p	0.006	0.003	0.253**	0.354**
β_s	0.019**	0.015**	0.446**	0.158**
β_f	−0.019*	−0.016*	0.047*	0.000
β_{f1}	0.010	0.001	0.092**	−0.002
β_{f2}	0.008	0.009	−0.024	−0.001
β_{f3}	0.004	0.003	−0.050	−0.017
β_{f4}	0.008	0.008	0.009	−0.017
R^2	0.078	0.013	0.439	0.098
n	28469	28469	55028	29782

three times higher than on the other four following days (0.139 compared to 0.047). IR remains unchanged: none of the following days shows intra-day volatility behavior different from unaffected days. The uncertainty caused by the spam disappears very quickly.

5.3. Successive spam days

In many cases, we observe a clustering of spam events, meaning that a certain stock is being targeted repeatedly on successive trading days. There are two possible motives for spammers to pursue this strategy: if the manipulation attempt on the first day was not successful, the spammer tries desperately to repeat his attempt on the following day. Alternatively, if the manipulation attempt was successful but the spammer could not sell all of his stocks due to liquidity problems, he will now try to liquidate the rest of his position before demand for the stock falls. To investigate this question, we define new dummy variables based on the position within a sequence of spam days and run the regression described in Eq. (7). Table 7 shows the results.

The first and third days in a row show positive excess returns, with β_{s3} more than twice as large as β_{s1}. For several days, the promises made in spam e-mails become self-fulfilling prophecies: as long as new spam victims enter the market, prices keep rising. Apparently, selling pressure from those who bought on the first few days quickly outweighs any demand from new entrants: there are no excess returns if spamming is maintained for more than three days.

Table 7
This table reports results for the regression from Eq. (7) (successive spam days). The top line shows the dependent variables, "robust" refers to the robustness check where AR_1 is omitted, **/* indicate statistical significance at the 1%/5% level (two-sided tests with period SUR corrections, one-sided for β_l, $l \in \{p, s\cdot, f\}$)

	RET	RET (robust)	TO	IR
α	−0.023**	−0.023**	0.111**	0.621**
β_{AR_1}	−0.260**		0.403**	0.212**
β_{AR_2}	−0.081**	−0.018**	0.151**	0.079**
β_{AR_3}	−0.051**	−0.035**	0.095**	0.057**
β_{AR_4}	−0.034**	−0.021**	0.088**	0.048**
β_{tue}	−0.001	0.005	0.011	0.026
β_{wed}	0.008*	0.015**	−0.002	−0.007
β_{thu}	0.009**	0.013**	−0.005	−0.018
β_{fri}	0.024**	0.028**	−0.015*	−0.034
β_p	0.006	0.003	0.252**	0.355**
β_{s1}	0.019**	0.018**	0.485**	0.204**
β_{s2}	0.022*	0.013	0.475**	0.130*
β_{s3}	0.045**	0.039**	0.374**	0.118
β_{s4}	−0.002	−0.009	0.337**	0.221*
β_{s5}	−0.009	−0.002	0.403**	0.268
β_f	−0.013**	−0.012**	0.054**	−0.006
R^2	0.078	0.013	0.439	0.099
n	28469	28469	55028	29782

Turnover is significantly higher on all spam days, regardless of whether or not they are successive. The turnover effect is also significantly positive for preceding and following days. The effect is much larger in magnitude for spam days than for preceding or following days. Although returns no longer increase after the third successive spam day, we conclude that successive spamming has the desired effect of sustaining excess demand for the target stock. This effectively gives the spammer more time to liquidate his position. For IR, we find positive coefficients for all spam days, although only three of them are statistically significant.

A closely related question is whether market participants learn from previous spams, i.e., whether observed effects on the dependent variables decrease with the number of times a particular stock is targeted (even if the spam events do not occur on successive days). Unfortunately, this question is very difficult to investigate, at least based on the information we have. Two requirements that we consider crucial to the feasibility of such an analysis are not met: first, we do not know how many investors actually received a particular spam e-mail (whether it was sent to a comparatively large or small group of recipients). An investor who receives only, say, every fifth spam e-mail about a certain stock cannot be expected to learn from experience without input from the four spams he did not receive. Second, we would need information about the absolute number of spam e-mails that previously targeted each stock in our sample. Recording a spam e-mail in January 2005 does not tell us anything about the number of spams prior to that time (in 2004, 2003,…). Therefore, our data does not provide an answer to this interesting question.

Boehme and Holz (2006) investigate whether the "stock spam trick is wearing out over time" by conducting the analysis described above for the effect on trading volume across all stocks, independent of how many spams there were about the stocks prior to the sample period. Put differently, they suggest that recipients should learn with every spam mail they receive, independent of the particular target stock. In their search for an answer to this somewhat modified question, they find no evidence of decreasing effects of spam mails. This supports our suggestions that additional spam e-mails find new victims.

5.4. Day-of-the-week dependence of spam effects

To test for a possible day-of-the-week dependence of spam effects, we run the regression described in Eq. (8). The results are given in Table 8. We find that excess returns caused by spam e-mails are significantly lower for spams initiated on Fridays, although (or perhaps because?) Fridays without spams show returns which are significantly higher compared to all other days of the week. Spammers seem to be aware of this: Fig. 2 shows that there is much less spam activity on Fridays than in the first half of the week.

Spams attributed to Mondays (β_s, which includes e-mails sent out during the weekend) result in a marked increase in turnover, whereas the effect of e-mails sent from Tuesday to Friday, while still present, is significantly smaller. One explanation could be that recipients of spam e-mails tend to ignore them during the week while they are busy working, but may have more time to read them (or generally tend to their investments) on weekends. The intra-day range does not show any weekday dependence.

Table 8
This table reports results for the regression from Eq. (8) (spam effects depending on weekday). "Robust" refers to the robustness check where AR_1 is omitted, **/* indicate statistical significance at the 1%/5% level (two-sided tests with period SUR corrections, one-sided for β_l, $l \in \{p, s, f\}$)

	RET	RET (robust)	TO	IR
α	−0.023**	−0.024**	0.108**	0.619**
β_{AR_1}	−0.260**		0.403**	0.212**
β_{AR_2}	−0.081**	−0.018**	0.151**	0.079**
β_{AR_3}	−0.051**	−0.035**	0.094**	0.057**
β_{AR_4}	−0.034**	−0.021**	0.088**	0.048**
β_{tue}	−0.001	0.005	0.015*	0.031
β_{wed}	0.008*	0.015**	0.003	−0.005
β_{thu}	0.010**	0.014**	−0.000	−0.016
β_{fri}	0.026**	0.030**	−0.010	−0.031
β_p	0.006	0.003	0.253**	0.354**
β_s	0.027**	0.028**	0.581**	0.231**
β_f	−0.013**	−0.012**	0.056**	−0.007
$\beta_{tue,s}$	0.003	−0.005	−0.125**	−0.153
$\beta_{wed,s}$	0.009	0.004	−0.193**	−0.082
$\beta_{thu,s}$	−0.013	−0.021	−0.195**	−0.045
$\beta_{fri,s}$	−0.050**	−0.057**	−0.209**	−0.091
R^2	0.078	0.013	0.440	0.098
n	28469	28469	55028	29782

5.5. Dependence on average turnover

Next, we examine a possible dependence of our results on average daily turnover. Turnover can be regarded as a rough proxy for liquidity. Table 9 shows the results for the subsamples of high-, medium- and low-turnover stocks, respectively, using \$20.000 and \$45.000 as boundaries.

As far as RET is concerned, spam day dummies are significant for medium- and low-turnover stocks, but not for high-turnover stocks. The dummy coefficient is largest for

Table 9
This table reports results on the dependence of spam effects on average daily turnover. The top line shows the dependent variables ("robust" refers to the robustness check where AR_1 is omitted); the second line indicates the respective subsample, **/* indicate statistical significance at the 1%/5% level (two-sided tests with period SUR corrections, one-sided for β_l, $l \in \{p, s, f\}$)

	RET			RET (robust)		
	High TO	Med. TO	Low TO	High TO	Med. TO	Low TO
α	−0.012**	−0.025**	−0.032**	−0.012**	−0.026**	−0.032**
β_{AR_1}	−0.117**	−0.199**	−0.346**			
β_{AR_2}	−0.038**	−0.076**	−0.117**	−0.026**	−0.038**	−0.006
β_{AR_3}	−0.028**	−0.066**	−0.061**	−0.024*	−0.055**	−0.028**
β_{AR_4}	−0.008	−0.046**	−0.041**	−0.006	−0.038**	−0.018
β_{tue}	0.000	0.003	−0.005	0.001	0.008	0.005
β_{wed}	0.004	0.014**	0.009	0.005	0.019**	0.021**
β_{thu}	0.001	0.016**	0.012	0.003	0.019**	0.018**
β_{fri}	0.012**	0.030**	0.032**	0.013**	0.032**	0.039**
β_p	−0.006	0.025	0.005	−0.008	0.026	−0.005
β_s	0.003	0.026**	0.032**	0.004	0.020*	0.022*
β_f	0.001	−0.016**	−0.023**	0.000	−0.016**	−0.018**
$R2$	0.0201	0.0562	0.1255	0.0062	0.0183	0.0144
n	10016	8401	10052	10016	8401	10052

	TO			IR		
	High TO	Med. TO	Low TO	High TO	Med. TO	Low TO
α	0.087**	0.101**	0.128**	0.608**	0.543**	0.689**
β_{AR_1}	0.574**	0.415**	0.351**	0.211**	0.247**	0.192**
β_{AR_2}	0.108**	0.149**	0.149**	0.100**	0.086**	0.059**
β_{AR_3}	0.061**	0.095**	0.096**	0.055**	0.064**	0.053**
β_{AR_4}	0.091**	0.098**	0.080**	0.043**	0.075**	0.034**
β_{tue}	0.004	0.012	0.012	−0.011	0.024	0.057
β_{wed}	−0.009	0.007	−0.007	0.020	−0.014	−0.032
β_{thu}	−0.015	0.007	−0.009	−0.008	0.010	−0.052
β_{fri}	−0.038**	−0.000	−0.013	−0.061*	−0.025	−0.018
β_p	0.081*	0.202**	0.358**	0.420**	0.087	0.512**
β_s	0.246**	0.416**	0.599**	0.119*	0.124*	0.225**
β_f	0.031*	0.014	0.098**	−0.016	−0.053	0.037
R^2	0.6163	0.4668	0.3701	0.1000	0.1265	0.0851
n	13066	13225	28737	10169	8717	10896

low-turnover stocks, and a bit smaller for medium-turnover stocks. On the following day, a significant counter-reaction can only be found for medium- and low-turnover stocks, pointing to liquidity as the most likely cause. As already noted in Section 5.1.1, this counter-reaction exceeds the initial positive impact caused by the spam.

One explanation for the absence of spam-related effects for high-turnover stocks could be that the larger a stock's turnover, the more of the spam's return impact is digested the same day (i.e., wiped out by the counter-reaction which sets in earlier for high-turnover stocks). This would depress return effects visible in the daily data on both the spam day and the following day, while the same process might take longer for low turnover (low liquidity) stocks. Alternatively, high-turnover stocks generally react less to stock spam because it would take more victims to move their prices noticeably. Our results for the intra-day range seem to favor the latter explanation: if both the positive spam effect and the counter-reaction set in on the same day, this should increase the intra-day range on spam days for high-turnover stocks more than for the other subsamples. In contrast, the spam day dummy coefficient of the IR regression is smallest for high-turnover stocks. This indicates that high-turnover stocks are less suitable targets for spammers.

The "Friday effect" that occurs independently of spam events also depends on turnover and is about three times larger for low-turnover stocks than for their high-turnover counterparts. Different alphas for our subsamples indicate that average returns differ depending on turnover. Returns for less liquid stocks were generally much lower in 2005.[14]

For turnover, all dummies for spam days and previous days are positive and significant, while following days are only significant for high- and low-turnover stocks. Coefficients are largest in magnitude for low-turnover stocks. This is again most likely a liquidity effect.

5.6. *Price level of target stock*

Given that almost all stocks in our sample trade *very* cheaply, we want to find out whether *cheaper* means *better for the spammer*, in the sense that cheaper stocks react more strongly to spam e-mails. To this end, we split our sample into two parts, using ten cents as the boundary. The results of the regression described in Eq. (1) applied to these subsamples are shown in Table 10.

For returns, large differences in α hint at large differences in average returns between the subsamples. The return impact on spam days is larger for low-priced stocks, and so is the counter-reaction on following days: the positive impact of 0.011 for stocks trading above ten cents is followed by a negative effect of 0.016 ($1.78 \cdot 0.009$). This shows less over-reaction than low-priced stocks with a positive spam effect of 0.016, followed by a counter-reaction of 0.034. Given that stocks trading very cheaply are on average less liquid, these numbers support our explanation that the over-reaction we observe on following days is mainly driven by a lack of liquidity.

Turnover is markedly higher for higher-priced stocks, independent of spam events. This indicates generally higher liquidity for stocks trading at higher price levels. The impact on spam day turnover is significantly higher for low-priced stocks, which is again consistent with our interpretation that the effects are mainly driven by liquidity. Higher (relative) bid-ask spreads for stocks trading very cheaply amplify this effect. Compared to the high-price

[14]Of course, our time frame is too short to generalize these results. According to standard finance theory, we would expect higher returns for stocks with lower liquidity.

Table 10
This table reports results on the dependence of spam effects on the price level of the target. The top line shows the dependent variables ("robust" refers to the robustness check where AR_1 is omitted); the second line indicates the respective subsample, **/* indicate statistical significance at the 1%/5% level (two-sided tests with period SUR corrections, one-sided for β_l, $l \in \{p,s,f\}$)

	RET		RET (robust)		TO		IR	
	High price	Low price	High price	Low price	High price	Low price	High price	Low price
α	−0.009**	−0.041**	−0.009**	−0.041**	0.274**	0.196**	0.569**	0.722**
β_{AR_1}	−0.211**	−0.306**			0.403**	0.369**	0.219**	0.191**
β_{AR_2}	−0.064**	−0.107**	−0.025**	−0.020*	0.120**	0.127**	0.099**	0.049**
β_{AR_3}	−0.066**	−0.053**	−0.055**	−0.029**	0.066**	0.087**	0.063**	0.042**
β_{AR_4}	−0.022**	−0.047**	−0.010	−0.032**	0.057**	0.078**	0.047**	0.039**
β_{tue}	−0.004	0.003	−0.002	0.013	0.011	0.007	0.050*	−0.005
β_{wed}	0.004	0.014*	0.007*	0.025**	−0.007	−0.007	0.026	−0.050
β_{thu}	0.006*	0.013*	0.008*	0.020**	−0.002	−0.019	0.001	−0.041
β_{fri}	0.013**	0.038**	0.014**	0.046**	−0.025	−0.024*	−0.045*	−0.023
β_p	0.007	−0.001	0.006	−0.007	0.215**	0.230**	0.320**	0.444**
β_s	0.014**	0.021*	0.011*	0.016*	0.359**	0.470**	0.098**	0.301**
β_f	−0.009*	−0.023**	−0.009**	−0.019**	0.001	0.054**	−0.018	0.054
R^2	0.0649	0.1101	0.0209	0.0235	0.4199	0.4210	0.1279	0.0861
n	16106	12363	16106	12363	22609	18547	16837	12945

group, the intraday volatility of low-priced stocks reacts much more strongly on spam days. The negative image of these "penny stocks" is supported by our data: Low turnover, high volatility and large negative average returns raise the question why any rational investor would want to invest in these stocks.

6. Conclusions and areas for further research

Using a large sample of spam e-mails and corresponding market data, we find significant effects of stock spam on various market statistics (returns, turnover, intra-day price range). Our economic story, formulated from casual observation and basic economic reasoning, is supported by the data: spammers build up their positions shortly before sending out their e-mails. On spam days, we observe positive abnormal excess returns at high volumes and volatilities, but only for low- and medium-turnover (less liquid) stocks. In the week following a spam event, excess returns are negative, more than offsetting the price increase observed on the spam day. This shows that in contrast to information spread via discussion forums on the web, the "information" contained in spam e-mails is completely false. The magnitude of the observed over-reaction is larger for stocks trading very cheaply at low to medium turnover. The best explanation we can provide for this effect is the liquidity problems spam victims face when they realize that prices are not rising as promised and try to get out of their positions quickly. All the return effects we find are robust to possible bias from bid-ask bounce.

Uncertainty, proxied by the intra-day range, reverts back to normal very quickly, while turnover remains high for a number of days following a spam event. Sequential spamming on successive days does have a sustaining effect on both turnover and excess returns and should allow the spammer to liquidate larger positions even in illiquid stocks over a three- to five-day period. Excess returns on spam days are highest for low to medium turnover stocks trading below ten cents. Interestingly, we also find a significant day-of-the-week effect even in the absence of any spam activity: returns of target stocks are higher on Fridays (the lower the turnover, the more pronounced the effect), while volatility—proxied by the intra-day price range—is lower on Fridays. Spamming on Fridays has much smaller effects than spamming on other days of the week.

Open-high-low-close price data (as available in our database) are insufficient to analyze the success of trading strategies based on spam e-mails. This would require intra-day transactions data and bid-ask quotes instead of prices. The same holds for an analysis of possible trading gains of other investors trying to "ride the wave." Bid-ask quotes would also allow us to eliminate any bid-ask bounce a priori. As soon as we obtain intra-day quotes for the stocks in our sample, we will further investigate the economic significance of our results by exploring the success of various trading strategies based on stock spam e-mails.

References

Aggarwal, R., Wu, G., 2006. Stock market manipulation. The Journal of Business 79, 1915–1953.

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. The Journal of Finance 57, 1047–1091.

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

Allen, F., Gorton, G., 1992. Stock-price manipulation, market microstructure and asymmetric information. European Economic Review 36, 624–630.

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

Beck, N., Katz, J.N., 1995. What to do (and not to do) with time-series cross-section data. American Political Science Review 89, 634–647.

Boehme, R., Holz, T., 2006. The effect of stock spam on financial markets, 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.

Cherian, J.A., Jarrow, R., 1995. Market manipulation. In: Jarrow, R.A., Maksimovic, V., Ziemba, W.T. (Eds.), North-Holland Handbooks in Operations Research and Management Science, vol. 9. Elsevier, Amsterdam, pp. 611–630 (Chapter 9).

Dewally, M., 2003. Internet investment advice: investing with a rock of salt. Financial Analysts Journal 59, 65–77.

Frieder, L.L., Zittrain, J.L., 2006. Spam works: evidence from stock touts and corresponding market activity. SSRN Working Paper No. 920553.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. The Journal of Finance 45, 881–898.

Jiang, G., Mahoney, P.G., Mei, J., 2005. Market manipulation: a comprehensive study of stock pools. Journal of Financial Economics 77, 147–170.

Khwaja, A., Mihan, A., 2005. Unchecked intermediaries: price manipulation in an emerging stock market. Journal of Financial Economics 78, 203–241.

Larson, S., Luft, C., Levine, L.M., 2001. Over the counter bulletin board exchange: market structure, risk, and return. Journal of Alternative Investments 4, 33–42.

Lease, R.C., Mahulis, R.W., Page, J.R., 1991. An investigation of market microstructure impacts on event study returns. The Journal of Finance 46, 1523–1536.

Luft, C., Levine, L.M., 2004. Over the counter bulletin board exchange: the impact of liquidity and size to return, volatility, and bid/ask spread. Journal of Alternative Investments 7, 95–106.

Mei, J., Wu, G., Zhou, C., 2004. Behavior based manipulation: theory and prosecution evidence. SSRN Working Paper No. 457880.

Ritter, J.R., Welch, I., 2002. A review of IPO activity, pricing, and allocations. Journal of Finance 57, 1795–1828.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. The Journal of Finance 39, 1127–1139.

Stoll, H.R., 1978. The supply of dealer services in securities markets. The Journal of Finance 33, 1133–1151.

Tumarkin, R., Whitelaw, R.F., 2001. News or noise? Internet postings and stock prices. Financial Analysts Journal 57, 41–51.

van Bommel, J., 2003. Rumors. The Journal of Finance 58, 1499–1519.

Wall Street Journal, 2000. Teenage trader runs afoul of the SEC as stock touting draws charges of fraud. 21.09.2000.

Wall Street Journal, 2005a. SEC steps up effort to fight stock fraud; agency begins to halt trading when it suspects spammers are touting shares illegally. 02.02.2005.

Wall Street Journal, 2005b. Scam artists tap text messaging to lure folks seeking hot stocks. 27.09.2005.

Wooldridge, J.M., 2006. Introductory Econometrics—A Modern Approach, third ed. Thomson.