Preliminary Draft: Please do not distribute or cite without permission.
Comments welcome.

e-Information

by Sanjiv Das, Asís Martínez-Jerez and Peter Tufano

Sanjiv R. Das
Srdas@scu.edu
Santa Clara University
500 El Camino Real
Santa Clara, CA 95053

Asís Martínez-Jerez
asis@hbs.edu
Harvard Business School
Soldiers Field Road
Boston, MA 02163

Peter Tufano
Ptufano@hbs.edu
Harvard Business School
Soldiers Field Road
Boston, MA 02163

Draft dated: November 5, 2001

We would like to thank executives at The Motley Fool, Yahoo, Silicon Investor and Raging Bull for their cooperation; Sarah Woolverton, Sarah Erikson, John Kosmidis and Jose Camacho for assistance with data collection; Mike Chen and John Sheridan for assistance with developing some of the computer code; and David Leinweber and seminar participants at the HBS Brown Bag, the University of Texas at Austin , Yale, Ohio State and Northwestern for their comments. This research does not represent the views of any of the firms from which we obtained chat board information nor the views of the firms studied. This work has been funded in part by the Dean Witter Foundation (Das), the Bank of Spain (Martínez-Jerez) and the Division of Research at the Harvard Business School (Tufano and Martínez-Jerez).

e-Information

In this clinical study, we create a new micro-database of much of the information flow about four stocks for a period of six months, including a novel database of e-Information. We define e-Information as estimates of the intensity and dispersion of investor sentiment extracted from four major stock chat message boards. We combine this e-Information with other components of the traditional information set (formal press releases by the firms, filings, analyst revisions, and news stories available electronically) to create an intensive database of as close to the full information flow as is practical. We document certain patterns in the release of information and show that there are suggestions of links between our measure of e-Information and contemporaneous stock returns and implied volatility. While e-Information does not seem to predict subsequent stock price returns, it may be related to the implied volatilities taken from options prices.

> *It is a tale*
> *Told by an idiot, full of sound and fury,*
> *Signifying nothing.*
> William Shakespeare, **Macbeth**, Act 5, Scene 5.

1. Introduction

For at least thirty-five years, financial economists have formally considered the notion of informationally-efficient markets, which instantaneously process all available information and impound it into security prices (Samuelson (1965), Fama (1965 and 1975)). Extensive literatures examine the degree to which markets are informationally-efficient, including studies of the performance of filter-rules, information reactions to events, informational anomalies, etc. Accounting scholars have studied how quickly and fully information (typically earnings releases) is incorporated into security prices (Bernard and Thomas (1990), Kothari (2000)). Market-microstructure scholars have modeled how new information, the dispersion of beliefs, and the intensity of private beliefs should affect the volumes, volatilities and spreads of stock prices (O'Hara (1995), Hasbrouck (1996)). By our rough estimate, hundreds of articles on the informational efficiency of financial markets have been published over the past quarter century in major finance and accounting journals.

In light of all of this outstanding research, there seems little left to learn from empirical examination of the information in markets and the relationship of this information to the price formation process. However, the past few years have witnessed the creation of new means by which information, opinions and analyses can be shared among investors in the form of stock message boards (or chat rooms) posted in cyberspace. These message boards allow investors to share legitimate information and opinions, unscrupulous rogues to manipulate prices, and foul-mouthed people to rant about almost anything. These boards provide a real-time window into the minds of some individual investors in a way that has heretofore not been possible. For example, when Apple Computer announced lower than expected earnings on September 28, 2000, its stock lost half its value in an hour. Between 4 p.m., when the announcement was made, and midnight that day, investors posted more than 2,000 messages to the Yahoo! Apple stock message board, or over 250 messages an hour, to discuss the implications of Apple's announcement. Our work is part of a small new literature that seeks to understand the possible relevance of this technology.[1] It is also related to research on how changes in technology affect financial markets; for instance, Garbade and Silber (1978) demonstrate that the opening of a Trans-Atlantic telegraph cable in the 19th century led to immediate adjustments of prices between London and New York.

Concurrent with the evolution of new data sources, advances in artificial intelligence and statistical language processing have begun to provide tools that can extract information from these types of massive databases. We use these new technologies and new data to infer crude measures of investor sentiment (bullishness or bearishness) and differences of opinion among investors. We label this extracted information "e-Information."

We have three goals in this paper, which is a preliminary study into what we believe is a new and important potential field of research. First, we define the

1 There are two papers similar in spirit to this paper: Antweiler and Frank (2001) and Turmarkin and Whitelaw (2001). Each was produced independently of the others at approximately the same time.

concept of e-Information and characterize the e-Information flow for four stocks for a six-month period in 1998. Second, we relate e-Information to more traditional and widely studied forms of information such as company filings, press releases, announcements by security analysts, and news stories. This section of our work is meant to describe how information flows through the economy. We do this by examining autocorrelations of information flow, cross-correlations of information flow, and by studying a few "tracer events" that show the movement of specific pieces of information through the information flow. Finally, we will explore the usefulness of e-Information by examining its predictive power. In particular, we examine e-Information's usefulness in explaining returns and implied volatility . e-Information may be the "tale told by an idiot," to use Shakespeare's language, or it may help us understand the price formation process.

2. The concept and measurement of e-Information

2.1 Definition and motivation of e-Information

We would like to measure the flow of information in the economy. The easiest characterizations of this information are "activity" measures: simple counts of the numbers of news stories or postings, or the length of news story or posting. These type of metrics have been used by Mitchell and Mulherin (1994; number of new stories) and Wysocki (1999; number of postings). These activity measures indicate the level of interest, excitement, puzzlement or "buzz" about the information set, similar to the measure of the decibels of noise in trading pits used by Coval and Shumway (2001). Activity measures are based on the notion that discussion (whether in person, by electronic posting, or news stories) is correlated with the salience and newness of information releases.

Capturing the *content* of the information is a more complicated matter. If one assumes informationally-efficient markets, then the salience of a component of the information flow is captured by its impact on security prices. This approach provides the underpinning of the event study literature. This technique is most useful if the information is discrete, the release date is clear, and the information release is not confounded by other potential news (MacKinlay (1997)). As a result, event studies are used primarily around identifiable corporate announcements (Fama et al. (1969)), government pronouncements (e.g. Leftwich (1981)), or issuance of "professional" statements such as by ratings agencies or equity analysts (e.g. Mendenhall (1991)).

Capturing the information content of the nearly continuous flow of information in news or electronic message board chat is more difficult than in the discrete case, due to the sheer volume of information and the time required to classify each one. In this paper, we extract a subjective measure of the meaning of the information using five computer algorithms that read and categorize the content of each individual message. The algorithms parse from each message a single measure: the degree to which the message conveys a buy, sell or neutral sentiment about a stock. By aggregating these messages over some time period, we can gauge the average sentiment as well as the distribution of sentiment manifested by the stock message board information flow. We call the combination of activity measures and content measures (distribution of sentiment indices) "e-Information." [2] We also use this same technology to classify the "sentiment" of news stories.

Our content measure keys off of investor sentiment because this intangible quality is critical to many of models used in financial economics. In behavioral finance, investor sentiment (or noise trader sentiment) is used to explain deviations in prices from "rational" levels (see Delong, Shleifer, Summers and Waldman (1990)). To measure sentiment, academics have used the closed end fund discount (Lee, Shleifer and Thaler (1991)) and flows into mutual funds (Goetzman, Massa and Rouwenhorst (2000)). We create a new measure of investor sentiment by examining whether a posting could be read as suggesting that others buy or sell a stock.

Deciphering meaning from anonymous stock board postings that average 437 characters each is not a precise task. It strains the limits of modern artificial intelligence and statistical language processing. We use a variety of procedures to provide the classification scheme, and benchmark the results against results of human interpretation. We acknowledge that our measure of sentiment is likely to be quite noisy and will likely evolve over time. (**Appendix A** contains a description of the methodology we use to classify messages, summarized from Das and Chen (2000).) Nevertheless, we are hopeful that this new measure of

[2] Tumarkin and Whitelaw study a subset of postings from one message board, which permits posters to voluntarily classify their short-term opinion about each stock. While these voluntary disclosures are convenient for study, only less than a quarter of posters choose to reveal a "short term opinion," and the board that permits this disclosure accounts for less than 5% of the total postings in our sample.

sentiment will be useful for four reasons:

(a) It reflects the widely available opinions of a set of retail investors, albeit a self-selected group of posters. The information is freely available on the Internet and the web sites that offer this information are widely visited. We recognize that posters may not be a random sampling of investors, but rather vocal individuals who hold extreme views.[3]

(b) The disaggregated observation of sentiment allows us to calculate various distribution measures. We can not only calculate an average measure (net bulls less bears), but also we can calculate measures of the degree to which investors disagree, which is a variable that theoretical research suggests will correlate with both trading volumes and volatility (Kim and Verrecchia (1991)).

(c) The stock-by-stock observation of sentiment is more granular than broad market-wide sentiment indices.

(d) The high frequency of the data allows us to plot changes in sentiment over time.

2.2 Sample selection

Many existing studies of information and security markets study a single information source across a large number of companies, for example, all of the

[3] An interesting issue we do not treat in this paper is the motivation for voluntary postings on stock chat message boards. The theoretical justification for posting might be related to many existing models of information disclosure. For example, suppose we posit that each investor receives a noisy signal about future stock price (e.g., their opinion as to the importance of a new product announcement) . By sharing their signals with others, they can verify the information (or part of it) *before* trading, or can share the signal with others *after* trading with the hopes that their interpretation will lead to the desired movement in share prices. On a more mundane level, stock chat boards can be locations where disgruntled shareholders, customers, employees and former employees can share their experiences with others. Modeling the many reasons why people may care to share their opinions with others, and why readers might put any stock on these opinions is a complex question that is beyond the scope of this paper. Admati and Pfeider (2001) have begun to address these questions in a recent working paper.

press announcements of a particular type of security offering. We seek to examine a wide range of information sources, each of which is incredibly rich, about a wide range of different types of information. Therefore, in this preliminary paper, we chose to study four firms over a period of seven months.[4] Our goal was to use these four different firms to serve as archetypes for firms in different information environments where the traditional information and the e-Information flows varied. We have consciously not selected pathological examples where stock message boards were used to explicitly manipulate prices (Leinweber and Madhavan (2001)).

To select the four firms for our preliminary study, we first collected information on the 3724 firms that had at least one posting on The Motley Fool (TMF) stock message board in the period July 1, 1998 through January 31, 1999.[5] Then we classify each by number of TMF messages. For the 504 firms in the TMF list that had at least 25 posts in the period July 1, 1998 through January 31, 1999, we collected the number of major news stories from Dow Jones News Retrieval. We define a "major news story" as one in which the name of the company either be in the headline *or* be mentioned in the lead paragraph and have at least three occurrences in the body of the article. We stratified the 504 firms into quintiles along two dimensions (number of posts and stories) and selected one firm from each of the four extreme categories of the joint

[4] The two comparable papers also choose to use quite small samples: Tumarkin and Whitelaw (2001) study 73 stocks and Antweiler and Franks (2001) study 45 stocks. We believe that none of these samples, including ours, are large enough to make definitive statements about message board activity, so we decided to exploit the advantages of a clinical study in this paper. However, we are assembling a database of approximately 3000 stocks to permit a more robust econometric study of the hypotheses developed here.

[5] The Motley Fool graciously provided us with the data to perform this screening, but subsequent posting information for this and other boards was collected through a proprietary web crawler program.

distribution. See **Figure 1** for a plot of the joint distribution and for the positioning of our four firms. The four firms are shown below:

		Traditional Information Environment (number of news stories)	
		Rich (High)	Poor (Low)
E-Information Environment (number of posts)	Rich (High)	Amazon.com	General Magic
	Poor (Low)	Delta Air Lines	Geoworks

These four stocks were *not* selected to be representative of the average stock, but rather to help us to understand the extremes of information flow. **Table 1** provides summary statistics on the four firms. Delta Airlines is an old-economy company with a large work force, substantial institutional ownership, and positive earnings. Amazon is (was) a flagship new-economy company. General Magic and Geoworks are small, not very profitable firms attempting to serve the new economy, but General Magic—founded by former Apple Computer executives—has an extremely high level of posting activity for a firm its size.

2.3 The calculation of e-Information measures

Electronic message boards or chat rooms allow users (posters and readers) to share their opinions, analyses or information with others. The four major boards (Yahoo, The Motley Fool, Silicon Investor and Raging Bull) allow users to post anonymous messages, which are available to others in real time.

Messages can reflect a relatively intense discussion about the firm, including new "information" about the firm, strongly held beliefs about the future of the company (and its immediate stock price), and probing questions They can also wander off into discussions that are wholly unrelated to the firm.

Wysocki's (1999) study of posting activity on the web provided the first evidence about chat room activity. In particular, he studied the *number* of postings on 3,478 of the 8,011 firms on the Yahoo! Message boards. He found that certain types of firms are more likely to generate high posting activity. These firms had high short-seller activity, high market values relative to fundamentals, low institutional holdings, high trading volume, extreme performance, or extensive analyst following. In addition, he studied a panel of 50 firms to understand the time series properties of the number of postings. He found that posting volume varied by day of week, increased near earnings announcements, and was related to the volume of trading activity and abnormal stock returns.

Like Wysocki, one of our analyses looks at the number of posts. **Table 2** provides summary statistics on the posting activity for our four firms, which together received over 170,000 posts over seven months. However, our primary interest is in trying to extract some meaning from the messages, in particular the "bullishness and bearishness" of the posts and the degree to which posters seem to agree or disagree.

The Appendix describes the methodology we use to classify messages (and news stories). Briefly, we adopt a voting system.[6] The raw message is read by the program, and the five language processing routines use different rules to classify whether the message is a buy, a sell or neutral. We then count the number of "votes" across the five different measures. Messages that receive at least 3 bullish (bearish or neutral) votes are categorized in that category; otherwise they are not categorized (nc). This methodology bears some resemblance to "fuzzy logic" routines that use the notion of statistical closeness.[7]

While the unit of observation for message classification is the message, we also create *daily* measures of the e-Information (as well as measures for smaller time intervals). The primary measure is a *sentiment index*, defined as the number of buy messages less sell messages (excluding null and not classified messages). The sentiment index picks up the net bullish sentiment and is scaled by the net number of messages. In addition, we calculate a number of other related measures:

- *Sentiment sign*: 1 if sentiment > 0; -1 if sentiment <0, 0 otherwise.
- *Sentiment percentage*: Sentiment index divided by all messages for the day.
- *Opinion index*: Fraction of all messages that are classified as either buy or sell messages. One interpretation of the inverse of the opinion index is

[6] Anweiler and Frank (2001) use two classifiers similar to two of the five we use. We find that no one classifier is terribly powerful, so we adopt a voting rule to determine content. They analyze the two separately, rather than using a voting rule.

[7] We have other information that is not currently incorporated in our analysis, including the length of the post and the stated identity of the poster. As posters use self-chosen nicknames that can be changed at will, we cannot be assured that two posts by different named posters represent two different individuals.

that it represents the extent of questions, non-directional comments and noise in the discussion.

- *Disagreement index*: This index is defined as

$$\left| \frac{|BUYS - SELLS|}{BUYS + SELLS} - 1 \right|$$

or na if Buys+Sells = 0).

This measure is intended to capture whether the opinionated posters have the same view, or whether there is dispersion of belief. If everyone is on the buy or sell side of the market, this index is 0, but the index is 1if half the signed posts are buys and half are sells.[8]

Table 2 reports on the overall categorization of the messages for the four stocks over the entire sample period. For the four firms, we can classify about 40-50% of the messages as either buys or sells. About 6-8% of the total posts are net buys. This low level of sentiment reflects the fact that there tend to be large numbers of both buy and sell messages, shown by the disagreement index of 80% or more (100% would mean that the buys and sells are equally split.) Looking at the daily averages, there is fairly substantial time series variation in each of these measures, which is also shown in **Figure 2** that graphs various e-Information variables for the four firms over our sample period.

[8] To the extent that total messages are highly correlated with signed messages, the disgreement index is like an unsigned version of the sentiment percentage.

2.4 Calculation of traditional information measures

In addition to collecting information from stock chat message boards, we also collect data that would constitute the information set available to investors. This information includes the following:

- *Company-released information*: Press releases and filing information were collected from Dow Jones Interactive, Edgar, and Global Access.
- *Professional monitor information*: New analyst reports and earnings revisions were collected from Investext and from IBES.
- *Electronically-available news media*: Major news stories (as defined above) collected from Dow Jones News Retrieval.
- Intraday stock prices, including bid and ask information, and trading volumes, collected from TAQ.
- Implied volatilities on short-term traded options collected from Bloomberg.

Our sample includes information that can be more-or-less easily obtained by a *retail* investor *without real-time monitoring*. We do not include information that is not stored, such as transcripts of TV and radio broadcasts. In the current analysis, we also exclude information that might be available only to large institutional investors (e.g., conference call proceedings prior to web-broadcasting or private communications with management; see Bushee, Matsumoto and Miller (2001)).

Table 3 breaks down the information events for each of the four companies. In total over the seven month period, there were 168 press releases, 58 filings, 207 analyst forecast revisions, 2,346 major news stories and 170,953 stock chat posts. The dispersion in the information releases is intentionally large as we were attempting to capture four different types of firms. For example, there are 73 times more postings at Amazon than at Delta, but 19 times more stories about Delta than about either General Magic or Geoworks.

3. Describing the information flow

In this section we describe the patterns of information flows for the four securities we study. These patterns include the following: temporal patterns (hour of day, day of week, month of year); autocorrelations of information measures; and cross-correlations among the various information measures. We also report on a few "tracer events" where we trace the flow of particular pieces of information across the various information sources.

3.1 Temporal patterns of activity measures.

We begin by highlighting a few temporal patterns in the information flow for the four stocks with respect to day-of-week and month-of-year effects. See **Table 4**. For these four stocks, we find the following

- Company filings are most likely on Fridays, but press releases are more likely made in the middle of the week. Analyst revisions tend to be issued in the second half of the week. Since both company announcements and professional announcements are relatively infrequent, these patterns probably reflect conscious public relations strategies. News stories and postings also peak mid-week. Unlike company filings, press releases and analyst revisions, which are virtually absent on weekends, press and posting activity continues through the weekend, although at lower levels of activity. The chat board activity is the most active of the information sources, with about 16% of weekly activity taking place on Saturdays and Sundays (or 23% from close of market on Friday to market open on Monday.)

- With respect to monthly patterns, company filings seem to have been reduced in July and December, probably reflecting summer vacations and the holidays. Analysts release few new earnings projections in August and December--again probably due to summer and winter holidays. News stories and postings show less pronounced monthly trends.

While these temporal patterns are not themselves interesting, it is helpful to understand them when controlling for other effects.

3.2 Autocorrelations.

Table 5 reports day-to-day autocorrelations among the information variables after correcting for day of week effects. As would be expected, there is little autocorrelation in the periodic release of information in the form of press releases, and some negative autocorrelation in filings (firms filing yesterday are less likely to file today.) There is some autocorrelation or bunching in release of analyst reports.

However there is strong autocorrelation in news story volume and postings volume, as well as levels of sentiment and disagreement. These high autocorrelations suggest that these news sources may reflect analysis--the slow and persistent digesting of information. This is not to say that there are not shocks to posting volume, sentiment or disagreement, as shown in **Figure 2**. Rather these innovations must be understood in light of the persistence of the information flow variables.

3.3 Cross correlations

By looking at contemporaneous correlations among the various information sources, we get a sense of how the parts of the information flow interact. **Table 6** reports the median correlations for the four firms as well as the number of these correlations that are significant. Not surprisingly, there are fairly strong contemporaneous correlations between company press releases and news stories. There is also a strong correlation between news stories and posting volume. Sentiment is positively correlated with both the number of news stories and even more so with the level of posting activity, possibly suggesting that any news is good news. Finally, there is a positive correlation between increased posting activity and disagreement, suggesting that increased discussion and disagreement go hand in hand (If everyone agreed, why post?).

In **Table 7**, we report on the number of postings and news stories around three information "events": company press releases, company filings and analyst earnings revisions. We observe that there is more press activity and posting activity not only on days when press releases and analyst reports are issued, but also on subsequent days. This analysis differs from the data in Table 6, which showed no contemporaneous correlation between these events and posting activity; here we see an increase in next day activity, especially after firm press releases.

3.4 Tracer events

One advantage of clinical studies is that they permit some types of analyses that do not easily lend themselves to large-scale econometric test. We sought to understand the leads and lags in information flow. To get a sense of this flow, we select 16 seemingly material press releases by the four companies, and for each trace the “electron trail” of how the “news” were communicated to investors through news stories and postings. (**Table 8**, **panel C** lists the events). For each, we sought to understand how the press or message boards provided advance information of the event, and how they responded to the event. With respect to response, we measured the speed from the press release to the first discussion on the message boards, the time series of subsequent discussion, and the nature of the discussion. The evidence in Table 8 is anecdotal, but nonetheless helps us to understand the nature of the manner in which information is shared. We found the following patterns:

1. *Message boards provided some factual foreshadowing of subsequent press releases*. In quite a few instances, posters provided readers with advance warning of subsequent news events. For example, one of our tracer events is the spin-off off the DataRover division by General Magic. Nine days before the

DataRover spin-off, someone reported on the Yahoo message board that they had found a new DataRover URL that did NOT mention General Magic. This site was apparently taken off line in a few hours by General Magic, which was probably testing the site for the imminent spin-off. Readers of the board could not only read the message, but also confirm it by going to the site. In another example, one day before General Magic announced an agreement with Microsoft, someone posted that the two looked like they would share a booth at the Consumer Electronics Show, a tip-off to some closer relationship. A third General Magic poster alerted readers to a local radio broadcast which suggested that the firm would enter into an agreement with Intuit, which was not publicly announced until a few hours later. In other instances, posters speculated about upcoming splits and bond issues. In two others, we see advance discussion of upcoming earnings numbers, or so-called "whisper numbers" as studied by Bagnoli et al. (1999). Put together, this anecdotal evidence suggests that posters provide active surveillance, especially of smaller companies, before any press reports pick up news events.

2. Rapid postings disseminate company information quickly. Companies tend to issue press releases either before markets open (sometimes in the middle of the night) or after markets close. **Table 8, panel C** shows the number of minutes between the time-stamped on each press release and the first posting of the news on one of the stock chat boards. For many, but not all, announcements, the first post is incredibly soon after the news is posted, and in a few cases, prior to the first major news wire story. First posts tend to contain a short notice of the news, often with a URL directing readers to the press release. The boards are apparently serving to disseminate information to interested investors quite quickly.

3. Extensive on-point discussion is sustained for eight hours after news releases. Table 8, panels A and B characterize the postings that followed these 16 events. Panel A shows that in the few hours immediately after a news event, posting volume rises, but then tails off over time. Panel B displays the composition of the posts over the first eight post-news hours. The nature of the discussion was measured by categorizing each subsequent post into one or more of five possible categories (asks question, offers alleged fact, shares opinion, comment unrelated to news event and spam/garbage). The first three categories of posts can be thought of as on-point postings, i.e., ones that relate to the news at hand. We see that for the first four hours after a news event, over two-thirds of all posts are on-point, and even eight hours later, about half are still discussing the news (as opposed to other issues or spam).

4. The on-line discussion is mix of questions, answers and opinions. We categorize on-point posts as asking a question, offering an alleged fact, or proposing an opinion about the meaning of the news. For the first hour, we see more of a question-and-answer pattern, with a quarter of all posts and a third of the on-point posts either asking a question or supplying a fact. Over time, the discussion tends toward more analysis: What does this news event mean for the company and its stock price? Much of this later discussion does not focus on facts, but rather on interpretation of facts.

These clinical observations about this handful of news events are not conclusive, but they are supportive of the notion that chat boards may contain kernels of new information, fairly rapid information dissemination, and sustained discussions of interpretation of news—features which could lead some investors to read them and base opinions on the arguments advanced in them. Unfortunately, this wheat comes with a large measure of chaff, and whether the boards influence opinions is ultimately an empirical question.

4. e-Information and the price formation process

Market efficiency implies that public information is immediately embedded in the stock price (Fama (1965 and 1975)). If message boards just rehash old information, then e-Information would not make markets any more or less efficient nor predict stock returns. However, message boards and e-Information would be useful if the boards provide new information about the firm (whether new facts, new analysis of public facts) or about the state of mind of noise traders whose actions might affect prices. We use simple specifications to explore whether stock returns are related to our measure of sentiment and other information state variables. We would expect that this sentiment index would be most salient for firms where the traditional information environment is poor and the chat board information environment is rich (General Magic). To be clear, our null hypothesis is that stock returns are NOT predictable using these measures.

We remind readers that with a clinical study of four stocks (or even with 45 or 73 stocks), one cannot reliably test these propositions, rather only suggest that the measures we have constructed may have enough merit to warrant large-scale study. Furthermore, our four-firm research design is designed to investigate the time series of price formation process. However, there are likely large cross sectional differences among firms. Most importantly, we would expect that any effect of the e-Information on volume or volatility should be greatest for firms like General Magic which receive little media attention, and smallest for firms like Delta which are well covered in the traditional media.

4.1 Correlations.

We report in **Table 9** the contemporaneous correlations between our information variables and the market variables (returns, excess returns, share turnover, implied volatility and intraday volatility and average bid-ask spreads) as well as the autocorrelations of the market variables.

For the two small firms, there is a significant contemporaneous correlation between the number of news stories and market returns, but not so for the two larger firms. For the two firms with substantial chat board activity, there is a contemporaneous correlation between the sentiment measure and returns, and a negative correlation between disagreement and returns (when people disagree, returns tend to be lower). This might suggest that e-Information becomes salient (with respect to volume, volatility or spreads) when it is quite extensive.

While the contemporaneous correlation between information variables and returns is modest, the contemporaneous correlation between the e-Information variables and turnover, volatility, bid-ask spreads and jumps is more robust, especially for the most actively discussed firms in our sample (Amazon and General Magic). These non-return aspects of the financial markets are also correlated with one another, as shown on the right hand part of Panel A, and highly autocorrelated, as shown in Panel B.

4.2 Are returns for these four stocks explicable using e-Information?

For the four stocks we study, we examine whether the e-Information and other information variables help to explain returns. If e-Information variables are meaningful, at a minimum we should observe contemporaneous correlations between them and returns. If they contain truly new and novel information, perhaps we might observe e-Information predicting returns.

For stock returns, we use not only close-to-close returns, but also open-to-close returns. We use this latter measure because we seek to understand whether reading posts from the day before and *the night/early morning prior to market opening* would permit a trader to predict subsequent returns. Because a bleary-eyed trader could at best transact at the open price (not the prior night's close), the traditional close-to-close return could not be executed.

In this draft, we focus on tradable quantities (stock returns and option implied volatilities) versus non-tradable aspects of the price formation process (bid-ask spreads and volume), which we are currently studying.

We use a simple specification, similar to the ones used by Wysocki (1999) and Mitchell and Mulherin (1994). **Table 10** defines the variables we use, which include the following:

- Announcements, including
  - Press releases by the firm
  - SEC filings by the firm
  - Analyst revisions

- News story information
  - Abnormal measures of news stories are created by calculating the residual when the measures are regressed against their lagged values as well as day of week and month dummies.
  - Sentiment indices from news stories.

- Posting information, including
  - Abnormal number of market posts in the period 9:30 am to 4:00 p.m. on a trading day, as well as the abnormal number of "premarket" posts from market close the prior trading day through 9:30 am the current trading day.
  - Sentiment using the same time conventions.

- Interactions: The first three events (press releases, filings and analyst reports) are probably most salient if they generate news, chat or change in sentiment. Therefore we create interaction variables by multiplying them by the abnormal news levels, news sentiment, abnormal posting levels and posting sentiment measures.[9]

We include both contemporaneous and lagged information variables. We want to see if there is a relationship with contemporaneous variables to verify if our constructed measures are sheer noise or whether they are picking up signals that confirm the current state of the market. We use lagged information sources to determine whether the information seems to be useful in predicting returns, or whether it is probably already impounded in stock prices (or irrelevant).

Table 11 provides the results of this inquiry. Panel A looks at the relationship between contemporaneous information measures and close-to-close market-adjusted returns. Panel B repeats this analysis using close-to-close market adjusted returns. Both of these panels show whether returns are related to the *same day's* announcements, news activity, and posting activity. Evidence

[9] The specifications we show only include the press release interaction terms, as the others tended not to be material.

of significant relationships would not suggest market inefficiency; to the contrary, it would be consistent with information production being impounded quickly into prices.

Panels C and D repeat the analysis in the first two panels, except that all of the independent variables are lagged. Announcements and news levels are from the prior day, postings from the period from the last market close to the current market open of the market. Relationships here would suggest that a trader could observe data today (until the open of the market) and profit from it.

Panels A and B, which examine contemporaneous information flows and returns, hint that the information environment has a rich set of interactions with the price formation process. For these four stocks, there are suggestions that stock returns may be higher on days when:

- There are more news stories (AMZN, GMGC);
- This news conveys more positive sentiment, as measured by our sentiment algorithm applied to the text of news stories (GMGC, GWRX);
- The message board postings reflect greater positive sentiment (especially for stocks with more active postings; AMZN, GMGC); and
- There is a company press release *combined with* more news, more positive news stories, and especially more positive posting sentiment (AMZN, GMGC, GWRX).

Separately, stock returns seem to be lower on days when firms issue press releases.

Our sentiment index is more closely related to contemporaneous prices than is sheer numbers of postings, the measure used by Wysocki (2000). These results are encouraging, in that they suggest that our sentiment index, applied to either short messages or longer news stories, apparently captures the tone of the text. Furthermore, news and sentiment measures can help us to “sign” various news events, in this case, company press releases. Not surprisingly, press releases seem to gain importance as they are interpreted by the news and by board posters.

Panels C and D are also encouraging, at least for those who are proponents of informationally-efficient markets. Using information that arrives prior to the opening of the market, including overnight posting activity, we find that no "information" variable is consistently informative to a trader who will transact over the course of the day. Of the 43 firm-coefficients in each panel (4 stocks x 10 or 11 variables), 3 or 4 are significant at the 10% level, just as chance would predict. For close-to-close returns, overnight sentiment is predictive in two cases, but one could not read the posts in the evening and then trade at the earlier price, so this information is of little use to a trader.

Our results are consistent with those found by Tumarkin and Whitelaw (2000) and by Antweiler and Frank (2001). Our three papers were all independently produced and use different samples and different methodologies for coding the information content of the message boards. Nevertheless, all three of the papers show no predictive power for the message boards to explain subsequent stock returns. In a word, all three of these small sample papers suggest that people trade first and talk later, so that returns precede postings, rather than the other way around. However, we recognize that while overall, there may not be predictive power of postings to explain returns, in a large sample, we may be able to detect that certain classes of postings, certain patterns of postings, or certain types of posters may be more informative and predict returns.

4.3 Implied volatility

In **Table 12**, we examine whether implied volatility is explicable using the information variables we study. Unlike other papers that have looked at historical volatilities (standard deviations of returns), we focus on *implied* volatilities, which are measures of the traded prices of options. An options trader who gained an "edge" on predicting tomorrow's implied volatility would likely profit.[10]

Panel A asks whether contemporaneous information is related to option volatilities, and Panel B asks whether lagged information has any predictive power. As is predictable, the prior day's implied volatility is a strong predictor of today's volatility.

The remainder of the results are less strong, but not easily dismissable. For Amazon, there are strong and persistent contemporaneous relationships between information variables and implied volatilies, and a few relationships for General Magic. The results for lagged data are suggestive: Six of the eight announcement dummies are negative (two significantly so), hinting that announcements tend to reduce uncertainty and volatility. All of the abnormal news story variables are negative (one barely significant), again consistent with the notion that new information may reduce volatility. For Amazon and General

[10] Our measure of implied volatility comes from Bloomberg. It is a weighted average of mostly near-dated call options, as calculated by Bloomberg. A more complete test would use the implied volatilities (or prices) of individual options.

Magic, the two firms with substantial posting activity, a number of the lagged press release, news, and interaction terms are close to levels of statistical significance. We are reluctant to draw conclusions from this clinical study, but it leads us to hypothesize that while the market may anticipate news, the actual release of it may reduce uncertainty, thus affecting options prices. We are collecting a large sample with appropriate option data to test this hypothesis.

5. Preliminary findings

Our goal in this paper is to introduce the concept of e-Information, which includes the distribution of investor sentiment information as extracted from stock chat message boards. We define this concept, briefly describe its calculation, and relate it to other forms of information that might affect stock prices. We document certain patterns in the release of information, in particular, we were surprised to find that chat boards were often quicker to disseminate information to readers than traditional news media. We characterize chat board traffic, and were surprised to find that after news events, posters stayed "on-point," and the boards' function changed from a forum for asking and answering questions to one of opinion sharing.

We find suggestions of links between our measure of e-Information and contemporaneous stock returns. However, our work is consistent with other small sample papers that fail to find that message board activity predicts stock returns. However, we remain agnostic about the relationships among e-information, more traditional information and implied volatilities from options prices. In subsequent work using a comprehensive dataset, we intend to probe these relationships, as well as relationships among e-information, volume, volatility and bid-ask spreads.

Appendix: Classifier Algorithms

The classification algorithms in this paper were developed from several different ideas in the field of linear algebra and statistical theory. The algorithms in this paper are decribed in Das and Chen (2000). Earlier work in a different text classification domain comes from the work of Koller and Sahami (1997) and Dom, Chakrabarti, Agrawal and Raghavan (1998).

Instead of a single classification algorithm, the sentiment index comes from the majority vote of five distinct algorithms. Messages that are ambiguous tend to be characterized by low agreement amongst diverse classifiers, and hence are better discarded as noise, a feature that is implemented by the voting structure. Keeping messages with high agreement amongst classifiers results in better signal detection. The voting algorithm discards on average about 20% of the data. The accuracy of the algorithms tends to be in the 65% range. Human agreement on message classification is slightly higher at about 72%.

The algorithms classify messages into 3 categories: (a) bullish, (b) bearish and (c) neutral/null content. The algorithms are based on a set of keywords carried in a “lexicon”. The lexicon contains words that have meaningful finance implications, and are hand picked or chosen via the use of a discriminant function. Each word in the lexicon is tagged with a bullish or bearish flag.

The classification process also uses a training data set or “grammar.” We pick a set of messages from internet boards and hand classify them. These messages form the basis for further classification of messages. We may also think of these messages as being a base set which reflects the types of messages that we may see. For example, we may assess the “closeness” (using some metric) of messages to this base grammar, and use this measure as a basis for assigning a classification.

In the following paragraphs, we describe the five algorithms in brief. Technical discussion is avoided and the reader is referred to the original papers for more details.

Naive classifier: This algorithm is the simplest of the ones we used. It consists of simply undertaking a count of the words in the message that appear in the lexicon. Each word that appears is assigned a value of –1 if it is signed as bearish in the lexicon and is assigned a value of +1 if it is signed as bullish. The net count for the message is made, and if it exceeds a given positive threshold T, we assign the message a buy classification. If the message count is less than a lower threshold $-T$, it is assigned a sell classification.

Vector Distance Classifier: Each classified message in the training set is coded as a vector of word counts. If the training set is of size M, then the number of word vectors is M too. The total length of each word vector is the size of the lexicon, denoted *L*. The matrix of training messages is denoted *X*, and is of dimension (*L* x *M*). Every incoming message is also coded as a vector of words, which may be denoted Y, which is of dimension (*L* x 1). The vector distance between any message in the training set, say X_i and the message vector *Y* is given by the angle between the two vectors, given by the standard formula for vectors, i.e. $\cos(\theta)=X.Y/(|X||Y|)$, which is a number between 0 and 1. ($|X|$ stands for the norm of vector *X*, and *X.Y* is the dot product of vectors *X* and *Y*). We classify the message *Y* as having the same class as the message from the training set to which it is closest in the training set, i.e. the one with the highest value of $\cos(\theta)$.

Discriminant classifier: In this algorithm, we also undertake a word count, but weight each word in the message by its discriminant value from the training set. The discriminant for a word *w* is given by the ratio of the across-class variance to its within-class variance. This is a simple way to find words that give better classificatory performance on the training data set. To fix intuition, lets suppose that the word "bullish" always appears exactly twice in every buy message and never appears in a message in any other category. The discriminant function for this word would be infinity as its numerator, the across class variance, would be a finite number, and the denominator, the within class variation, would be zero. Once a weighted average word count is taken, the classifier scores the message and accordingly determines its buy, sell or null category.

Adjective-Adverb Phrase classifier: This classifier is based on the simple notion that the parts of a message that contain more emphasis are those that have adverbs and adjectives. To identify the parts of speech of each word in the message, we use a dictionary for parsing each word. Whenever we come across an adjective or adverb, we take triplets of words around the message and submit them for lexical analysis. The score for the message determines its classification.

Bayesian classifier: This classifier uses the training set as a prior and finds the posterior probabilities for each message to undertake classification in a Bayesian setting. The classifier has three elements: words (*w*), messages (*m*) and categories (*c*). Our final goal is to obtain the conditional probability of the category given a message, i.e. $P(c|m)$. Using Bayes theorem, this is equal to:

$$P(c\,|\,m)=\frac{P(m\,|\,c)P(c)}{\sum P(m\,|\,c)P(c)}$$

The probability $P(c)$ is simply the proportion of messages in category *c*. The probability $P(m|c)$ is multinomial over the words in the message. This is given by:

$$P(m\,|\,c)=\binom{n(m)}{n(m,w)}\prod_{k=1}^{W}\theta(c,w)^{n(m,w)}$$

where $\theta(c,w)$ is the probability of seeing word *w* in category *c*. *n(m,w)* is the number of words *w* in message *m*. These basic statistics are developed via word counts on the training set.

Implementation: We used these five classifiers, and the results from these classifiers was fed into a voting scheme where we choose to undertake classification of a message when 3 or more of the 5 classifiers agreed on a classification. If we do not get this simple majority, the message is discarded.

We created a training data set on which the classifiers were trained. The classifiers with the voting scheme were then tested out of sample on a separately classified data set. The following is the brief summary of tests performed using the classification technology:

1. Our tests showed that this approach provided an accuracy of attempted classification of around 63%, i.e. the methodology was correct about 2 out of 3 times, compared with random choice accuracy levels of 33%.
2. The voting scheme results in discarding messages which do not attain a simple majority amongst classifiers. While this is a good way to remove noise and retain the signals, it may result in throwing away too many messages in case the level of disagreement amongst classifiers is too high. However, we found this not to be the case. A simple majority voting scheme results in a loss of approximately 20% of the messages.
3. Of the messages classified incorrectly, we are more concerned with messages that were actually buys and were classified as sells, and vice versa. There is less concern that neutral messages were classified as buys or sells and vice versa. We found that the level of error in classification where buys and sells were mistaken for each other was lower than the average error, and is about 20%, compared to about 33% on average. This is reassuring, since it means that the sentiment index is less likely to be erroneous in an extreme way.
4. The test set is hand classified and then compared to the machine classification to give an approximate accuracy rate of 2/3. However, the high level of ambiguity of the messages implies that even human agreement on the messages may be reasonably less than 100%. Hence, we ran some tests to see what the level of human-human agreement was. On average, we found this to be 72%, not substantially higher than that of machine-human agreement.

In the following paragraphs we examine a few messages and report their classifications. These messages were posted to the Yahoo message board for Delta Airlines (ticker: DAL). The message and its classification are provided below. The messages are reported verbatim, including spelling and grammatical errors.

Message	NC	VDC	DISCR	AAPC	BC	Vote
Curious communication: does the management of this company often use apublic message board to communicate matters ofemployement issues? when has any manager of this company everresponded to sniper attacks on this silly board? there havecertainly been many arguements in the 700+ messages to thisboard about mechanics vs pilots, ground workers vsmanagement and so on. i do not believe you are the real donburkett. i think you play with these people.	0	0	0	0	1	0
contact if you are who you say you are contact one of the schedulers by company e-mail for any meeting you want to call, not some suspected ringleader you want to crucify.	0	0	0	0	0	0
Skippy's informed: it looks like im going to have to buy a lap topto keep hawkster off my a___.i was in orlando thisweekend and was told that dal express is kickingbutt,exceeding all expectations.the only problem is that theycan not get enough airplanes fast enough.does anybodyknow what dal got for the sale of the greenbriarcenter?i hope the guys out in dfw get their problems takencare of.as of today june 29 the system load factor isnear 80 percent,looks like a great quarter to me.	3	3	3	3	1	3
Month loads how do loads look compared to june 97? from what else i see in the industry, i am guessing flat. july bookings, however, look terrific. i'd also bet that june yields were strong.	3	1	3	3	0	3
So are you guys cutting commissions tomo rumor, now reported by holly hegeman, says dal and ual are cutting agency commission to 4% tomorrow. i am skeptical of this report, but i am glad i am not a travel agent.	1	1	1	1	1	1
Res Commissions: if it is true, then that is part of the fullmeaning of the proposed "alliances"--reduce cost to enhanceshare holder value..and do not forget, deregulationmade all this possible.. maybe all the ramprats should vote for a union, so they would have anofficial mouth piece, rather than boring us withtheir personal bitching--this is an investment board--buy lowand sell high or short the stuff and leave yourpersonal bitching at home..that is whatwife/husbands/friends are for... maybe a stock split is in theoffering to further dilute the esopholdings... cheers..	0	3	0	0	3	0
Employee Griping employee griping remember one thing about this board and employeegriping, employee moral does play a big part in companyearnings. if employees are not happy then that reflectsattitude towards the customer. i agree, i do not like thebitchin either to an extent but i am glad to know all thefacts about the company not just somebody telling mehow great delta airlines is, (which by the way is nottrue anymore) because they let go some of the greatestpeople they ever had with 7.5.	1	1	0	1	1	1

Sources

Admati, Anat R. and Paul C. Pfleiderer, 2001, Disclosing Information on the Internet: Is It Noise or Is It News? Unpublished manuscript.

Antweiler, W. and Murray Frank, 2001, Is all that talk just noise? The information content of internet stock message boards, unpublished manuscript.

Bagnoli, Mark, Messod Beneish, and Susan G. Watts. "Whisper Forecasts of Quarterly Earnings per Share." *Journal of Accounting and Economics* 28.1 (1999)

Bernard, Victor L. and Jacob K. Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics* 13, 305-340.

Brooks, Chris, 1998, Predicting stock market volatility: Can market volume help? *Journal of Forecasting* 17 (1), 59-80.

Bushee, Brian, Dawn Matsumoto and Gregory Miller, 2001, Open versus closed conference calls: The determinants and effects of broadening access to disclosure, Unpublished manuscript.

Chakrabarti, S., B. Dom, R. Agrawal and P. Raghavan, 1998, Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies, *The VLDB Journal*, Springer-Verlag.

Coval, Joshua and Tyler Shumway, 2001, Is sound just noise? *Journal of Finance* 56 (5), 1887-1910.

Das, Sanjiv and Mike Y. Chen, 2000, Sentiment parsing from small talk on the Web, Unpublished manuscript.

Delong, J. Bradford, Andrei Shleifer, Lawrence Summers and Robert J. Waldman, 1990, Positive feedback investment strategies and destabilizing rational speculation, *Journal of Finance* 45 (2), 379-395.

Fama, Eugene, 1965, The behavior of stock market prices, *Journal of Business*, 38, 34-105.

Fama, Eugene, 1975, *Foundations of Finance*, Basic Books, New York.

Garbade, K. and W. Silber, 1978, Technology, Communication and the Performance of Financial Markets: 1840-1975, Journal of Finance 33, 819-832.

Glosten, Lawrence R., 1987, Components of the bid-ask spread and the statistical properties of transaction prices, *Journal of Finance* 42, 1293-1307.

Glosten, Lawrence R. and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.

Goetzman, William, Massimo Massa and Geert Rouwenhorst, 2000, Behavioral Factors in Mutual Fund Flows, unpublished manuscript.

Hasbrouck, Joel, 1996, Modeling market microstructure time series, Unpublished manuscript.

Ho, Thomas and Hans R. Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47-73.

Holthausen, Robert W. and Robert E. Verrecchia, 1988, The effect of sequential information releases on the variance of price changes in an intertemporal multi-asset market, *Journal of Accounting Research* 26 (1), 82-106.

Karpoff, Jonathan M., 1987, The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis* 22 (1), 109-126.

Kim, Oliver and Robert E. Verrecchia, 1991, Trading volume and price reactions to public announcements, *Journal of Accounting Research* 29 (2), 302-321.

Koller, D. and M. Sahami, 1997, Hierarchically classifying documents using very few words, International Conference on Machine Learning 14, Morgan-Kaufman, San Mateo, CA.

Kothari, S.P., 2000, Capital markets research in accounting, *Journal of Accounting and Economics*, forthcoming.

Lee, Charles, Andrei Shleifer and Richard Thaler, 1991, Investor Sentiment and the Closed-end fund puzzle, *Journal of Finance* 46 (1), 75-109.

Leftwich, Richard, 1981, Evidence of the impact of mandatory changes in accounting principles on corporate loan agreements, *Journal of Accounting and Economics* 3, 3-36.

Leinweber, David and Anath Madhavan, 2001, Three hundred years of stock market manipulation, *The Journal of Investing* 10 (7/Summer), 7-16.

MacKinlay, A. Craig, 1997, Event studies in economics and finance, *Journal of Economic Literature* 35, 13-39.

Mendenhall, Richard R., 1991, Evidence on the possible underweighting of earnings-related information, *Journal of Accounting Research* 29 (1), 170-179.

Mitchell, Mark and J. Harold Mulherin, 1994, The impact of public information on the stock market, *Journal of Finance* 49, 923-950.

O'Hara, Maureen, 1995, *Market Microstructure Theory*, Blackwell Publishers, Cambridge, Mass.

Samuelson, Paul, 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review*, 6, 41-49.

Stickel, Scott E. and Robert E. Verrecchia, 1994, Evidence that trading volume sustains stock price changes, *Financial Analysts Journal* , November-December, 57-67.

Stoll, Hans R., 1978, The supply of dealer services in security markets, *Journal of Finance* 33, 1133-1151.

Tumarkin, Robert and Robert Whitelaw, 2001, "News or Noise? Internet message board activity and stock prices," *Financial Analysts Journal* 57, 41-51.

Wysocki, Peter, 1999, Cheap talk on the Web: The determinants of postings on internet stock message boards, unpublished manuscript.

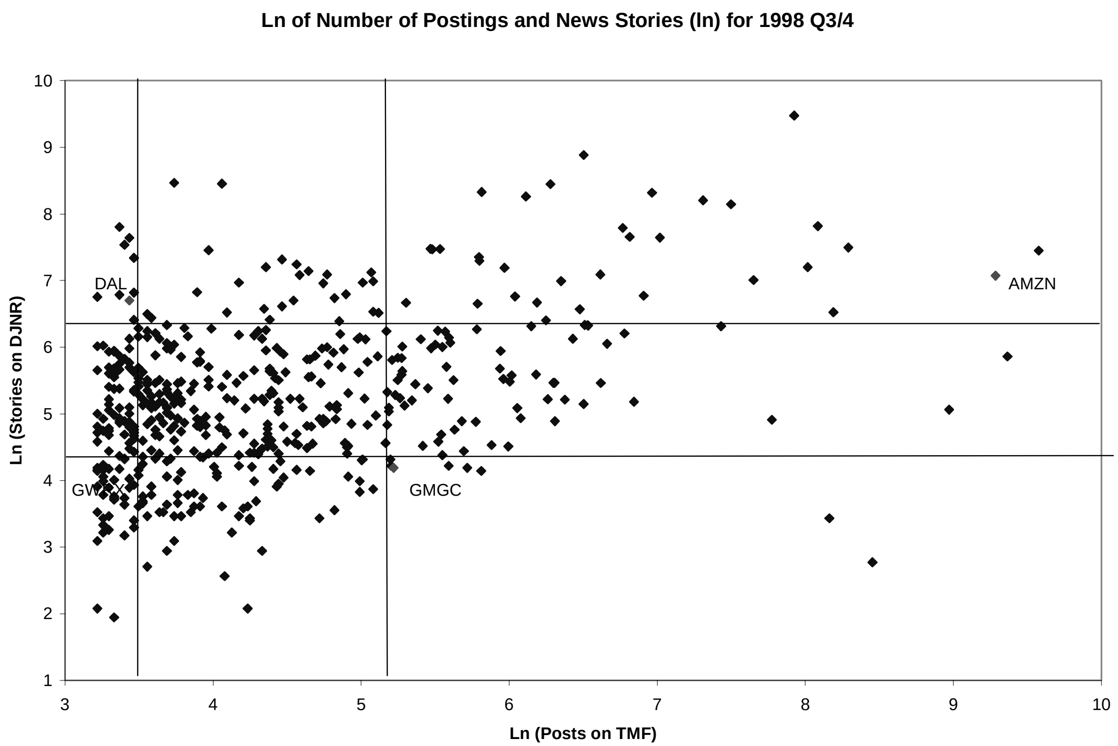


Figure 1: Distribution of posting volume and number of news stories, 1998 Q3/4.

This figure plots for each of 504 firms that received at least 25 posts on The Motley Fool over the period 7/1/98-1/31/99 against the number of major news stories recorded in Dow Jones News Retreival. A "major news story" is defined as one in which the firm's name is either in the title of the story or in the lead paragraph (plus three additional mentions.) The horizontal and vertical bands represent the top and bottom quintiles along each dimension and the four sample firms are identified (DAL = Delta Air Lines, AMZN = Amazon.com, GWRX = Geoworks, and GMGC = General Magic).

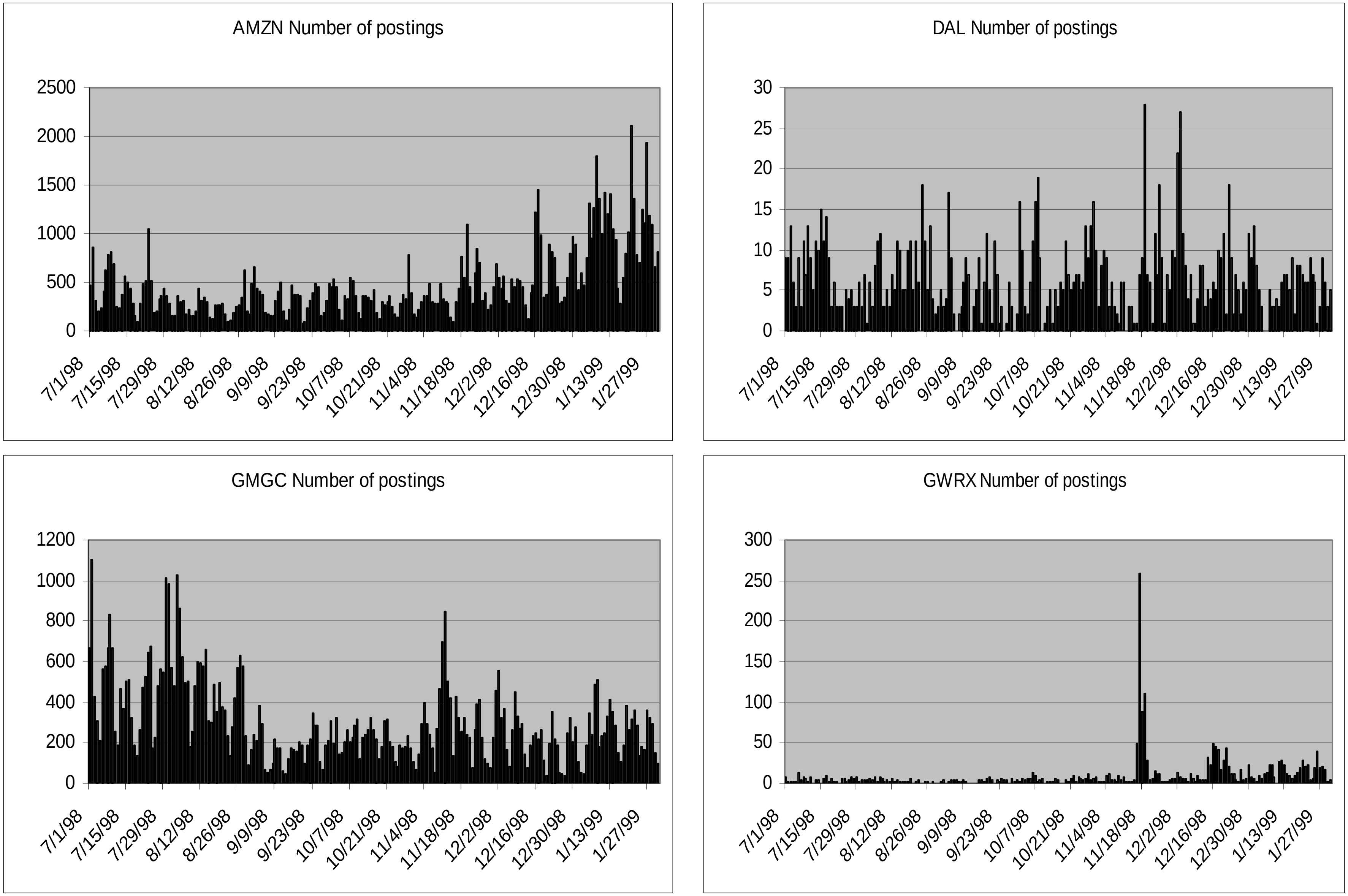


Table 2a: Numbers of Postings by Day for the four sample companies

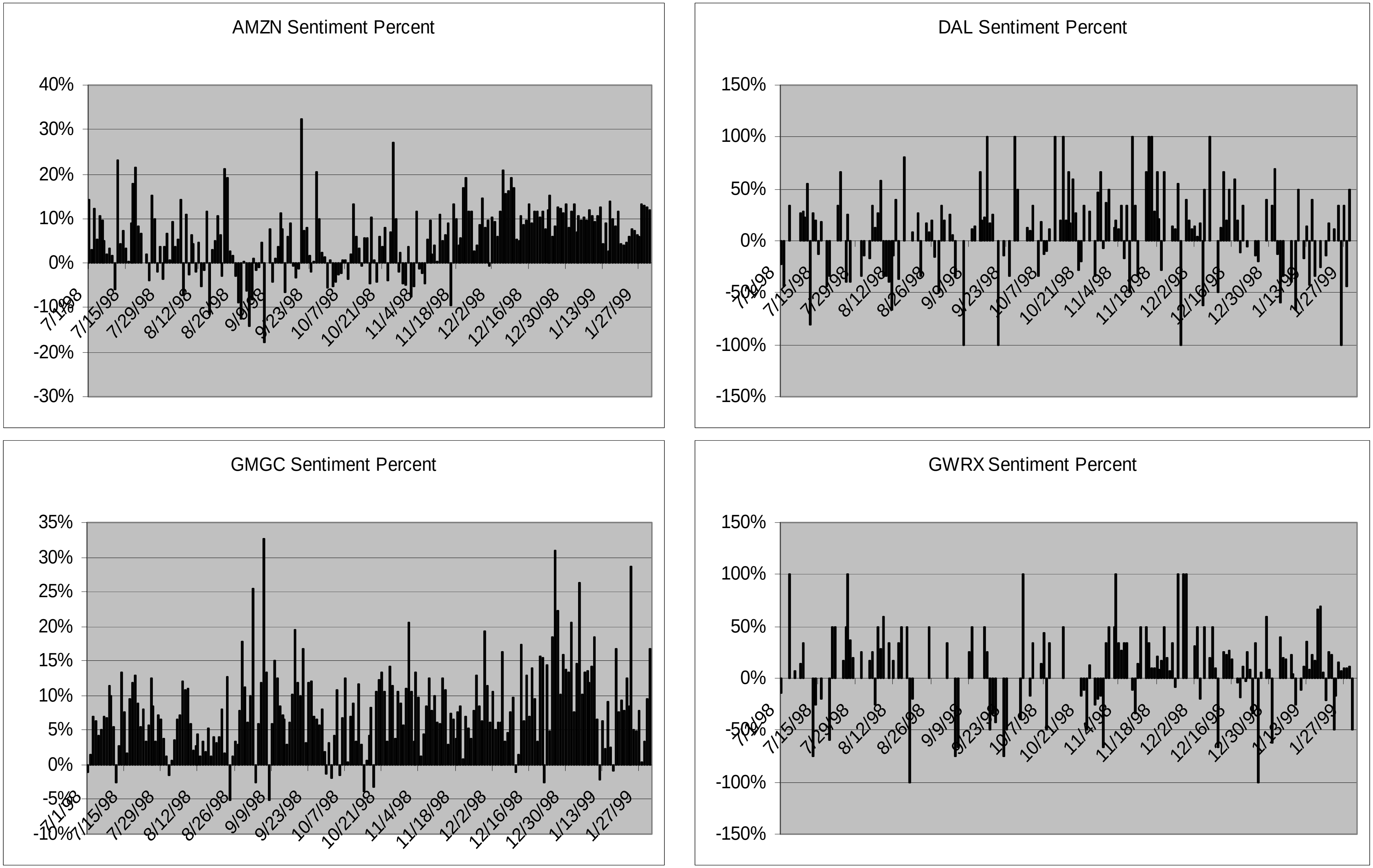


Table 2b: Sentiment Percentage by Day for the four sample companies
Sentiment percentage is calculated as (Buy-Sell Messages)/Total number of postings.

Table 1
Characteristics of the four companies studied

The table below provides basic business and financial information for the four companies we study. The data sources and dates are given in the table. The information for this table comes from onesource.com, Hoovers, Bloomberg and public filings. All $ in millions. Financial figures are as of the end of each company's fiscal year: 12/31/98 for Amazon.com and General Magic, 6/30/99 for Delta Air Lines, and 3/31/1999 for Geoworks.

Firm	**Amazon.com, Inc.**	**Delta Air Lines, Inc**	**General Magic, Inc.**	**Geoworks Corporation**
Business	On-line retailer	Major air carrier	Voice Application Service Provider to telecom and Internet companies	Provider of wireless software solutions
Web site	amazon.com	delta-air.com	genmagic.com	geoworks.com
Industry	Retail (specialty; non apparel)	Airlines	Sofware and programming	Communications services
Stock listing (ticker)	NASD (AMZN)	NYSE (DAL)	NASD (GMGC)	NASD (GWRX)
Market value (Year end)	17054	7984	168	56
Year founded	1995	1924	1990	1983
Total assets	2471.6	16750	36.3	18.2
Total sales	1639.8	14597	2.3	8.8
Net income	-720	1101	-38.9	-15.8
Institutional ownership	30%	75%	10%	15%
Number of institutions	442	776	61	54
Bond rating	B	BBB	not covered	not covered
Number of analysts (equity + fixed income)	26 + 2	14 + 11	4+ 0	5+0
Number of employees	2100	74000	169	110
Avg trading volume (M shares / day)	27.75	1.19	1.04	0.28
Avg Volume (% Outstanding)	9.37%	0.83%	3.50%	1.73%
Average $ value of trades/day ($ mil)	$555	$65	$8	$0.6

Table 2
e-Information Variables for the four companies studied

The table below shows the classification of the messages posted on the four major stock message boards (Yahoo!, The Motley Fool, Silicon Investor and Raging Bull) for the period July 1, 1998 through January 31, 1999 for the four stocks. The top panel shows the message classification for the entire period, using the algorithm described in the Appendix. "Opinion" is defined as the percentage of all messages which are either buys or sells. "Sentiment" is defined as the net number of buy minus sell messages. "Sentiment %" divides the sentiment measure by the total number of messages. "Disagreement" is defined as ||Sentiment|/(Buy + Sell Messages) -1|.

Company	AMZN		DAL		GMGC		GWRX	
TOTAL PERIOD								
MESSAGES								
Buy	29367	29%	404	29%	15276	24%	557	28%
Sell	23017	22%	293	21%	10949	17%	372	19%
Neutral	36916	36%	535	38%	31854	49%	813	41%
Nonclassified	13363	13%	166	12%	6835	11%	236	12%
Total	102663	100%	1398	100%	64914	100%	1978	100%
OPINION	51%		50%		40%		47%	
SENTIMENT	6350		111		4327		185	
SENTIMENT %	6%		8%		7%		9%	
DISAGREEMENT	88%		84%		84%		80%	
DAILY AVERAGES								
No message days	0		10		0		20	
OPINION								
Mean	52%		45%		41%		45%	
Median	52%		50%		41%		50%	
Std Deviation	5%		25%		5%		27%	
SENTIMENT								
Mean	30		1		20		1	
Median	20		0		17		0	
Std Deviation	43		2		17		3	
DISAGREEMENT								
Mean	85%		43%		80%		43%	
Median	87%		50%		82%		50%	
Std Deviation	10%		41%		13%		40%	

Table 3
Information Events for the Firms in the Sample

The table below shows the total number of information events (press releases, filings, analyst reports/revisions, major news stories and posts) for the four sample firms over the time period July 1, 1998 through January 31, 1999.

Firm	**Amazon**	**Delta**	**General Magic**	**Geoworks**	**Total**
News quintile	High	High	Low	Low	
Chat quintile	High	Low	High	Low	
Press releases	22	109	20	17	168
Filings	26	10	12	10	58
Analyst reports/revisions	135	68	-	4	207
Major news stories	987	549	66	65	1,667
Postings	102,663	1,398	64,914	1,978	170,953
Total	103,833	2,134	65,012	2,074	173,053

Table 4
Temporal Patterns in Activity Measures

The table below shows the total number of information events (press releases, filings, analyst reports/revisions, major news stories and posts) for the four sample firms, and the breakdown by day and month for each of the types of information. For postings, we also report time-of-day where "weekday mkt open" is M-F from 9:30 am to 4:00 p.m., weekend is F 4:00 p.m. to M 9:30 am, and weekday mkt close is the other times. These numbers are equally-weighted, giving equal weight to each firm-information event, rather than firm weighted. There are differences among the four firms, for example, the dip in posting volume for September and October is attributable to a large decline in General Magic postings.

	Company Generated			Revisions	News	Postings
	Press Releases	Filings	Total	to Analysts' Reports	Stories	
Time of day (postings)						
Weekday mkt open	--	--	--	--	--	36%
Weekday mkt close	--	--	--	--	--	41%
Weekend	--	--	--	--	--	23%
Day of Week						
Monday	17%	16%	17%	13%	17%	13%
Tueday	30%	21%	27%	16%	21%	17%
Wednesday	23%	16%	21%	23%	22%	18%
Thursday	20%	17%	19%	25%	21%	19%
Friday	9%	28%	14%	24%	13%	17%
Saturday	0%	2%	0%	0%	3%	9%
Sunday	1%	2%	1%	0%	4%	7%
Month of Year						
July	16%	7%	14%	19%	14%	24%
August	14%	17%	15%	8%	12%	17%
September	10%	19%	12%	10%	10%	13%
October	18%	17%	18%	20%	12%	9%
November	11%	16%	12%	9%	14%	10%
December	13%	5%	11%	5%	16%	13%
January	19%	19%	19%	29%	23%	15%

Table 5
Autocorrelation in Information Events

For each firm, we measure the autocorrelation in the level of information event activities by estimating the following regression with daily data:

$$I_{j,t} = \sum_{d=1}^{7} \alpha_d D_d + \beta \ I_{j,t-1}$$

where $I_{j,t}$ is an information event of type j (number of press releases, filings, analyst reports, news stories and postings; as well as the average sentiment index and the disagreement index)at time t and D_d represents day fixed-effects. The table below reports the coefficient β by firm and by type of information.

Firm	**Amazon**		**Delta**		**General Magic**		**Geoworks**	
News quintile	High		High		Low		Low	
Chat quintile	High		Low		High		Low	
	coeff	p-value	coeff	p-value	coeff	p-value	coeff	p-value
Press releases	-0.031	0.636	0.027	0.713	-0.015	0.840	0.096	0.316
Filings	0.099	0.205	-0.043	0.064	-0.060	0.013	-0.046	0.032
Analyst revisions	0.429	0.000	0.124	0.097	na	na	0.000	0.080
Major news stories	0.448	0.000	0.257	0.001	0.179	0.040	0.266	0.000
Postings	0.787	0.000	0.336	0.000	0.717	0.000	0.523	0.003
Sentiment	0.728	0.000	0.044	0.609	0.400	0.000	0.493	0.000
Disagreement index	0.223	0.003	-0.006	0.933	0.236	0.001	0.152	0.112
News Sentiment	0.325	0.007	0.185	0.023	0.022	0.791	0.146	0.475

Table 6
Correlations between Information Events

The table below shows the correlations between various pairs of information events. For each firm, we calculate the correlation between the stated events on a daily basis (excluding weekends.[11]) Each cell represents the median correlation among the four firms, the median p-value and the identity of the firms for which the p-value is .05 or better where a=Amazon, d=Delta, m= General Magic and x = Geoworks.

Median correlation Median p-value Firms sign at 5%	**Press Release**	**Company filing**	**Revision in analyst forecast**	**Number of major news**	**News Sentiment**	**Number of postings**	**Analyst sentiment**	**Posting sentiment**
Company filing	0.024 0.768							
Revision in analyst forecast	-0.043 0.477	-0.032 0.680 a						
Number of major news stories	**0.621** **0.000** **admx**	-0.004 0.951	0.079 0.460 a					
News Sentiment	**0.327** **0.000** **dmx**	-0.004 0.779	0.013 0.905	**0.390** **0.000** **admx**				
Number of postings	0.058 0.485 m	-0.013 0.519	0.150 0.482 a	**0.284** **0.010** **amx**	-0.021 0.209 ax			
Analyst sentiment	-0.089 0.642 a	0.000 0.718 a	-0.518 0.500 ad	-0.101 0.612 a	-0.012 0.945 a	0.108 0.129 a		
Posting sentiment	0.055 0.571 m	0.014 0.364	0.026 0.870	**0.251** **0.006** **amx**	**0.096** **0.014** **adx**	**0.562** **0.000** **admx**	0.080 0.201 a	
Disagreement index	0.023 0.785	0.019 0.367	0.100 0.609	-0.075 0.248 a	-0.107 0.252 a	**0.269** **0.001** **admx**	-0.028 0.436	-0.341 0.120 am

[11] Including weekends tends to increase the correlations as the level of all activity on weekends is reduced.

Table 7
News and Posting Activity Around Company and Analyst Announcements

The table below shows the number of "excess" news stories and postings for the three days around company press releases, company SEC filings and analyst revisions. The three days represent the day prior to, of, and following the date of these announcements. The number of excess stories and postings represents the difference between the observed levels on these days and the mean daily level of all news stories and postings for the period 7/1/98-1/31/99. For each firm, each information event is equally weighted in the calculation of these excess numbers of stories and posting. The average shown on the table represents the firm-equal-weighted average.

Number of New Stories around Information Events

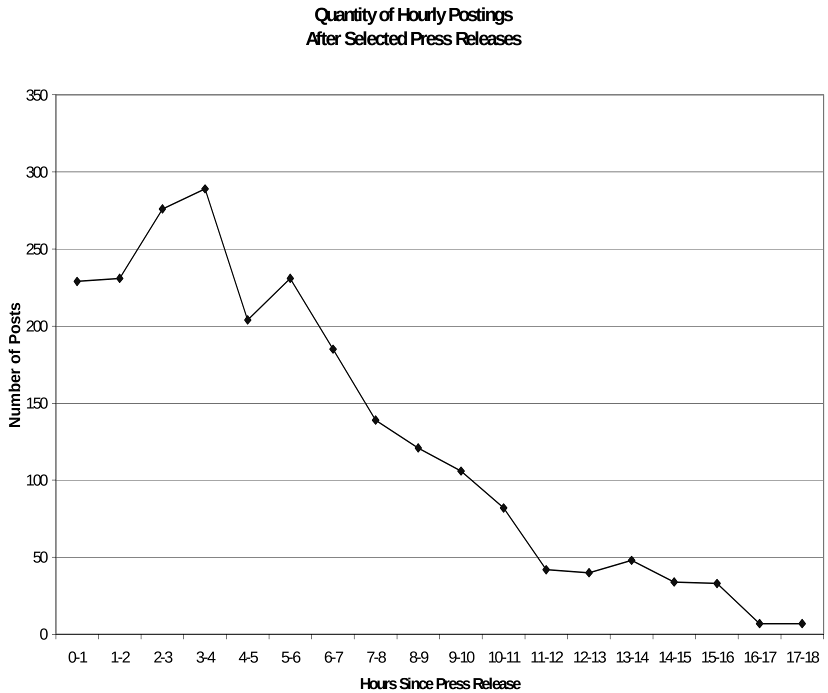
News Event	Press releases	SEC filings	Analysts' revisions	All
Newsday	1.3	0.2	0.4	0.7
Newsday-1	0.2	-0.5	0.5	0.2
Newsday+1	1.1	-0.0	0.1	0.2
NoNewsday	-0.3	0.0	-0.1	-0.2

Number of Postings around Information Events

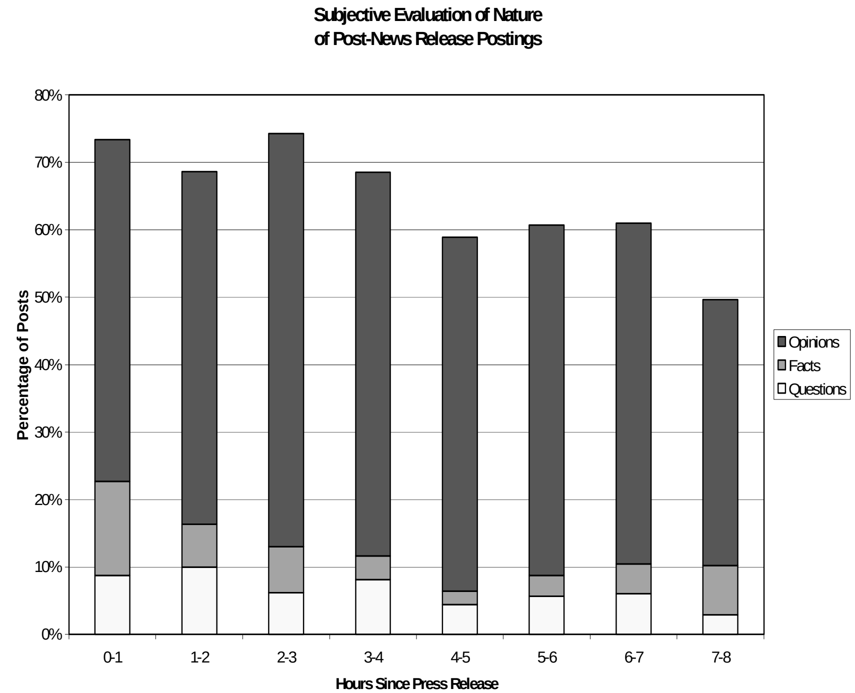
News Event	Press releases	SEC filings	Analysts' revisions	All
Newsday	11.7	23.1	4.4	17.7
Newsday-1	-6.5	-12.4	-4.1	-7.7
Newsday+1	47.2	1.7	0.6	15.8
NoNewsday	5.4	-1.3	1.8	-4.7

Table 8
Sixteen "Tracer Events"

We report on sixteen press releases made by the four companies which seemed to indicate a material change in the firm's business or financing. For each event, we summarize the speed of information dissemination by traditional and chat room sources. Panels A shows the aggregate number of posts after the time-stamp of the press release. For each news release, we manually categorized the postings in the subsequent eight hours into on-point (relating to the news event) or off-point (unrelated to the news event or spam), and we also categorized the on-point postings into ones that asked a question, offered an alleged fact, and proposed an opinion. Panel B shows the distribution of the on-point postings. Panel C provides detailed information on the nature of the events, the first posting activity, and the first press activity.



Panel A: Number of postings by hour after 16 selected corporate press releases.



Panel B: Distribution of type of posting by hour after 16 selected corporate press releases. Postings are classified as on-point if related to the news story, and off-point otherwise. The histogram shows the percentage of on-point posts (the height of each bar) and the nature of the on-point posts (asks question, provides alleged fact, proposes opinion.)

Panel C: Data on sixteen tracer events.

Company	Event	Release Date and Time	First Press Notice	Minutes to First Press Activity	Subsequent Press Activity	First Posting	Minutes to First Post	Subsequent Related Postings	Foreshadowing
Geoworks	Poor 1st Quarter Results	January 26, 1999, 8:02 AM EST	Dow Jones NewsWire 1/26; 9:34 AM EST	52	Warning of Price drop 1/25/99	1/26/1999 8:45 AM (Yahoo!)	43	15 Posts before market closes; 8 more before market opens next day	Discussion of poor earnings possibilities given large price drop 1/25/99
Geoworks	GWRX announces Cooperation with Optimay	January 20, 1999, 6:08 AM EST	Dow Jones NewsWire 1/21; 3:53 AM EST	1313	Warning of Price Jump 1/19/99	1/20/1999 7:07 AM (Yahoo!)	67	18 Posts before market closes, 9 more before market opens next day; 4 before close	
Geoworks	Dave Grannan Named as CEO	January 11, 1999, 8:05 AM EST	Dow Jones NewsWire 1/11; 8:03 AM EST (preliminary information)	-2	1/11: 2 Dow Jones Articles 1/12: Wall Street Journal	1/11/99 9:15 AM (Yahoo!)	70	16 posts before market closes; 4 more before opening next day; 5 before close	
Geoworks	Debuts Enhanced Phone with Mitsubishi	November 16, 1998, 6:08 AM EST	Dow Jones NewsWire 11/16, 11:10 AM EST(preliminary)	310	11/16: 7 Dow Jones Articles, 4 Price Alerts; 1/17 Wall Street Journal	11/16/98 10:08 AM (Yahoo!)	246	35 posts before market closes; over 80 before next day market opening	
Delta	Alliance With Korean Air	August 6, 1998, 8:00 PM EST	Dow Jones NewsWire 8/6 8:25 PM EST	25	8/6: 1 Dow Jones Article; 8/10: M2 Presswire release	N/A	NEVER	N/A	
Delta	Finances Beat Consensus	July 16, 1998, 9:14 AM EST	Dow Jones NewsWire 7/16 8:59 AM EST (preliminary)	-15	7/16: 4 Dow Jones Articles, 7/17: 5 (4 Newspaper, 1 BusinessWire)	7/16/98 9:25 AM (Yahoo!)	11	2 before market closes	
Delta	Announces Strategic Management Reorganization in Int'l Markets	December 9, 1998, 7:01 AM EST	Dow Jones NewsWire 12/9 7:03 AM EST (preliminary)	2	12/9: 3 Dow Jones Articles, 12/10: M2 Presswire	N/A	NEVER	N/A	
Delta	2 for 1 Stock Split Approved	October 22, 1998 1:54 PM EST	Dow Jones NewsWire 10/22 1:55 PM EST (preliminary)	1	10/22: 2 Dow Jones Articles, 10/23: M2 PressWire	10/22/98 4:01 PM (Yahoo!)	127	1 after market closes 10/22, 3 on 10/27	Questions about split timing 10:46am
General Magic	Agreement with Intuit for voice access to financial info	November 9, 1998, 4:04 AM EST	Dow Jones NewsWire 11/9 4:09 AM EST (preliminary)	5	11/9: 4 Dow Jones Articles, 11/10: 1 Dow Jones Article	11/9/98 12:39 AM (Yahoo!)	515	54 before market opens; 112 more before close, over 90 before open next day	Info of radio broadcast of Quicken news before press release
General Magic	2nd Quarter Results	July 29, 1998, 4:02 PM EST	Dow Jones NewsWire 7/29 4:06 PM EST (preliminary)	4	7/29: 2 (Dow Jones, BusinessWire), 7/30: Wall Street Journal	7/29/98 4:11 PM (Yahoo!)	9	130 between release and 9pm that night	Discussion of potential results beforehand
General Magic	Alliance With Microsoft for Auto-Enabled PC	January 7, 1999, 7:34 AM EST	Dow Jones NewsWire 1/7 2:23 PM EST (preliminary)	409	1/7: 2 Dow Jones Articles	1/7/99 7:53 AM (Yahoo!)	19	59 before market opens, 179 more before market close	Rumor of shared booth at CES between MSFT and GMGC (1/6 10:12AM)
General Magic	Spin off of DataRover Division as an independent company	October 28, 1998, 8:10 PM EST	LA Times and New York Times, 10/30	>1500	10/30: 3 (Bloomberg, LA Times, ComputerWire) 11/2: 1 (Electronics News) <two subsequent mentions in articles referencing Q3 Results>	10/28/1998 8:41 PM (Yahoo!)	31	32 Additional Posts on boards before market opens the following morning	Posting of suspicious independent DataRover URL nine days before announcement
Amazon	Acquisition of Junglee	August 4, 1998, 7:30 AM EST Note: Filing on 8/5/98 (8-K)	Dow Jones NewsWire 8/4; 9:06 AM EST	96	8/6: 17 8/8: 69 8/10: 3 8/11-18: 27	8/4/98 7:47 AM (Silicon Investor), 7:58 AM (Yahoo!)	17	5 additional related posts before the opening of the market (9:30 am), 21 more by the close of the market (4:00 pm), 22 additional before the next morning's market open	
Amazon	$500 million debt issue	January 28, 1999, 7:25 AM EST	Dow Jones Newswire 1/28 7:41 AM EST	16	1/29: 1 Dow Jones Article, 1/29: 6, 2/1: 4	1/28/99 7:40 AM (Yahoo!)	15	34 before market open, over 300 more before market closes	Bond Discussion in days before (no facts, though)
Amazon	Enters European Book Market	October 15, 1998, 5:00 AM EST	Dow Jones Newswire 10/15 5:51 AM EST	51	10/17: 3 Newspaper Articles	10/15/98 6:18 AM (Motley Fool)	78	6 before market opens, 6 more before closes	
Amazon	Amazon Announces 3 for 1 Split	November 19, 1998, 5:45 PM EST	Dow Jones NewsWire 11/19 5:47 PM EST	2	11/20: 9; 11/21: 3; 11/22: 1; 11/23: 1	11/19/98 5:51 PM (Yahoo!)	6	Over 300 more posts before midnight 11/19	Split speculation in days before announcement

Table 9
Correlations between information and financial markets

Panel A reports the same-day contemporaneous correlations between information variables and financial market variables (return, excess return, volume, implied volatility, intraday volatility, and bid-ask spreads.) Panel B reports the autocorrelations for the financial market variables. Each cell represents the median correlation among the four firms, the median p-value among the four firms, and the identity of the firms for which the p-value is .05 or better where a=Amazon, d=Delta, m= General Magic and x = Geoworks.

Panel A: Contemporaneous correlations: Information and financial markets

Median correlation p-value firms signif.	**Press Release**	**Filing**	**Analyst Revision**	**No of Major News Stories**	**News Sentiment**	**Number of Postings**	**Analyst Sentiment**	**Sentiment**	**Disagree-ment**	**Open-to-close return**	**Close-to-close return**	**Share turnover**	**Intraday volatility**	**Implied Volatility**	**Bid-ask %**
Open-to-close	0.03	-0.01	-0.04	0.18	0.15	-0.04	0.12	0.10	-0.05						
return	0.52	0.56	0.81	0.05	0.30	0.60	0.54	0.22	0.58						
	x			mx	mx			d	m						
Close-to-close	0.09	-0.02	0.02	**0.25**	**0.27**	-0.07	0.03	**0.23**	-0.09	**0.84**					
return	0.37	0.73	0.83	**0.01**	**0.02**	0.52	0.71	**0.01**	0.07	**0.00**					
	mx			**amx**	**amx**			**admx**	am	**admx**					
Share turnover	0.20	0.04	0.00	**0.41**	0.21	**0.50**	-0.03	0.17	-0.02	0.16	0.16				
	0.04	0.45	0.99	**0.00**	0.15	**0.00**	0.64	0.19	0.35	0.08	0.08				
	mx	m		**amx**	am	**amx**		am	a	a	am				
Intraday volatility	0.03	0.01	0.07	0.17	0.12	0.12	-0.07	0.03	-0.05	0.05	-0.05	**0.44**			
	0.74	0.56	0.67	0.04	0.25	0.19	0.62	0.35	0.55	0.30	0.11	**0.00**			
				am	am	am		ax	x	m		**admx**			
Implied Volatility	0.02	-0.02	0.07	0.05	-0.02	0.03	-0.21	-0.03	-0.06	-0.15	-0.19	0.19	**0.45**		
	0.08	0.79	0.70	0.55	0.63	0.23	0.15	0.33	0.49	0.07	0.02	0.02	**0.00**		
				a		a	a	a		m	am	ad	**adm**		
Bid-ask %	-0.09	-0.09	-0.01	**-0.26**	-0.06	**-0.37**	-0.01	**-0.15**	0.14	-0.08	-0.13	**-0.07**	**0.25**	**0.11**	
	0.18	0.26	0.89	**0.00**	0.44	**0.00**	0.93	**0.00**	0.19	0.10	0.02	**0.00**	**0.00**	**0.00**	
	m	x		**amx**	a	**admx**		**amx**	ax	a	am	**amx**	**admx**	**dmx**	
Number of Jumps	0.09	-0.01	0.07	0.14	0.13	0.17	-0.15	0.03	0.00	0.02	0.05	-0.09	-0.14	0.06	**-0.17**
	0.31	0.53	0.65	0.11	0.18	0.08	0.15	0.67	0.69	0.45	0.52	0.06	0.04	0.21	**0.04**
	x			ax	am	mx		ax	x	x	x	mx	dm	x	**admx**

Panel B: Autocorrelations for financial markets variables

Autocorrelations	**AMAZON**		**DELTA**		**GENERAL MAGIC**		**GEOWORKS**	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Open/close return	-0.045	0.624	0.249	0.005	-0.098	0.368	-0.033	0.728
Close/closre return	0.102	0.279	0.223	0.008	0.026	0.831	-0.019	0.820
Share turnover	0.544	0.000	0.412	0.000	0.542	0.000	0.515	0.000
Intraday volatility	0.602	0.000	0.508	0.000	0.486	0.000	0.265	0.042
Implied volatility	0.702	0.000	0.866	0.000	0.529	0.000	n/a	n/a
Bid/ask spread	0.820	0.000	0.795	0.000	0.724	0.000	0.687	0.000
Number of jumps	0.049	0.571	0.249	0.037	0.215	0.040	0.315	0.019

Table 10
Variable Definitions for Regressions

Variable	Definition	SUMMARY STATISTICS: Mean (Std. Deviation) All firm-periods	AMZN	DAL	GMGC	GWRX
Implied volatility	Implied volatility on call options, as taken from Bloomberg	0.880 (0.354)	0.948 (0.168)	0.450 (0.074)	1.243 (0.140)	n/a n/a
Share turnover	Number of shares traded that day divided by number of shares outstanding	0.040 (0.054)	0.096 (0.051)	0.009 (0.004)	0.036 (0.038)	0.020 (0.057)
<u>Announcements</u>						
Press release	Dummy variable. Equals 1 if the company has made a press release on that day	0.209 (0.374)	0.142 (0.350)	0.459 (0.500)	0.135 (0.343)	0.101 (0.303)
Filing	Dummy variable. Equals 1 if the company has made a SEC filing on that day	0.090 (0.281)	0.149 (0.357)	0.061 (0.240)	0.081 (0.274)	0.068 (0.252)
Analyst revision	Dummy variable. Equals 1 if any analyst has issued some sort of earnings revision on that day	0.122 (0.327)	0.250 (0.434)	0.216 (0.413)	0.000 (0.000)	0.020 (0.141)
<u>News activity</u>						
Abnormal news stories	Residual from regression of news stories on prior day news stories, day of week, and month of year dummy variables.	0.000 (2.996)	0.000 (4.765)	0.000 (5.445)	0.000 (0.822)	0.000 (0.951)
Lagged abnormal stories	Note: lag here determined by last calendar day, where the lag of Monday includes news published in the three previous days.					
<u>Posting activity</u>						
Abnormal number of market posts	Residual from regression on posts on trading day from 9:30 am to 4 pm on its lag, and day of week, and month of year dummy variables.	0.000 (39.423)	0.000 (82.28)	0.000 (2.293)	0.000 (63.86)	0.000 (9.257)
Abnormal number of pre-market posts	Residual from regression on posts from 4 p.m. prior day to 9:30 am on trading day on its lag, and day of week, and month of year dummy variables.	0.000 (55.703)	0.000 (133.7)	0.000 (3.031)	0.000 (76.96)	0.000 (9.085)
Sentiment level during market hours	Number of buy messages - number of sell messages from 9:30 am to 4:00 pm	4.973 (13.465)	9.378 (23.25)	0.284 (1.195)	9.655 (10.01)	0.574 (1.885)
Sentiment level during pre-market hours	Number of buy messages - number of sell messages from 4:00 pm prior day to 9:30 am trading day. For Monday it includes all posting activity from 4:00 pm Friday to 9:30 am Monday	8.855 (19.205)	21.345 (31.22)	0.297 (1.459)	13.378 (13.22)	0.399 (2.861)
Interaction terms	Note: the information events are dummy variables indicating a press release (etc.) on the current day. The four interaction terms indicate the abnormal level of news stories or posts, or the level of news sentiment or posting sentiment.					
Lags	Note: lags are determined by previous trading day, not by previous calendar day					

Table 11
Explaining returns using information variables

The dependent variables in this table are the market adjusted return (raw return less S&P 500) earned by each of the four firms during the period 7/1/98-1/31/99. For Panel A and C, this return is calculated from the open of the market to the close of the market, while for Panels B and D, it represents the return from the close of the market the prior trading day to close of the market the current day. Panels A and B use contemporaneous information, i.e., information from the same day as the return. This includes all filings, press releases and news stories with that date and postings before the close of the market on that day. Panels C and D use lagged information, i.e., filings, press releases and news stories from the prior day, and postings that preceded the opening of the market.

Panel A

Dependent variable: market adjusted returns, open-to-close	AMZN		DAL		GMGC		GWRX	
All independent variables measured the same "day."	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.107	(0.002)	0.009	(0.501)	0.159	(0.000)	-0.006	(0.786)
Volatility	-0.101	(0.014)	-0.019	(0.483)	-0.136	(0.000)	0.003	(0.653)
<u>Announcements</u>								
Press Release	0.002	(0.932)	-0.001	(0.933)	-0.114	(0.000)	-0.345	(0.000)
Filing	-0.014	(0.188)	-0.009	(0.191)	-4.5E-04	(0.980)	0.045	(0.245)
Revision	-0.013	(0.178)	-0.003	(0.529)			-2.5E-04	(0.993)
<u>News Activity</u>								
Abnormal number of stories	0.002	(0.101)	0.001	(0.452)	0.023	(0.028)	0.021	(0.277)
News sentiment	-0.002	(0.157)	-0.001	(0.671)	0.006	(0.626)	0.049	(0.003)
<u>Posting Activity</u>								
Abnormal number of posts	-7.1E-06	(0.848)	9.4E-05	(0.881)	-3.3E-05	(0.489)	-1.9E-04	(0.675)
Sentiment level	2.2E-04	(0.271)	0.001	(0.477)	1.3E-04	(0.671)	-0.004	(0.260)
<u>Interaction terms: Press release dummy interacted with</u>								
...abnormal number of stories	0.001	(0.727)	-0.002	(0.341)	0.013	(0.443)	0.133	(0.000)
...sentiment of stories	-0.009	(0.120)	0.001	(0.635)	0.019	(0.170)	-0.018	(0.575)
...abnormal number of posts	2.1E-04	(0.124)	-4.4E-04	(0.639)	0.000	(0.352)	0.005	(0.405)
...sentiment of posts	0.001	(0.006)	0.002	(0.329)	0.001	(0.148)	0.024	(0.057)
Number of observations	146		147		147		105	
Adjusted R-squared	0.137		0.063		0.309		0.730	
F-statistic	4.740		0.940		4.010		23.890	

Panel B

Dependent variable: market adjusted returns, close-to-close	AMZN		DAL		GMGC		GWRX	
All independent variables measured the same "day."	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.142	(0.000)	0.023	(0.066)	0.175	(0.000)	0.009	(0.646)
Volatility	-0.152	(0.000)	-0.054	(0.043)	-0.153	(0.000)	0.000	(0.991)
<u>Announcements</u>								
Press Release	0.037	(0.136)	-0.001	(0.891)	-0.115	(0.000)	-0.342	(0.000)
Filing	-0.002	(0.877)	-0.011	(0.110)	0.003	(0.890)	-0.012	(0.733)
Revision	-0.002	(0.815)	0.001	(0.899)			0.004	(0.856)
<u>News Activity</u>								
Abnormal number of stories	0.002	(0.113)	0.001	(0.548)	0.028	(0.008)	0.016	(0.381)
News sentiment	-0.001	(0.537)	-0.002	(0.576)	0.020	(0.076)	0.056	(0.006)
<u>Posting Activity</u>								
Abnormal number of posts	-3.0E-06	(0.934)	-0.001	(0.365)	-1.0E-04	(0.067)	6.5E-05	(0.889)
Sentiment level	5.0E-04	(0.006)	0.002	(0.174)	6.2E-04	(0.071)	-0.004	(0.249)
<u>Interaction terms: Press release dummy interacted with</u>								
…abnormal number of stories	-4.3E-04	(0.906)	-0.002	(0.447)	0.018	(0.299)	0.142	(0.000)
…sentiment of stories	-0.016	(0.007)	0.003	(0.392)	0.017	(0.249)	-0.030	(0.398)
…abnormal number of posts	2.0E-04	(0.145)	-0.001	(0.474)	-9.0E-05	(0.336)	0.008	(0.194)
…sentiment of posts	0.001	(0.008)	0.001	(0.581)	0.001	(0.061)	0.022	(0.087)
Number of observations	146		147		147		105	
Adjusted R-squared	0.1892		0.1009		0.4425		0.7541	
F-statistic	3.42		1.82		8.23		16.13	

Panel C

Dependent variable: market adjusted returns, open-to-close All independent variables measued the prior "day".	AMZN Coeff	 p-value	DAL Coeff	 p-value	GMGC Coeff	 p-value	GWRX Coeff	 p-value
Constant	0.036	(0.366)	-0.019	(0.164)	-0.078	(0.064)	-0.029	(0.342)
Volatility	-0.027	(0.570)	0.041	(0.120)	0.058	(0.071)	0.017	(0.300)
Announcements								
Press Release	0.005	(0.791)	-0.001	(0.808)	0.003	(0.898)	0.020	(0.686)
Filing	-0.019	(0.167)	0.002	(0.756)	-0.010	(0.477)	-0.106	(0.108)
Revision	-0.012	(0.359)	-0.002	(0.710)			-0.019	(0.488)
News Activity								
Abnormal number of stories	0.001	(0.596)	-0.001	(0.566)	-4.5E-04	(0.966)	-0.020	(0.388)
News sentiment	0.001	(0.634)	0.000	(0.975)	0.008	(0.633)	0.041	(0.008)
Posting Activity								
Abnormal number of posts	4.2E-05	(0.479)	0.001	(0.465)	1.0E-05	(0.920)	-0.003	(0.571)
Sentiment level	9.9E-06	(0.967)	0.002	(0.367)	7.1E-05	(0.889)	-0.005	(0.349)
Interaction terms: Press release dummy interacted with								
…abnormal number of stories	0.003	(0.354)	0.000	(0.814)	-0.025	(0.045)	-0.014	(0.735)
…sentiment of stories	-0.001	(0.849)	-0.002	(0.431)	-0.009	(0.608)	-0.003	(0.929)
…abnormal number of posts	-3.0E-05	(0.664)	-0.001	(0.378)	0.000	(0.400)	0.003	(0.642)
…sentiment of posts	-0.001	(0.224)	0.002	(0.563)	0.000	(0.995)	3.6E-04	(0.981)
Number of observations	145		146		146		104	
Adjusted R-squared	0.069		0.082		0.118		0.083	
F-statistic	1.950		1.290		4.170		37.840	

Panel D

Dependent variable: market adjusted returns, close-to-close All independent variables measued the prior "day".	AMZN Coeff	 p-value	DAL Coeff	 p-value	GMGC Coeff	 p-value	GWRX Coeff	 p-value
Constant	0.042	(0.200)	-0.010	(0.468)	-0.069	(0.161)	-0.020	(0.537)
Volatility	-0.040	(0.286)	0.022	(0.442)	0.051	(0.189)	0.017	(0.317)
Announcements								
Press Release	-0.018	(0.454)	-0.001	(0.898)	-0.015	(0.530)	0.058	(0.342)
Filing	-0.005	(0.728)	-1.6E-05	(0.998)	-0.011	(0.519)	-0.091	(0.205)
Revision	-0.010	(0.452)	-0.002	(0.757)			0.013	(0.650)
News Activity								
Abnormal number of stories	0.001	(0.301)	-1.3E-04	(0.940)	0.008	(0.564)	-0.008	(0.714)
News sentiment	6.5E-05	(0.962)	2.4E-04	(0.931)	0.006	(0.738)	0.029	(0.113)
Posting Activity								
Abnormal number of posts	-6.4E-05	(0.362)	-0.001	(0.303)	-2.6E-05	(0.831)	-0.003	(0.505)
Sentiment level	3.4E-04	(0.241)	0.005	(0.025)	0.001	(0.088)	-0.009	(0.142)
Interaction terms: Press release dummy interacted with								
…abnormal number of stories	-0.001	(0.869)	-0.001	(0.729)	-0.030	(0.066)	-0.039	(0.370)
…sentiment of stories	0.008	(0.190)	-0.003	(0.402)	0.004	(0.833)	0.025	(0.433)
…abnormal number of posts	2.0E-04	(0.032)	-3.2E-04	(0.798)	9.4E-05	(0.498)	0.009	(0.185)
…sentiment of posts	-0.001	(0.269)	-0.001	(0.730)	-0.001	(0.385)	-0.012	(0.411)
Number of observations	145		146		146		104	
Adjusted R-squared	0.076		0.103		0.125		0.0737	
F-statistic	1.030		2.460		2.010		23.94	

Table 12
Implied volatility and information variables

The dependent variables in this table are the implied volatility of the three firms with listed options during the period 7/1/98-1/31/99. Implied volatilities are taken from Bloomberg. The information variables in Panel A are contemporaneous, taken from the same day as the implied volatility, while the variables in Panel B are from the prior day (or in the case of postings, from the postings through the time of the market open.)

Panel A

Dependent variable: implied volatility from Bloomberg	**AMZN**		**DAL**		**GMGC**	
All independent variables measured the same "day."	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.225	(0.000)	0.057	(0.005)	0.555	(0.000)
Lagged implied volatility	0.757	(0.000)	0.883	(0.000)	0.548	(0.000)
Announcements						
Press Release	-0.020	(0.666)	-0.007	(0.390)	-0.109	(0.048)
Filing	-2.4E-04	(0.986)	0.008	(0.543)	-0.032	(0.379)
Revision	0.003	(0.814)	-0.008	(0.212)		
News Activity						
Abnormal number of stories	0.001	(0.464)	0.001	(0.581)	0.050	(0.020)
News sentiment	-0.005	(0.308)	-0.002	(0.388)	0.035	(0.378)
Posting Activity						
Abnormal number of posts	1.5E-04	(0.034)	3.4E-04	(0.614)	-1.1E-04	(0.237)
Sentiment level	0.001	(0.004)	0.002	(0.205)	0.001	(0.124)
Interaction terms: Press release dummy interacted with						
…abnormal number of stories	0.013	(0.085)	-2.7E-04	(0.916)	-0.004	(0.913)
…sentiment of stories	0.007	(0.526)	0.002	(0.465)	-0.009	(0.829)
…abnormal number of posts	3.9E-04	(0.089)	-0.002	(0.179)	0.000	(0.661)
…sentiment of posts	0.002	(0.031)	-0.003	(0.321)	0.000	(0.836)
Number of observations	145		147		147	
Adjusted R-squared	0.675		0.823		0.471	
F-statistic	39.750		51.060		11.190	

Panel B

Dependent variable: implied volatility from Bloomberg	AMZN		DAL		GMGC	
All independent variables measured the prior "day."	**Coeff**	**p-value**	**Coeff**	**p-value**	**Coeff**	**p-value**
Constant	0.317	(0.094)	0.055	(0.007)	0.634	(0.000)
Lagged implied volatility	0.647	(0.003)	0.871	(0.000)	0.485	(0.000)
Announcements						
Press Release	-0.014	(0.674)	0.008	(0.348)	-0.080	(0.044)
Filing	-0.020	(0.178)	-0.013	(0.071)	0.049	(0.102)
Revision	-0.006	(0.691)	-0.004	(0.652)		
News Activity						
Abnormal number of stories	-0.003	(0.101)	-0.003	(0.171)	-0.016	(0.347)
News sentiment	0.006	(0.129)	0.002	(0.495)	0.015	(0.577)
Posting Activity						
Abnormal number of posts	2.6E-04	(0.126)	0.001	(0.517)	4.0E-05	(0.810)
Sentiment level	0.001	(0.524)	-0.001	(0.817)	-3.1E-04	(0.757)
Interaction terms: Press release dummy interacted with						
…abnormal number of stories	0.002	(0.681)	0.003	(0.343)	0.070	(0.002)
…sentiment of stories	-0.006	(0.455)	-0.001	(0.785)	0.024	(0.413)
…abnormal number of posts	0.000	(0.295)	0.001	(0.698)	-3.4E-04	(0.108)
…sentiment of posts	0.001	(0.448)	-0.002	(0.715)	-4.7E-04	(0.772)
Number of observations	144		146		146	
Adjusted R-squared	0.5476		0.7688		0.3576	
F-statistic	36.07		39.17		8.99	