

# On the Effects of Stock Spam E-mails\*

Michael Hanke,<sup>†</sup> Florian Hauser<sup>‡</sup>

August 30, 2006

## Abstract

A rising number of unsolicited e-mails recommends buying certain stocks, pretending that the sender has private information that will boost these stocks' prices when it becomes publicly available. We first describe the common characteristics of stocks pushed by such e-mails. Then, we investigate the effect of stock spam e-mails on returns, volatility, intraday spread, and volume. We find a significant impact of spam mails on all of these variables. As a second contribution, we characterize features of stocks that are particularly easy to manipulate, and we investigate dependencies between these features.

JEL classification: D82, G14, G24

Keywords: stock spam, market manipulation

## 1 Introduction

Almost everyone with an e-mail account has already received e-mails recommending to buy certain stocks. The principal purpose of such mails obviously is to manipulate the prices of the stocks they push. The senders of such e-mails spread forged “news” about the stocks, trying to artificially increase demand, which in turn leads to price increases. Activity of this kind has recently received increasing attention by the media (Wall Street Journal (2000, 2005a,b)).

Manipulation of stock markets is a phenomenon that has been around for a long time, possibly for as long as stock markets. There is a host of analytical literature on this topic (e.g., van Bommel (2003), Jianga et al. (2005)). In earlier days, manipulators had to work much more subtly. Manipulation attempts camouflaged as analyst's reports were quite common. In contrast to their modern counterparts initiated via the internet, their success was highly dependent on the analysts' reputation, which was put on the line by repeatedly issuing “cooked” reports. Precise timing was much more difficult, if not impossible with traditional media such as newspapers which are read by some people in the morning, by others in the evening, and might be accessed by a few “insiders” already the day before they go to press.

---

\*We thank participants of the 20th workshop of the Austrian Working Group on Banking and Finance (Graz 2005) for helpful comments and discussions.

<sup>†</sup>Innsbruck University School of Management, Department of Banking and Finance, Universitaetsstrasse 15, 6020 Innsbruck, Austria. e-mail: michael.hanke@uibk.ac.at

<sup>‡</sup>Corresponding author. Innsbruck University School of Management, Department of Banking and Finance, Universitaetsstrasse 15, 6020 Innsbruck, Austria. e-mail: florian.hauser@uibk.ac.at

For would-be manipulators, the internet offers the ideal framework, allowing simultaneous contacts to large numbers of potential investors. Apart from e-mail, electronic newsletters and internet discussion forums are popular tools used by manipulators. While electronic newsletters can be viewed as just the good old print newsletters transferred to a new medium, discussion forums in the internet are a new framework for exchanging information on stocks. There are already several studies confirming 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)).

Our goal in this paper is to explore the effects of stock spam e-mails that can be found in market data. Using panel regression, we investigate their effects on return, volatility, volume, and intraday spread measures of the target stocks. Our results indicate that spam e-mails show marked effects on all of these variables. Some limitations of our study are caused by the unavailability of intra-day data. These would have been necessary to provide a clearer picture of potential profits of trading strategies. Nevertheless, we can learn a lot from the daily data we use here.

When we started this project, to the best of our knowledge, there had been no empirical study yet on stock spam e-mails. Only very recently, we have become aware of a competing paper (Boehme and Holz, 2006). Although they retrieve their data from the same database, there are several major differences between their paper and ours (more details on these differences will be provided in the corresponding sections below):

- Our sample is much more comprehensive,<sup>1</sup> covering more stocks (235 compared to 93) and more spam events (1241 compared to 526 (volume) and 152 (returns) in Boehme and Holz (2006)).
- Our market data come from Datastream, whereas Boehme and Holz (2006) use data from Yahoo Finance. We strongly believe that our data are more reliable.
- We use panel regression instead of pooled OLS.
- Whereas Boehme and Holz (2006) focus only on returns and volume, we also describe effects on volatility and intraday spread.
- Our study provides a more comprehensive analysis of the dependencies between spam success and characteristics of target stocks.

The paper is organized as follows: Section 2 provides additional information about stock spam e-mails and discusses possible motivations and trading strategies followed by spammers. Moreover, we formulate a number of interesting questions that will be answered through 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.

---

<sup>1</sup>This is despite the fact that we use only data from 2005, whereas Boehme and Holz (2006) cover the period from November 2004 to February 2006.

## 2 Stock Spam E-mails

We define stock spam e-mails as unsolicited e-mails that

- explicitly recommend to buy a certain stock (the *target*) and/or
- contain “information” about events that investors will most certainly interpret as a buy recommendation.

Stock spam e-mails represent one of the contemporary variants of strategies known under the names “pump-and-dump” or “scalping”. 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 the technological possibilities brought about by the internet, the execution of such strategies was fraught with problems. The strategies only work well if the forged information can be spread quite quickly among a larger audience. Ideally, all recipients should get the information at the same time. The fact that such practices are illegal in many jurisdictions leads to a strong desire of the manipulator to remain anonymous.

Among all e-mails pretending to contain information relevant for a stock’s price that we have seen, there was not a single one with either negative information or an outright sell recommendation. Another common feature of stock spam e-mails is that the information provided is incorrect, be it because it is outdated, exaggerated, or – in most cases – purely fictitious. The only purpose of such e-mails seems to be the manipulation of the target’s stock price. Assuming this is the case, the question naturally arises who benefits from spam e-mails, and therefore has an incentive to initiate them. A related question is what strategies stock spammers use in order to benefit financially from their activities.

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 OTC-BB or Pink Sheets<sup>2</sup> 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 smaller,
- traded quite irregularly, with frequent gaps in daily price series,
- 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 main possible motives for spamming a stock:

---

<sup>2</sup>These platforms will be described in more detail in Section 3.2.

1. The spammer intends to sell a certain stock he already owns and wants to boost the selling price.
2. The spammer buys the target stock, initiates the spam shortly 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, initiating stock spam e-mails on a regular basis. The negative average return of target stocks (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.

Putting ourselves in the position of a spammer of the latter type, many of the characteristics of popular target stocks described above make good sense. It certainly is much easier to manipulate the price of a stock with low turnover. Small movements in absolute terms translate into large percentage changes for so-called *penny stocks*, which makes cheap stocks attractive targets. Given that stocks with these characteristics usually cannot be shorted,<sup>3</sup> 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:

- When stocks are too thinly traded, the spammers might have too large a price impact. He drives the price upwards when building up his position, and drives it downwards again when liquidating the position, thus reducing his returns. Is high or low turnover better from the spammers' point of view in the sense that stocks within a certain turnover (liquidity) range are more desirable targets than others?
- The spammers' success depends crucially on other investors' willingness to invest in their target stocks. Given that stocks trading at very low stock prices usually carry a negative image, is the spammers' rule really "the cheaper, the better"?
- Certain stocks are frequently targeted during our observation period, while others receive only one or two spam mails during the year 2005. A possible reason for this is that stocks that are easier to manipulate are targeted more often. Do stocks that are targeted frequently react differently to such manipulation attempts?
- At times, we observe stock spam e-mails targeting the same stocks on successive days. Is sequential spamming more effective than one-shot spams? Do the other effects observed differ between one-off spams and sequential spamming?
- The spammers' success depends crucially on other investors' willingness to invest in their target stocks. Given that stocks trading at very low stock prices usually carry a negative image, is the spammers' rule really "the cheaper, the better"?

---

<sup>3</sup>At least, not at reasonable transaction costs.

### 3 Data

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 occurred before the closing of business. Spam events are retrieved from a publicly available database.<sup>4</sup> The database allows an exact matching of spam events to trading days. Daily stock prices were 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 more detail. The author of the database maintains a large number of so-called *trap accounts* – e-mail accounts whose sole purpose is the reception of spam e-mails. Every day in the afternoon, when trading in the U.S. closes, automatic scripts evaluate the e-mails received for all trap accounts, classify the subset of stock spam e-mails according to the target stock, and time-stamp them. We checked a large sample of e-mails from the database for possible errors in the automated classification scripts. About 20 ticker symbols that turned out to be incorrectly classified in the database have been excluded from our sample. We should point out that there is absolutely no way to ensure that the database is exhaustive. There may (and will!) be stock spam e-mails during 2005 which have not been recorded in the database. We strongly believe, however, that this in no way casts doubt on the validity of our results. On the contrary, more spam events in our sample should rather improve the statistical significance of our results.

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

- Availability of daily 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.
- Trading at a U.S. marketplace (ensuring adequate matching of spam events to trading days).

#### 3.1 Descriptive Statistics

Out of about 360 stocks in the database, 235 matched our criteria. For these stocks included in the sample, we have 1241 spam events in 2005. Figure 1 shows that the stocks are affected quite differently by spam e-mails: While 130 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 of our study, an analysis of the number of spams associated with certain trading days provides a good indicator for the spam activity over time. Figure 2 shows that the spam events in our sample are quite evenly distributed over time with a maximum of 12 spam events for a single trading day. We found no evidence of calendar effects.

A preliminary analysis of our data confirms the characterization of target stocks provided above:

---

<sup>4</sup>The Crummy database, <http://www.crummy.com/features/StockSpam/reports/>

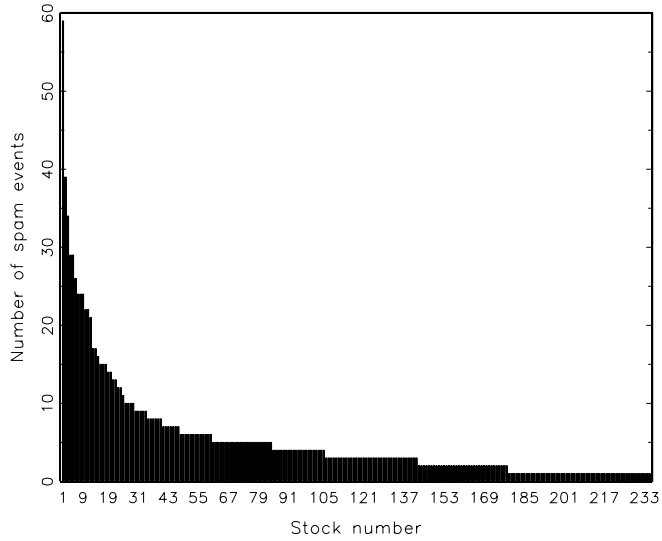


Figure 1: Number of spam events per stock (stocks sorted by spam event frequency)

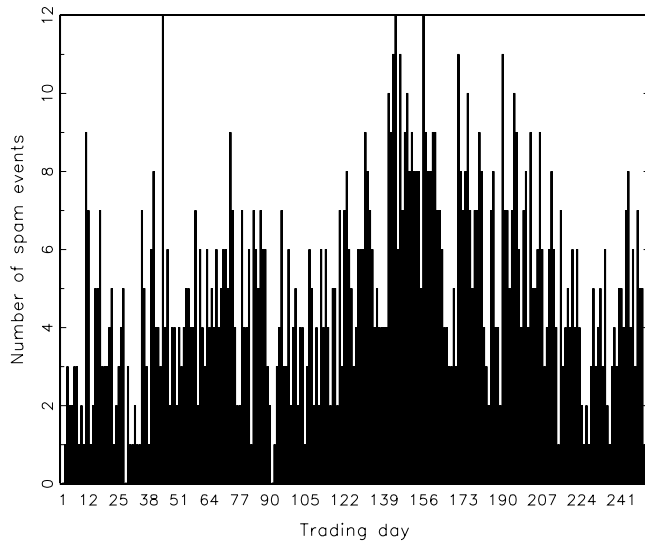


Figure 2: Number of spam events per trading day

- Trading in those stocks was quite illiquid: For 59220 stock trading days in total (252 trading days per year times 235 stocks), we recorded only 41764 quotes.
- Penny stocks are the predominant targets: 36309 of the 41764 unadjusted quotes were below 1\$. Figure 3 gives detailed information about the dis-

tribution of quotes in our sample.

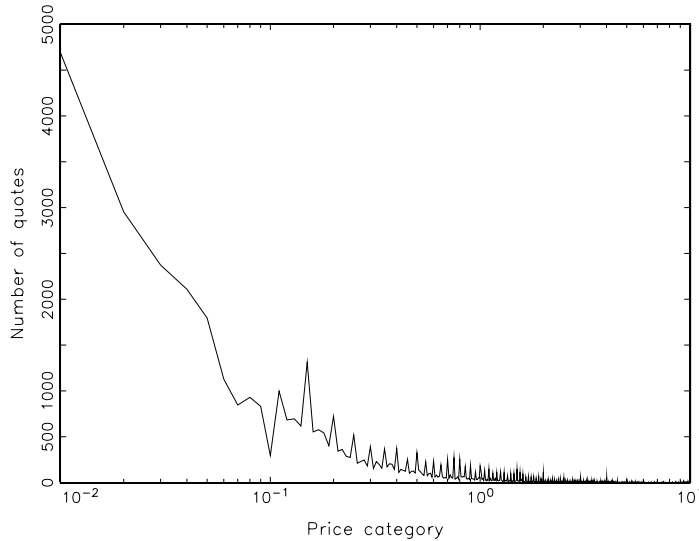


Figure 3: Number of quotes in categories (0.01\$ steps)

After several warnings issued by the SEC<sup>5</sup>, investors should already be aware that investing in penny stocks is a quite risky business. Nevertheless it is remarkable that even a portfolio of 235 stocks targeted by spam mails in 2005 does not provide protection against suffering huge losses. Figure 4 shows the value of a hypothetical portfolio, starting at 100 on January 2 and growing at compounded average daily returns of the stocks in our sample (this can roughly be thought of the performance of a naive-diversification-strategy). The portfolio loses almost 90% over one year. During the same period, the S&P 500 gained about 3.8%.

Figure 5 shows that daily returns of the stocks in our sample are leptokurtic. The red 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 returns on spam days (blue) than for those on unaffected days (for the exact definition of unaffected days see Section 4).

### 3.2 Electronic Quotation Platforms for OTC-traded Stocks

Most of the stocks in our sample are not traded on organized exchanges, but quoted on platforms like Pink Sheets and OTCBB (Over-the-Counter Bulletin Board). In contrast to many stock exchanges, OTC markets are dealer markets, meaning that securities are traded by market makers, providing bid- and ask prices for those stocks. To inform potential traders about current quotes, such platforms process quotes of eligible market makers who ensure the liquidity for the listed stocks. Companies quoted on those platforms usually cannot

<sup>5</sup><http://www.sec.gov/investor/pubs/microcapstock.htm>

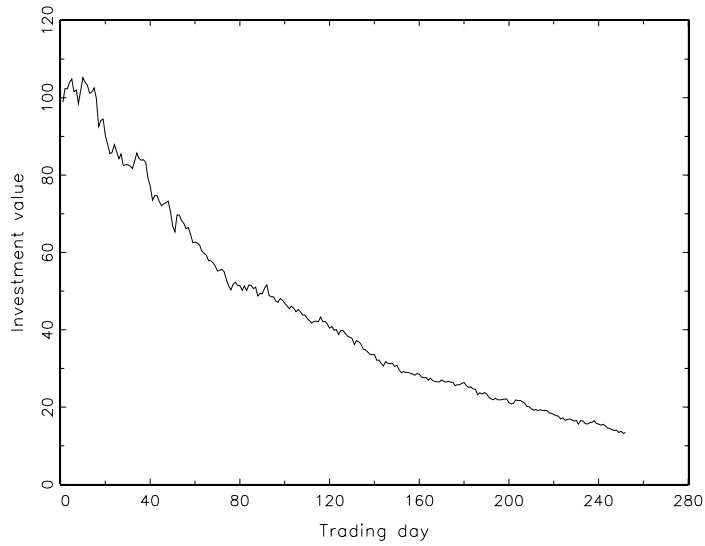


Figure 4: Value of a hypothetical portfolio growing at compounded average daily returns of all stocks in our sample

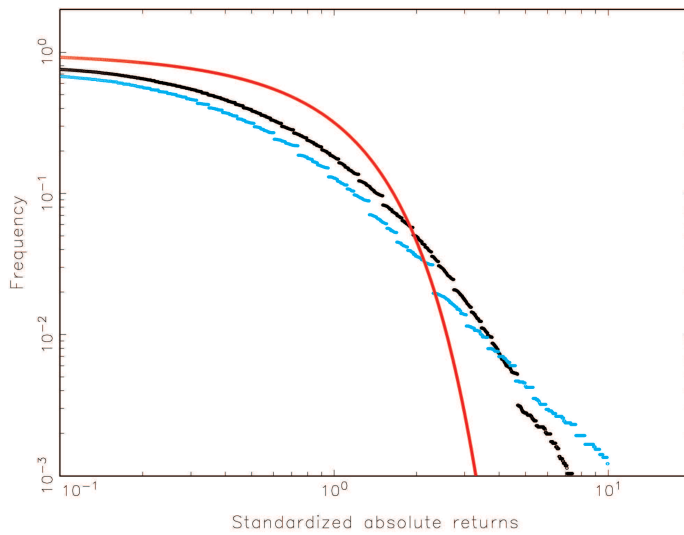


Figure 5: Cumulated frequency of standardized absolute returns of the stocks in our sample. Blue line: returns on spam days, black line: returns on unaffected days, red line: normal distribution.

meet the requirements of reputable exchanges like NYSE or NASDAQ. Listing requirements on OTC platforms are very lenient: There are no minimum financial standards, and in some cases not even any action on the part of the issuing



company itself is necessary for a stock to be listed on those platforms. A minor difference in “regulation” is the requirement for companies quoted on OTCBB to publish periodic financial reports (this is not necessary for stocks quoted at 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 et al. (2001); Luft and Levine (2004).

|                                       | 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 Volume (bill.)                 | 4.2   | 8.7         |
| Exclusive Securities                  | 3269  | 4788        |

Table 1: Comparison of OTCBB and Pink Sheets for January 2006 (data obtained from <http://www.otcqx.com/docs/OTCQXBrochure.pdf>)

## 4 Method

### 4.1 Basic Setup

We look for answers to the questions formulated in Section 2 using panel regression with dummy variables. At first sight, our problem seems to lend itself well to the standard event study methodology. However, 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 above in Section 3, each spam mail is time-stamped and easily associated with the corresponding trading day.

We take advantage of our large cross-section to mitigate the problem of observing only a small number of spam days for many of the stocks in our sample by using panel regression with dummies. We have both a large cross-section and a large number of observations in time, so many econometric problems discussed in the panel regression literature which are due to a small number of observations in time are not a concern for us.

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 allow for unobserved effects in only one of the domains, but not both. A preliminary analysis allowing for unobserved time effects showed that there are no such 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)). Trying to remove any correlation from the residuals, we find that including AR-terms up to order four yields coefficients that are both economically and statistically significant. Sometimes, AR-coefficients of higher order are statistically significant (due to the size of our sample), but very small (close to zero).

In the basic model, we include different dummies for the days preceding a spam day ( $D_{i,p,t}$ ,  $i = 1, \dots, 235$ : index for the stocks in our sample), spam days ( $D_{i,s,t}$ ), and the days following a spam day ( $D_{i,f,t}$ ). A further condition we imposed for a day to qualify as a preceding day was that at least two non-spam-event days must lie between this day and the most recent following day. All days that are neither spam days, nor within a one-day time interval around a spam day, will be called *unaffected days* (returns on such days will be referred to as *unaffected returns*).

Thus, the basic regression equation reads as follows:

$$\ddot{y}_{i,t} = \sum_{j=1}^4 \beta_{AR_j} \ddot{y}_{i,t-j} + \sum_{k \in \{p,s,f\}} \beta_k \ddot{D}_{i,k,t} + \ddot{u}_{i,t}, \quad (1)$$

where  $\ddot{y}_{i,t} = y_{i,t} - \bar{y}_i$ , the time-demeaned data on  $y$  (this is usually called the *fixed-effects transformation*, similarly for  $D$  and  $u$ ). Analyses of residuals from our regressions show no significant autocorrelation, but some (time-) heteroskedasticity within cross-sections. We account for this by using a period SUR specification (Wooldridge, 2002, p. 144) for the covariances.

Boehme and Holz (2006) use a log-linearized version of a multiplicative regression model, estimated by standard pooled OLS. Our approach allows for greater flexibility in variable transformations described in the following subsections.

## 4.2 Dependent Variables

The following variables are used as dependent variables:

- RET, representing returns,
- SIG, a measure of volatility,
- SPR, a measure of intraday volatility, and
- VOL, measuring daily turnover.

Various standardization procedures are applied to the raw data to facilitate a meaningful aggregation across stocks.

To calculate RET, we start from the price series in DATASTREAM, which is already corrected for dividends and stock splits. Denoting these prices by  $S_{i,t}$ , we calculate raw returns  $R_{i,t}$  using

$$R_{i,t} = \ln S_{i,t} - \ln S_{i,t-1}. \quad (2)$$

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. For this reason, we employ a constant mean return correction which is handled implicitly within our model through the fixed-effects transformation (as described in Section 4.1).

SIG is calculated from the daily variance (squared raw return  $R_{i,t}$ ) by standardizing over the average variance on unaffected days and then taking the square root:

$$\text{SIG}_{i,t} = \sqrt{\frac{R_{i,t}^2}{\bar{R}_i^2}}, \quad (3)$$

where  $\overline{R}_i^2$  denotes the arithmetic mean of  $R_{i,t}^2$  for unaffected days.

To generate a measure for intra-day volatility, we start by calculating

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

where  $\text{high}_t(\cdot)$  and  $\text{low}_t(\cdot)$  denote the intra-day high and low of the respective stock on day  $t$  (from DATASTREAM). SPR is then calculated from

$$\text{SPR}_{i,t} = \frac{\varsigma_{i,t}}{\overline{\varsigma}_i}, \quad (5)$$

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

VOL is calculated from the daily turnover by standardizing over the turnover on unaffected days, and then applying a log transformation:

$$\text{VOL}_{i,t} = \ln \left[ 1 + \frac{\text{to}_{i,t}}{\overline{\text{to}}_i} \right], \quad (6)$$

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

### 4.3 Explanatory Variables

In the base case, we use  $D_{i,k,t}$ ,  $k \in \{p, s, f\}$  together with AR( $j$ )-terms ( $j=1, \dots, 4$ ) as independent variables.

For analyzing many of the questions discussed in Section 2, the dummy methodology is not applicable together with the fixed effects specification, because for many cross-sections the dummies would be constant in time. In these cases, we split our sample into sub-samples and repeat the regressions on each of the sub-samples.

To investigate a possible dependence on turnover, we split our sample into two: one contains all stocks with “high” average daily turnover (more than 20.000\$ per day), and the other all stocks with “low” turnover (less than or equal to 20.000\$ per day).

To distinguish between stocks that are frequently targeted by spammers and those that receive only a few spam mails per year, we split our sample according to the number of spam events recorded in 2005. One sub-sample contains all stocks with five or more spam events during the observation period, and the other all stocks with four spam events or fewer.

The question whether repeated spamming on successive days shows greater or smaller effects is investigated by defining separate spam day dummies for the first, second, third, and fourth spam day in a row. These dummies will be denoted by  $D_{i,s,t,1}$ ,  $D_{i,s,t,2}, \dots, D_{i,s,t,4}$ . The regression equation then reads

$$\ddot{y}_{i,t} = \sum_{j=1}^4 \beta_{\text{AR}_j} \ddot{y}_{i,t-j} + \sum_{k \in \{p, f\}} \beta_k \ddot{D}_{i,k,t} + \sum_{l=1}^4 \beta_{s,l} \ddot{D}_{i,s,t,l} + \ddot{u}_{i,t}. \quad (7)$$

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, using a stock price of 10 cents as the boundary.

## 5 Results

In the following subsections, we present and discuss the results for the regressions described in Section 2. In preliminary runs, we also included a coefficient for linear time dependence, which was zero in all the 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 (smaller) sample.

### 5.1 Base Case

#### 5.1.1 Return

The base case panel regression is given by equation (1). Results for the first regression, using RET as the dependent variable, are given in Table 2. As discussed in Section 4,  $\alpha$  denotes the constant implied by the fixed-effects transformation. We find significant negative autocorrelation up to lag 4, which is a

| Coeff.        | Est.   | <i>p</i> -val. |
|---------------|--------|----------------|
| $\alpha$      | -0.015 | 0.000          |
| $\beta_{AR1}$ | -0.260 | 0.000          |
| $\beta_{AR2}$ | -0.082 | 0.000          |
| $\beta_{AR3}$ | -0.053 | 0.000          |
| $\beta_{AR4}$ | -0.035 | 0.000          |
| $\beta_p$     | 0.010  | 0.273          |
| $\beta_s$     | 0.019  | 0.000          |
| $\beta_f$     | -0.008 | 0.235          |
| $R^2$         | 0.0744 |                |
| $n$           | 28469  |                |

Table 2: Results for the base case, dependent variable: RET. *p*-values for two-sided test with period SUR corrections.

common sign of illiquidity (bid-ask-bounce) and therefore to be expected given the characteristics of our stocks. The spam day dummy coefficient is highly significant and positive: On spam days, returns are indeed markedly higher than on unaffected days. In contrast, there is no significant return effect on days preceding or following spam days. The corresponding coefficients are only half the size of the spam day dummy coefficient. Their signs, however, support our economic story.

One problem with the interpretation of our returns is that they are based on closing prices. Unfortunately, we do not have information on intra-day price developments (apart from spreads). For relatively liquid stocks, it might well be the case that prices rise after spreading the spam mail, the spammer takes his profits, and prices are already coming back to “normal” levels before the close of trading on the day of the spam event. In these cases, our results would be affected twofold: Returns on spam days would underestimate the true return impact of the spam mails, and returns on following days would underestimate the market’s correction after the spam event. Therefore, our coefficients should provide rather conservative estimates of the respective return impacts.

### 5.1.2 Volatility

Results for the base case with SIG as the dependent variable are given in Table 3. We find significant positive autocorrelation in SIG up to lag 4 with the first-

| Coeff.        | Est.  | <i>p</i> -val. |
|---------------|-------|----------------|
| $\alpha$      | 0.404 | 0.000          |
| $\beta_{AR1}$ | 0.205 | 0.000          |
| $\beta_{AR2}$ | 0.063 | 0.000          |
| $\beta_{AR3}$ | 0.041 | 0.000          |
| $\beta_{AR4}$ | 0.047 | 0.000          |
| $\beta_p$     | 0.037 | 0.351          |
| $\beta_s$     | 0.082 | 0.000          |
| $\beta_f$     | 0.007 | 0.818          |
| $R^2$         | 0.087 |                |
| <i>n</i>      | 28469 |                |

Table 3: Results for the base case, dependent variable: SIG. *p*-values for two-sided test with period SUR corrections.

order term clearly dominating in magnitude, which points to the well-known phenomenon of volatility clustering. On spam days, SIG is significantly larger than usual.

### 5.1.3 Intraday Spread

Table 4 shows the results for SPR used as the dependent variable. Here, again, autocorrelation is positive and highly significant up to lag 4 with the first lag clearly dominating. Days with large intraday price changes tend to cluster.  $\beta_p$

| Coeff.        | Est.   | <i>p</i> -val. |
|---------------|--------|----------------|
| $\alpha$      | 0.614  | 0.000          |
| $\beta_{AR1}$ | 0.212  | 0.000          |
| $\beta_{AR2}$ | 0.079  | 0.000          |
| $\beta_{AR3}$ | 0.057  | 0.000          |
| $\beta_{AR4}$ | 0.048  | 0.000          |
| $\beta_p$     | 0.279  | 0.000          |
| $\beta_s$     | 0.164  | 0.000          |
| $\beta_f$     | 0.033  | 0.457          |
| $R^2$         | 0.0977 |                |
| <i>n</i>      | 29782  |                |

Table 4: Results for the base case, dependent variable: SPR. *p*-values for two-sided test with period SUR corrections.

and  $\beta_s$  are highly significant with positive sign, indicating larger spreads on spam days and days preceding spam days compared to unaffected days. Interestingly,  $\beta_p$  is about two thirds larger than  $\beta_s$ . This may be viewed as supporting our story of the spammer building up his position shortly before he initiates the spam. Taking this result together with the corresponding coefficient of the volatility regression implies that the spammer’s build-up orders are issued on the day preceding the spam event.

### 5.1.4 Volume

Table 5 shows the results using VOL as the dependent variable. All AR coeffi-

| Coeff.        | Est.  | $p$ -val. |
|---------------|-------|-----------|
| $\alpha$      | 0.110 | 0.000     |
| $\beta_{AR1}$ | 0.400 | 0.000     |
| $\beta_{AR2}$ | 0.151 | 0.000     |
| $\beta_{AR3}$ | 0.095 | 0.000     |
| $\beta_{AR4}$ | 0.089 | 0.000     |
| $\beta_p$     | 0.228 | 0.000     |
| $\beta_s$     | 0.448 | 0.000     |
| $\beta_f$     | 0.165 | 0.000     |
| $R^2$         | 0.439 |           |
| $n$           | 55028 |           |

Table 5: Results for the base case, dependent variable: VOL.  $p$ -values for two-sided test with period SUR corrections.

icients up to lag 4 are highly significant. Strong positive autocorrelation indicates a clustering of high- and low-volume days. The effect on VOL is about twice as large on spam days compared to preceding days and following days. Again, this indicates that the spammers’ activities leave clear traces in market data. Regardless of whether the manipulation attempts really cause prices to move *in the desired direction*, trading volume in spammed stocks is clearly higher on and around spam days. The fact that volume is already up on the preceding day indicates that it is not uncommon for the spammer to build up his position quite shortly before he spreads his e-mails.

Boehme and Holz (2006) also find a positive dependence between volume and spam events, but they do not consider effects on preceding and following days.

## 5.2 Dependence on Turnover

Next, we examine a possible dependence of our results on turnover. Table 6 shows the results for the sub-samples of high-turnover and low-turnover stocks, respectively. As far as RET is concerned, spam day dummies are significant for both high- and low-turnover stocks. The coefficient for low-turnover-stocks, however, is more than twice as large as that for high-turnover stocks. The coefficient for days following spam days is zero for high-turnover-stocks, but significantly negative for low-turnover stocks. A possible explanation is that much of the return impact of the spam is already digested during the spam day for high-turnover stocks, depressing 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.

For SIG, we find an interesting result. Volatility is only significantly increased on spam days for high-turnover stocks, while the corresponding coefficient for low-turnover stocks shows a  $p$ -value of around 0.11 (with the expected positive sign).

For the intraday spread, the same coefficients are significant as in the base case (spam day and preceding day). The two groups differ, however, as far as the magnitude of the effect is concerned: The total effect of  $\beta_p$  (calculated as

| Coeff.        | RET           |                |              |                | SIG           |                |              |                |
|---------------|---------------|----------------|--------------|----------------|---------------|----------------|--------------|----------------|
|               | high turnover |                | low turnover |                | high turnover |                | low turnover |                |
|               | Est.          | <i>p</i> -val. | Est.         | <i>p</i> -val. | Est.          | <i>p</i> -val. | Est.         | <i>p</i> -val. |
| $\alpha$      | -0.011        | 0.000          | -0.023       | 0.000          | 0.413         | 0.000          | 0.386        | 0.000          |
| $\beta_{AR1}$ | -0.159        | 0.000          | -0.345       | 0.000          | 0.205         | 0.000          | 0.203        | 0.000          |
| $\beta_{AR2}$ | -0.056        | 0.000          | -0.118       | 0.000          | 0.070         | 0.000          | 0.049        | 0.000          |
| $\beta_{AR3}$ | -0.048        | 0.000          | -0.063       | 0.000          | 0.034         | 0.000          | 0.055        | 0.000          |
| $\beta_{AR4}$ | -0.029        | 0.003          | -0.043       | 0.000          | 0.044         | 0.000          | 0.053        | 0.000          |
| $\beta_p$     | 0.013         | 0.185          | 0.006        | 0.738          | 0.075         | 0.128          | -0.038       | 0.553          |
| $\beta_s$     | 0.013         | 0.028          | 0.031        | 0.004          | 0.092         | 0.001          | 0.062        | 0.109          |
| $\beta_f$     | 0.005         | 0.528          | -0.031       | 0.028          | 0.017         | 0.664          | -0.010       | 0.831          |
| $R^2$         | 0.0328        |                | 0.121        |                | 0.0825        |                | 0.097        |                |
| $n$           | 18417         |                | 10052        |                | 18417         |                | 10052        |                |

| Coeff.        | SPR           |                |              |                | VOL           |                |              |                |
|---------------|---------------|----------------|--------------|----------------|---------------|----------------|--------------|----------------|
|               | high turnover |                | low turnover |                | high turnover |                | low turnover |                |
|               | Est.          | <i>p</i> -val. | Est.         | <i>p</i> -val. | Est.          | <i>p</i> -val. | Est.         | <i>p</i> -val. |
| $\alpha$      | 0.570         | 0.000          | 0.681        | 0.000          | 0.091         | 0.000          | 0.124        | 0.000          |
| $\beta_{AR1}$ | 0.227         | 0.000          | 0.191        | 0.000          | 0.479         | 0.000          | 0.349        | 0.000          |
| $\beta_{AR2}$ | 0.093         | 0.000          | 0.058        | 0.000          | 0.139         | 0.000          | 0.149        | 0.000          |
| $\beta_{AR3}$ | 0.059         | 0.000          | 0.052        | 0.000          | 0.081         | 0.000          | 0.098        | 0.000          |
| $\beta_{AR4}$ | 0.058         | 0.000          | 0.034        | 0.000          | 0.096         | 0.000          | 0.082        | 0.000          |
| $\beta_p$     | 0.190         | 0.004          | 0.441        | 0.000          | 0.116         | 0.000          | 0.342        | 0.000          |
| $\beta_s$     | 0.126         | 0.001          | 0.229        | 0.000          | 0.328         | 0.000          | 0.598        | 0.000          |
| $\beta_f$     | 0.018         | 0.739          | 0.056        | 0.464          | 0.119         | 0.000          | 0.214        | 0.000          |
| $R^2$         | 0.11          |                | 0.0837       |                | 0.535         |                | 0.37         |                |
| $n$           | 18886         |                | 10896        |                | 26291         |                | 28737        |                |

Table 6: Dependence on turnover. Top line shows dependent variables, second line indicates sub-sample. *p*-values for two-sided test with period SUR corrections.

$\beta_p/(1 - \beta_{AR1} - \beta_{AR2} - \beta_{AR3} - \beta_{AR4})$  is twice as large for preceding days and 50% larger on spam days for low-turnover-stocks than for high-turnover-stocks. This makes sense from an economic point of view: The lower the liquidity, the larger the price impact.

Volume shows similar characteristics as spreads: The same coefficients are significant as in the base case (here, coefficients for spam days, preceding and following days are highly significant). The impact on volume is markedly higher for low-turnover-stocks, which can again be attributed to liquidity.

### 5.3 Popularity as a Target

Interestingly, some stocks are spammed very frequently, while others are only targeted once or twice during the year 2005. An interesting question is what might be the reason for the popularity of some targets. Maybe the stocks that are more popular among spammers react “better” (from the viewpoint of the spammer) to manipulation attempts than others? To answer this question, we divide our sample into two groups: One sub-sample contains all stocks with 5 spam events or more during 2005, the other sub-sample all stocks with less than 5 spam events. We run the regression described in equation (1). The results are shown in Table 7. The effects on returns are about the same in magnitude for both groups. For frequently spammed stocks, the effect is highly significant on spam days, while for less frequently spammed stocks, it has a *p*-value of 6.8%.

| Coeff.        | RET                  |           |                   |           | SIG                  |           |                   |           |
|---------------|----------------------|-----------|-------------------|-----------|----------------------|-----------|-------------------|-----------|
|               | $\geq 5$ spam events |           | $< 5$ spam events |           | $\geq 5$ spam events |           | $< 5$ spam events |           |
|               | Est.                 | $p$ -val. | Est.              | $p$ -val. | Est.                 | $p$ -val. | Est.              | $p$ -val. |
| $\alpha$      | -0.016               | 0.000     | -0.015            | 0.000     | 0.396                | 0.000     | 0.410             | 0.000     |
| $\beta_{AR1}$ | -0.231               | 0.000     | -0.279            | 0.000     | 0.214                | 0.000     | 0.198             | 0.000     |
| $\beta_{AR2}$ | -0.102               | 0.000     | -0.070            | 0.000     | 0.067                | 0.000     | 0.061             | 0.000     |
| $\beta_{AR3}$ | -0.069               | 0.000     | -0.040            | 0.000     | 0.044                | 0.000     | 0.040             | 0.000     |
| $\beta_{AR4}$ | -0.048               | 0.000     | -0.026            | 0.001     | 0.044                | 0.000     | 0.048             | 0.000     |
| $\beta_p$     | 0.008                | 0.478     | 0.014             | 0.347     | -0.018               | 0.714     | 0.127             | 0.053     |
| $\beta_s$     | 0.019                | 0.002     | 0.020             | 0.068     | 0.049                | 0.054     | 0.189             | 0.000     |
| $\beta_f$     | -0.003               | 0.704     | -0.020            | 0.120     | 0.014                | 0.691     | -0.017            | 0.749     |
| $R^2$         | 0.0628               |           | 0.0843            |           | 0.0916               |           | 0.0859            |           |
| $n$           | 11501                |           | 16968             |           | 11501                |           | 16968             |           |

| Coeff.        | SPR                  |           |                   |           | VOL                  |           |                   |           |
|---------------|----------------------|-----------|-------------------|-----------|----------------------|-----------|-------------------|-----------|
|               | $\geq 5$ spam events |           | $< 5$ spam events |           | $\geq 5$ spam events |           | $< 5$ spam events |           |
|               | Est.                 | $p$ -val. | Est.              | $p$ -val. | Est.                 | $p$ -val. | Est.              | $p$ -val. |
| $\alpha$      | 0.595                | 0.000     | 0.626             | 0.000     | 0.104                | 0.000     | 0.113             | 0.000     |
| $\beta_{AR1}$ | 0.211                | 0.000     | 0.212             | 0.000     | 0.433                | 0.000     | 0.381             | 0.000     |
| $\beta_{AR2}$ | 0.096                | 0.000     | 0.070             | 0.000     | 0.145                | 0.000     | 0.152             | 0.000     |
| $\beta_{AR3}$ | 0.058                | 0.000     | 0.056             | 0.000     | 0.103                | 0.000     | 0.091             | 0.000     |
| $\beta_{AR4}$ | 0.047                | 0.000     | 0.049             | 0.000     | 0.072                | 0.000     | 0.099             | 0.000     |
| $\beta_p$     | 0.244                | 0.001     | 0.330             | 0.000     | 0.108                | 0.000     | 0.384             | 0.000     |
| $\beta_s$     | 0.099                | 0.004     | 0.362             | 0.000     | 0.391                | 0.000     | 0.566             | 0.000     |
| $\beta_f$     | 0.010                | 0.835     | 0.067             | 0.404     | 0.103                | 0.000     | 0.248             | 0.000     |
| $R^2$         | 0.101                |           | 0.0966            |           | 0.498                |           | 0.403             |           |
| $n$           | 11936                |           | 17846             |           | 19759                |           | 35269             |           |

Table 7: Dependence on target popularity. Top line shows dependent variables, second line indicates sub-sample.  $p$ -values for two-sided test with period SUR corrections.

When looking at our volatility measure SIG, we find that volatility of frequently spammed stocks is much less affected compared to infrequently spammed stocks. The popularity of a target is inversely related to its volatility reaction on spam days. Unaffected volatility, however, is no different between our subsamples volatility: Stocks with high spread increases on spam days and preceding days are less frequently spammed.

VOL also reacts less sensitively for frequently spammed stocks compared to the other sub-sample. The difference in effects is particularly high on preceding days.

A closely related question is whether market participants learn from previous spams, i.e., do observed effects on the dependent variables decrease with the number of times a particular stock is targeted? Unfortunately, this question is very difficult to investigate, at least based on the information we have. Two requirements that we consider crucial for the feasibility of the analysis are not met in practice: First, we do not know who actually received a particular spam e-mail (whether it was sent to a comparatively large group of recipients, or rather a smaller one). An investor who receives only, say, every fifth spam e-mail concerning a certain stock cannot be expected to learn anything from experience with the four spams in between. Second, we would need information about the absolute number of spam e-mails by which each stock in our sample has been targeted. Recording a spam e-mail in January 2005 does not tell



us anything about how many spams occurred prior to that point in time (in 2004, 2003, . . .). Therefore, we cannot provide an answer to this question based on our data. Boehme and Holz (2006) investigate whether the “stock spam trick is wearing out over time”, conducting the analysis described above across all stocks, independently of how many spams the stocks received prior to the sample period. They find no evidence of decreasing effects of spam mails.

## 5.4 Successive Spam Days

From time to time, we observe a clustering of spam events, meaning that a certain stock is targeted on successive trading days. We can only speculate about possible motives of spammers for that: Maybe the manipulation attempt on the first day was not successful, and the spammer desperately tries to repeat his attempt on the following day. Or it was successful, but due to liquidity problems the spammer could not liquidate his entire position, and he wants to make sure that demand for the stock does not decrease while he is trying to liquidate the remainder of his position. To investigate this question, we define new dummy variables based on the number of spam days in a row, and run the regression described in equation (7). Table 8 shows the results of this regression. For RET, the spam day coefficient is significantly positive for the first day in a

| Coeff.        | RET    |                | SIG    |                | SPR   |                | VOL   |                |
|---------------|--------|----------------|--------|----------------|-------|----------------|-------|----------------|
|               | Est.   | <i>p</i> -val. | Est.   | <i>p</i> -val. | Est.  | <i>p</i> -val. | Est.  | <i>p</i> -val. |
| $\alpha$      | -0.006 | 0.000          | 0.395  | 0.000          | 0.575 | 0.000          | 0.270 | 0.000          |
| $\beta_{AR1}$ | -0.210 | 0.000          | 0.203  | 0.000          | 0.219 | 0.000          | 0.399 | 0.000          |
| $\beta_{AR2}$ | -0.065 | 0.000          | 0.057  | 0.000          | 0.098 | 0.000          | 0.121 | 0.000          |
| $\beta_{AR3}$ | -0.067 | 0.000          | 0.040  | 0.000          | 0.062 | 0.000          | 0.066 | 0.000          |
| $\beta_{AR4}$ | -0.023 | 0.003          | 0.049  | 0.000          | 0.048 | 0.000          | 0.057 | 0.000          |
| $\beta_p$     | 0.009  | 0.330          | 0.045  | 0.331          | 0.228 | 0.000          | 0.190 | 0.000          |
| $\beta_{s,1}$ | 0.019  | 0.008          | 0.087  | 0.022          | 0.148 | 0.004          | 0.392 | 0.000          |
| $\beta_{s,2}$ | 0.017  | 0.102          | 0.037  | 0.512          | 0.038 | 0.595          | 0.417 | 0.000          |
| $\beta_{s,3}$ | 0.017  | 0.260          | -0.087 | 0.255          | 0.101 | 0.299          | 0.257 | 0.000          |
| $\beta_{s,4}$ | 0.001  | 0.977          | 0.184  | 0.119          | 0.080 | 0.603          | 0.302 | 0.001          |
| $\beta_{s,5}$ | -0.052 | 0.072          | 0.135  | 0.369          | 0.405 | 0.053          | 0.343 | 0.004          |
| $\beta_f$     | -0.008 | 0.283          | 0.052  | 0.161          | 0.072 | 0.159          | 0.142 | 0.000          |
| $R^2$         | 0.0629 |                | 0.105  |                | 0.127 |                | 0.42  |                |
| $n$           | 16106  |                | 16106  |                | 16837 |                | 22609 |                |

Table 8: Successive spam days – results for the regression from equation (7). Top line shows dependent variables. *p*-values for two-sided test with period SUR corrections.

row of successive spam days only. Repeating spam e-mails on successive days does not seem to be a profitable strategy.

For SIG and SPR, we also find a significantly positive coefficient for the first spam day only. While intraday spreads are significantly larger than usual on preceding days, no such effect can be detected for volatility. Neither volatility, nor intraday spreads are significantly affected on following days.

Volume is significantly higher on all spam days, regardless of whether they occur in a row or not. The volume effect is also significantly positive for preceding and following days. The effect on spam days is much larger in magnitude for spam days than for preceding or following days.

## 5.5 Price Level of Target Stock

Given that almost all of the stocks in our sample trade *very* cheap, we want to find out whether “cheaper” means “better for the spammer”, in the sense that cheaper stocks are easier to manipulate (react more sensitively to manipulation attempts). To this end, we split our sample into two, taking a price of ten cents as the boundary. The results of the regression described in 1 applied to these sub-samples are shown in Table 9. For returns, we find a marked influence of

| Coeff.        | RET                   |                |                  |                | SIG                   |                |                  |                |
|---------------|-----------------------|----------------|------------------|----------------|-----------------------|----------------|------------------|----------------|
|               | price $\geq$ 10 cents |                | price < 10 cents |                | price $\geq$ 10 cents |                | price < 10 cents |                |
|               | Est.                  | <i>p</i> -val. | Est.             | <i>p</i> -val. | Est.                  | <i>p</i> -val. | Est.             | <i>p</i> -val. |
| $\alpha$      | -0.006                | 0.000          | -0.028           | 0.000          | 0.395                 | 0.000          | 0.440            | 0.000          |
| $\beta_{AR1}$ | -0.210                | 0.000          | -0.307           | 0.000          | 0.203                 | 0.000          | 0.193            | 0.000          |
| $\beta_{AR2}$ | -0.065                | 0.000          | -0.109           | 0.000          | 0.057                 | 0.000          | 0.059            | 0.000          |
| $\beta_{AR3}$ | -0.067                | 0.000          | -0.055           | 0.000          | 0.040                 | 0.000          | 0.036            | 0.000          |
| $\beta_{AR4}$ | -0.023                | 0.002          | -0.049           | 0.001          | 0.049                 | 0.000          | 0.035            | 0.000          |
| $\beta_p$     | 0.009                 | 0.336          | 0.006            | 0.753          | 0.045                 | 0.332          | 0.043            | 0.522          |
| $\beta_s$     | 0.014                 | 0.012          | 0.022            | 0.038          | 0.058                 | 0.041          | 0.140            | 0.000          |
| $\beta_f$     | -0.008                | 0.276          | -0.012           | 0.385          | 0.052                 | 0.162          | -0.065           | 0.233          |
| $R^2$         | 0.0625                |                | 0.105            |                | 0.104                 |                | 0.085            |                |
| $n$           | 16106                 |                | 12363            |                | 16106                 |                | 12363            |                |

| Coeff.        | SPR                   |                |                  |                | VOL                   |                |                  |                |
|---------------|-----------------------|----------------|------------------|----------------|-----------------------|----------------|------------------|----------------|
|               | price $\geq$ 10 cents |                | price < 10 cents |                | price $\geq$ 10 cents |                | price < 10 cents |                |
|               | Est.                  | <i>p</i> -val. | Est.             | <i>p</i> -val. | Est.                  | <i>p</i> -val. | Est.             | <i>p</i> -val. |
| $\alpha$      | 0.575                 | 0.000          | 0.700            | 0.000          | 0.270                 | 0.000          | 0.071            | 0.000          |
| $\beta_{AR1}$ | 0.219                 | 0.000          | 0.192            | 0.000          | 0.399                 | 0.000          | 0.327            | 0.000          |
| $\beta_{AR2}$ | 0.098                 | 0.000          | 0.049            | 0.000          | 0.120                 | 0.000          | 0.141            | 0.000          |
| $\beta_{AR3}$ | 0.062                 | 0.000          | 0.042            | 0.000          | 0.066                 | 0.000          | 0.093            | 0.000          |
| $\beta_{AR4}$ | 0.048                 | 0.000          | 0.040            | 0.000          | 0.057                 | 0.000          | 0.085            | 0.000          |
| $\beta_p$     | 0.228                 | 0.000          | 0.398            | 0.000          | 0.190                 | 0.000          | 0.213            | 0.000          |
| $\beta_s$     | 0.109                 | 0.004          | 0.296            | 0.000          | 0.370                 | 0.000          | 0.509            | 0.000          |
| $\beta_f$     | 0.072                 | 0.162          | 0.007            | 0.933          | 0.143                 | 0.000          | 0.155            | 0.000          |
| $R^2$         | 0.127                 |                | 0.0856           |                | 0.42                  |                | 0.442            |                |
| $n$           | 16837                 |                | 12945            |                | 22609                 |                | 32419            |                |

Table 9: Dependence on the price level of the target. Top line shows dependent variables, second line indicates sub-sample. *p*-values for two-sided test with period SUR corrections.

price level, independently of spam events. Stocks in the lower price group show much higher negative returns than the higher-priced stocks.

The increase in volatility on spam days is significantly larger for lower-priced stocks. Spreads are generally higher for lower-priced stocks, as is the effect on spam days and preceding days.

Volume is markedly higher for higher-priced stocks, indicating generally higher liquidity. The impact on volume on spam days is significantly higher for low-priced stocks, which is consistent with our interpretation of the effects being mainly driven by liquidity.

## 6 Conclusions and Directions for Further Research

Using a large sample of spam e-mails, we find significant effects of stock spam on a number of market statistics (price, volatility, intra-day spread, trading volume). This, in itself, does not provide evidence for the success of spammers. However, there is still a lot of information that we can extract from the analysis of our data: We can characterize stocks that are particularly vulnerable (i.e., easy to manipulate). Stocks that are frequently targeted react – based on closing prices! – less sensitively to stock spam than others. Sequential spamming on successive days has no additional effect. Return reacts strongest for stocks trading below 10 cents at low volumes.

Our database is insufficient to analyze the success of certain trading strategies. Instead of open/high/low/close data, intraday transactions data would be required. The same holds for an analysis of possible trading gains of other investors trying to “ride the wave”. As soon as we get intra-day price data for the stocks in our sample, we will investigate also the economic significance of our results by exploring the success of various trading strategies based on stock spam e-mails.

## References

- Antweiler, Werner, Murray Z. Frank. 2004. Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance* **59/3** 1259–1294.
- Boehme, Rainer, Thorsten Holz. 2006. The effect of stock spam on financial markets.
- Dewally, Michael. 2003. Internet investment advice: Investing with a rock of salt. *Financial Analyst Journal* **59/4** 65.
- Jianga, Guolin, Paul G. Mahoney, Jianping Mei. 2005. Market manipulation: A comprehensive study of stock pools. *Journal of Financial Economics* **77/1** 147–170.
- Larson, Scott, Carl Luft, Lawrence M. Levine. 2001. Over the counter bulletin board exchange: Market structure, risk, and return. *Journal of Alternative Investments* **4/2** 33–42.
- Luft, Carl, Lawrence M. Levine. 2004. Over the counter bulletin board exchange: The impact of liquidity and size to return, volatility, and bid/ask spread. *Journal of Alternative Investments* **7/3** 95–106.
- Tumarkin, Robert, Robert F. Whitelaw. 2001. News or noise? internet postings and stock prices. *Financial Analyst Journal* **57/3** 41–51.
- van Bommel, J. 2003. Rumors. *The Journal of Finance* **LVIII/4** 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. Scam artists tap text messaging to lure folks seeking hot stocks **27.09.2005**.

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

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics - A Modern Approach*. 3rd ed. Thomson.