The Role of the Media in the Internet IPO Bubble
(*Forthcoming: Journal of Financial and Quantitative Analysis)*

Utpal Bhattacharya, Neal Galpin, Rina Ray, and Xiaoyun Yu*

Abstract

We read all news items that came out between 1996 and 2000 on 458 internet IPOs and a matching sample of 458 non-internet IPOs – a total of 171,488 news items – and classify each news item as good news, neutral news, or bad news. We first document that the media was more positive for internet IPOs in the period of the dramatic rise in share prices, and was more negative for internet IPOs in the period of the dramatic fall in share prices. We then document that media hype is unable to explain the internet bubble: there was a 1646% difference in returns between internet stocks and non-internet stocks from January 1, 1997 through March 24, 2000 (the market peak), and the media can explain only 2.9% of that.

---

* Bhattacharya, ubhattac@indiana.edu, and Yu, xiyu@indiana.edu, Kelley School of Business, Indiana University, Bloomington, IN 47408; Galpin, ngalpin@mays.tamu.edu, Mays Business School, Texas A&M University, 360 Wehner Building, College Station, TX 77845; and Ray, rina.ray@nhh.no, Norwegian School of Economics and Business Administration, Helleveien 30, 5045 Bergen, Norway. We are grateful for suggestions from Stephen Brown (the editor), Laura Field, Diego Garcia, Mariassunta Giannetti, Chris Lundblad, Guojun Wu, Wei Xiong, an anonymous referee, and seminar participants at Indiana University, London Business School, McGill University, Penn State University, Singapore Management University, Swedish Institute for Financial Research, University of Minnesota, University of New Orleans and at the following conferences: China International Finance Conference 2005, European Finance Association Annual Meeting 2005, and Financial Intermediation Conference on Banking, Corporate Finance and Intermediation 2006. We thank Jay Ritter for providing the list of internet IPO firms. All mistakes remain our own responsibility.

I. Introduction

The 1996-2000 period is an interesting period in the history of US stock prices. A subset of stocks, called internet stocks, went up by 1000% from January 1996 through March 24, 2000 (we call this period the "bubble" period), only to crash by 45% in the next few months (we call this period the "post-bubble period"). Many believe that this dramatic rise and fall in internet share prices cannot be explained by fundamentals, though a growing number of papers argue that they can.[1]

Though the debate about whether the dramatic rise and fall in internet share prices in the 1996-2000 period had rational or behavioral causes is fascinating, our goal in this paper is more modest: we want to formally explore whether Shiller (2000) was correct when he claimed that, first, the media hyped internet stocks,[2] and second, this media hype was a major factor responsible for the dramatic rise and fall of internet

[1] Cutler, Poterba and Summers (1989) is one of the early papers that show stock return variances cannot be explained by fundamental news such as macroeconomic news. A sample of papers that question whether fundamentals could explain the dramatic rise and fall of internet share prices in the late 1990s: Cooper, Dimitrov, and Rao (2001), Ofek and Richardson (2002, 2003), Lamont and Thaler (2003), Brunnermeier and Nagel (2004), Loughran and Ritter (2004) and Griffin, Harris, and Topaloglu (2005). These papers provide various other explanations, but some of these explanations have themselves been questioned. For example, Battalio and Schultz (2006) contest the explanation that short sale restrictions prevented traders from arbitraging. A sample of papers that argue that fundamentals could explain the dramatic rise and fall of internet share prices in the late 1990s: DeMarzo, Kaniel, and Kremer (2007) and Pastor and Veronesi (2006, 2007).

[2] A positive feedback hypothesis figures prominently in Shiller's (2000) description of media hype during the internet bubble. For example, he writes, "…stock price increases in the late 1990s were driven by irrational euphoria among individual investors, fed by an emphatic media, which maximized TV ratings and catered to investor demand for pseudo-news." Further, when discussing the role of news media in "speculative bubbles", he writes, "…the news media are fundamental propagators of speculative price movements through their efforts to make news interesting to their audience. They sometimes strive to enhance such interest by attaching news stories to price movements that the public has already observed, thereby enhancing the salience of these movements and focusing greater attention on

stocks. So we ask and answer the following two questions: was the overall media coverage for internet IPOs in the years 1996 through 2000 different from a matching sample of non-internet IPOs and, if yes, did this difference in the media coverage have any effect on the difference in risk-adjusted returns between internet stocks and non-internet stocks.[3]

Was the overall media coverage different for internet IPOs? We read all news items that came out between 1996 and 2000 on 458 internet IPOs and a matching sample of 458 non-internet IPOs – a total of 171,488 news items – and classify each news item as good news, neutral news or bad news. We find, not surprisingly, that the overall media coverage was more intense for internet IPOs. There were more good, more bad, and more neutral news for internet IPOs than for non-internet IPOs in both the bubble period and the post-bubble period. Next, we use net news (good news minus bad news) to proxy for media sentiment and find that, compared to the matching sample, net news was more positive for internet IPOs in the bubble period and more negative for internet IPOs post-bubble. Third, we document that net news increased after a positive stock return and decreased after a negative stock return. Interestingly, the increase in net news after a positive stock return was *larger* for internet IPOs than for non-internet IPOs during the bubble, and the decrease in net news after a negative stock return was *larger* for internet IPOs post-bubble. This result remained after controlling for liquidity, size, variance of returns, past news, and past abnormal returns. Our paper thus provides evidence in favor of Shiller's (2000) first hypothesis: media hype existed, especially for internet stocks during the bubble period.

---

them. Or they may remind the public of past market episodes, or of the likely trading strategies of others. Thus the media can sometimes foster stronger feedback from past price changes to further price changes…"

[3] In this paper we look at stock returns *after* the first day of trading. In Bhattacharya, Galpin, Ray, and Yu (2006), we look at stock returns *on* the first day of trading. The reason we separated our analysis is because information dissemination during the pre-IPO book-building stage is very different from information dissemination during the post-IPO stage. In the pre-IPO stage, institutions disseminate information and, therefore, the types of these institutions are the significant control variables. In the post-IPO stage, the main source of information is the traded price itself. This, therefore, becomes the paramount control variable in this paper.

Did this differential media coverage have any effect on the difference in risk-adjusted returns between internet stocks and non-internet stocks during this period? We check whether news in the media, measured by numbers and type, affected contemporary and future abnormal returns, where abnormal return is the error term of a Fama-French (1993) three-factor model.

After controlling for lagged abnormal returns, market capitalization, contemporary market trading conditions, and contemporary and past news, we find that net news today is positively related to today's and tomorrow's risk-adjusted returns. However, the net news effect dies out after these two trading days for both internet and non-internet IPOs. We find that the effect of today's net news on *today's* risk-adjusted return is lower for internet IPOs than for non-internet IPOs during both the bubble and post-bubble periods. The effect of today's net news on *tomorrow's* risk-adjusted return is also lower for internet IPOs, especially during the bubble period. In addition, since net news on average is positive during the bubble period and is negative post-bubble, this implies today's good (bad) news matters less for today's and tomorrow's risk-adjusted returns for internet IPOs during the bubble (post-bubble) period. Our results are robust to whether we risk-adjust individual stocks, or whether we risk-adjust a portfolio consisting of either internet or non-internet stocks.

While the above results indicate that the media had a lower *marginal* effect on the returns of internet IPOs than on the returns of non-internet IPOs, they do not suggest that the media had a lower *total* effect on the returns of internet IPOs. Given that our data shows that a typical internet stock has about 6 media stories in a week whereas a typical non-internet stock has about 2 media stories in a week, it is possible that, because of diminishing returns, the 7th story about the internet stock has less impact on returns than the 3rd story about the non-internet stock. However, cumulatively, the 6 stories about the internet stock could have more impact on returns than the 2 stories about the non-internet stock. So the *total* effect of the media could still explain the difference in risk-adjusted returns between internet and non-internet IPOs.

We test this alternative hypothesis in the following two ways. First, to allow for the possibility of a non-linear effect of news on returns, we control for the cumulative news in the past. This takes care not just of possible diminishing returns – each successive news item has less and less effect on returns[4] – but also of possible increasing returns – each successive news report increases the credibility of the news in the mind of the public and so has more and more effect on returns. Our results remain unchanged. Second, we estimate how much of the difference in cumulative returns between internet IPOs and non-internet IPOs from January 1, 1997 to March 24, 2000 (the day Nasdaq 100 reached its peak) – an incredible 1646% – can be explained by the media. We find that the addition of media news as a control variable improves the explanatory power of an augmented Fama-French three-factor model by only 2.9%. So the media can explain only 2.9% of the difference in cumulative returns between internet IPOs and non-internet IPOs during the bubble period.

Our paper thus provides evidence against Shiller's (2000) second hypothesis: the differential media coverage of internet stocks compared to non-internet stocks was not able to explain the huge difference in stock returns between the two groups. Our results, which are robust to a number of different alternative explanations (described in the internet appendices to this paper[5]), suggest that the media was not a major determinant of the bubble and its crash. So we have to look for other factors to explain the dramatic rise and fall of the stock market in the period of 1996 to 2000.

The rest of the paper is organized as follows. In Section II, we discuss the related literature on media. Section III discusses how we obtained our data. Section IV documents the differential media coverage of internet IPOs as opposed to a matching sample of non-internet IPOs. Section V examines the *marginal* effect of the media coverage on returns. Section VI answers whether the differential media coverage affected the difference in *cumulative* returns between the two samples. Section VII concludes.

---

[4] If the effect of news is broken up over several news items, then each news item will have a small effect on returns. Brown and Warner (1980, 1985) show that the power of event studies is low in detecting such small events.

[5] The appendices are available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=606264. Appendix 1 describes in detail the Jorda (2005) empirical methodology we used in this paper. Appendix 2 describes an experiment and a simulation exercise which show that our results are not biased by mis-classifications of news items. Appendix 3 shows that our results are robust to different variable definitions and specifications.

II. The Related Literature on Media

There is a growing literature on bias in the media. How do the media choose which stories to cover? Of the stories they choose to cover, what is the slant given? And why is there a slant? Shiller (2000) writes, "The role of the news media in the stock market is not, as commonly believed, simply as a convenient tool for investors who are reacting directly to the economically significant news itself. The media actively shape public attention and categories of thought, and they create the environment within which the stock market events we see are played out." He believes that the media strive to enhance interest in news stories by covering stocks that have already experienced a dramatic price change. The extra media coverage, or hype, draws investors' attention to these stocks. This leads to a positive feedback effect, in which big returns follow big returns because of increased media coverage. Dyck and Zingales (2003), Mullainathan and Shleifer (2005), and Baron (2006) provide theoretical rationales for media bias.

Our research question – how much did the media explain the difference in returns between internet IPOs and non-internet IPOs – extends from a large literature on how media news affects returns. According to classical asset pricing models, news will affect returns if it affects expectations of future cash flows and/or expectations of the discount rate. Numerous empirical studies have documented a strong relationship between news and price movements in securities markets (see for example, Niederhoffer (1971), Mitchell and Mulherin (1994), Chan (2003), and Tetlock, Saar-Tsechansky, and Macskassy (2007)). Alternatively, Merton (1987) argues that investors will buy and hold only those securities which they are aware of. The most common way to facilitate investors' awareness is to promote the visibility of the firm through the media. Consistent with this view, empirical evidence shows that media exposure is associated with investors' stock picking and trading activities (Falkenstein (1996), Wysocki (1999), Huberman and Regev (2001), Antweiler and Frank (2004), Antunovich and Sarkar (2006), Chen, Noronha, and Singal (2006), Barber and Odean (2007), and Tetlock (2007)). The extent of media exposure has also been shown to be related to IPO underpricing, volumes, return volatility and post-offer performance (Reese (1998), Ducharme, Rajgopal, and Sefcik (2001), Demers and Lewellen (2003), and Johnson and Marietta-Westberg (2004)).

Our paper differs from this literature in that we investigate whether the media was responsible for the phenomenal rise and fall in the market value of internet stocks from 1996 to 2000. We do not focus on what drives the media coverage of IPO firms, or on the effect of pre-IPO media coverage on the first day's return. Second, unlike most of the above papers, we look at both the number of news items and their type (good, bad or neutral) from all media sources to capture the aggregate media effect.

III. Data

A. The IPO Sample

We start with a large sample of firms that went public between January 1996 and December 2000. After excluding unit offers, rights offers, closed-end mutual funds, REITs, and ADRs, our search of the Thomson Financial's SDC database yielded 2,603 completed issues.

We identify and extract 461 internet companies from this sample using the reference list from Loughran and Ritter (2004). We then remove one firm due to its lack of media coverage and one firm that went public twice and was, therefore, counted twice during our sample period. This gives us 458 internet IPOs. From the 2,142 non-internet IPOs, we exclude 15 firms due to misclassification, foreign offer, or missing offer and pricing information. We then extract a matching sample of non-internet IPOs based on offer size and offer date. Matches are formed without replacement.[6] Our final sample thus contains 458 internet IPOs and 458 non-internet IPOs.

Offer characteristics such as offer size, venture-capital backing, and the stock exchange in which the IPO first traded, are from SDC. Underwriter ranking is based on Carter and Manaster (1990) and Loughran and Ritter (2004). Stock prices and daily returns are from CRSP. Fama-French factors are obtained from French's website. We manually collect missing founding date for 415 issues from SEC filing prospectuses, subsequent 10-Ks, or news sources.

[6] We use the matched firms in two ways. For the main analysis, we ignore the individual matches, drawing conclusions based only on whether internet firms differ from non-internet firms. For robustness, we also do paired matchings (see Appendix 3 of the internet appendices).

Table 1 reports summary statistics of offer characteristics obtained from SDC and CRSP, broken down by internet and non-internet IPOs. Internet firms average a stunning 84% initial return during our sample period, more than twice the return for non-internet firms.[7] The internet IPO sample also had a cumulative return of 2016% from January 1, 1997 to March 24, 2000, whereas the non-internet IPO sample had a return of only 370%. The difference is an astonishing 1646%.

Surprisingly, the two samples do share many similarities. Because of our method of construction, the average gross proceeds are around $88 million for both samples. Both sample firms have similar *ex ante* uncertainty about their values (proxied by the filing price range) and are on average underwritten by investment banks of similar rank. About two third of the firms in each sample come from the "high-tech" industries, and most of the firms in both samples trade on Nasdaq.[8] These facts reflect the "high-tech" industry clustering in the sample period of 1996-2000.

B. The News Sample

We define the media based on the Dow Jones Interactive Publications Library's archive (DJI) of newspapers, periodicals, and newswires. After DJI's conversion to Factiva in June, 2003, we create a customized list that includes major news and business publication sources worldwide to be consistent with the news sources in DJI prior to the conversion.[9] We choose DJI and Factiva because this source does not suffer from gaps in coverage, and is the best approximation of public news for general investors (Chan

[7] We calculate initial returns as the percentage change between the final offer price and the first-day closing price. We take the first day closing price from CRSP if available within seven days of the offer date (see, for example, Lowry and Schwert (2002)).

[8] Following Benveniste, Ljungqvist, Wilhelm, and Yu (2003) and "Hi Tech Industry Group" definition in SDC, we define high-tech firms as those with three-digit SIC codes 283, 357, 366, 367, 381, 382, 383, 384, 737, 873, or 874. This definition includes industries such as pharmaceuticals, computing, computer equipment, electronics, medical and measurement equipment, software, and biotech.

[9] The resulting list of data sources includes Dow Jones Asia, Europe, Africa, North America, South America, Australia and New Zealand, and contains all the English language sources of daily news.

(2003)). We do not include magazines, since it is difficult for us to pin down precisely when the information is publicly available. We also exclude investment newsletters, analyst reports, and other sources that are not available to the general public.[10] There are more sources in Factiva towards the end of our sample period. However, the difference will not be crucial to our results because all the econometric analyses are benchmarked with the non-internet sample during the same period.

For each IPO in our sample, we "search by name" in Factiva for the period between 90 days prior to its public offer and the end of December, 2000.[11] We hand-collect all the news articles in which the IPO firm was mentioned. We do not limit our news articles only to those news items where the firm is mentioned in the headline or in the lead paragraph, because doing so could potentially exclude a large volume of news reports that actually cover the firm.

There are a total of 171,488 news items. Two co-authors read and classify each news item into one of three categories: "good", "bad", or "neutral".[12] The reading and classification began in the fall of 2002 and were completed at the end of 2004. Good (bad) news items are defined as news items which carry positive (negative) statements or implications about the firm. Neutral news items are news items that

---

[10] We exclude the following: Factiva Aviation Insurance Digest, Factiva Marine Insurance Digest, Dow Jones Emerging Market Reports, Dow Jones Commodities Service, Dow Jones Money Management Alert, and Dow Jones Professional Investor Report.

[11] "Search by name" generates news articles that are more related to the firm and therefore more focused, while "search by keyword" returns all news articles that at least mentioned the name of the firm once, which could be noisy. Prior to March 2004, Factiva re-indexed all existing news reports about target firms to the acquirer in the case of a merger. "Search by name" in this case only returns news items where both the target and the acquirer are reported. After March 2004, this problem was resolved as Factiva introduced an updated version of its database.

[12] There are two ways of classifying news items: mechanically using content analysis software or using human judgment. Mechanical classification is faster, more consistent, and less expensive. However, human judgment makes better use of context. For example, if software is programmed to classify a news item as good when positive words exceed negative words, it misclassifies news items that contain many good words about a competitor, and few bad words about the firm. Therefore, we chose human judgment for classification.

cannot be classified as good or bad. We do not classify news based on previous returns as in Chan (2003), because doing so automatically assumes the signed direction of causality from returns to news. Our judgment is based on the content of each individual news item, without forming a new expectation after each piece of news. This method of human judgment has obvious drawbacks, the most important of which is lack of consistency. To reduce possible time-varying judgment errors, we have one author start from the last firm that went public in the internet IPO sample and read the news in reverse chronological order, while the second author starts from the first firm that went public in the non-internet IPO sample and read the news in chronological order.

However, even with the above approach, we could still face possible judgment error as the same piece of news may be categorized differently by different human beings. We conduct an experiment and a simulation test to address this issue. Our simulation indicates that measurement errors, if they exist, cannot explain our results. Details of both are available online in Appendix 2 to this paper.

We define the intensity of media coverage as the number of news items about a sample firm during a specific period. For the post-IPO period, news items are classified and counted on a daily basis. We base our analysis on the publication date of the news to reflect the time when a news report is available to the public. For any given day, we aggregate news items about the same firm from multiple media sources, without delving into which media reported what. There are two reasons for this. First, the criterion to categorize media by influence is ambiguous. Very often the same contents are covered by various media sources. Second, our research design is created with the intent to investigate the impact of the intensity of the media coverage, and is based on the fact that different types of media may reach different types of investors. We also do not distinguish between “real news” and “spin news” that exist for a small subset of our stocks. This is because this distinction is ambiguous as well.

IV. Media Coverage

Did the media cover internet IPOs differently than it covered non-internet IPOs? We explore this question for the entire sample period of 1996-2000, and for two sub-periods: before and after the price peaked. The first peak definition is a market-wide definition. On March 24, 2000, the Nasdaq 100 index reached its highest point in our sample period. We follow the traditional literature and take this date as the market peak. This definition of a peak gives the sub-periods in *calendar time*. The second definition of a peak is stock-specific, and is defined as the date at which the firm's market capitalization reaches the highest point in the sample period. This definition of a peak gives the sub-periods in *event time*. Throughout this paper, we use *before the peak* and *bubble period* interchangeably, and *after the peak* and *post-bubble period* interchangeably.

A. Unconditional Media Coverage

First, we examine the unconditional media coverage of the internet sample and the non-internet sample. Figures 1-a through 1-c provide a visual presentation of the news items per day per firm over various periods of time. Figure 1-a covers the entire sample period. Figure 1-b (1-c) covers the period before and after the peak in event time (calendar time). Compared to the non-internet sample, the internet sample had significantly higher media coverage in terms of all three measures (total number of news, good news, and bad news), during all time periods (pre-peak and post-peak, event or calendar time).

Next, we use net news, defined as the difference between the number of good and bad news items, to proxy for media sentiment. During the bubble period, internet IPO firms have more positive net news than their matching sample (Figures 1-b and 1-c). This suggests that the media generally not only provided more coverage but also had a more optimistic view, whether rational or not, about internet firms in the bubble period. Figures 1-b and 1-c also reveal that post-peak, there was a dramatic shift in media sentiment. Internet IPO firms have more negative net news than their matching sample after the bubble burst. This indicates that the media had a more pessimistic view, whether rational or not, about internet firms in the post-bubble period. Finally, note the following asymmetry: the relative pessimism on internet firms over non-internet firms in the post-bubble period was *higher* than the relative optimism on internet firms over non-internet firms in the bubble period.[13]

[13] In untabulated results, our above observations of the difference in media coverage between the two samples hold, regardless of the size of the offer, the technological nature of the firm, or whether or not the issue is backed by venture capitalists.

B. Conditional Media Coverage

We now explore news coverage conditional on price movements. Specifically, we ask whether media coverage is more positive following a price increase and more negative following a price decrease, as discussed in Shiller's (2000) positive feedback hypothesis.

Figures 2-a through 2-d provide graphical evidence that media optimism and total coverage move with firm value in both calendar and event time. This result is confirmed by the daily contemporaneous and lead-lag correlations in Table 2. We report internet firm correlations above the diagonal in Table 2, and non-internet IPO firm correlations below the diagonal. Further, Table 2 reveals the difference in correlation between internet and non-internet samples: the contemporaneous daily correlation between net news and market capitalization is 0.124 and significant (0.013 and insignificant) for internet firms (non-internet firms) pre-peak, and is 0.250 and significant (0.068 and insignificant) for internet firms (non-internet firms) post-peak. This suggests that the positive feedback discussed by Shiller (2000) exists especially for internet stocks. Although not reported, we find the same positive contemporaneous correlation at weekly and monthly frequencies.

We also observe that the net news per firm spiked before March 24, 2000 (Figure 2-a), or before the firm reached its maximum value (Figure 2-b), suggesting that media sentiment turned before the market peaked. Table 2 shows that for both internet and non-internet firms during the bubble and post-bubble periods, the magnitude of daily correlation between market capitalization and lagged net news is *always larger* than the corresponding magnitude of daily correlation between net news and lagged market capitalization. This suggests that net news leads market capitalization rather than the other way around. Although not reported, we observe the same contemporaneous and lead-lag results at weekly and monthly levels. We explore this tantalizing result formally in the next two sections, where we ask whether media sentiment affects contemporary and future risk-adjusted returns.

In Panel A of Table 3, we explore whether news follows returns, as the positive feedback hypothesis suggests. We report the results based on two arbitrarily selected cutoff points about the degree of price movement: price increases or decreases more than 0% and 1% from previous day. Panel A of

Table 3 reveals that during the bubble, if prices increased in the previous day, then net news today was much more positive for internet stocks (0.186 per day) than for non-internet stocks (0.089 per day). Post-bubble, if prices decreased in the previous day, then net news today was more negative for internet stocks (-0.146 per day) than for non-internet stocks (-0.022 per day). This means that Shiller's (2000) hypothesis of media hype exists especially for internet shares. Using alternative cutoff points, or monthly instead of daily frequencies, or abnormal returns instead of raw returns do not change our results.

Panel A of Table 3 also shows that in the bubble stage, even if prices decreased yesterday, net news today is still positive. In the post-bubble stage, even if prices increased yesterday, net news today is still negative. This finding suggests that the media ignored bad information during the bubble period and ignored good information in the post-bubble period.[14]

It is possible that the observed differential media coverage is not driven by the distinction between "internet" and "non-internet" IPOs but by some other distinguishing features. While the two samples are matched by size at the beginning of each firm's public trading period, market capitalizations differ greatly as time progresses because cumulative returns from the beginning differ greatly between these samples. Therefore we control for contemporary market capitalization, which also controls for cumulative returns. News coverage could also be affected by past market conditions (trading volumes, bid-ask spreads, abnormal returns, and variance of stock returns in the past twenty trading days) and past news coverage, where past news coverage is measured by both sentiment (net news) and intensity (total news). To take into account the effects of these differences between internet and non-internet firms on media coverage, we regress tomorrow's net news against today's abnormal returns, a dummy variable for internet IPOs, the

[14] The results documented in Table 3 cannot be explained by the fact that news disseminates throughout the day (some newspapers publish in the afternoon and therefore can capture the event occurring during the same day, and other newspapers publish in the morning and therefore could delay the news about the market or firm's price movement to the next day). This is because we observe the same pattern using monthly data.

interaction between today's abnormal returns and the dummy variable, and the control variables described above.[15] We run this regression once for the pre-peak period and once for the post-peak period.

Panel B of Table 3 presents the multivariate regression results. According to the intercept term, news coverage for internet firms is more optimistic than news coverage for non-internet firms pre-peak (0.070 pieces of net news versus 0.049 pieces of net news), and is more pessimistic than news coverage for non-internet firms post-peak (-0.063 pieces of net news versus -0.029 pieces of net news). More importantly, we observe that today's abnormal returns are positively related to tomorrow's net news. In addition, the same increase in today's abnormal return leads to 0.001 pieces of net news tomorrow for non-internet stocks, but leads to 0.004 pieces of net news tomorrow for internet stocks during the bubble. This indicates that Shiller's (2000) hypothesis of media hype exists especially for internet shares, which suggests a potential role for the media in the internet IPO bubble. Post-peak, the effect of today's abnormal return on tomorrow's net news is similar in economic magnitude for internet stocks and for non-internet stocks, albeit statistically insignificant for both groups.

Therefore, even after addition of controls that may explain the differential media coverage between internet stocks and non-internet stocks, our basic conclusion from Figures 1 and 2 and Panel A of Table 3 does not change: the overall media coverage was more positive about internet IPOs in the bubble period and was more negative about internet IPOs post-bubble. This suggests that media hype was particularly elevated for internet IPOs.

Many researchers have pointed out the 1996-2000 bubble was not just an internet bubble but also a high-tech bubble. However, Table 1 shows that our internet sample does not contain a significantly higher proportion of high-tech stocks than our non-internet sample. So why, in Figure 2 and Table 3, do we observe the differential media coverage between internet stocks and non-internet stocks? The only

[15] So we include the following control variables: day $t-1$ abnormal close-to-close return obtained by fitting a Fama-French three-factor model for each firm ($ABRET_{i,t-1}$), day t number of net news ($NN_{i,t}$) and total news ($TN_{i,t}$) per firm, day t bid-ask spread ($Bid\text{-}Ask_{i,t}$), number of shares traded ($Vol_{i,t}$), market value of equity ($MVE_{i,t}$), and $VAR\_RET_{i,t}$, the variance of firm's stock returns in the past twenty days (from day $t-19$ to day t).

resolution to this puzzle is to acknowledge that there was something unique about the label “internet” that made the media cover this group of stocks differently. We explore this issue further in Appendix 3 of the internet appendices of this paper.

V. The Marginal Effect of Media Coverage

In the previous section, we documented the differences in aggregate media coverage between internet firms and non-internet firms. This is a necessary, but not sufficient, condition to conclude that the media had a role in the meteoric rise and fall of internet stocks in the late 1990s. We now explicitly examine the impact of this differential media coverage on stock prices. In this section, we investigate the *marginal* impact of the media on abnormal returns at the firm level and then at the portfolio level. In the next section, we investigate the *total* impact of the media on abnormal returns.

A. Firm-level Analysis

We examine the effect of today’s news on today’s return and on future returns, with the future extending to twenty trading days after the news comes out (roughly a month in calendar time). We adopt the Jorda (2005) least squares estimation of the impulse response function.

The purpose of Jorda’s (2005) empirical methodology is to estimate the effect of an independent variable on contemporary and future values of a dependent variable. It consists of estimating a set of linear OLS regressions, where the first regression regresses the contemporary value of the dependent variable on the contemporary value of the independent variable, the second regression regresses the lead (where the lead is by one period) value of the dependent variable on the contemporary value of the independent variable, the third regression regresses the lead (where the lead is by two periods) value of the dependent variable on the contemporary value of the independent variable, and so forth. Jorda’s approach is intuitively appealing, especially in our case. First, it allows us to test if today’s news items (the independent variable) affect future abnormal returns (the dependent variable) and, if yes, to find out how long into the future the effect lasts. Second, it allows us to aggregate both the direct effect of news –

contemporary news affects contemporary and future returns – as well as the indirect effect of news ("piggyback effect") – contemporary news affects future news which in turn affects future returns.[16]

Our dependent variable is the daily abnormal return of a firm's stock estimated by fitting a Fama-French (1993) three-factor model.[17] Our primary independent variable is net news, which is the difference between the number of daily good news items and the number of daily bad news items for each firm. To apply the Jorda (2005) approach, we run the following twenty-one OLS regressions one at a time in the period 1996 through 2000:

(1)
$$
\begin{aligned}
ABRET_{i,t} &= \alpha_0 + \beta_0 \times NN_{i,t} + \rho_0 \times ABRET_{i,t-1} + \mathbf{\Theta_0^T \Omega_{i,t}} + \varepsilon_{i,t} \\
ABRET_{i,t+1} &= \alpha_1 + \beta_1 \times NN_{i,t} + \rho_1 \times ABRET_{i,t} + \mathbf{\Theta_1^T \Omega_{i,t}} + \varepsilon_{i,t+1} \\
&\vdots \\
ABRET_{i,t+20} &= \alpha_{20} + \beta_{20} \times NN_{i,t} + \rho_{20} \times ABRET_{i,t} + \mathbf{\Theta_{20}^T \Omega_{i,t}} + \varepsilon_{i,t+20}
\end{aligned}
$$

where $ABRET_{i,t+s}$ is the abnormal stock return for firm i on day $t + s$ obtained by fitting the contemporaneous Fama-French (1993) three factors, and $NN_{i,t}$ is the net news for firm i on day t.

We include $ABRET_{i,t}$ as a control for daily reversals or daily momentum and for bid-ask bounce (Roll 1984). Vector $\mathbf{\Omega_{i,t}}$ contains other firm-specific control variables. These are market microstructure variables that control for transaction costs – the day t bid-ask spread $Bid\text{-}Ask_{i,t}$ (closing bid-ask spread scaled by the price at the end of the day) – and control for liquidity – the day t trading volume $log(1+Vol_{i,t})$ for each firm i, where $Vol_{i,t}$ is the number of shares traded at time t. We include a control for size, $log(MVE_{i,t})$, and a control for return's volatility, $VAR\_RET_{i,t}$, where $MVE_{i,t}$ is the market value of equity on day t and $VAR\_RET_{i,t}$ is the variance of firm i's returns in the past twenty days (from day t - 19 to day t).

We include two control variables for news coverage. One is $log(1 + TN_{i,t})$, where $TN_{i,t}$ is the total news for firm i on day t. As we notice a vast difference in the amount of news between the internet and

[16] For a detailed discussion of Jorda's (2005) approach, please see Appendix 1 in the internet appendices.

[17] To ensure the conservatism of our news analysis, we use contemporaneous Fama-French factors to control for the most recent market-wide information, though our results are almost identical with respect to both economic and statistical significance if lagged factors are selected.

non-internet firms in the previous section, we add $log(1 + TN_{i,t})$ as a scaling variable to normalize the net news variable. Since large firms tend to receive more media attention than small firms, $log(1 + TN_{i,t})$ can also be viewed as a control for size. Finally, if there is an attention effect, cumulative amount of news till that point in time can drive returns as well as net news. So our last control variable is a measure of cumulative news, $CUMNN_{i,t}$, where $CUMNN_{i,t}$ is net news cumulated from day t - 5 to day t - 1 (a week in calendar time). In the robustness analysis (see Appendix 3 of the internet appendices), we use another measure of cumulative news: the total news cumulated from the day the firm went public till time t.

Table 4 reports the results of the marginal effects of the media coverage. Panel A presents the effect of today's news on today's and tomorrow's abnormal returns during the bubble and post-bubble periods. Panel B reports the coefficients associated with today's news on returns extending up to five trading days. As cross-sectional correlation exists in the standard errors, p-values are corrected for clustering of errors by date. We also use the Fama-Macbeth (1973) methodology to address the problem of cross-sectional correlation of errors, and our results are similar.

Columns 1 and 2 of Panel A in Table 4 illustrate the *contemporaneous* media impact: net news in the bubble period generates an extra 110.2 and 136.5 basis points in the same day returns for an internet firm and a non-internet firm, respectively. This implies that, after controlling for total news coverage and market trading conditions, the marginal impact of net news on the same day risk-adjusted returns is 26.3 basis points smaller for internet firms than for non-internet firms. This difference is also statistically significant ($p = 0.01$). Columns 5 and 6 of Panel A in Table 4 show that post-bubble, the marginal impact of contemporaneous net news on the same day risk-adjusted returns is 58.3 basis points smaller for internet firms than for non-internet firms. The difference is again statistically significant ($p = 0.00$).

The above observations apply not only for the contemporaneous abnormal returns ($s = 0$), but also for the *next day's* abnormal returns ($s = 1$). Columns 3 and 4 of Panel A in Table 4 show that the marginal impact of today's net news on tomorrow's abnormal returns during the bubble period is 21.2 basis points smaller for internet firms than for non-internet firms. The difference is again highly statistically significant ($p = 0.00$). Columns 7 and 8 of Panel A in Table 4 show that the marginal impact of today's net news on tomorrow's abnormal returns is 11.2 basis points smaller for internet firms post-bubble. The difference is not statistically significant. These results suggest that even though net news today leads to higher abnormal returns tomorrow, this effect is lower for internet firms, *especially* during the bubble period.

The results in Panel B of Table 4 indicate that it takes approximately two days for the effect of net news on abnormal returns to die down. Although we have run these tests for up to twenty days ahead, we note that for abnormal returns two days and later, the coefficients associated with net news variable are not significantly different from zero most of the time. So in Panel B we report only the coefficients up to five days in the future. It should be mentioned here that though Busse and Green (2002) have shown that stock prices react within seconds to CNBC news, we find that the effect of news lasts more than one day. Our findings are also consistent with Antweiler and Frank (2005).

To summarize, Table 4 indicates that net news positively affects *contemporaneous* abnormal returns for internet IPOs, but the effect is lower in all periods compared to non-internet IPOs. Net news is also positively related to the *next day's* abnormal returns for internet IPOs, but the effect is again lower compared to non-internet IPOs, *especially* during the bubble period.[18]

Recall that Figures 1-a through 1-c show that for all IPOs, there was more good news than bad news in the bubble period, and more bad news than good news in the post-bubble period. However, the difference was higher for the internet firms. The above evidence, coupled with our finding that the market discounted the net news from internet firms in both the bubble and post-bubble periods, suggests that the market perhaps downplays the good news about internet firms during the bubble period and downplays the bad news about internet firms post-bubble.

Our results implicitly assumed that media coverage immediately causes returns (Table 4), and that media coverage responds to returns only slowly (after one day, Panel B of Table 3). If instead, media coverage responds immediately to returns but returns do not respond immediately to media coverage, then we should ignore the contemporaneous relationship between media coverage and returns when estimating

[18] In a study whose results complement ours, Agrawal and Chen (2007) find that investors appear to have discounted analyst opinions more during the stock market bubble.

the marginal effect of media coverage, which is equivalent to ignoring the result of $s = 0$ from Table 4. However, the results of $s = 1$ from Panel A of Table 4 suggests that our main finding – the market discounts media coverage for internet stocks relatively to non-internet stocks – still holds.[19]

However, within a given day, we still cannot say whether news causes returns or whether returns cause news. The only way to definitively resolve this conundrum, or more generally, address Granger-causality, is to analyze higher frequency data than daily data. We do not have this higher frequency data for news. In an attempt to explore the timing issue of news coverage, we re-estimate our tests in Table 4 using open-to-close returns instead of close-to-close returns.

Table 5 reports the results. Notice that news today does not affect open-to-close returns tomorrow. This suggests that prices reflect news in less than a day. Notice also that contemporaneous news affects returns less for internet stocks than for non-internet stocks, which is the same conclusion we drew from Table 4. However, without higher frequency data, we cannot fully resolve whether news causes returns or whether returns cause news within a given day.

B. Portfolio-level Analysis

Table 4 examined the marginal impact of the media on individual firm returns. We now examine the effect of the media on the following four portfolios – equally-weighted internet and non-internet portfolios, and value-weighted internet and non-internet portfolios.

[19] It is possible that the timing of news may differ between internet stocks and non-internet stocks. For example, today's news for internet firms may come out in the evening while today's news for non-internet firms comes out in the afternoon. If news causes returns, we would expect a smaller contemporaneous effect of media on returns for internet firms. The following day, however, we would expect a larger effect, as the news about internet firms is incorporated into prices. Instead, we observe a smaller effect for internet stocks on both days. Therefore, this type of timing difference cannot explain our results. The flip side of the above example is that it is also possible for today's news about internet firms to arrive in the afternoon, but today's news about non-internet firms to arrive in the evening. If news causes returns, then we would expect to see a smaller contemporaneous effect for non-internet stocks than for internet stocks. We do not observe this. Again, timing difference cannot explain our results.

In Table 6, we present portfolio-level results using the same Jorda (2005) empirical methodology as that in Table 4.[20] Value-weighted portfolio results are the same as the firm-level results. The impact of daily news on contemporary abnormal returns is positive for all IPOs, but it is less positive for the internet IPOs. This is especially true during the bubble period. However, now the effect lasts less than a day – there seems to be no effect of today's news on tomorrow's returns. For the equally-weighted portfolio, the effect of news on today's abnormal returns is not significantly different between internet stocks and non-internet stocks, and these effects die out the next day.

The difference in results between the value-weighted portfolio and the equally-weighted portfolio suggests that the market perhaps discounts media sentiment for large internet firms. Since large firms attract more media attention, value-weighting may be a more relevant weighting method than equal-weighting in our analysis. So we will abide by the results for the value-weighted portfolio, but we will note the insignificance of the results from the equally-weighted portfolio (the coefficients of net news are not significantly different between the two samples during the bubble for $s = 0$).

We conclude the following from our portfolio analysis: when examined at the portfolio level, net news about internet firms was less credible than net news about non-internet firms, especially in the bubble period. This conclusion is weaker than our conclusion from the firm analysis in Table 4, where we had an economically and statistically significant difference in both the bubble and the post-bubble periods.

VI. The Cumulative Effect of Media Coverage

In the previous section we showed that the *marginal* effect of the media on risk-adjusted returns of internet firms was lower than its effect on the risk-adjusted returns of non-internet firms. However, as internet firms had more news items than non-internet firms (see Figures 1-a through 2-d), we could argue

[20] Control variables $VAR\_RET_{i,t}$, bid-ask spreads, and trading volumes are included in the individual IPO tests, but are excluded from the portfolio-level analysis. This is because these are control variables for firm-level return volatility and liquidity measures, respectively, and cannot be averaged to capture the volatility and liquidity of the portfolio. We also remove $log(MVE_{i,t})$ as we are constructing value-weighted portfolios. Again, *p*-values are corrected for clustering of errors by date.

that the media had a higher *overall* effect on the risk-adjusted returns of internet firms. We now turn to the analysis of the impact of the media on *cumulative* difference in returns between the internet sample and the non-internet sample. [21]

We start with the difference in the *actual* cumulative returns between the two sample firms. For a given day, we compute the mean one-day actual return for the internet stocks. We cumulate these average returns over time from the first trading day of 1997 to the last trading day of December 2000. The same process is repeated for the non-internet stocks. We then take the difference between the two cumulative returns, which we call "Internet Sample – Non-Internet Sample: Actual." This is the top curve in Figure 3 (the red solid line). We notice that the returns of the internet IPOs dramatically outperform the returns on the non-internet IPOs till March 24, 2000, and then there is a dramatic reversal. It is this sudden rise and fall that is popularly known as the "internet bubble." The difference in cumulative returns between internet firms and non-internet firms is 1646% at its March 24, 2000 peak.

Next, we compute three sets of differences in *predicted* cumulative returns between the two sample firms over the same period. First, we obtain the predicted daily return of an internet stock from the following regression:

(2) $$RET_{i,t} = \alpha_0 + \mathbf{\Phi_0^T FF}_{i,t} + \varepsilon_{i,t}$$

where $RET_{i,t}$ is the stock return for firm i on day t and $\mathbf{FF}_{i,t}$ is the contemporary Fama-French three factors. We then compute the average daily predicted returns for the internet stocks and cumulate over the 1997-2000 period. We do the same for the non-internet firms. We then take the difference, which we call "Internet Sample – Non-Internet Sample: Predicted (FF)" to capture the difference in predicted cumulative returns (by Fama-French three-factor model) between internet and non-internet firms. The difference is 352% on March 24, 2000. Note that while the Fama-French three risk factors explain 21.4% (=352%/1646%) of the difference in cumulative returns between internet stocks and non-internet stocks,

[21] Throughout 1996, there are only 18 firms in our non-internet sample that went public. So, unlike the other tests in this paper which begin the sample in 1996, we construct cumulative returns in this section starting from 1997.

most of the difference is still left unexplained. This is the reason why the "internet bubble" remains a puzzle to researchers attempting to explain it using rational asset pricing models.

Second, we compute the predicted return to an average internet stock from both contemporaneous Fama-French three factors and a set of control variables. The control variables used in this analysis are the independent variables that we used in Table 4, *except* the news variables and market capitalization. For each day and each internet firm, we regress the abnormal return from fitting Fama-French three factors in regression (2) against the above set of control variables to obtain the predicted abnormal return based on these control variables. Then we add the predicted return from the Fama-French three factors. This process generates the return predicted by both the Fama-French three factors and the above control variables for a single internet firm. We then compute the average daily predicted returns and cumulate over the 1997-2000 period. The above step is repeated for non-internet firms. We then take the difference, which we call "Internet Sample – Non-Internet Sample: Predicted (FF + controls, news variables excluded)." This is the bottom curve in Figure 3 (the blue dashed line). The difference in cumulative predicted returns is 482% on March 24, 2000. The control variables improve the fit of the Fama-French three-factor model. The model improves from explaining 21.4% (=352%/1646%) of the difference in cumulative returns between internet stocks and non-internet stocks to explaining 29.3% (=482%/1646%) of the difference. Notice again that most of the bubble is left unexplained.

We now *add* the news variables to the above set of control variables and repeat the above process. The difference is now called the "Internet Sample – Non-Internet Sample: Predicted (FF + controls, news variables included)." This is the curve above the bottom curve in Figure 3 (the pink dashed line). The difference in cumulative predicted returns is 530% on March 24, 2000. We observe that the addition of the news variables has hardly any effect on the fit of the model. The model improves from explaining 29.3% (=482%/1646%) of the difference in cumulative returns between the two sample firms to 32.2% (=530%/1646%). Thus, news variables explain only 2.9% of the difference in cumulative returns between internet and non-internet IPOs in the January 1, 1997 to March 24, 2000 period. This implies that though net news does improve our ability to explain internet stock prices, the improvement is negligible. In other words, we can make the same point about internet IPOs that Roll (1988) made for all stocks: news does not explain much of realized returns even *ex post*.

We next conducted a battery of robustness tests to find out whether our results could be explained by other reasons like biases in news classification using human judgment, change in information environment such as the implementation of Regulation Fair Disclosure law, alternative specifications of key variables, price-driven news, survivorship bias, investor boredom, lock-up expiration, investor learning, high-tech nature of internet stocks, different credibility of news sources, spin new versus real news, and scaled news versus unscaled news. We found that they cannot explain our results. Appendices 2 and 3 in the internet version of the paper discuss these tests.

VII. Conclusion

ISDEX, an authoritative and widely cited internet stock index, rose from 100 in January 1996 to 1100 in February 2000 – an incredible increase of about 1000% in four years – only to fall to 600 in May 2000 – an incredible decrease of about 45% in four months. Of all the bubbles in history, this internet bubble ranks amongst the most spectacular.

Though there is some agreement that such a spectacular rise and fall of internet stock prices cannot be explained by fundamentals, there is less agreement on what can explain it. One explanation is given by Shiller (2000), who argued that (1) the media hyped internet stocks, and (2) this hype was a major factor responsible for the internet bubble.

We first show that media coverage for internet IPOs was different than a matching sample over the 1996-2000 period. There was more media coverage for internet firms than for non-internet firms. However, coverage was more positive for internet IPOs in the bubble period and more negative post-bubble. Consistent with Shiller's (2000) first hypothesis on media hype, we find higher abnormal returns lead to more positive media coverage the following day. The effect is stronger for internet stocks.

We next show that the media coverage was not a significant factor in the internet bubble. The media explains only 2.9% of the difference in between internet and non-internet firm returns from January 1, 1997 to March 24, 2000 (the day the Nasdaq peaked). This is because the market downplayed media sentiment: though today's net news affected today's and tomorrow's risk-adjusted returns for both groups of IPOs, the effect was lower for internet IPOs, especially in the bubble period. Overall, these results provide evidence against Shiller's (2000) second hypothesis – the media was not a significant factor in the dramatic rise and fall of internet shares in the late 1990s.

References

Agrawal, A. and M. Chen. "Do Analyst Conflicts Matter? Evidence From Stock Recommendations." *Journal of Law and Economics*, forthcoming (2007).

Antunovich, P., and A. Sarkar. "Fifteen Minutes of Fame? The Market Impact of Internet Stock Picks." *Journal of Business*, 79 (2006), 3209–3251.

Antweiler, W., and M. Frank. "Is All that Talk Just Noise? The Information Content of Internet Stock Message Boards." *Journal of Finance,* 59 (2004), 1259-1294.

Antweiler, W., and M. Frank. "Do US Stock Markets Typically Overreact to Corporate News Stories?" Working Paper, University of Minnesota (2005).

Battalio, R., and P. Schultz. "Options and the Bubble" *Journal of Finance,* 61 (2006), 2071-2102.

Barber, B., and T. Odean. "All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors." *Review of Financial Studies*, forthcoming (2007).

Baron, D. "Persistent Media Bias." *Journal of Public Economics,* 90 (2006), 1-36.

Benveniste, L.; A. Ljungqvist; W. Wilhelm Jr.; and X. Yu. "Evidence of Information Spillovers in the Production of Investment Banking Services." *Journal of Finance,* 58 (2003), 577-608.

Bhattacharya, U.; N. Galpin; R. Ray; and X. Yu. "The Role of the Media in the Launch of Internet IPOs." *Betriebswirtschaftliche Forschung und Praxis,* 58 (2006), 442-456.

Brown, S., and J. Warner. "Measuring Security Price Performance." *Journal of Financial Economics,* 8 (1980), 205-258.

Brown, S., and J. Warner. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics,* 14 (1985), 3-31.

Brunnermeier, M., and S. Nagel. “Hedge Funds and the Technology Bubble.” *Journal of Finance,* 59 (2004), 2013-2040.

Busse, J., and T. Green. “Market Efficiency in Real Time.” *Journal of Financial Economics,* 65 (2002), 415-437.

Carter, R., and S. Manaster. “Initial Public Offerings and Underwriter Reputation.” *Journal of Finance,* 45 (1990), 1045-1068.

Chan, W. “Stock Price Reaction to News and No-News: Drift and Reversal after Headlines.” *Journal of Financial Economics,* 70 (2003), 223-260.

Chen, H.; G. Noronha; and V. Singal. “Investor Awareness and S&P 500 Index Changes.” *Financial Analysts Journal*, 62 (2006), 31-47.

Cooper, M.; O. Dimitrov; and P. Rau. “A Rose.com by Any Other Name.” *Journal of Finance,* 56 (2001), 2371-2388.

Cutler, D.; J. Poterba; and L. Summers. “What moves stock prices?” *Journal of Portfolio Management,* 15 (1989), 4-12.

DeMarzo, P.; R. Kaniel; and I. Kremer. “Relative Wealth Concerns and Financial Bubbles.” *Review of Financial Studies,* forthcoming (2007).

Demers, E., and K. Lewellen. “The Marketing Role of IPOs: Evidence from Internet Stocks.” *Journal of Financial Economics,* 68 (2003), 413-437.

Ducharme, L.; S. Rajgopal; and S. Sefcik. “Why was Internet IPO Underpricing So Severe?” Working Paper, University of Washington (2001).

Dyck, A., and L. Zingales. “The Bubble and the Media.” in *Governance and Capital Flows in a Global Economy*, Chapter 7, P. Cornelius, and B. Kogut, eds. New York, NY: *Corporate* Oxford University Press (2003).

Falkenstein, E. “Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings.” *Journal of Finance,* 51 (1996), 111-135.

Fama, E., and J. MacBeth. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy,* 81 (1973), 607-636.

Fama, E., and K. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics,* 33 (1993), 3-56.

Griffin, J.; J. Harris; and S. Topaloglu. "Who Drove and Burst the Tech Bubble?" Working Paper, University of Texas (2005).

Huberman, G., and T. Regev. "Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar." *Journal of Finance,* 56 (2001), 387-396.

Johnson, W., and J. Marietta-Westberg. "The Effect of News on Volatility: A Study of IPOs." Working Paper, Michigan State University (2004).

Jorda, O. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review,* 95 (2005), 161-182.

Lamont, O., and R. Thaler "Can the Market Add and Subtract? Mispricing in Tech Stock Carve-outs." *Journal of Political Economy,* 111 (2003), 227-268.

Loughran, T., and J. Ritter. "Why has IPO Underpricing Changed over Time?" *Financial Management,* 33 (2004), 5-37.

Lowry, M., and G. Schwert. "IPO Market Cycles: Bubbles or Sequential Learning?" *Journal of Finance,* 57 (2002), 1171-1200.

Merton, R. "Presidential Address: A Simple Model of Capital Market Equilibrium with Incomplete Information." *Journal of Finance,* 42 (1987), 483-510.

Mitchell, M., and J. Mulherin. "The Impact of Public Information on the Stock Market." *Journal of Finance,* 49 (1994), 923-950.

Mullainathan, S., and A. Shleifer. "The Market for News." *American Economic Review,* 94 (2005), 1031-1053.

Niederhoffer, V. "The Analysis of World Events and Stock Prices." *Journal of Business,* 44 (1971), 193-219.

Ofek, E., and M. Richardson. "The Valuation and Market Rationality of Internet Stock Prices." *Oxford Review of Economic Policy*, 18 (2002), 265-287.

Ofek, E., and M. Richardson. "DotCom Mania: The Rise and Fall of Internet Stock Prices." *Journal of Finance,* 58 (2003), 1113-1137.

Pastor, L., and P. Veronesi. "Was There a NASDAQ Bubble in the Late 1990s?" *Journal of Financial Economics,* 81 (2006), 61-100.

Pastor, L., and P. Veronesi. "Technological Revolutions and Stock Prices." Working Paper, University of Chicago (2007).

Reese, W., Jr. "IPO Underpricing, Trading Volume, and Investor Interest." Working Paper, Tulane University (1998).

Roll, R. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *Journal of Finance*, 39 (1984), 1127-1139.

Roll, R. "R-Squared", *Journal of Finance,* 43 (1988), 541-566.

Shiller, R. *Irrational Exuberance*. Princeton University Press, Princeton, NJ (2000).

Tetlock, P. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance,* 62 (2007), 1139-1168.

Tetlock, P.; S. Macskassy, S.; and M. Saar-Tsechansky. "More than Words: Quantifying Language to Measure Firms' Fundamentals." *Journal of Finance*, forthcoming (2007).

Wysocki, P. "Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards." Working Paper, University of Michigan (1999).

Figure 1. Daily News Coverage

The figures below compare the daily average numbers of good, bad, total and net new items per firm over various periods of time between internet and non-internet IPOs. Net news is defined as the difference between the number of good news and the number of bad news. Figure 1-a covers the entire sample period (1996-2000). Figures 1-b covers the period before and after the peak in event time, where the peak is the day that the firm's stock price peaked. Figures 1-c covers the period before and after the peak in calendar time, where the peak is March 24, 2000.

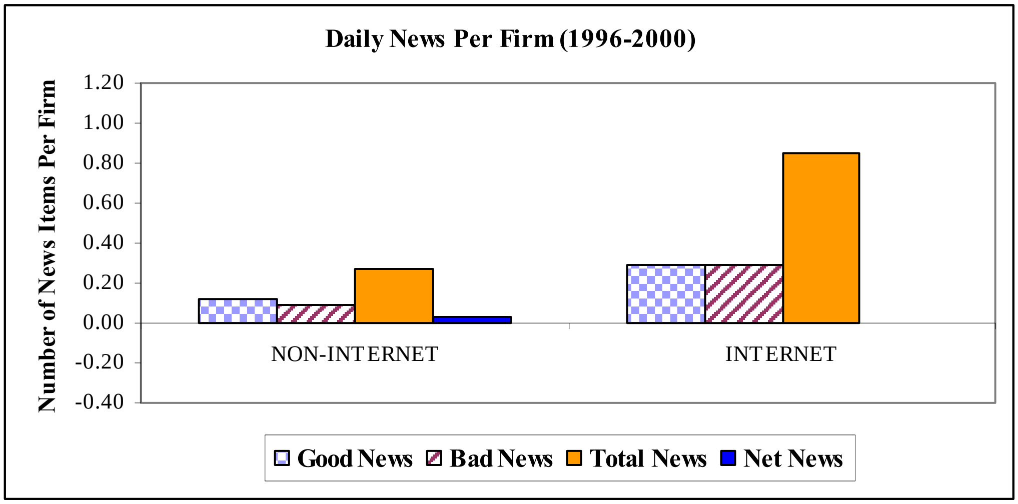


Figure 1-a. Daily average news coverage per firm for the entire sample period (1996-2000)

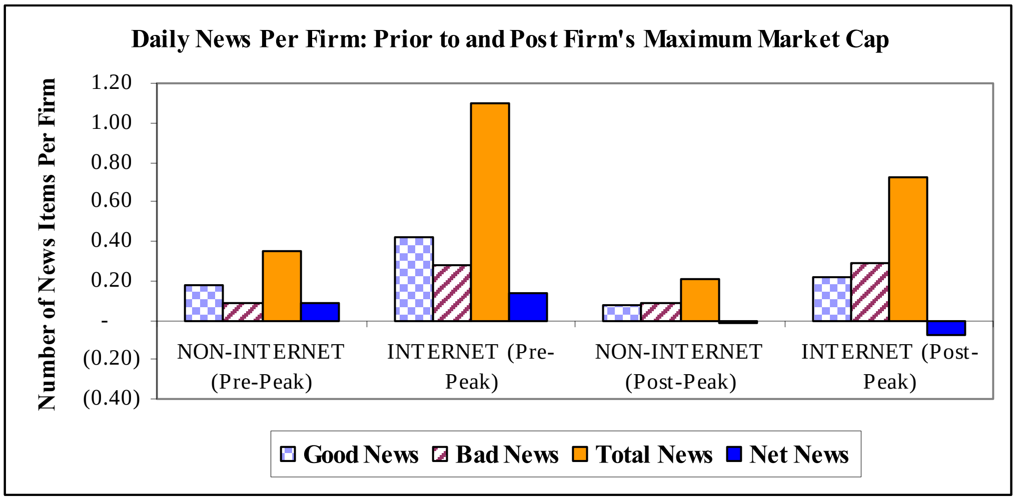


Figure 1-b. Daily average news coverage per firm before and after firm's maximum market cap

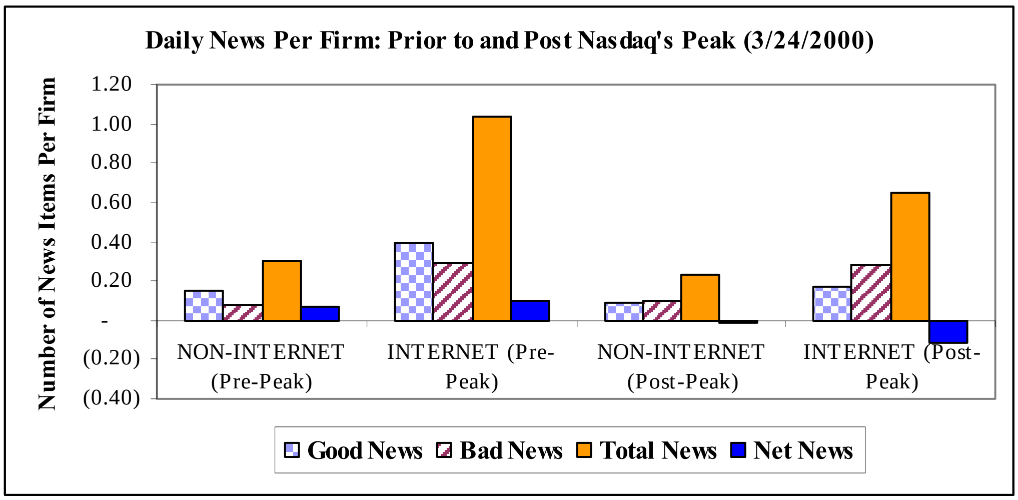
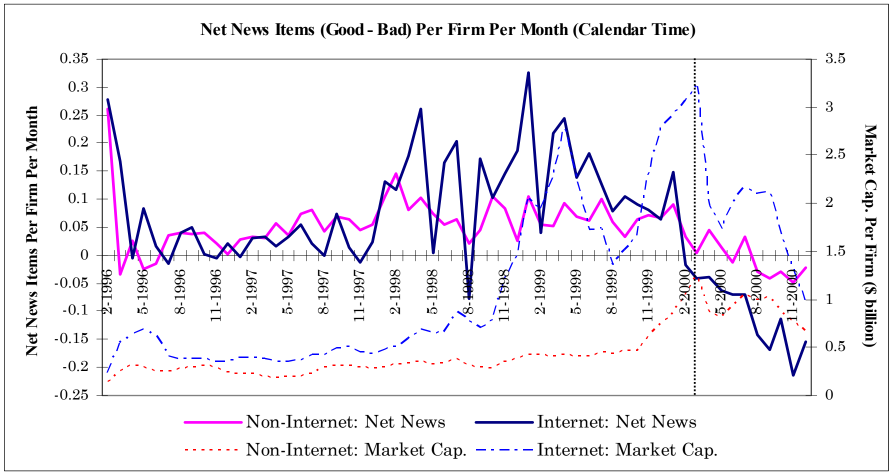


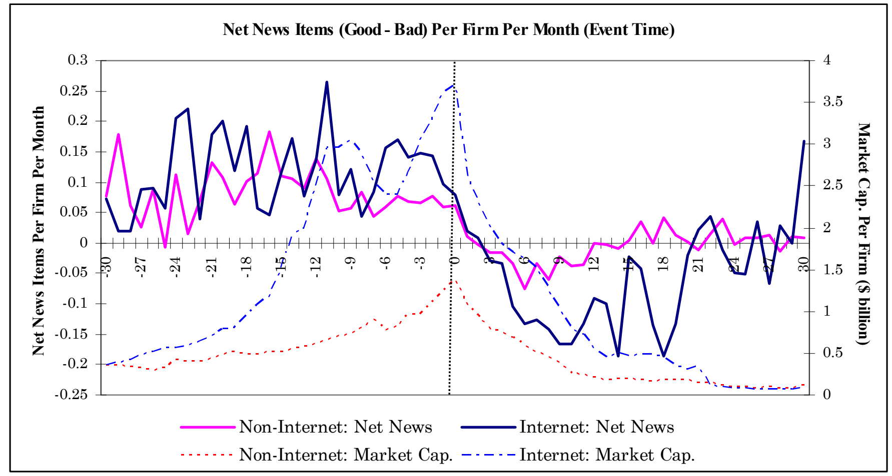
Figure 1-c. Daily average news coverage per firm prior to and post Nasdaq's Peak (March 24, 2000)

Figure 2. News Coverage and Market Capitalization

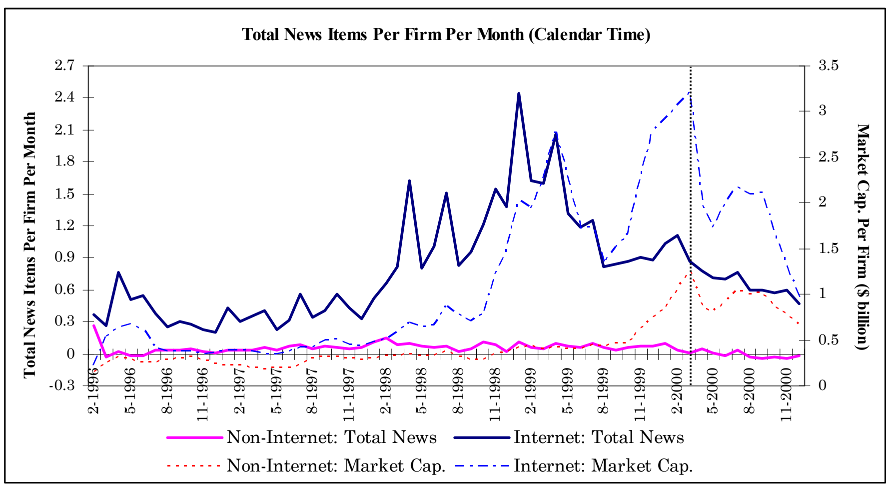
The figures below link the numbers of news items per firm per month for internet IPOs and non-internet IPOs to the respective monthly average market capitalization (in billions of U.S. dollars) between 1996 and 2000. Figures 2-a and 2-b present the linkage of monthly net news to monthly average market capitalization, where net news is the difference between the number of good news and the number of bad news. Figures 2-c and 2-d do the same for monthly total news. Vertical dotted line in each figure indicates the day of the peak, where the peak, in calendar time, is March 24, 2000, and the peak, in event time, is the day the market cap of the individual stock peaked.



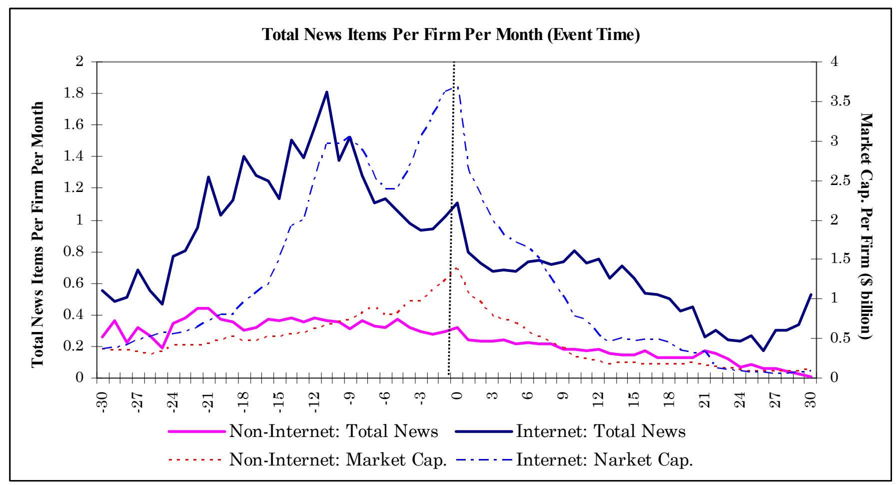
2-a



2-b



2-c



2-d

Figure 3. The Overall News Effect on Stock Returns

The figure below presents three curves of differences in cumulative returns between internet firms and non-internet firms from January 1997 to December 2000. The red solid line is the difference in the actual cumulative returns between the two sample firms. The blue dashed line is the difference in the predicted cumulative returns (by Fama-French three factors and control variables described in Section VI, excluding news variables) between the two sample firms. The pink dashed line is the difference in the predicted cumulative returns (by Fama-French three factors and control variables described in Section VI, including news variables) between the two sample firms.

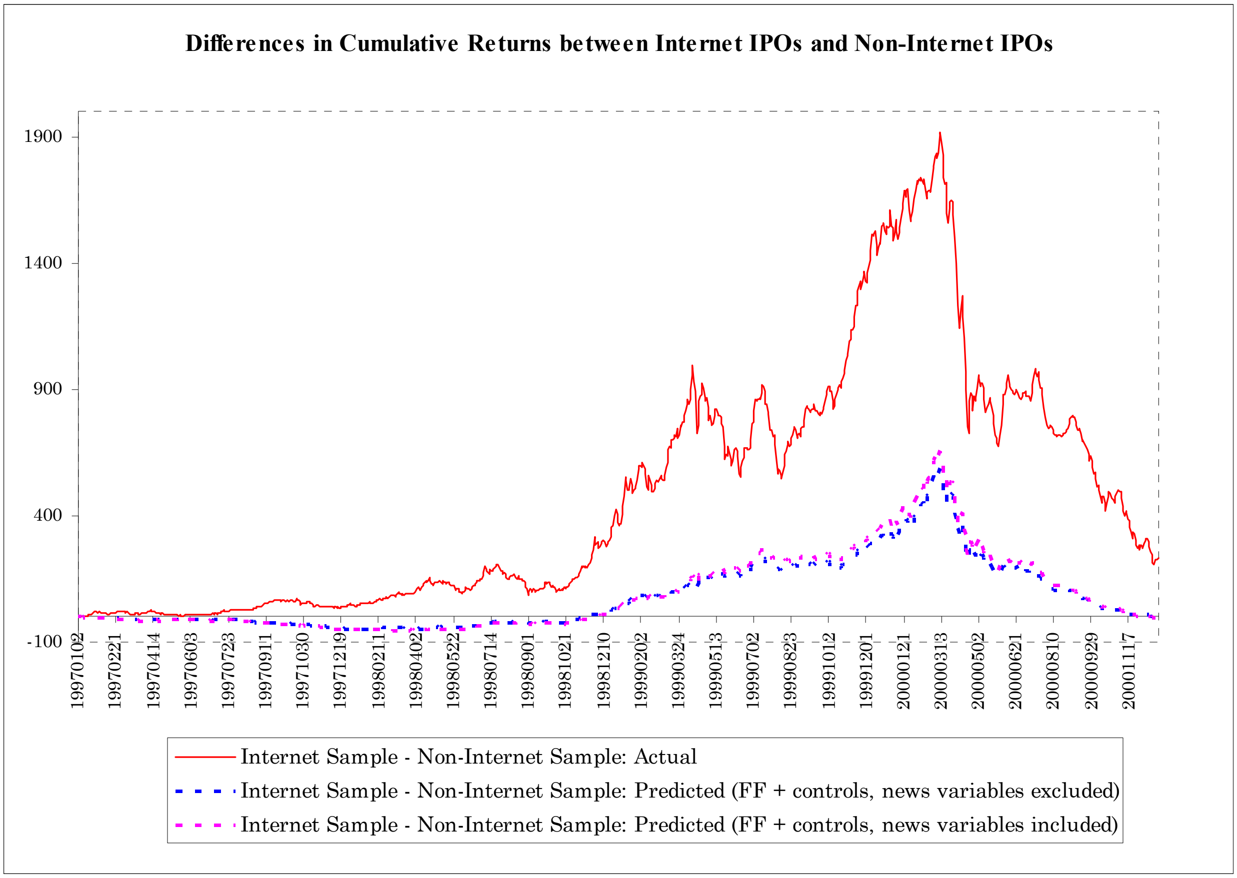


Table 1. Descriptive Statistics of Sample Firms

The sample period is 1996-2000. Internet IPOs are identified as in Loughran and Ritter (2004). Gross proceeds do not include the over-allotment option. Filing price range is the difference between the high and low prices suggested in the preliminary prospectus filed with SEC. Price revisions are the percentage change between the expected and final offer prices over the book-building period. Initial return is the first-day close price over the offer price, minus one. Length of the book building period is the pre-IPO stage between the filing day (when a company files a preliminary prospectus with the SEC) and the pricing day (when the final offer price is set). Age is IPO year minus founding year. We manually collect missing founding date for 193 issues within the non-internet sample and 222 issues within the internet sample from SEC prospectuses, subsequent 10-Ks, or news sources. Information regarding venture capital backing is from Securities Data Corporation (SDC). Underwriter rankings are investment banker rankings based on the Loughran-Ritter (2004) update of the Carter-Manaster (1990) tombstone measure. "High-tech" industries are classified by the first three-digit SIC codes 283, 357, 366, 367, 381, 382, 383, 384, 737, 873 and 874, covering industries such as pharmaceuticals, computing, computer equipment, electronics, medical and measurement equipment, biotech, and software industries. **, * represents the difference between the internet sample and the non-internet sample at 1% and 5% level (two-sided, Satterthwaite test for means and Wilcoxon signed rank test for medians), respectively.

	Mean	Median	Standard Deviation	N
Gross proceeds (in $MM)				
Internet IPO sample	88.22	61.05	124.54	458
Non-Internet IPO sample	87.96	60.50	122.72	458
Filing price range				
Internet IPO sample	1.96	2	0.57	456
Non-Internet IPO sample	1.98	2	0.46	458
Price revisions				
Internet IPO sample	23.00%	18.18%	37.29%	456
Non-Internet IPO sample	4.15%**	0.00%**	27.66%	458
Initial returns				
Internet IPO sample	83.72%	49.17%	100.57%	458
Non-Internet IPO sample	41.09%**	17.68%**	68.07%	458
Length of book-building period (in days)				
Internet IPO sample	91.46	77	47.65	455
Non-Internet IPO sample	104.57	77	94.93	457
Firm age (in Years)				
Internet IPO sample	4.84	3	4.64	444
Non-Internet IPO sample	9.56**	5**	12.49	434
Fraction of venture capital backed				
Internet IPO sample	69.65%			453
Non-Internet IPO sample	54.59%**			444
Underwriter ranking				
Internet IPO sample	8.05			458
Non-Internet IPO sample	8.07			458
Fraction of high-tech issues				
Internet IPO sample	67.47%			458
Non-Internet IPO sample	62.45%			458
Fraction of issues traded at Nasdaq				
Internet IPO sample	95.63%			458
Non-Internet IPO sample	87.12%			458
Cumulative return from 1/1/97 to 3/24/00				
Internet IPO sample	2016%			
Non-Internet IPO sample	370%**			

Table 2. Daily Correlations in Calendar Time

The news item data is hand-collected from Dow Jones Interactive and Factiva for both the internet and the non-internet IPO samples. We read and classify each news item as good, bad, or neutral news. Net news NN_t is the average difference between the number of good and bad news items on day t. TN_t is the average number of total news items on day t. $MKCAP_t$ is the average market capitalization at the end of day t. Nasdaq Market Peak is March 24, 2000, the day when Nasdaq 100 index reached its highest level during the sample period. **, * = significant at 1% and 5%, respectively.

Panel A: Prior to Nasdaq Market Peak

		Internet IPO sample					
		$MKCAP_t$	TN_t	NN_t	$MKCAP_{t-1}$	TN_{t-1}	NN_{t-1}
Non-Internet IPO sample	$MKCAP_t$		0.521**	0.124**	0.998**	0.521**	0.127**
	TN_t	0.264**		0.336**	0.518**	0.718**	0.223**
	NN_t	0.013	0.208**		0.103**	0.199**	0.254**
	$MKCAP_{t-1}$	0.998**	0.291**	0.005		0.523**	0.126**
	TN_{t-1}	0.265**	0.388**	0.055	0.267**		0.336**
	NN_{t-1}	0.016	0.144**	0.151**	0.014	0.207**	

Panel B: Post Nasdaq Market Peak

		Internet IPO sample					
		$MKCAP_t$	TN_t	NN_t	$MKCAP_{t-1}$	TN_{t-1}	NN_{t-1}
Non-Internet IPO sample	$MKCAP_t$		0.344**	0.250**	0.982**	0.331**	0.272**
	TN_t	0.250**		0.143*	0.351**	0.539**	0.207**
	NN_t	0.068	0.304**		0.198**	0.187**	0.466**
	$MKCAP_{t-1}$	0.974**	0.282**	0.048		0.339**	0.249**
	TN_{t-1}	0.204**	0.565**	0.311**	0.227**		0.141*
	NN_{t-1}	0.073	0.247**	0.451**	0.077	0.311**	

Table 3. The Effect of Stock Returns on Media Coverage

Panel A of the table reports the difference in media coverage between the internet and the non-internet sample firms, following a price increase or decrease from previous date. $NN_{i,t}$ is the difference between the number of good news items and bad news items for firm i in period t (day or month). $RET_{i,t\text{-}1}$ is firm i's stock return on day t-1 for daily analysis, and month t-1 for monthly analysis. To compute the conditional net news, we average only over observations that have positive returns, negative returns, returns greater than 1% on previous day, or less than -1% on previous day. Nasdaq Market Peak is March 24, 2000, the day when Nasdaq 100 index reached its highest level during the sample period. Firm Market Cap. Peak is the firm-specific day when a firm achieved the highest market capitalization during the sample period. **, * = significant at 1% and 5%, respectively, based on p-values obtained from testing the equality of the conditional mean number of net news between the two samples and from Satterthwaite standard errors. Panel B reports the results from the OLS regressions of abnormal return and firm characteristics on news reports. The dependent variable is the net news variable $NN_{i,t+1}$. $D_{internet}$ is a dummy variable equal to 1 if a firm is an internet IPO and 0 otherwise. $ABRET_{i,t}$ is the day t abnormal close-to-close return. We estimate abnormal returns using the residual from the Fama-French three-factor model, which we fit for each firm separately. Control variables include $TN_{i,t}$ (the number of total news for firm i on day t), $Bid\text{-}Ask_{i,t}$ (the bid-ask spread), $Vol_{i,t}$ (the number of shares traded for firm i on day t), $MVE_{i,t}$ (the market value of equity for firm i on day t), and $VAR\_RET_{i,t}$ (the variance of firm i's stock returns from day t – 19 to day t). We run two pooled regressions, one prior to Nasdaq's peak and one after Nasdaq's peak. For brevity, we do not report the coefficients for the control variables. **, * = significant at 1% and 5%, respectively, based on p-values corrected for clustering by date.

Panel A: The Double Standard of Media Coverage

	Mean $NN_{i,t}$			
	Internet IPOs	**Non-Internet IPOs**	**Internet IPOs**	**Non-Internet IPOs**
	Prior to Nasdaq Market Peak		**Post Nasdaq Market Peak**	
$RET_{i,t-1}$>0	0.186	0.089 **	-0.071	-0.001 **
$RET_{i,t-1}$<0	0.026	0.050 **	-0.146	-0.022 **
$RET_{i,t-1}$> 1%	0.200	0.094 **	-0.072	0.003 **
$RET_{i,t-1}$< -1%	0.026	0.047 *	-0.151	-0.022 **

Panel B: Determinants of Media Coverage

	Dependent Variable = $NN_{i,t+1}$	
	Prior to Nasdaq Market Peak	**Post Nasdaq Market Peak**
Variables of Interest		
Intercept	0.049**	-0.029**
$D_{internet}$	0.021**	-0.034**
$ABRET_{i,t}$	0.001*	0.000
$ABRET_{i,t} \times D_{internet}$	0.003**	0.001
Control Variables	Yes	Yes
No. of observations	78,231	69,846

Table 4. The Marginal Impact of the Media on Abnormal Close-to-Close Returns of Individual Stocks

The table below reports a Jorda (2005) estimation of the impulse response function for individual IPO stocks. For each s where s = 0, 1, …, 20, we run a pooled regression of $ABRET_{i,t+s}$ (the abnormal return of firm i on day $t+s$) on independent variables measured at time t. We estimate daily close-to-close abnormal return, $ABRET_{i,t+s}$, using the residual from the Fama-French three-factor model, which we fit to each firm separately. The main variable of interest, Net news ($NN_{i,t}$) is the difference between the numbers of good and bad news items for firm i on day t. $CUMNN_{i,t}$ is the cumulative net news for firm i from day t - 5 to day t - 1. $TN_{i,t}$ is the number of total news for firm i on day t. $Bid\text{-}Ask_{i,t}$ is the bid-ask spread and $Vol_{i,t}$ is the number of shares traded for firm i on day t. $MVE_{i,t}$ is the market value of equity for firm i on day t. $VAR\_RET_{i,t}$ is the variance of firm i's stock returns in the past twenty days (from day t – 19 to day t). Nasdaq Market Peak is March 24, 2000, the day when Nasdaq 100 index reached its highest level during the sample period. In Panel A, we provide estimates for all independent variables for s = 0 and 1. In Panel B, we report only the coefficients associated with the net news variable for s =2,3,4,5. **, * = significant at 1% and 5%, respectively, based on p-values corrected for clustering by date.

Panel A: Full regression results												
	Prior to Nasdaq Market Peak						**Post Nasdaq Market Peak**					
Dependent Variable	$ABRET_{i,t}$			$ABRET_{i,t+1}$			$ABRET_{i,t}$			$ABRET_{i,t+1}$		
	Internet IPOs	**Non-Internet IPOs**	**Difference**	**Internet IPOs**	**Non-Internet IPOs**	**Difference**	**Internet IPOs**	**Non-Internet IPOs**	**Difference**	**Internet IPOs**	**Non-Internet IPOs**	**Difference**
	(1)	(2)	(1) - (2)	(3)	(4)	(3)-(4)	(5)	(6)	(5)-(6)	(7)	(8)	(7)-(8)
Intercept	-0.152**	0.266**		0.127	0.097		-0.274**	-0.057		-0.282**	-0.054	
$NN_{i,t}$	1.102**	1.365**	-0.263**	0.148**	0.360**	-0.212**	0.898**	1.481**	-0.583**	0.354**	0.466**	-0.112
$CUMNN_{i,t}$	-0.078**	-0.087**		-0.014*	-0.035**		-0.026*	-0.075**		-0.011	-0.028	
$log(1+TN_{i,t})$	0.222**	0.197*		-0.289**	-0.209*		0.251**	-0.070		-0.168	-0.349**	
$Bid\text{-}Ask_{i,t}$	0.194**	0.114**		0.003	0.016		0.227**	0.120**		0.054	0.042*	
$log(1+Vol_{i,t})$	0.677**	0.119**		0.265**	0.181**		-0.056	-0.049		0.289**	0.217**	
$log(MVE_{i,t})$	-0.264**	0.063*		-0.263**	-0.213**		0.509**	0.338**		-0.155**	-0.163**	
$VAR\_RET_{i,t}$	0.005**	0.006**		-0.001	0.000		0.008**	0.008**		0.000	0.000	
$ABRET_{i,t-1}$	-0.009	-0.028**		-0.050**	-0.033**		-0.082**	-0.051**		-0.052**	-0.035**	
$ABRET_{i,t}$				0.010	-0.027**					-0.084**	-0.051**	
Observations	79,335	69,870		79,313	69,847		73,810	67,789		73,387	67,358	
Panel B: Net News coefficients												
$ABRET_{i,t+s}$	**Internet IPOs**	**Non-Internet IPOs**	**Difference**				**Internet IPOs**	**Non-Internet IPOs**	**Difference**			
s = 2	0.040	-0.017	0.057				0.054	-0.043	0.097			
s = 3	0.010	0.020	-0.010				0.083**	0.055	0.028			
s = 4	0.025	-0.002	0.027				0.015	0.037	-0.022			
s = 5	-0.022	0.011	-0.033				0.053	0.027	0.026			

Table 5. The Marginal Impact of the Media on Abnormal Open-to-Close Returns of Individual Stocks

The table below reports the coefficients in a Jorda (2005) estimation of the impulse response function for individual IPO stocks. For each s where s = 0, 1, …, 20, we run a pooled regression of $ABRET_{i\ t+s}$ (the abnormal return of firm i on day $t+s$) on independent variables measured at time t. We estimate daily open-to-close abnormal returns using the residual from the Fama-French three-factor model, which we fit for each firm separately. Net news variable $NN_{i,t}$ is the difference between the numbers of good news and the number of bad news on day t. Control variables $ABRET_{i,t-1}$ (for s = 0 and 1) and $ABRET_{i,t}$ (for s = 1) are calculated as open-to-close abnormal returns by fitting the Fama-French three factors. All other control variables are as in Table 4. For brevity, we do not report the coefficients for the control variables and results for s > 1. Nasdaq Market Peak is March 24, 2000, the day when Nasdaq 100 index reached its highest level during the sample period. We fit these impulse response functions using pooled data from two time periods, one prior to Nasdaq's peak and one after Nasdaq's peak. **, * = significant at 1% and 5%, respectively, based on p-values corrected for clustering by date.

Dependent Variable	$ABRET_{i,t}$			$ABRET_{i,t+1}$		
	Internet IPOs	**Non-Internet IPOs**	**Difference**	**Internet IPOs**	**Non-Internet IPOs**	**Difference**
	(1)	(2)	(1) - (2)	(3)	(4)	(3) - (4)
Panel A: Prior to Nasdaq Market Peak						
Variables of interest						
$NN_{i,t}$	0.592**	0.697**	-0.105	0.014	0.057	-0.043
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	79,005	68,870		78,983	68,847	
Panel B: Post Nasdaq Market Peak						
Variables of interest						
$NN_{i,t}$	0.328**	0.619**	-0.291**	0.088	0.062	0.027
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	73,720	67,321		73,297	66,890	

Table 6. The Marginal Impact of the Media on Abnormal Returns of Portfolios

The table below reports the coefficients associated with the net news variable in a Jorda (2005) estimation of the impulse response function for portfolios of IPO stocks. For each s where s = 0, 1, …, 20, we run a pooled regression of $ABRET_{j,t+s}$ (the abnormal return for portfolio j on day $t+s$) on the net news variable and a set of control variables measured at time t. We estimate daily close-to-close abnormal returns, $ABRET_{j,t+s}$, using the residual from the Fama-French three-factor model, which we fit for each portfolio separately. Net news variable $NN_{j,t}$ is the daily difference between the sum of good news items and the sum of bad news items for firms in portfolio j. Control variables are defined as follows: $CUMNN_{j,t}$ is the cumulative net news for portfolio j from day t - 5 to day t – 1. $TN_{j,t}$ is the total number of news items for all firms in portfolio j on day t. $MVE_{j,t}$ is the market value of equity for portfolio j on day t. $VAR\_RET_{j,t}$ is the variance of portfolio j's returns in the past twenty days (from day t – 19 to day t). For brevity, we do not report the coefficients associated with the control variables and results for s > 5. Nasdaq Market Peak is March 24, 2000, the day when Nasdaq 100 index reached its highest level during the sample period. **, * = significant at 1% and 5%, respectively, based on p-values corrected for clustering by date.

	Prior to Nasdaq Market Peak			Post Nasdaq Market Peak		
	Internet IPOs	**Non-Internet IPOs**	**Difference**	**Internet IPOs**	**Non-Internet IPOs**	**Difference**
	(1)	(2)	(1) - (2)	(3)	(4)	(3) - (4)
Panel A: Value Weighted Portfolio						
s = 0	0.535**	1.554**	-1.019**	0.917**	1.733**	-0.816
s = 1	0.034	0.195	-0.161	0.825**	0.669	0.155
s = 2	-0.163	0.088	-0.252	-0.250	-0.147	-0.103
s = 3	0.034	-0.381*	0.416	-0.246	0.964	-1.210
s = 4	0.101	0.006	0.095	0.269	2.008**	-1.739
s = 5	-0.267**	-0.190	-0.077	0.135	1.640*	-1.505
Panel B: Equal Weighted Portfolio						
s = 0	3.157**	2.965**	0.192	3.119**	3.344*	-0.225
s = 1	0.004	0.459	-0.454	4.089**	1.774	2.315
s = 2	0.446	0.644*	-0.197	2.005	0.532	1.473
s = 3	0.104	-0.079	0.183	1.567	3.132*	-1.565
s = 4	0.375	0.574	-0.199	0.841	4.052**	-3.211
s = 5	-0.126	0.136	-0.262	1.337	4.089**	-2.753

APPENDIX 1

Using the Jorda (2005) Approach to Measure the Marginal Impact of the Media on Returns

There are two approaches to estimate the marginal effect of the media on current and future returns. One is the Vector Auto Regression (VAR) approach, which requires specification of an underlying dynamic model. The VAR provides us the best possible estimates of the short-term effects of news on returns. The estimates for longer-run effects, however, come only indirectly from these short-term effects. Misspecifications – for example, moving average error terms – corrupt the estimates for longer lead-lags in a VAR. This problem does not exist in Jorda's (2005) local projection approach, because Jorda estimates the longer-term effects of news directly without regard to the underlying model.

To appreciate the above point, let us examine how news is generated. News reports typically have different publication dates for the same underlying event. Given this nature of media coverage, it is unavoidable to observe positive serial correlation in news. However, this serial correlation will *not* affect our inferences obtained from the Jords (2005) approach.. To illustrate, suppose that "piggybacking" exists only for one period,

$$NEWS_{i,t+1} = g_0 + g_1 NEWS_{i,t} + g_2 RET_{i,t} + \upsilon_{i,t+1}$$

and news affects today's return and tomorrow's return (as we showed),

$$RET_{i,t+1} = h_0 + h_1 NEWS_{i,t} + h_2 RET_{i,t} + h_3 NEWS_{i,t+1} + \kappa_{i,t+1}$$

This gives us:

$$\begin{aligned} RET_{i,t+1} &= h_0 + h_3 \times g_0 + \left(h_1 + h_3 \times g_1\right) NEWS_{i,t} + \left(h_2 + h_3 \times g_2\right) RET_{i,t} + h_3 \upsilon_{i,t+1} + \kappa_{i,t+1} \\ &= \eta_0 + \eta_1 NEWS_{i,t} + \eta_2 RET_{i,t} + \sigma_{i,t+1} \end{aligned}$$

Our methodology – regressing today's news on today's returns, tomorrow's returns, day after's returns (up to twenty days) – estimates η_1, which contains the direct effect of news, h_1, and the indirect effect of news, $h_3 \times g_1$. So our methodology gives an estimate of η_1 that aggregates both the indirect effect and the direct effect. As we run the regressions for twenty days into the future, we can tackle piggybacking for up to twenty days.

APPENDIX 2

Robustness of Our Results to News Classification Using Human Judgment

We made the critical choice in our research design to use human beings instead of content analysis software to classify news items. Human beings disagree. That could potentially corrupt our study. Therefore, we conduct an experiment to find out how much disagreement really existed between human beings when classifying news items. The details of this experiment are given in Appendix 2.1. The question we pose in our experiment is: when classifying the same news items, how much agreement exists between seven human beings in general and the two co-authors who coded the news items in particular. The answer from our study is: the two co-authors agree 65% of the time (71% of the time) for internet IPOs (non-internet IPOs). Agreement increases to more than 90% if the neutral news classification is removed. There is less agreement with the other human subjects, but it is still, on average, above 50%.[1]

Though it is comforting that human classifications agree on average – the correlations are positive – this could still lead to an error-in-variables problem, which would bias the magnitude of our coefficients downward. A particular concern is that since there was more disagreement in internet IPOs than non-internet IPOs between the two co-authors, the magnitude of the coefficients are more biased downward for internet IPOs, and that is what explains our results.

To deal with the above concern, we conducted a simulation exercise to ask how disagreement can affect our conclusions. The details of this simulation exercise are given in Appendix 2.2. The question we pose in our simulation is: how much disagreement in classification is needed to explain our results. The answer from our simulation is: a lot of

[1] Our consistency results are in line with Niederhoffer (1971), who documents an 80% agreement between the average of two news coders and ten college students when classifying thirty randomly picked New York Times headline news items per person. In a parallel comparison, the maximum precision level of electronic filters in the artificial intelligence literature when deployed for mainstream news has been achieved at 50% (see, for example, Shepherd, Watters and Marath (2002) for a detailed discussion).

disagreement is needed, but our actual disagreement is much less than that. So disagreement cannot explain our results.

2.1. Human Subject Experiment

To ensure that our categorization is consistent, we conducted a small experiment (Human Subject Study # 04-9087, approved on April 22, 2004). To utilize news from all stages of the internet boom and the subsequent bust, we selected two firms among the earliest IPOs in the combined sample: Yahoo! from internet stocks and Sapient Inc. from non-internet stocks. We then recruited seven undergraduate students to participate in the experiment and divided them into two groups.[2] The three students in the first group were each given 100 random news items about Yahoo! and the four students in the second group were each given 100 random news items about Sapient Inc. The undergraduates were instructed to use their own judgment to categorize each news article into good, bad, or neutral, except in cases of news about insider trading (sells were automatically bad, buys were automatically good) or news about analyst recommendations.

The experiment occurred on April 23, 2004 and lasted about two hours. Each student received a payment of $50 for his or her participation in this experiment. The resulting number of instances of agreement between the authors and each undergraduate is presented in Table A2.1.

From Panel A of Table A2.1, the non-internet firm's news results in relatively few disagreements. The authors agree in 71% of cases. Though unreported, that number jumps to over 97% if neutral classifications are ignored. Undergraduates 1, 2, and 4, agree to a similar degree, suggesting that anyone reading news about non-internet firms comes to roughly the same conclusion. Only undergraduate number 3 appears to classify news differently.

Panel B of Table A2.1 demonstrates that news on internet firms is harder to interpret. The authors agree in 65% of cases, though that number jumps to 90% when neither chooses a neutral classification. This difference arises mainly from the fact that Author 1 is less conservative in

[2] We originally recruited eight undergraduates, but one did not show up at the time of the experiment.

assigning classifications. Interestingly, since Author 2 was responsible for much of the internet data collection, this fact suggests that the effect of news on internet returns is actually slightly upward biased. Even with this potential bias, we are able to determine that the returns of internet firms have a significantly lower marginal response to news items than the returns of control firms.

Table A2.1. Agreement in news classification

Seven undergraduates each read and classified one hundred news items into good, bad, or neutral news. Four undergraduates (U1 – U4) each read one hundred pieces of news from Sapient, Inc., a control firm; three undergraduates (U5 – U7) each read one hundred pieces of news from Yahoo!, an internet firm. This table shows the pairwise incidence of agreement between individuals; that is, the percent of one hundred news items to which two individuals both assign a value of good, bad, or neutral.

	Panel A: Sapient					**Panel B: Yahoo!**			
	Author 2	U1	U2	U3	U4	Author 2	U5	U6	U7
Author 1	71%	84%	71%	53%	70%	65%	44%	56%	71%
Author 2		71%	60%	48%	77%		36%	38%	52%
U1			67%	55%	72%				
U2				48%	70%				
U3					52%				
U4									
U5								55%	50%
U6									55%

2.2. Simulation Analysis

We first estimate a Vector Auto Regression of order 3 (VAR(3)) news and abnormal returns for internet firms, and then for non-internet firms. We then simulate 80,000 samples of firm abnormal returns and news for the internet firms (1,000 firms × 800 trading days) assuming that they follow the same VAR(3) process of non-internet firms, but with the same error variances as in the VAR(3) for the internet firms.

If a human being studies this simulated data, and classifies good news and bad news correctly, then net news is classified without error. We ask what would be the coefficients at s = 0, s = 1 and s = 2 of the Jorda (2005) approach, if the regressions are run on the simulated data for this perfect human being. Our answers, given in the last row in Table A2.2, are 1.352, 0.208, and -0.050, respectively. Note that these are coefficients from a data set that has been generated from a VAR(3) process for non-internet firms. As the coefficients from the actual data for the non-internet sample in Table 5 for s = 0, s = 1 and s = 2 are 1.365, 0.360, and -0.017, respectively, we are pleased to note that our method works quite well.

We ask now what happens to the estimated coefficients if a second human being makes errors in news classification. For example, if this second human being studies the simulated data and classifies good news incorrectly as neutral news 10% of the time and bad news incorrectly as neutral news 10% of the time, then net news will be classified with error 18% of the time (1 - (1 - 0.1)×(1 - 0.1) + 0.1×0.1) if good news and bad news arrives simultaneously, and the two misclassifications offset each other. However, net news will be classified with error 19% of the time (1 - (1 - 0.1)×(1 - 0.1)) if good news and bad news do not arrive simultaneously and the two misclassifications do not offset each other. In the simulation, good news and bad news rarely arrive together. So 19% better reflects the error in classifying net news for the second human being. As the first human being is perfect, this implies that the agreement between the two human beings is 81%.

As noted, if the errors in classifying good and bad news are both 10%, the agreement between the two human beings is 81%. This case is given in the third to last row in Table A2.2. This implies, if we use the Jorda (2005) approach with the mis-measured news classifications, the effect of net news on abnormal return at s = 0, s = 1 and s = 2 is 1.342, 0.203, and -0.048, respectively.

We repeat the above experiment assuming different levels of agreement in classification. Table A2.2 gives us the results for each level of agreement. Note that even when the two human beings agree only 64% of the time, which is approximately the agreement on news about internet firm between our two real co-authors in Table A2.1, the coefficients are 1.334, 0.200, and -0.048, respectively. Though biased downward, these estimates still do not approach the estimates in Table 5, which were 1.102, 0.148 and 0.040, respectively. Even when the two human beings agree only 25% of the time, the coefficients are 1.307, 0.188, and -0.045, which are still far from the estimates we found in Table 5.

Table A2.2. Simulation analysis

This table gives the effect of net news (which is the estimate of the coefficient of NN_t) on today's abnormal returns ($ABRET_t$), on tomorrow's abnormal returns ($ABRET_{t+1}$), and on the day after tomorrow's abnormal returns ($ABRET_{t+2}$), for various values of agreement between two human beings. Agreement between two human beings is the percent of one hundred news items to which human being 1 (assumed to be always right) and human being 2 (assumed to be making errors in classification) both assign a value of good, bad, or neutral. The abnormal return of a firm's stock is estimated by fitting a Fama-French (1993) three-factor model for each firm.

Agreement between two human beings	$ABRET_t$		$ABRET_{t+1}$		$ABRET_{t+2}$	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
25.00%	1.307	0.018	0.188	0.020	-0.045	0.019
30.25%	1.312	0.018	0.190	0.018	-0.046	0.019
36.00%	1.316	0.017	0.192	0.018	-0.045	0.018
42.26%	1.320	0.016	0.193	0.018	-0.046	0.017
49.00%	1.324	0.016	0.196	0.017	-0.046	0.017
56.25%	1.329	0.015	0.197	0.017	-0.047	0.017
64.00%	1.334	0.015	0.200	0.016	-0.048	0.016
72.25%	1.338	0.014	0.201	0.016	-0.047	0.016
81.00%	1.342	0.014	0.203	0.016	-0.048	0.015
90.25%	1.347	0.014	0.205	0.015	-0.049	0.015
100.00%	1.352	0.013	0.208	0.014	-0.050	0.014

Additional Reference:

Shepherd, M., C. Watters, and A. T. Marath. "Adaptive user modeling for filtering electronic news." Working Paper, Dalhousie University (2002).

APPENDIX 3

Other Robustness Checks

A. Change of information environment within the sample period

On October 23, 2000, the SEC implemented Regulation Fair Disclosure, requiring that material disclosures by publicly traded companies be disseminated so that the disclosures are simultaneously accessible to all concerned. Prior to the adoption of this law, selective disclosure such as disclosing important nonpublic information to securities analysts or selected institutional investors or both was permissible. By enforcing this law, the SEC intended to eliminate differential informational advantages. We take into account this regime change in our data set by running two regressions. The first regression uses data between March 24, 2000 and October 23, 2000, whereas the second regression uses data between October 23, 2000 and December 31, 2000. For contemporaneous abnormal returns, the difference in coefficients associated with net news between internet sample and non-internet sample is -0.530 (p = 0.001) for the first regression and -0.737 (p = 0.001) for the second regression. For the next day's abnormal returns, the difference is -0.072 (p = 0.440) in the first regression and -0.221 (p = 0.221) in the second.

B. Alternative variable specifications

We substitute the raw returns with abnormal returns to find out whether our method of risk-adjusting made a difference. Our results remain similar. For contemporaneous abnormal returns, the difference in coefficients associated with net news between internet sample and non-internet sample is -0.211 (p = 0.03) during the bubble and -0.562 (p = 0.000) post-bubble. For the next day's abnormal returns, the difference is also -0.211 (p = 0.004) during the bubble but falls to -0.068 (p = 0.438) post-bubble.

We also add news at time $t - 1$ as an additional control variable to control for the momentum in news. Our results stay virtually the same. For contemporaneous abnormal returns, the difference in coefficients associated with net news between internet sample and non-internet sample is -0.233 (p = 0.01) during the bubble and -0.579 (p = 0.000) post-bubble. For the next day's abnormal returns, the difference becomes -0.220 (p = 0.003) during the bubble and -0.115 (p = 0.193) post-bubble.

C. Alternative sample specifications

C.1. Exclusion of price-driven news items

Most of our news is about economic fundamentals, but some of our news is about the previous period's price movement. Through the process of news collection and classification, we have observed that the second type of news occurs usually during the first month after a firm goes public. So we re-estimate the results in Tables 4 by excluding the first-month data of each firm. Again, our results remain qualitatively unchanged. For contemporaneous abnormal returns, the difference in coefficients associated with net news between internet sample and non-internet sample is -0.295 (p = 0.002) during the bubble and -0.618 (p = 0.000) post-bubble. For the next day's abnormal returns, the difference is -0.210 (p = 0.005) during the bubble but falls to -0.096 (p = 0.277) post-bubble.

C.2. Paired matching: Early termination due to mergers, liquidations, bankruptcy and delisting

In this paper, we investigate the impact of the media on IPO firms' returns during the bubble and post-bubble periods by comparing the media coverage between internet IPO firms and a matching sample of non-internet IPO firms. Among the 458 internet firms, 31 were acquired later during the sample period, 19 of which were acquired by firms outside the internet sample. In addition, two internet firms got liquidated, one went into bankruptcy, and two de-listed. It is possible that an internet IPO firm is terminated earlier from our internet sample while the returns of its matched non-internet firm are still included in the non-internet sample.

To check the robustness of our results, we re-estimate the results in Table 4 using a sub-sample where all the firms terminated prior to the end of the sample period are dropped. For contemporaneous abnormal returns, the difference in coefficients associated with net news between internet sample and non-internet sample is -0.254 (p = 0.012) during the bubble and -

0.405 (p = 0.000) post-bubble. For the next day's abnormal returns, the difference is -0.205 (p = 0.013) during the bubble but falls to -0.089 (p = 0.322) post-bubble.

Next, to ensure a perfect paired-matching in our analysis, we use a sub-sample where we throw out firm-days whose matched firm-day contains missing data. We re-estimate the results in Table 4. For contemporaneous abnormal returns, the difference in coefficients associated with net news between internet sample and non-internet sample is -0.207 (p = 0.058) during the bubble and -0.543 (p = 0.000) post-bubble. For the next day's abnormal returns, the difference is -0.125 (p = 0.062) during the bubble but falls to -0.013 (p = 0.892) post-bubble.

D. Boredom: Are investors over-exposed to the news?

The key result in this paper, that the market discounts the media coverage for internet IPO firms, especially during the bubble period, may also be explained by investors' limited attention (Daniel, Hirshleifer and Teoh, 2002, and Hirshleifer and Teoh, 2003) and over-exposure to the news reports at that time. Given the substantially high volume of media coverage on internet firms during the bubble period, investors who have been surrounded by the news about the same type of firms in the past may eventually "grow tired" of any reports about internet stocks, and hence may discount the impact of the news.

To check this claim, we examine the effect of cumulative news exposure by re-estimating our regression model in Table 4. We calculate the cumulative total number of news of a firm till date t, and interact that with the net news variable. This additional control variable allows the impact of news on abnormal returns to decrease as boredom sets in, where boredom is defined as the cumulative total number of news items till then.

Controlling for boredom, we find that the difference in coefficients associated with net news between internet sample and non-internet sample is still negative (-0.265 and -0.226 for today's and tomorrow's abnormal returns) and significant pre-bubble (p = 0.025 and 0.027, respectively). Post-bubble, the difference remains negative (-0.562 and -0.098 for today's and tomorrow’s abnormal returns) with p = 0.000 and 0.371, respectively. This means that boredom is not affecting the main results.

E. Lockup expiration

Most IPOs feature lockup agreements, which prevent insiders from selling their shares to the market over a specified period, typically 180 days. Field and Hanka (2001) show that the popular press has interest in lockup expiration and find significant negative abnormal return around the scheduled unlock day.

To correct for the possible unusual impact of news coverage around lockup expiration, we re-estimate the regression model in Table 4 by removing the returns and news data of each firm in our sample in its sixth month after the offer date. The difference in coefficients associated with net news between internet sample and non-internet sample is still negative, -0.181 (p = 0.121) and -0.138 (p = 0.054) for today’s and tomorrow’s abnormal returns. Post-bubble, the difference remains negative (-0.572 and -0.170 for today’s and tomorrow’s abnormal returns) with p = 0.000 and 0.110, respectively. Thus, lockup expiration cannot explain the main results.

F. Learning

It can be argued that our main result – the market “discounts” news coverage for internet firms – is driven by the learning process of investors about them. We know that learning curves flatten out over time, and as many internet firms went public later in the sample period than earlier, the media impact on the returns of internet firms may have been lower than the media impact on the returns of non-internet firms.

We re-estimate our regressions using only the data between 1996 and 1998. If the argument about learning holds, then the difference in the return sensitivity to news between internet firms and non-internet firms should be less significant at this early stage of the sample period. We find that the difference in coefficients associated with net news between internet sample and non-internet sample is still negative, -0.301 (p = 0.009) and -0.294 (p = 0.001) for today’s and tomorrow’s abnormal returns.

G. Non-high-tech IPOs as the benchmark

The purpose of this paper was to analyze whether the media treated internet IPOs differently than non-internet IPOs. That is why our control sample was non-internet IPOs. It may be argued that the bubble in the late 1990s happened really in technology stocks, and so the correct benchmark should be non-high-tech IPOs. To check this, we restrict our sample to that sub-sample where the match is a non-internet non-high-tech IPO (172 internet IPO firms versus 172 non-internet-non-high-tech IPO firms). Interestingly, our results weaken a little. We find that the difference in coefficients associated with today's return becomes insignificantly negative during the bubble period, -0.087 (p = 0.393), while the coefficient associate with tomorrow's return remains negative and significant and -0.170 (p = 0.006). Post-bubble, the coefficient associated with today's return is -0.428 (p = 0.000) and tomorrow's return is -0.055 (p = 0.542). This suggests that the largest differences are between internet and high-tech non-internet firms, which further imply that the internet/non-internet classification is more important than the high-tech/low-tech classification.

H. The difference in the impact of media sources

Notice that all our results are benchmarked with respect to the non-internet IPO sample. *Ex ante*, there is no reason to believe that a newspaper such as the Wall Street Journal would systematically treat internet IPO firms and non-internet IPO firms differently from, say, the Indianapolis Star. If they do, there may be a bias. As the Wall Street Journal is more influential than the Indianapolis Star, the bias may explain our results if there is more positive sentiment in the bubble period for internet firms in the Indianapolis Star, and it would go against our results if there is more positive sentiment in the bubble period for internet firms in the Wall Street Journal.

It is not possible for us to formally check this, as we did not record the media source for each of our 171,488 news items. So this is what we did. We first identified the top 10 media sources by circulation (Associated Press News Wire, Chicago Tribune, Daily News (NY), Dow Jones News Service, Houston Chronicle, LA Times, Reuters News, New York Times, Wall Street

Journal, and USA Today). We find, surprisingly, that the top 10 media sources covered internet firms with more statistically significant intensity than non-internet firms (the top 10 media sources accounted for 58.74% of the media coverage for internet firms and only 55.66% of the media coverage for non-internet firms).

Second, we identified the wire services (Associated Press News Wire, Dow Jones News Service, and Reuters News). We find, surprisingly, that the wire services showed no statistically differential preferences between their coverage of internet firms and non-internet firms (the wire services accounted for 53.28% of the media coverage for internet firms and 52.17% of the media coverage for non-internet firms). Third, we ran a small experiment. The biggest internet IPO is Yahoo!, and its non-internet match is Sapient. As we have identified the top 10 media sources by circulation in our previous test, we can classify the rest of the media sources into the non-top 10. We then ask whether the non-top 10 was more optimistic about Yahoo! than Sapient in the bubble period. Our answer, after analyzing a random sample of news, is “no.”

I. Real news versus spin news

One type of “real news” versus “spin news” is related to the timing of the news release. For example, a news report such as “Netscape today announced a new browser function” should cause a price movement today, whereas a news report such as “Analysts yesterday gushed about Netscape” describes information that has already been incorporated into yesterday’s price movement. Giving equal importance in the news count to both types of news effectively introduces an error-in-variables problem in the small subset of our stocks that have both these types of news. In our small experiment on media coverage of Yahoo! and Sapient, we find that the proportion of this particular type of “real news” relative to the total news coverage is statistically and economically similar between the two firms for the top 10 media sources (70% for Yahoo! and 68% for Sapient, and $p = 0.53$).

J. Scaling net news

Till now all our tests are run with the objective of finding out the effect of net news on returns. We find that the marginal effect of net news on returns is lower for internet stocks than for non-internet stocks. Since the number of net news for internet stocks is much higher than non-internet stocks, we also estimated the total effect by multiplying the marginal effect of net news by the number of net news. We find that the total effect of net news is higher for internet stocks than for non-internet stocks, but not by much: net news can only explain 2.9% of the 1646% difference in cumulative returns between internet stocks and non-internet stocks from January 1, 1997 through March 24, 2000.

We now re-run our tests using *Scaled Net News*, defined as $NN_{i,t}/(1+TN_{i,t})$. Since total news per day per firm ($TN_{i,t}$) is often 0, we scale net news ($NN_{i,t}$) by 1 plus total news to prevent division by the number 0. We find that the marginal effect of *Scaled Net News* is higher for internet stocks than for non-internet stocks. It is important to point out here that the interpretations of the marginal effects are different in the two cases. The marginal effect of net news is the effect on returns that an extra unit of net news causes, whereas the marginal effect of *Scaled Net News* is the effect that an extra unit of the ratio of net news to $(1+TN_{i,t})$ causes. For the ratio to increase one extra unit, the numerator – net news – has to increase more than the denominator – $(1+TN_{i,t})$. To arrive at the total effect, which is what we are after, we multiply the marginal effect of *Scaled Net News* by the number of *Scaled Net News*. We find a similar result as before: the total effect of *Scaled Net News* is higher for internet stocks than for non-internet stocks, but not by much. Predictability improves from being able to explain 29.3% of the difference in cumulative returns between internet stocks and non-internet stocks to explaining 35.6%. That is, this time it can explain 6.3% of the 1646% difference in cumulative returns between internet stocks and non-internet stocks from January 1, 1997 to March 24, 2000.

Additional References:

Daniel, K.; D. Hirshleifer; and S. Teoh. “Investor Psychology in Capital Markets: Evidence and Policy Implications.” *Journal of Monetary Economics*, 49 (2002), 139-209.

Field, L., and G. Hanka. “The Expiration of IPO Shares Lockups.” *Journal of Finance,* 56 (2001), 471-500.

Hirshleifer, D., and S. Teoh. “Limited Attention, Information Disclosure, and Financial Reporting.” *Journal of Accounting and Economics,* 36 (2003), 337-386.

Niederhoffer, V. “The Analysis of World Events and Stock Prices.” *Journal of Business,* 44 (1971), 193-219.