Is All That Talk Just Noise?  
The Information Content of Internet Stock Message Boards

Werner Antweiler and Murray Frank  
University of British Columbia

Presentation at the Northern Finance Association, September 28, 2002

This paper is available on the web at 
Background

“Internet message boards have come of age. [...] even investment pros are watching the message boards closely and are profiting from it. With ‘posts’ running in the millions, Internet message boards have become an essential part of the savvy investor’s arsenal. [...] Internet messages really do move markets, for better or worse.”

— G. Weiss, Business Week, May 22, 2000

- Yahoo! Finance, Raging Bull, Silicon Investor, ...
- News reports of impact on financial markets
- Securities Exchange Commission has prosecuted people for internet messages.
Message Posting Activity

Yahoo! Finance Message Boards - Posting Volume

• 1 million messages per month on Yahoo Finance
• 45 stocks in our sample: 30 DIA and 15 XLK companies covering both the “old” and “new” economy
• trading activity is U-shaped between 09:30 and 16:00
• most of the posting activity is during the business day; many people must be accessing the message boards from work
• who are the message posters? day traders?
An Example

BUY ON DIPS – This is the opportunity
by: plainfielder

to make $$$ when IBM will be going up again following this profit taking bout by Abbey Cohen and her brokerage firm.

IBM shall go up again after today.

Posted as a reply to: Msg 1 by YahooFinance
Research Questions

1. Do the messages help predict returns?
2. Is disagreement among the messages associated with more trades?
   - Karpoff (1986) and Harris and Raviv (1993): Disagreement induces trading
   - No-trade theorem of Milgrom and Stokey (1982): Disagreement leads to revision of beliefs, not trades
3. Do the messages help predict volatility [and trading volume]?

Work related to ours:
- Wysocki (1999) is first study of internet stock message boards.
- Tumarkin and Whitelaw (2001) find no predictive ability for returns (sample of 9 firms).
- Das and Chen (2001) develop new natural language algorithm to classify stock messages.
Preparing the Research Data Set

Extracting ⇒ Classifying ⇒ Aggregating ⇒ Merging

Data Sources • Computer Linguistics

1.5 million messages for 30 DIA and 15 XLK stocks downloaded from Yahoo! Finance and Raging Bull

We classify training set of 1,000 messages BUY / HOLD / SELL

Train Naive Bayes Text Classification Engine (‘rainbow’)  

Computer classifies all 1.5m messages into BUY / HOLD / SELL

TAQ Quotes/Trades NYSE/NASDAQ

filter & aggregate into 26 15-minute intervals [09:30 - 16:00]

aggregate into 96 15-minute intervals [00:00-24:00]

RESEARCH DATA SET:
Message Board & Trading Activity  
251 days x 26 int. x 45 companies

BULLISHNESS & AGREEMENT

weights for messages: none; length; inverse contribution frequency
Computer Linguistics

- How can one machine-classify large amounts of text? statistical methods vs. content analysis
- Methods are crude but remarkably accurate
- We use Naive Bayes:
  - each word is associated with an odds ratio for being ‘bullish’ relative to ‘bearish’
  - multiply odds ratio for each word to prior odds ratio in order to obtain posterior odds ratio for entire text.
  - algorithm can be fooled easily (e.g., negation) but is remarkably robust for longer messages.
- We also obtain results from another classification method: Support Vector Machine
- General problem: stock message boards have a low signal-to-noise ratio.
- Specific problem: SVM tends to overfit.
Bullishness and Agreement

- Have \( M_t = M_t^{\text{BUY}} + M_t^{\text{SELL}} \) messages in time period \( t \)
- Which measure of bullishness?
  \[
  B_t \equiv \frac{M_t^{\text{BUY}} - M_t^{\text{SELL}}}{M_t^{\text{BUY}} + M_t^{\text{SELL}}} \in [-1, +1]
  \]
  \[
  B_t^* \equiv \ln \left[ \frac{1 + M_t^{\text{BUY}}}{1 + M_t^{\text{SELL}}} \right] \approx B_t \ln(1 + M_t)
  \]
  \[
  B_t^{**} \equiv M_t^{\text{BUY}} - M_t^{\text{SELL}} = B_t M_t
  \]
- We choose intermediate measure \( B^* \) which reflects direction and size of sentiment, but
- We ignore neutral/hold messages
- A measure of agreement
  \[
  A_t \equiv 1 - \sqrt{1 - B_t^2} \in [0, 1]
  \]
  is standard deviation of bullishness index \( B_t \).
Contemporaneous Effects: 15-minute intervals

Empirical Strategy

- Regress financial variables on message board variables

Findings

- Increased message posting activity is associated with higher trading volatility and higher trading volume.
- Greater agreement is associated with reduced market volatility.
- There is a contemporaneous (ie, non-predictive) link between bullishness and price returns.
Time Sequencing Tests

Empirical Strategy

- Vector Auto Regression of pooled panel: (return, volatility, trades by size, volume, spread) against (messages, words, bullishness, agreement)
- we also include a market variable (price/return of SPY)
- we conduct Granger causality tests
- using daily data and 15-minute data

Findings

- Bullishness does not predict returns (but number of trades)
- Greater agreement today predicts more trades tomorrow, in contrast to the contemporaneous effect
- Predictive ability works both ways: message board activity predicts some financial variables, and financial variables predict some message board variables.
Predictive Ability: Wall Street Journal

Empirical Strategy

- Try to predict whether message posting today predicts articles in the WSJ tomorrow (or day after tomorrow)
- We use a logistic regression of daily data with company and day-of-week fixed effects

Findings

- Message boards predict news stories in WSJ
  - one day before publication: significant predictive ability
  - two days before publication: somewhat weaker but still significant
Volatility

Empirical Strategy

• Predict realized volatility $v_{i,t}$ as in Andersen et.al. (2001), using fractionally-integrated time series with $d=0.297$. We estimate (individually and as a panel)

$$
(1 - L)^d \ln(v_{i,t}) = \beta_i + \beta_{v,i} \ln(v_{i,t-1}) + \beta_{v(r,i} \ln(v_{i,t-1}) I(r_{i,t-1} < 0)
+ \beta_{A,i} A_{i,t} + \beta_{M,i} \ln(1 + M_{i,t-1})
+ \beta_{N,i} \ln(N_{t-1}) + \epsilon_{i,t}
$$

• GARCH, EGARCH, and GJR-TGARCH as robustness checks

Findings

• Individual regressions
  – Strong cross-sectional differences
  – Amazon, IBM, MP3.com and Microsoft all have significant effects from message posting to volatility.
  – Agreement appears to have insignificant effect.

• Panel regression: clear effect of message posting on volatility; smaller effect for DIA than XLK firms.
Trading Volume

Empirical Strategy

• There is a strong *contemporaneous* link between trading volume and message board activity.

• But does message board activity *predict* trading volume?

• We estimate a volume model along the lines of Chordia, Roll, and Subrahmanyam (2001)

• We control for market momentum and individual stock momentum, 5-day volatility, day-of-week effects, and WSJ news stories.

Findings

• Message board posting volume 4 hours prior to market opening helps predict day’s trading volume

• No effect when considering longer aggregation periods.
Conclusions

- There is useful information present on the stock message boards. All that talk is not just noise.
- Message boards do not successfully predict stock returns at short time horizons.
- We find support for Harris and Raviv (1993): disagreement induces trading.
- Message posting helps predict volatility both at daily frequency and within the trading day.
- Stock messages reflect public information rapidly. They may be useful in studies of insider trading and events. They may also be helpful in market microstructure studies.

New Working Paper

- We currently examine 1999-2001 data of >5000 messages boards
- Focus is on longer (monthly) time horizon
- Does posting activity predict returns? volatility?
- Fama-French type of analysis