The Impact of Manipulation in Internet Stock Message Boards

Jean-Yves Delort∗,a,b, Bavani Arunasalama, Maria Milosavljevica, Henry Leungc

∗Capital Markets CRC Limited, Sydney, NSW 2001, Australia
aDepartment of Computing, Macquarie University, Sydney, NSW 2109, Australia
bDiscipline of Finance, Faculty of Business and Economics, University of Sydney, Sydney, NSW 2006, Australia

Abstract

Internet message boards are often used to spread information in order to manipulate financial markets. Although this hypothesis is supported by many cases reported in the literature and in the media, the real impact of manipulation in online forums on financial markets remains an open question. This work analyses the effect of manipulation in internet stock message boards on financial markets. We employ a unique corpus of moderated messages to investigate market manipulation. Internet message boards administrators use the process of moderation to restrict market manipulation. From the data we find that manual supervision of stock message boards by moderators does not effectively protect users against manipulation. Furthermore, by focusing on messages that have been moderated as manipulative due to ramping we show that ramping is positively related to market returns, volatility and volume. We also demonstrate that stocks with higher turnover, lower price level, lower market capitalization and higher volatility are more common targets of ramping.

Key words: Ramping, Market Manipulation, Internet Stock Message Boards, Event Study

JEL classification: D.82; G.14; G.22

∗Corresponding author
Email addresses: jydelort@cmcrc.com (Jean-Yves Delort), bavani@cmcrc.com (Bavani Arunasalam), maria@cmcrc.com (Maria Milosavljevic), henry@cmcrc.com (Henry Leung)

Preprint submitted to Elsevier  October 30, 2009

1. Introduction

Internet stock message boards provide an effective medium for investors to communicate, to disseminate and discover new information. However, the value and integrity of internet stock message boards are often criticised and investors are warned not to trade on the information provided (Fraser, 2007). In order to address these issues, regulators often dictate that message board providers conform to specific guidelines\(^1\). Nevertheless, people continue to use forums for their own purposes, and financial surveillance analysts at regulators and exchanges are known to use message boards for investigating suspicious trading activities.

HotCopper\(^2\) is the most popular internet stock message board in Australia and forms the focus of our study. This site operates under a strict code of conduct which prohibits unethical or illegal use of the forum. Administrators moderate posts which do not conform to the guidelines and can revoke access for users who consistently disregard the rules. Moderated posts are labeled with the reason for moderation, for example ramping, flaming or profanity. These posts serve to proxy manipulative behaviour. Using an event study methodology, we investigate whether ramping is related to stock price volatility, trade volume and returns. Further, we study the firm characteristics targeted by rampers.

We first find that manual moderation of messages cannot effectively prevent ramping to take place in internet stock messages. This result is supported by the high number of posts that are moderated due to ramping every day which indicates that ramping is a common phenomenon. In addition, we find that the lead time to moderation is approximately 8 hours, allowing sufficient time for these posts to impact the market. Our results also demonstrate that ramping is positively related to market returns, volatility and volume. Firms with higher turnover, lower price level, lower market capitalization and higher volatility also receive a higher proportion of ramping posts suggesting that they are more common targets of ramping.

The remainder of this study is organised as follows. In Section 2 we

---

\(^1\)For example, in Australia, see Australian Securities & Investments Commission (ASIC) Regulatory Guide 162, http://www.asic.gov.au

\(^2\)http://www.hotcopper.com.au
present an overview of previous work. This serves as motivation for the
development of our hypotheses in Section 3. In Section 4 we describe the data
and empirical methodology used in our analysis. The results are reported
and discussed in Section 5, and Section 6 concludes and provides some areas
for further work.

2. Literature review

Stock market manipulation is generally defined as the creation and exp-
loitation of arbitrage opportunities (Zigrand, 2006). Manipulation of stock
prices may occur through direct trading strategies or indirectly via the dis-
semination of distorted price sensitive information. The former may be ob-
served when trades are intentionally placed in the wrong direction or short
term losses are being undertaken to move prices in the desired direction
(Chakraborty and Ylmaz, 2004). The latter may occur through mediums
such as investment blogs, spam emails and online stock message forums
(Antweiler and Frank, 2004; Das and Sisk, 2005; Aggarwal and Wu, 2006).
However, the detection of both types of manipulative behaviour remains dif-
ficult due to the lack of proxy for the occurrence of manipulation.

The impact of trade strategy manipulation can be observed in the market
response to such strategies. For example, Aggarwal and Wu (2006) demon-
strate that stock market manipulation cases pursued by the U.S. Securities
and Exchange Commission (SEC) from January 1990 through October 2001
were associated with greater stock volatility, greater liquidity and higher
returns. Other evidence of stock market reaction to manipulation is doc-
umented by Hillion and Suominen (2004), who highlight a general rise in
volatility, trading volume and one-minute returns in the closing price of Paris
Bourse stocks between January and April, 1995. Further, the proportion of
partially hidden orders increased during this period.

Evidence of manipulation has also been found in certain types of deriva-
tives markets such as the futures market. It is shown that uninformed in-
vestors earn positive profits by creating a futures position and simultaneously
trading in the spot market. As a result, arbitrage free profit opportunities are
exploited to derive profitable cash settlement at the time of delivery (Kumar
and Seppi, 1992; Merrick et al., 2005).

Additional evidence of trade strategy manipulation include those found
by Guo et al. (2008), who uses low earnings quality firms to proxy for firms’
intention to manipulate. They find that these firms tend to use stock splits
to manipulate equity values before acquisition announcements. On the other hand, Eom et al. (2009) find that investors strategically place spoofing intraday orders in the Korean Exchange. Spoofing orders have little chance of being executed and are used to mislead other traders of an order book imbalance and thereby influence the direction of stock price movement. They conclude that stocks targeted for manipulation had higher return volatility, lower market capitalization, lower price level, and lower managerial transparency. Finally, large traders have also been found to carry out manipulation due to its scale of operation and its ability to better sequence and time trades (Gastineau and Jarrow, 1991; Allen et al., 2006).

Manipulative behaviour involving the intentional mis-interpretation and dissemination of price sensitive information is more difficult to document. Bagnoli and Lipman (1996) address the problem of market manipulation through the announcement of a takeover bid. They develop a model where the manipulator takes a position, announces a takeover bid, and then unwinds his position. In their attempt to elucidate market manipulation from the 1600s to the internet technology frauds of today, Leinweber and Madhavan (2001) show that through technology, information can be quickly disseminated to influence investor perception on particular stocks. Stock spam emails and posting on internet message boards are two major avenues manipulators can affect investor sentiment because of their global reach and high degree of accessibility to investors. In particular, significant reactions of trading volume and returns have been shown to respond to spam campaigns (Bohme and Holz, 2006). Hanke and Hauser (2008) then extends the works of Leinweber and Madhavan (2001) and Bohme and Holz (2006) to provide corroborating evidence that stock spam e-mails significantly and positively impact volatility and intraday spread. Similar evidence is noted by Hu et al. (2009), who found that pump and dump email campaigns show a statistically significant decline in stock price from the peak spam day to the following day.

Evidence of online message board posts’ impact on stock activity is well documented in prior literature (Antweiler and Frank (2004); Das and Sisk (2005); Aggarwal and Wu (2006); Cook and Lu (2009)). Antweiler and Frank (2004) use computational linguistics methods to analyze the effect of message sentiment (buy, sell or hold) in forums to determine whether the information content of posts influence the stock market. They find that stock messages help predict volatility and disagreement between messages to be associated with increased trading volume. However, they find no correlation of mes-
sage sentiments with the direction of returns. By improving sample selection and removing noise caused by program generated sentiments, Cook and Xu (Cook and Lu, 2009) find that the bullishness of board messages positively and significantly predict abnormal stock return up to 2 days ahead. When taking poster’s credibility into account, they also find that the board messages’ predictive power over stock returns becomes much stronger in terms of both economic magnitude and significance. These studies demonstrate that stock markets are likely to be sensitive to internet message board activities. Moreover, many anecdotal evidences of market manipulation in Internet stock message boards have been reported in the financial literature and in the media (Leinweber and Madhavan, 2001; Harris, 2002). However, the phenomenon has not been investigated through a rigorous approach. To the best of our knowledge, this work provides the first quantitative analysis of market manipulation in Internet stock message boards.

3. Hypotheses development

This paper employs moderated posts unique to the Australian HotCopper online message forum. Improving upon prior literature such as Das and Sisk (2005), Chemmanur and Yan (2009) and Brown and Cliff (2004), HotCopper moderated posts allow us to proxy for manipulative behaviour directly.

Stock ramping through forums occurs when a ramper recommends a stock and highlights its huge potential to rise in share price in the very near future. The ramper buys a large quantity of a stock, and posts to influence the masses in to buying the stock with the aim to drive up the share price. By implying that readers of these posts are in possession of privileged information - such as insider knowledge of an impending takeover offer - rampers seek to persuade the gullible into purchasing a particular stock. If a significant enough number of easily-led individuals invest in the touted stock, a ramper can “ramp up” the share price so that he or she could sell their shares at a quick profit. Ramping can be both up or down.

Stock message board administrators use the moderation process to restrict the manipulations of stock prices through ramping posts. However, the wide accessibility of internet technology and the high speed at which investors absorb news may diminish the effect of moderation. Therefore we want to test the following hypothesis:

**Hypothesis 1.** Manual moderation of stock message boards, as a monitoring strategy, does not protect users against manipulation.
The direct measure of ramping through HotCopper moderated posts allow to investigate the impact of ramping on stock market activity. It has been shown in the literature that well advertised stocks are related to higher stock returns (Das and Sisk, 2005; Chemmanur and Yan, 2009; Brown and Cliff, 2004). Based on the investor attention hypothesis proposed in Chemmanur and Yan (2009) which implies that high levels of stock advertising is associated with larger stock returns, stocks heavily promoted to investors via ramping posts may also generate higher returns. One possible explanation proposed by Barber and Odean (2008) may be investors' need to select stocks out of the many available and ones that attract their attention will be the ones most likely to be included in their investment portfolio.

Furthermore, extreme news announcements have been shown to produce high trade volume (Tetlock, 2007). Similarly, ramping posts, as a type of extreme news announcements, may potentially have the same effect. Both stock spam e-mails and online message forum messages rely on the internet technology to increase its reach to investors (Leinweber and Madhavan, 2001). Drawing upon the works of (Bohme and Holz, 2006; Hanke and Hauser, 2008; Hu et al., 2009) on the impact of stock spam e-mails on trade volatility and intraday spread, we hypothesize that manipulators may employ online message forum message ramping to influence investor sentiment in a similar fashion. If ramping posts does indeed have an impact on investor sentiment prior moderation, we may form the following hypothesis to test the relation between moderated ramping posts and stock market response:

**Hypothesis 2.** Ramping in internet stock message boards affects stock price volatility, volume and returns.

Finally, we attempt to describe the determinants of target firms by rampers. Since manipulation is more effective when there is greater uncertainty about the value of a firm, we expect that stocks with a higher return volatility would attract more ramping attempts (Eom et al., 2009). Firm size has also been shown to be significantly positively associated with the level of disclosure (Alsaed, 2006). This can explained by the fact that larger publicly listed firms receive higher analyst coverage. On the other hand, smaller firms receive less coverage and thus are more susceptible to price movements driven by private information release. To that end, we expect smaller firms to be more prone to manipulation through ramping. Hanke and Hauser (2008) reports that liquidity is a significant characteristic of companies targeted by
spammers. The lower the liquidity, the larger the price impact. In addition, the impact of spamming on trade volume is markedly higher for low-turnover stocks. The reaction of high-turnover stocks to spamming is generally lower because it takes a greater number of participants to move prices. To similar effect, we expect rampers to target low-turnover stocks (low liquidity).

Therefore, in order to investigate the characteristics of firms targeted by rampers, we form the following hypothesis.

**Hypothesis 3.** *Stocks targeted for manipulation have higher return volatility, lower market capitalization, lower price level, and lower turnover.*

4. **Data and methodology**

4.1. **Data**

The data for this study comes from HotCopper, Australia’s largest internet stock message board. The time period for our study runs from January 2008 to December 2008 inclusive. Messages were downloaded from the HotCopper website using software written by the authors, and we restricted our collection of data to only include messages that have a ticker symbol field representing an ASX listed company. Each message contains the following fields: date, time, author, ticker symbol and content. In total, the data set contains 1,146,223 messages for 1,825 firms listed on the Australian Securities Exchange (ASX).

An interesting feature of the HotCopper forum is that the moderated posts are labeled with the main reason of their moderation. In accordance with Australian law regarding all information available in the securities market, HotCopper users are expected to comply with a set of strict usage guidelines. Messages that do not comply may be moderated by a moderation volunteer or by an administrator. Such messages are not removed from the forum, but their content is replaced by a message which contains the following information: moderation time, moderation type, and a comment specifying the reasons for moderation. The full set of moderation types with their respective percentage composition are listed in Table 1; each type is a simple phrase which describes the primary reason for moderation. Some examples of textual comments are provided below:

Table 1: Moderation types found on HotCopper in 2008

<table>
<thead>
<tr>
<th>Moderation cause</th>
<th>Total</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>6,379</td>
<td>45.1%</td>
</tr>
<tr>
<td>Flaming</td>
<td>3,583</td>
<td>25.3%</td>
</tr>
<tr>
<td>Ramping</td>
<td>1,519</td>
<td>10.7%</td>
</tr>
<tr>
<td>Profanity</td>
<td>1,207</td>
<td>8.5%</td>
</tr>
<tr>
<td>Spamming</td>
<td>389</td>
<td>2.8%</td>
</tr>
<tr>
<td>Defamatory</td>
<td>351</td>
<td>2.5%</td>
</tr>
<tr>
<td>Unknown Reason</td>
<td>159</td>
<td>1.1%</td>
</tr>
<tr>
<td>Advertising</td>
<td>152</td>
<td>1.1%</td>
</tr>
<tr>
<td>Unlicensed Advice</td>
<td>72</td>
<td>0.5%</td>
</tr>
<tr>
<td>Duplicate</td>
<td>73</td>
<td>0.5%</td>
</tr>
<tr>
<td>Insider Trading</td>
<td>73</td>
<td>0.5%</td>
</tr>
<tr>
<td>Copyright</td>
<td>63</td>
<td>0.4%</td>
</tr>
<tr>
<td>Sexist</td>
<td>43</td>
<td>0.3%</td>
</tr>
<tr>
<td>Racist</td>
<td>43</td>
<td>0.3%</td>
</tr>
<tr>
<td>Blasphemous</td>
<td>33</td>
<td>0.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14,139</strong></td>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

- All you seem to be doing is plaguing threads with your downramping on non held stocks. Find something else to do with your time instead of wasting others. Snide remarks on stocks are not considered helpful or useful posts.

- Rumours aren’t required thanks. They usually lead to misinformation on a public forum. No more like this thanks.

- This post is being moderated because of the unsubstantiated information particularly “Every deal done falls through.” This just isn’t true. Please refrain from flaming other posters too.
In addition to the HotCopper data, we obtained the daily closing prices, high and low intraday prices and volume data for all companies listed on the ASX in 2008 from Reuters. We excluded all firms that had fewer than 50 posts during 2008 from our data set. Our final sample size consists of 1,083,913 posts for 938 firms. Note that results presented in 1 are based on the final sample size.

4.2. Methodology

**Figure 1: Dependent variables defined by Hanke and Hauser (2008)**

**Stock Return:**
\[
\text{RET}_{i,t} = \ln S_{i,t} - \ln S_{i,t-1}
\]
where \( S_{i,t} \) is the closing price of stock \( i \) on day \( t \).

**Volatility:**
\[
\text{SIG}_{i,t} = \sqrt{\frac{(\text{RET}_{i,t} - \text{RET}_i)^2}{\sigma_i^2}}
\]
where \( \text{RET}_i \) is the average \( \text{RET}_{i,t} \) over the sample period; and \( \sigma_i^2 \) is the average variance of stock \( i \) over the sample period.

**Intraday volatility:**
\[
\text{SPR}_{i,t} = \frac{\varsigma_{i,t}}{\bar{\varsigma}_i}
\]
where \( \varsigma_i \) is the average intraday price range stock \( i \) over the sample period; and \( \varsigma_{i,t} \) is the intraday price range of stock \( i \) on day \( t \) given by:
\[
\varsigma_{i,t} = \ln(\text{intraday high}) - \ln(\text{intraday low})
\]

**Turnover:**
\[
\text{VOL}_{i,t} = \ln \left[ 1 + \frac{t_{o_i,t}}{\bar{t}_o} \right]
\]
where \( t_{o_i,t} \) is the turnover of stock \( i \) on day \( t \) and \( \bar{t}_o \) is the average turnover of stock \( i \) over the sample period.

We apply an event study methodology to analyze the market impact of ramping in internet stock message boards. We define a *trade event* for a certain stock as a day on which the market is open for trading in that stock. We further define a *posting event* for a certain stock as a day on which at least one message has been posted in the message board discussing this stock. Similarly, we define a *moderation event* for a certain stock as a day on which
Figure 2: Regression model for the impact study

\[ Y_{i,t} = \beta_0 + \beta_1 Vol_{i,t} + \beta_2 \text{Market}_t + \beta_3 \text{Comments}_{i,t} + \beta_4 \text{Info}_{i,t} + \sum_{m \in L} \beta_m \text{Label}_{i,m,t} + \epsilon_{i,t} \]  \hspace{1cm} (5)

where:

- \( Y_{i,t} \) - the dependent variable that takes the values of \( \text{RET}_{i,t}, \text{SIG}_{i,t}, \text{SPR}_{i,t} \) or \( \text{VOL}_{i,t} \)
- \( Vol_{i,t} \) - traded volume of stock \( i \) on day \( t \)
- \( \text{Market}_t \) - price of S&P/ASX200 on day \( t \) - used as a control variable for the market effect
- \( \text{Comments}_{i,t} \) - number of moderated posts regarding stock \( i \) on day \( t \) that mention ‘ramp’ or ‘unlicensed advice’ in the moderator’s comment
- \( \text{Info}_{i,t} \) - a dummy variable that indicates the release of a company announcement for stock \( i \) on day \( t \)
- \( L \) - the set of moderation types given by
  - \{Advertising, Blasphemous, Copyright, Defamatory, Duplicate, Flaming, Insider Trading, Other, Profanity, Racist, Ramping, Sexist, Spamming, Unknown Reason, Unlicensed Advice\}
- \( \text{Label}_{i,m,t} \) - number of posts regarding stock \( i \) that were moderated with the label \( m \in L \) on day \( t \)

and \( \beta_0 \) and \( \epsilon_{i,t} \) are the intercept and a random error term respectively.

at least one message discussing this stock has been moderated. Finally, we define a ramping event for a certain stock as a day on which at least one message has been moderated because of an attempt to ramp this stock.

The performance of a stock is measured in terms of the dependent variables defined by Hanke and Hauser (2008) and also used by Hu et al. (2009), that is, stock return, volatility, intraday volatility and turnover. Their definitions of these variables are shown in Figure 1.

As discussed in Section 4.1, HotCopper posts may be moderated due to several reasons, but each moderated post can only be labeled with one of the moderation types given in Table 1. This restriction and the close similarity between some of the moderation types may result in the ramping posts being labeled differently. For example, a post which could be labeled as both ‘Unlicensed Advice’ and ‘Ramping’ may only be labeled as ‘Unlicensed Advice’. Posts that are moderated due to multiple reasons are often labeled as ‘Other’ which explains the high percentage of moderation under ‘Other’ (45.1%) as given in Table 1.
We found a significant number of ambiguous moderated posts in which the moderators assign one label but mention several causes for moderation in the comment field. Since our focus is on investigating the impact of ramping, it is important that we incorporate all posts that could potentially be seen as ramping. As a result, we include posts that mention the keyword ‘ramp’ or ‘unlicensed advice’ in the moderator’s comment as an additional explanatory variable in our regression analysis.

Figure 3: Regression model for determinants of target firms

\[
Y_i = \beta_0 + \beta_1 Vol_i + \beta_2 MarketCap_i + \beta_3 Price_i + \beta_4 SIG_i + \epsilon_i
\]  

(6)

where the explanatory variables are:

- \( Vol_i \) - average traded volume of stock \( i \) during the sample period.
- \( Price_i \) - average closing price of stock \( i \) during the sample period.
- \( MarketCap_i \) - market capitalization of stock \( i \) at the end of the sample period.
- \( SIG_i \) - average volatility of stock \( i \) during the sample period calculated as described in Figure 1

and \( \beta_0 \) and \( \epsilon_i \) are the intercept and a random error term respectively.

The dependent variable is the proportion of ramping posts, and is calculated as:

\[
Y_i = \frac{Ramp_i}{Mod_i}
\]  

(7)

where

- \( Ramp_i \) - the number of ramping posts of stock \( i \) during the sample period.
- \( Mod_i \) - the number of moderated posts of stock \( i \) during the sample period.

The regression equation for this impact study is given in Figure 2. The market and volume are used as added control variables. We also included in our model a dummy variable that takes the value 1 if a company announcement for stock \( i \) was released on day \( t \) and 0 otherwise. We include this in our model in order to control for the impact of company announcements on the performance of stocks.

As outlined in Section 3, particular types of firms are more likely to be targeted for manipulation than others. We use the regression model shown
in Figure 3 in order to investigate the characteristics of the firms targeted by rampers. In this regression, the dependent variable is the probability for a post to be moderated because of ramping. This probability proxies the probability for the stock of being ramped.

5. Results

5.1. Descriptive statistics

5.1.1. Moderation activity

We define moderation delay as the difference between publication time and moderation time for a particular post on an internet message board. In 2008, for the 938 stocks included in the sample, the mean moderation delay was 7.95 hours ($\sigma = 15.93$). In addition, 40.78\% of moderated posts where moderated after two hours. This is sufficient time for ramping on forums to be effective, which supports Hypothesis 1.

We found that in general, inappropriate posts tend to be published outside of normal trading hours. This is demonstrated by the graph shown in Figure 4 which represents the proportion of moderated posts and with respect to the hour of day. What is most interesting here is the daily pattern of posts which are moderated because of ramping. The graph identifies quite clearly that ramping is more active during trading hours which is quite different to the general pattern for moderated posts. We believe that the reason for this is that ramping messages created outside of normal trading hours are less likely to be effective since they are more likely to be moderated before market opening.

5.1.2. Dependent variables

Table 2 presents descriptive statistics for our data sample. In 2008, for the 938 stocks included in the sample, we have 238,252 trade events, 100,695 posting events, 4,870 moderation events and 768 ramping events. The number of ramping events reported is only a lower bound to the actual number of ramping events. With the increasingly high volume of messages that are posted on internet stock forums, it is likely that some of the inappropriate posts will not be discovered in the manual moderation process. However, these posts are not taken into account in our analysis. For example, the following unmoderated posts could be considered as ramping:

- *I heard a rumor that the plant is still being repaired and may be operating again later today.*
This figure plots the proportions of moderated posts and ramping posts divided by the proportion of posts with respect to the hour of the day.

- *I’m focusing on oil stocks now, not affected by all this, and also I heard a rumor even General Motors now admit our oil has run out on what we can produce in one day and now in decline which is good news to the oil stocks.*

Table 2 also shows that ramping events have the highest mean for returns, volatility, intra-day volatility and volume which supports Hypothesis 2. Ramping events have the lowest mean price of the four types of events. It
Table 2: Mean and stddev for all dependent variables

<table>
<thead>
<tr>
<th>Panel</th>
<th>Nb. events</th>
<th>Price</th>
<th>RET</th>
<th>SIG</th>
<th>SPR</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trade event</strong></td>
<td>238252</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.9820</td>
<td>-0.0052</td>
<td>0.0785</td>
<td>0.3747</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.24</td>
<td>0</td>
<td>0.3747</td>
<td>1</td>
<td>1.4717</td>
<td></td>
</tr>
<tr>
<td>St.dev.</td>
<td>7.7096</td>
<td>0.0785</td>
<td>0.7776</td>
<td>0.7116</td>
<td>0.518</td>
<td></td>
</tr>
<tr>
<td><strong>Posting events</strong></td>
<td>100695</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.3754</td>
<td>-0.0022</td>
<td>0.0878</td>
<td>0.4817</td>
<td>1.1994</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.265</td>
<td>0</td>
<td>0.4817</td>
<td>1.1994</td>
<td>1.6403</td>
<td></td>
</tr>
<tr>
<td>St.dev.</td>
<td>9.7972</td>
<td>0.0878</td>
<td>0.8489</td>
<td>0.8886</td>
<td>0.6476</td>
<td></td>
</tr>
<tr>
<td><strong>Moderation events</strong></td>
<td>4870</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.7410</td>
<td>-0.0053</td>
<td>0.1151</td>
<td>0.5474</td>
<td>1.3636</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.245</td>
<td>0</td>
<td>0.5474</td>
<td>1.3636</td>
<td>2.6274</td>
<td></td>
</tr>
<tr>
<td>St.dev.</td>
<td>9.1831</td>
<td>0.1151</td>
<td>1.019</td>
<td>0.9844</td>
<td>0.7723</td>
<td></td>
</tr>
<tr>
<td><strong>Ramping events</strong></td>
<td>786</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.6275</td>
<td>0.0027</td>
<td>0.1395</td>
<td>0.6446</td>
<td>1.8063</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>0.19</td>
<td>0</td>
<td>0.6446</td>
<td>1.8063</td>
<td>5.5706</td>
<td></td>
</tr>
<tr>
<td>St.dev.</td>
<td>5.7589</td>
<td>0.1395</td>
<td>1.2379</td>
<td>1.1167</td>
<td>0.9636</td>
<td></td>
</tr>
</tbody>
</table>

This table presents the mean, median and standard deviation for the dependent variables namely, Return (RET), Volatility (SIG), Intraday volatility (SPR) and Turnover (VOL). Panel A presents the statistics for all the firms across all the days for January 2008 through December 2008. Panels B, C and D present the same statistics only for days with at least one post, days with at least one moderated post and days with at least one ramping post respectively.

indicates that the “pump and dump” interest is the highest for small-priced firms. Interestingly, the mean returns for panel A, B and C are negative. A possible explanation could be that an unprecedented global financial crisis hit the market in the last quarter of 2008. However, the mean returns for ramping events is positive, which shows that ramping co-occurs with price increase.

In section 5.3 we further investigate the impact of ramping on the four dependent variables taking into account the firm characteristics.

5.2. Impact of manipulation in IMS

Table 3 shows the results of the regression analysis for the model given in 2. The results show that Volume, Market, the release of company announcements and Ramping are significantly correlated to return, volatility, intra-day volatility and volume. The correlation coefficients are positive for the four variables, but the confidence is stronger for volatility, intra-day volatility and volume. Insider Trading is also is also significantly correlated to volume,
Table 3: Impact study results

<table>
<thead>
<tr>
<th>Variable</th>
<th>RET</th>
<th>SIG</th>
<th>SPR</th>
<th>VOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercepts</td>
<td>0.0166</td>
<td>6.9634</td>
<td>0.0992</td>
<td>-3.3960</td>
</tr>
<tr>
<td>Volume</td>
<td>0.0000</td>
<td>22.1030</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Market</td>
<td>0.0044</td>
<td>-85.6260</td>
<td>0.1428</td>
<td>-1.3150</td>
</tr>
<tr>
<td>Ramping</td>
<td>0.0131</td>
<td>12.2100</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Spamming</td>
<td>-0.0163</td>
<td>-85.6260</td>
<td>0.1428</td>
<td>-1.3150</td>
</tr>
<tr>
<td>Insider Trading</td>
<td>0.0129</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Other</td>
<td>0.0047</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Unknown Reason</td>
<td>0.0033</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Unlicensed Advice</td>
<td>0.0047</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Flaming</td>
<td>-0.0061</td>
<td>-85.6260</td>
<td>0.1428</td>
<td>-1.3150</td>
</tr>
<tr>
<td>Defamatory</td>
<td>-0.0175</td>
<td>-85.6260</td>
<td>0.1428</td>
<td>-1.3150</td>
</tr>
<tr>
<td>Duplicate</td>
<td>0.0211</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Advertising</td>
<td>0.0102</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Profanity</td>
<td>-0.0061</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Copyright</td>
<td>-0.0118</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Ramping in comment</td>
<td>-0.0004</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
<tr>
<td>Company info</td>
<td>0.0080</td>
<td>0.0982</td>
<td>0.0982</td>
<td>-3.5680</td>
</tr>
</tbody>
</table>

This table reports the results of the regression analysis with four dependent variables: Return (RET), Volatility (SIG), Intraday volatility (SPR) and Turnover (VOL) in turn. \( RET_{i,t} = \ln S_{i,t} - \ln S_{i,t-1} \) where \( S_{i,t} \) is the closing price of stock \( i \) on day \( t \). \( SIG_{i,t} = \sqrt{\frac{(RET_{i,t} - \overline{RET})^2}{\sigma_i^2}} \) where \( \overline{RET} \) is the average return of stock \( i \) over the sample period and \( \sigma_i^2 \) is the average variance of stock \( i \) over the sample period. \( SPR_{i,t} = \frac{\varsigma_i}{\overline{\varsigma}} \) where \( \varsigma_i \) is the intraday price range of stock \( i \) on day \( t \) given by \( \varsigma_i = \ln(\text{intraday high}) - \ln(\text{intraday low}) \) and \( \overline{\varsigma} \) is the average intraday price range stock \( i \) over the sample period. \( VOL_{i,t} = \ln \left[ 1 + \frac{t_{o,i,t}}{t_{i,t}} \right] \) where \( t_{o,i,t} \) is the turnover of stock \( i \) on day \( t \) and \( t_{i,t} \) is the average turnover of stock \( i \) over the sample period. The explanatory variables are as follows: Volume \( c_{i,t} \) is the traded volume of stock \( i \) on day \( t \). Market \( M_{i,t} \) is the price of S&P/ASX200 on day \( t \). Ramping, Spamming, Insider Trading, Other, Unknown Reason, Unlicensed Advice, Flaming, Defamatory, Duplicate, Advertising, Profanity and Copyright indicate the number of posts moderated under the respective category for the respective stock on a given day. Ramping in comment \( c_{i,t} \) is the number of moderated posts regarding stock \( i \) on day \( t \) that mention ‘ramping’ in the moderator’s comment. Company info \( \delta_{i,t} \) is a dummy variable that indicates the release of a company announcement for stock \( i \) on day \( t \). All data are for January 2008 through December 2008. All the explanatory variables are log transformed except for \( \delta_{i,t} \).
Table 4: Determinants of target firms

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.0992</td>
<td>0.1239</td>
<td>0.0704</td>
<td>0.0058</td>
<td>-0.0906</td>
</tr>
<tr>
<td></td>
<td>(-2.489)</td>
<td>(2.980)</td>
<td>(11.450)</td>
<td>(0.453)</td>
<td>(-1.341)</td>
</tr>
<tr>
<td>Volume</td>
<td>0.0125</td>
<td>0.0186</td>
<td>0.0186</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.025)</td>
<td>(3.545)</td>
<td>(3.545)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market cap</td>
<td>-0.0036</td>
<td></td>
<td>-0.0059</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.506)</td>
<td></td>
<td>(-1.243)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td></td>
<td>-0.0188</td>
<td>-0.0188</td>
<td>1.0479</td>
<td>0.4135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-2.856)</td>
<td>(-2.856)</td>
<td>(3.545)</td>
<td>(1.251)</td>
</tr>
<tr>
<td>Volatility</td>
<td></td>
<td>1.0479</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.545)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table presents the results of the regression analysis described in Figure 3. The dependent variable is the proportion of ramping posts given by $\frac{\text{Ramp}_i}{\text{Mod}_i}$, where Ramp$_i$ is the number of ramping posts of stock $i$ during the sampling period and Mod$_i$ is the number of moderated posts of stock $i$ during the sampling period. Volume$_i$ is the average turnover for stock $i$ during the sampling period, Market cap$_i$ is the market capitalization of stock $i$ measured at the end of 2008, Price$_i$ is the average closing price for stock $i$ during the sampling period and Volatility$_i$ is the average volatility of stock $i$ during the sampling period.

Intra-day volatility and return. We previously introduced the dummy variable ‘Ramping in comment’ which corresponds to moderated events where at least one moderation comment contain the word “ramp” or “unlicensed advice”. The variable is also correlated to the two volatility measures, and volume to a lower extent. For volume we find a positive correlation with Ramping, Insider Trading, Other, Advertising, Flaming and Profanity. Other is a general category that is used for posts moderated because of miscellaneous or mixed reasons (and so can include Ramping, Flaming and Profanity).

Our results are limited by the fact that we only have closing prices and do not currently have access to intra-day prices. This results in an inability to distinguish between different moderated posts, that is, we cannot identify the impact of a specific ramping post on the market. Indeed, a ramping post cannot be measured independently of other moderated posts (such as profanity) which occur on the same day. Hence, we cannot accurately identify whether a specific ramping post has been effective in its cause.

5.3. Determinants of target firms

In this section, we present our findings in investigating the characteristics of firms which are the most common targets for ramping in forums. Since some of the independent variables are significantly correlated, we implement
both simple and multiple regressions in order to minimize the multicollinearity problem. In the simple regressions (regressions (1) through (4)), stocks with higher turnover, lower price level and higher volatility tend to receive a higher proportion of ramping posts. Market capitalization is also negatively correlated to the proportion of ramping posts but with a low significant value. Regression (5), which controls for multicollinearity, confirms that firms with higher turnover, lower price level, lower market capitalization and higher volatility receive a higher proportion of ramping posts. However, the correlation is highly significant only for volume.

6. Conclusions and Directions for Future Work

To the best of our knowledge, this work provides the first quantitative analysis of market manipulation in Internet stock message boards. Employing a unique corpus of moderated messages, we are able to identify messages that have been moderated as manipulative due to ramping. We find that manual moderation of internet stock message boards does not effectively protect users against manipulation. Our experiments also show that ramping is positively related to market returns, volatility and volume. In addition, our results demonstrate that stocks with higher turnover, lower price level, lower market capitalization and higher volatility are more common targets of ramping.

In future work, we plan to use computational linguistics techniques to refine our proxy for ramping. As discussed in section 5.1.2, the number of posts moderated due to ramping is only a lower bound to the actual number of ramping posts. To identify ramping posts that have not been labeled by moderators, we need to examine the content of all unmoderated posts. We plan to build a text classifier to automatically identify posts with inappropriate content. Applying this classifier to unmoderated posts will enable us to discover posts that should have been moderated due to ramping and subsequently to refine our proxy for ramping.

References


