Internet Stock Message Boards and Stock Returns

Werner Antweiler and Murray Z. Frank*

November 7, 2002

Abstract

During 1999-2001 more than 35 million messages about public firms were posted on Yahoo! Finance. This paper examines whether stocks with high posting levels also have unusual subsequent returns and/or risk. They do. Stocks with the highest level of posting have unusually high realized volatility and unusually poor subsequent returns. This remains true after accounting for the effects of the market, firm size, value, momentum and liquidity factors. Consideration is given to market manipulation, differences of opinion, and anxiety reduction as possible explanations for the observed patterns. *(JEL: G12, G14)*

Keywords: Internet Message Boards, Stock Returns, Volatility, Difference of Opinion

*Both authors are at the Faculty of Commerce and Business Administration, University of British Columbia, 2053 Main Mall, Vancouver BC V6T 1Z2, Canada. E-mail correspondence can be sent to the corresponding author at murray.frank@commerce.ubc.ca. We would like to thank Sanjiv Das, David Hirshleifer and Lubos Pástor for helpful comments, and Bruce Lehmann for comments that helped stimulate us to undertake this study. We thank the SSHRC for financial support. The second author also thanks the B.I. Ghert Family Foundation for financial support. All errors are ours.

1 Introduction

A recent real world innovation has created an opportunity to study the stock market impact of the opinions of a significant class of individual investors—those who post messages on internet stock message boards. A great many of the stock message boards came into existence during 1998. Over the period 1999-2001 there were more than 35 million messages posted about American firms on Yahoo! Finance. Given the sheer magnitude of message posting activity, it seems plausible that the messages might reflect decisions that have an impact.

This study examines whether the level of stock message posting on Yahoo! Finance helps to account for stock returns. Firms are sorted into portfolios based on the number of messages posted about the firm in the previous month. Portfolios are rebalanced monthly. The returns for each portfolio are compared to each other, and they are also compared to the distribution of a large number of randomly constructed portfolios.

The riskiness of these portfolios is considered in several ways. Both the realized volatility (Andersen, Bollerslev, and Diebold, 2002) and the Sharpe ratios of the portfolios are computed. These notions of risk do not reflect popular conditioning factors. In standard asset pricing theory, stock returns are driven by common risk factors. Consideration is then given to a number of factors that are known to affect stock returns: market, size, value (Fama and French, 1993), momentum (Jegadeesh and Titman, 1993) and liquidity (Pástor and Stambaugh, 2002).

Stock message boards could affect expected stock returns through an aggregate risk channel if they provide a proxy for some form of aggregate uncertainty. This suggests looking at 'stock message board betas'; the sensitivity of a stock to the aggregate level of message posting. In order to construct an empirical proxy for a message board factor, a method patterned on Fama and French (1993) is used.

There are a number of ideas about how the message boards might affect the market. For example, difference of opinion might be important. Suppose that people post messages about the stocks for which there is considerable difference of opinion. Such stocks might tend to be temporarily overvalued due to costly short selling as in the theory of

Miller (1977). In that case subsequent negative returns might be expected for the stocks with unusually high message posting. In addition to differences of opinion, we consider message posting as market manipulation, message posters as noise traders, and message posting to reduce anxiety as candidate theories. The implications of these alternative hypotheses are discussed in section 2.

We find that portfolios with particularly high message posting have abnormally poor returns. The effect of the stock message boards is not proxying for a firm size effect. This is shown by conducting two-way sorts using message posting and market capitalization. The poor returns in the high message posting portfolios are accompanied by high volatility. The top message posting portfolio has abnormally poor returns given the level of volatility as measured by the Sharpe ratios.

Regressions explaining stock returns examine whether the effects of market, size, value, momentum, and liquidity factors can account for the observed role of the message boards. Since most of the stock message boards have only existed since 1998, we examine the relative magnitudes of the intercepts across portfolios, rather than focusing on whether the intercepts differ from zero. A monotonic pattern is found in which the highest intercept values are found for the highest message posting portfolios. This effect is robust to alternative specifications of the set of control factors.

It seemed possible that the message boards might be serving as a factor. When the role of a message board 'risk factor' is studied, it does prove to be statistically significant and economically large. The message board posting level contains information that affects returns, but is not captured by overall movements in the market, size, value, momentum or liquidity factors. The evidence is thus consistent with the idea that the number of messages posted on the stock message boards reflects some aspect of risk that is not reflected in the established factors.

Consistent with much of the asset pricing literature, this paper considers monthly stock returns. It is also possible to study higher frequency data. In contrast to this paper, Wysocki (1999), Das and Chen (2002), Das, Martinez-Jerez, and Tufano (2001), Tumarkin and Whitelaw (2001), and Antweiler and Frank (2002) study stock message boards and daily or intra-daily stock returns. While there are significant methodological differences among these studies, there is no reliable short-term connection between message posting and high frequency stock returns.

In section 2 the candidate hypotheses are discussed. Section 3 discusses the data construction procedures. Some basic empirical regularities are presented. The portfolio construction and returns are presented in section 4. It is shown that high message posting is associated with low average returns. Section 5 discusses risk-based interpretations of the portfolio results. The factor version of message board risk is studied in section 6. Conclusions are provided in 7.

2 Hypotheses

Messages posted on internet stock message boards are public information. If market prices fully reflect all public information, then internet stock messages might have no predictive ability for subsequent stock returns. Support for this hypothesis in the context of message boards has been found using daily data (Das and Chen 2001, Tumarkin and Whitelaw 2001). However, these studies do not consider longer period stock returns.

To go beyond the natural null hypothesis of 'no effect,' it is helpful to consider the reasons that people post messages on internet stock message boards. Since different people appear to be posting for different reasons, there may be more than a single effect at work.

Some people appear to be posting in an attempt to manipulate the market as part of an old-fashioned 'pump and dump' strategy. This kind of behavior has been a source of concern to the SEC, and has resulted in legal action in a number of cases.[1] To judge by the cases that the SEC has made public, this kind of behavior seems to be concentrated in smaller cap stocks.[2] If this is the dominant reason for message posting, then high message posting should be followed by the dump. The prediction is that negative stock returns would follow high message posting. This effect should be predominantly found among

[1]Related information about the SEC's Office of Internet Enforcement can be found on their web site (http://www.sec.gov/divisions/enforce/internetenforce.htm).

[2]Similar cases are also reported from Canada. The *Vancouver Sun* (Sept. 25, 2002) reports that the British Columbia Securities Commission took action against a Burnaby, BC man who posted hundreds of false messages on stock message boards in a pump-and-dump scam that targeted penny stocks, gaining US$41,753 in illicit profits. The offender was fined C$25,000 and will have to pay back his illicit trading profits.

small capitalization stocks.

Some people appear to be posting either questions or tentative answers that ask for further information. In these cases, people seem to be trying to make up their minds about what to believe about the stock in question. This seems likely to reflect periods in which there is significant difference of opinion regarding the stock. These differences of opinion sometimes turn into rather nasty 'flame wars.'

There is theory regarding financial markets with difference of opinion. Miller (1977) and Duffie, Garleanu and Pedersen (2002) analyze what happens if short selling is difficult. According to Miller (1977) this means that not all negative views get expressed in the market equilibrium. As a result, stock prices are upwardly biased while there are significant differences of opinion. Duffie, Garleanu and Pedersen (2002) show that stocks for which this short selling constraint is important can sell for more than the valuation of any investor. This happens because the optimistic investors not only expect returns from capital gains and from dividends, but also they expect to get extra fees from lending their stocks to short-sellers. This added benefit is of greatest significance when differences of opinion are particularly strong.

Antweiler and Frank (2002) have shown that message posting is positively correlated with differences of opinion. Accordingly the stocks for which message posting is currently high, might well be those that are on average temporarily overvalued. Negative subsequent returns would be expected. In contrast to the market manipulation hypothesis, this effect could equally well be found among large capitalization firms as among smaller firms. Stocks over which there are significant differences of opinion might naturally also be particularly volatile.

There is evidence that differences of opinion among stock analysts matters. Diether, Malloy and Scherbina (2002) have studied difference of opinion among stock analysts in the I/B/E/S database. Consistent with Miller (1977) and Duffie, Garleanu and Pedersen (2002), stocks with significant difference of opinion among the analysts had poor subsequent returns. Thus difference of opinion tend to be associated with overvaluation, rather than having higher risk and higher returns.

Some people may simply post messages about riskier stocks precisely because these stocks are riskier. Discussing the stocks might help alleviate their anxiety. In this case we expect positive loadings on established risk factors. But the established empirical models provide incomplete representations of risk as faced by investors (Fama and French, 1996). As a result, we should expect to find particularly high intercepts in the high message posting portfolios when we carry out the regression tests. This happens because the investors know about the risks that they are taking even though the financial econometricians do not.

The main ideas that we are investigating were stimulated by direct reading of a large number of the messages.[3] These hypotheses will be referred to as market manipulation, differences of opinion, and anxiety reduction. These hypotheses are not mutually exclusive. Accordingly, we are simply attempting to determine their relative helpfulness as ways to think about the message boards.

3 Message Board Data

The first step was to download message headers for the more than 35 million messages posted on the Yahoo! Finance message boards between January 1, 1999, and December 31, 2001. Yahoo Finance was inactive during two days during this period, 31 March 1999 and 4 February 2001, probably due to technical problems. Figure 1 shows the activity during this period. 1999 was characterized by a growing interest in the message board, peaking in early 2000 and settling on a monthly volume of about 1 million messages. By the end of 2001, 6,802 message boards were available, of which 6,463 were operative with at least a single message posted. A total of 5,911 boards had at least one week during which posting volume exceeded 10 messages. Posting activity is around 15,000 messages per day on weekends and around 35,000 messages per day on weekdays. Thursdays are the most active posting days. The most heavily discussed company averages almost 25,000

[3] Another hypothesis is that the people posting the messages are 'noise traders' as in Delong, Shleifer, Summers, and Waldmann (1990). In that model noise traders both create risk and get extra return in compensation. Accordingly, one might predict that high message board posting would be associated with both high risk and high return. Empirically this does not appear to be the right way to think about the role of the message boards.

messages per month, but the median company receives only 23 messages per month.

Much of the analysis involves sorting firms into quintiles by message posting volume, along with the creation of a portfolio of firms without message board. When firms are sorted in this manner, the top message posting portfolio has a much higher market capitalization ($10.66 billion) than does the bottom portfolio ($0.43 billion). The missing message board portfolio has an average market cap of $0.79 billion. Thus we know that market capitalization is not independent of message posting levels. In our analysis we control for this by the use of two-way sorted portfolios. We also use firm size as a factor in the regression analysis.

Some message boards are very active while others are much less active. Figure 2 plots the number of messages (on a logarithmic scale) against the rank of the board. The first firm is the company with the greatest number of messages, the second firm has the second largest number of messages, and so on. The resulting curve is convex with a bend in the function that is gradual, but centered at about firm rank 250. The curve becomes virtually linear for higher ranks.

The financial data are taken from the Center for Research in Security Prices at the University of Chicago (CRSP). Over the sample period that we study there were dramatic price changes among internet related firms. Thus we consider three indices for the market: CRSP(VW) is the CRSP value-weighted index, CRSP(EW) is the CRSP equal weighted index, and 'Internet Index' is the Dow Jones Internet index that is traded under the ticker symbol XLK.

These three market indices are plotted in Figure 3 and they follow quite distinct trajectories. The CRSP(VW) has a fairly mild 5% decline over the three years. The CRSP(EW) has a gain of 32% over the period. The fact that the equal-weighted index had a much better return indicates that small firms outperformed large firms during our sample period.

The Internet Index has both a huge run-up and a huge decline during our period. It is not clear whether the decline should be deemed to have started in March of 2000 (the peak), or August 2000. Since August 2000 is very close to the middle of our sample period, we use this as a defining date to distinguish the Internet collapse period. The movement in the Internet Index is remarkable. Over the first year and a quarter there was an 80% increase in market price. From the peak to the end the decline is more than 50%, which left the index at less than 80% of where it was at the start. The Internet index includes many of the most well-known Internet firms, making the decline particularly remarkable. These were not penny stocks to start with.

There are some problems matching CRSP data with data from the Yahoo! Finance message boards. Yahoo message boards are identified by ticker symbols that do not always match the common use of such symbols. In particular, Yahoo compresses the share class symbol and ticker symbol into a single code, which may lead to ambiguities when matching it to ticker symbols used in CRSP. The matching of ticker symbols was done based on the ticker symbols in effect on December 28, 2001. Despite hand coding and searching, we are unable to match 313 stocks. These include primarily stocks of foreign-owned companies such as Air Canada or Ballard Power Systems. There may also be some genuine mismatches due to ticker symbol errors.

A further problem that we encountered is that Yahoo seems to remove message boards when companies cease to exist as can easily happen due to merger, acquisition, or bankruptcy. As several of these stocks have ceased to exist by the end of 2001, we do not have any posting activity information for these. In order to quantify the magnitude of the problem, we created portfolio X which contains the stocks that exist in CRSP but for which we cannot find a message board.

When controlling for known factors, the data on the market, size, value, momentum, and risk-free interest rate are taken from Kenneth French's web page.[4] The method of constructing the liquidity factor follows Pástor and Stambaugh (2002) and is described in the appendix. Our method of constructing a message board factor is described in section (6).

[4]We are grateful to Kenneth French for making these time series available. They can be found on the web at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

4 Portfolios

4.1 Portfolio Construction

Portfolios are used to look for differences in returns.[5] First, we leave out firms with stock prices of less than \$1 at each portfolio formation date. We do not remove firms based on their market capitalization. Instead we examine the role of market capitalization as a factor in its own right.

Second, each month we sort the stocks into quintiles according to the number of messages posted about the firm on the stock message boards during the previous month. Portfolio A contains the stocks with the highest number of messages, and portfolio E has the smallest number of messages for that month. Deciles were also studied and gave very similar results. To save space, the decile results are not reported separately. Not all firms have message boards. Portfolio X contains the stocks for which we were not able to find a stock message board on Yahoo! Finance.

Third, \$1 is invested into each of the portfolios at the start of the month. The money is equally invested in each stock in the portfolio given the currently prevalent stock price. Results for value-weighted portfolios are also reported.

Fourth, at the start of the next month each portfolio is liquidated at current stock prices. A new sort is then done and the money from a given portfolio is reinvested in the portfolio defined by the same criterion at the start of the next month. This procedure is repeated each month from the start of 1999 until the end of 2001. Two-way sorts based on message posting and market capitalization are also considered.

While we can rank the results by eye, it is important to know if these final payoffs are unusual. In order to determine if the observed portfolio returns are 'unusual,' we employ a simple Monte Carlo procedure to generate a distribution of returns. We form 10,000 random portfolios in the same manner as portfolios A-E. Instead of ranking the stocks by message board activity, we use random numbers to rank the stocks. We examine the

[5]Portfolios are frequently used to study asset returns. Examples include studies of momentum (Jegadeesh and Titman, 1993), size and value (Fama and French, 1996), the probability of information-based trading (Easley, Hvidkjaer, and O'Hara, 2002), disagreements among stock analysts (Diether, Malloy, and Scherbina, 2002), and liquidity (Pástor and Stambaugh, 2002).

distribution of returns from the randomly formed portfolios. If a message board portfolio falls in the extreme tails of the random distribution, then we regard the return on that portfolio as being unusual.

A natural comparison to consider is the relation between our portfolio returns and the returns on the market. To do this we consider the performance of $1 invested in "the market" using the CRSP(EW) and CRSP(VW) as proxies. Since we know that CRSP(EW) outperformed CRSP(VW) in our period, we know that it is important to control for market capitalization. Smaller firms are often found to be riskier and to have higher returns. If the message boards focus disproportionately on firms of particular sizes, then one might misidentify a firm size effect as being a message board effect. In order to guard against this, firms are also sorted according to their size: small, medium, or large. Consequently, in addition to the basic message board ranked portfolios, results are also provided for portfolios that interact message posting volume and firm size.

4.2 Portfolio Returns

Table 1 Panel A, column 'all,' reports stock returns for the portfolios constructed by quintile sorts on message posting. Both high message posting (A), and missing message boards (X) have abnormally poor returns. Quiet portfolios such as D and E have abnormally good returns. Panel B constructs the same portfolios but weights them by value. Generally the value-weighted portfolios have similar results to the equally weighted, but the results tend to be noisier. As a result, the value-weighted portfolios are less often statistically significant.

It is important to consider the role of market capitalization. To do this we carry out two-way sorts. We form quintiles on message posting (A to E) plus the missing boards (X) and further divide each quintile by market capitalization (high, medium, small). Thus we construct 18 portfolios. As we expect from Figure 3, the low market cap portfolios perform best. Within each market cap category, high message posting exhibits poor returns. Portfolio X also has poor returns, so a missing-message board bias is not likely to account for the poor returns in portfolio A. On the other hand, depending on how the firms in portfolio X would otherwise have been distributed, it could have the effect of reducing the abnormally high returns in some other portfolios. The two-way sorts show that the poor returns for high message board posting volume is not being caused by an omitted firm size effect.

Beyond just looking at the portfolio returns, it is also of interest to consider the actual portfolio trajectories. Figure 4 plots the trajectories, along with the distributions of the random portfolio trajectories. The distribution is depicted by the shading. Darker shading means more trajectories. Panel A gives the equal-weighted case and panel B gives the value-weighted case. As expected, the market portfolios generally lie very close to the center of the distributions.

As expected from Figure 3, the value-weighted portfolios have a lower mean than do the equal-weighted portfolios. Beyond the differences in means, there is also a difference in variation. For the equally weighted distribution the 5% cutoff is at a (log) return of 0.279 and the 95% cutoff is at 0.393. For the value-weighted distribution these cutoffs are at –0.216 and 0.112 respectively. Given the greater variation in the value-weighted case, fewer returns are deemed abnormal. Portfolios D and E (low messages) are always abnormally good. Portfolio A is always the lowest return portfolio.

In Figure 4 the trajectory of portfolio A is noteworthy. From the start of 1999 until about the middle of 2000 portfolio A was always at, or near, the top of the portfolios. From the middle of 2000 until the end of 2001 portfolio A fell precipitously. By eye it seems that portfolio A is much like an exaggerated version of the market portfolio (M). This visual evidence suggests that portfolio A might be very risky in a high-beta CAPM sense. If so, then a risk-based interpretation might be promising.

4.3 Transaction Costs

Transactions costs are potentially important in any analysis of portfolio returns. However, transaction cost analysis is difficult because different traders pay different transactions costs. Some accounts have a fixed cost independent of the number of shares traded. Often there is a cost per share component. In the presence of fixed costs larger-sized trades fare better. But if trades become large, they can have unfavorable temporary market impact. We examined a much simpler version in which we assumed that the portfolio was worth \$1 million to start with and then we charged the portfolio \$0.05 for each net share traded during a portfolio adjustment.[6]

The most important effect of transactions costs is to reduce all returns. Thus the negative mean return observed in Table 1 for portfolio A becomes even more negative. However, not all portfolios generate the same number of trades. In particular, the low market cap portfolios require many more trades and thus face a much higher reduction from the added transaction cost. Portfolio A always faces less transactions costs than do portfolios B to E. High market cap portfolios always face less transactions costs than do medium or low market cap portfolios. With the numbers that we used, for equal-weighted portfolio A with high market cap, the annual decrease due to transactions costs is just 0.5%. The highest portfolio reduction was for equal-weighted portfolio C with low market capitalization. For this portfolio the annualized transactions cost is 11.5%. This is likely a reasonable upper bound calculation of the transactions cost impact. Naturally, value-weighted portfolios have considerably smaller transaction costs because they do not require rebalancing due to changes in the portfolio weights.

5 Does risk explain the poor returns?

5.1 Volatility

Two quite different approaches can be taken to risk. The simpler approach is to think about risk as volatility—the standard deviation of returns. The higher the standard deviation, the greater the risk. The idea that individual stock volatility matters cannot be lightly dismissed. Goyal and Santa-Clara (forthcoming) provide empirical evidence that

[6]It should be noted that the fee schedule that we have assumed had no fixed cost component, although most fee schedules indeed have a fixed cost. As of October 2, 2002 E-trade charged \$9.99 flat per trade for active traders. During parts of our sample period some of the aggressive discounters charged half that amount. Thus it was possible to reduce the transactions costs quite substantially through the choice of broker. With a well-chosen broker and large-valued trades, the transactions costs impact can be minimized considerably. Some firms even offered free trades as an inducement to setup an account. Obviously, if trades are free, then there is no transactions cost drag on the portfolio.

the average variance of the individual stocks is important for market returns.

There are conceptual problems with using volatility as the key risk measure. Typically, asset pricing theory does not measure risk in this way. More commonly risk is measured in terms of the sensitivity of returns to various risk factors. The classic risk factor is simply 'the market' as in the CAPM. However, multiple risk factors have become more popular as a result of evidence such as Fama and French (1993, 1996) that more than one factor seems to matter.

Both realized volatility and exposure to a number of standard risk factors are considered. In this subsection we examine the realized volatility (see Andersen, Bollerslev, and Diebold 2002) of portfolios constructed in the same manner as the Table 1 portfolios. Then we consider the associated Sharp ratios. In the next subsection we consider the use of standard risk factors.

In order to construct the realized volatility for a particular month we use daily data on a given stock within the month. Realized volatility is defined as the sum of absolute values of the daily log returns during a particular month. Table 2 reports the realized volatility for the same sets of portfolios considered in Table 1. The highest message posting portfolio (A) always has the highest realized volatility. This is true whether we construct equal-weighted portfolios, or value-weighted portfolios. This is also true within each market capitalization category. As is common, realized volatilities are more precisely estimated than are expected returns, and so more of the realized volatilities are considered to be abnormal.

The most volatile stocks are stocks with low market capitalization and high posting volume. Both with equal-weighted portfolios and with value-weighted portfolios, the results are fairly monotonic in both directions: higher posting has higher volatility, lower market capitalization has higher volatility. Stocks that have no message boards (portfolio X) are not all that extreme.

Tables 1 and 2 consider expected return and volatility separately. Table 3 considers these together by constructing Sharpe ratios for the same set of portfolios. Consistent with the earlier tables, the high message posting portfolio (A) has very bad Sharpe ratios. The highest Sharpe ratios are found in the low message posting and low market capitalization portfolios. It seems that our time period was not a good one for high profile firms, but it was a fairly good period for many small low profile firms. Table 3 reinforces the evidence from Tables 1 and 2. High message posting stocks are volatile and have low returns. Neither the missing message boards nor market capitalization are responsible for this bad performance.

5.2 Risk factors

The idea that low returns and high risk go together is troubling. One possibility is that realized volatility is not a good measure of risk as seen by investors. In order to get at this idea we turn next to the use of standard risk factors. Expected stock returns can be predicted in several ways. A particularly popular empirical approach is to use a set of standard 'factors.' The loading on each factor is a measure of how sensitive a particular asset is to the given risk factor. The interpretation of these 'factors' as 'risk factors' has itself been somewhat controversial, see Fama and French (1996).

Particularly popular factors to consider are: market, firm size, book to market ratio (Fama and French, 1993), and momentum (Jegadeesh and Titman, 1993). They are referred to as Beta, SMB (small minus big), HML (high minus low), and UMD (up minus down), respectively. Liquidity is often invoked in discussions of the stock market, and Pástor and Stambaugh (2002) have shown that a liquidity factor (LIQ) does appear to be important. We use these factors in regressions that are run on each portfolio.

If a given portfolio has an abnormal return relative to the set of factors, then the intercept will differ significantly from zero ('Jensen's alpha'). Given the relatively short time frame we do not place much weight on whether the alphas differ from zero. Instead we are interested in the *relative* sizes of these intercepts across the portfolios. If message posting is itself acting like a factor, then we might expect to see some monotonic pattern in the Jensen's alphas. These regressions also show the sensitivity of each portfolio to each of the factors.

We estimate the factor models using a Generalized Method of Moments (GMM) estimator with the Newey and West (1987) kernel. We also tried using OLS and panel regressions with firm-specific fixed effect and obtain very similar results in each case.

Table 4 provides the results of these regressions. Consistent with the visual impression in Figure 4, the high message posting portfolio A has a β that is greater than one and a high and significant effect from firm size (γ_1). Portfolio A also has significant negative parameter on value (γ_2), momentum (γ_3) and liquidity (γ_4).

All of the portfolios are significantly affected by the market (β) and by firm size (γ_1). The negative parameter on value (γ_2) is found in portfolios A and X. The other portfolios exhibit significant positive parameter estimates for the value factor. All portfolios have significant negative parameters on momentum and for liquidity factors. Thus portfolio A does have a somewhat different risk profile than do the other portfolios as reflected by these conventional factors.

The estimates of α in Tables 4 and 5 are surprisingly high on average. This reflects the divergence of equal-weighted and value-weighted returns during the 3-year period of our study, as shown in Figure 3. To understand the effect of this divergence, think of a simple CAPM model, $r_{it} - r_{ft} = \alpha + \beta(r_{mt} - r_{ft})$, where r_{it}, r_{mt} and r_{ft} denote, respectively, the return of company i in period t, the market return, and the return on the riskfree asset. Commonly a value-weighted index (typically the CRSP value-weighted index) is used as a proxy for r_{mt}. Sum over all companies i, and for simplicity assume that for the entire market $\beta = 1$. Let $\bar{r}_t = (1/N) \sum_i r_{it}$. Then it must be the case that $\bar{r}_t - r_{mt} = \alpha$. In other words, the intercept α is the difference between the equal-weighted and value-weighted average market returns.

Over long periods of time, the difference between the equal-weighted and value-weighted CRSP indices is statistically insignificant.[7] Accordingly, when the model is estimated the intercept tends to be close to zero no matter which index is used. However, during our sample period 1999-2001, there was a 1.1% gap between equal-weighted and value-weighted monthly returns. Compounded over the 3-year period, this amounts to a huge 48% gap that can be seen when comparing the M points in panels A and B of

[7] Comparing monthly log returns of the CRSP(EW) and CRSP(VW) indices over the 35-year period 1967 through 2001, there is a large but statistically insignificant difference between equal-weighted and value-weighted returns of 2.3% (annualized). Over subsample periods, sometimes one index has higher returns, while during other periods the other index has higher returns.

Figure 4. Since we follow common practice in using the value-weighted index, our estimates of α will tend to be positive.[8] As a result of this effect on the intercept, we are not interested in the issue of whether the intercepts differ from zero. Rather, the interesting question is the pattern of the intercepts across message posting portfolios as shown in Tables 4 and 5.

Message posting does identify abnormal returns relative to the 4-factor model. The intercept is significantly positive and quite large in portfolio A. More importantly the intercepts decline monotonically as one moves to portfolios with fewer messages. This monotonicity suggests that the message board posting volume might itself be serving as a factor. The possible role of the message boards serving as a factor is studied in section 6.

Since there is no consensus empirical model of asset returns, it is important to ask whether the patterns in the intercepts are an artifact of a particular specification. In order to address this concern, Table 5 provides alphas for the same set of portfolios relative to 4 different empirical models: a CAPM, the Fama-French three factor model (market, size, value), a four factor model that adds momentum, and a five factor model that also include liquidity. There is a remarkable degree of stability for portfolios A and B. Much less stability is found in portfolios with few messages E, and for in portfolio X.

6 Message Posting Levels as a Factor

The fact that the intercepts increase monotonically as the message postings increase suggests that message posting might be playing the role of a factor. To investigate whether that is true we construct two versions of a stock message board risk factor: NMQ and YYY. Our preferred measure NMQ ("noise-minus-quiet") is patterned on Fama and French's (1993) method of constructing a factor as the difference in portfolio returns. Concretely, we define NMQ_t as the difference in period t between the return on portfolio A minus

[8]We have also tried using the CRSP(EW) as a proxy for the market. When we do that the conventional factors such as SMB, HML, and UMD become statistically insignificant. The importance of the internet stock message boards remains highly significant for the top message posting firms. However, along with the other conventional factors, it lacks statistical significance in the broader population of firms when we use CRSP(EW).

the return on portfolio E. YYY is an alternative proxy for message board activity that does not exploit the cross-sectional information; it is the first difference across time periods in the total number of message posted on all of the message boards.

Table 6 shows that the correlation between these two proxies is 0.443. Given the results in Table 1, it is perhaps not surprising that NMQ is highly correlated with β and firm size, and strongly negatively correlated with HML. The alternative measure YYY has similar correlations, but they are a fair bit weaker.

Table 7 presents the regression results. As before these models are estimated using GMM and using the Newey and West (1987) kernel with 1 lag. As before the results are very similar when estimated using OLS or when estimated using a panel data method with firm-specific fixed effects.

In panel A of Table 7 four conventional models are presented: CAPM, Fama-French, 4-factor, Pástor-Stambaugh. In panel B the message board factor NMQ is added to each of these models. In panel C the results of using YYY in place of NMQ in the Pástor-Stambaugh model are presented along with robustness checks based on the Pástor-Stambaugh model.

It is natural to be concerned about whether the time period that we study is in some sense abnormal. Perhaps if we considered a longer time period the results might be different. This concern is reasonable. However, apart from waiting a decade or two, there is little that we can do about it. Many of the message boards were created during 1998 and so earlier data simply does not exist. Accordingly we consider whether our sample period appears to be unusual relative to the conventional factors as in Fama and French (1996). Panel A of Table 7 addresses this issue.

Most of the parameter estimates are not all that unusual relative to the findings of other studies. Finding a beta that is near 1 is standard, as is finding γ_1 in the neighborhood of 0.5. The lack of stability in HML is interesting. The negative sign on market momentum is not too surprising in light of the estimates reported by Pástor and Stambaugh (2002). They find that market liquidity is a significant factor for stock returns. For our sample period the sign on liquidity is negative and it is significant. Panel A shows that our sample period does not appear to be terribly unusual relative to these factors. The significant positive intercept, as explained above is due to the difference between the equal weighted and value-weighted market indices in our sample period.

In panel B we add NMQ as a factor. It is statistically significant in all specifications and has fairly stable coefficients as we move from one specification to another. The T-statistics are generally of about the same magnitude as the T-statistics on the other factors, with the exception of liquidity. Liquidity has lower T-statistics.

Two features of Panels A and B of Table 7 are notable. First, when momentum (UMD) is included in the analysis it has a dramatic impact on value (HML). Second, when NMQ is added to the analysis it has a significant effect on the other parameters. The biggest effects are on the importance of the market factor (β drops from 1.127 to 0.256), and on value (γ_2 rises from 0.049 to 0.783). While these other factors are sensitive to the presence or absence of NMQ, NMQ is fairly robust across alternative specifications given in lines 5 through 8 of Table 7.

In line 9 of panel C we replace NMQ with YYY—our alternative proxy for aggregate message posting activity. YYY performs more poorly than NMQ, although it is significant and has the same sign. As a result of the poor performance of YYY, HML is again seriously affected.

Figure 4 shows that prior to mid 2000 the trend was up, while afterwards the trend reversed itself. It is possible that different patterns would be found in up markets and in down markets. Table 7 Panel C investigates this hypothesis. When we split the sample into the normal/rising market (January 1999–July 2000) and the falling market (August 2000–December 2001) most of the coefficients are roughly similar. The notable exception is the coefficient on liquidity. Liquidity is not significant in the first time period, but it is powerfully significant in the falling market. The magnitude of β is also affected. For the current study our main concern is the robustness of the coefficient on NMQ. While the magnitude of γ_5 does increase in the falling market, the size of the coefficient does not change by an order of magnitude. Table 7 shows that the message board factor (NMQ) is routinely significant. The magnitude is reasonably stable across specifications.

There are a number of further robustness checks that can be considered. We tried using OLS and using panel regressions. We tried using deciles instead of quintiles. The results were not much different. Following Lo and MacKinlay (1990) and Lehmann (1990) we also tried using a weighted portfolio scheme using the message posting levels as the weights. The patterns in the results were very similar to those reported above. However the results were noisier and at times less significant. To save space we do not report these separately.

7 Conclusions

There are a variety of reasons why people post messages on Internet stock message boards. The main reason for posting messages could be an attempt to manipulate the market through a 'pump and dump' strategy. But it is hard to imagine manipulation of the market price of a Dow Jones Industrial Average stock with a message posted on Yahoo! Finance. Thus the poor subsequent returns ought to be concentrated in low market capitalization stock. But the poor returns are not concentrated in the small cap stocks. Thus this does not seem to be the main force.

It might be the case that high message posting reflects differences of opinion as documented by Antweiler and Frank (2002). If that is true, and short selling is difficult as in Miller (1977), then high volatility and poor returns should be associated with high message posting. This is observed empirically. In this approach it is unclear whether significant loadings on established risk factors should be observed. This is because the difference of opinion is being taken as given. Our findings are complementary to Diether, Malloy, and Scherbina (2002) who found similar results in a study of differences of opinion among stock analysts. Evidently, stock analysts are a very different sample of people than are those who post messages on internet stock message boards. The fact that the results are so similar enhances our confidence that differences of opinion are important.

It might be the case that people post messages about risks that they know themselves to be taking. They might do this in order to reassure themselves that they have made reasonable, even if risky, decisions. In this case high message posting should reflect established risk factors. Furthermore, established risk factors are not complete descriptions of the different types of risk that people know themselves to be facing. The message boards would then serve as a proxy for the risks that people are taking, but are not captured by existing factors. The evidence is consistent with this idea. Traditional risk factors affect message posting. Beyond the familiar factors, the messages seem to be capturing risk elements that affect the market.

Since the message boards are a new phenomenon, we are necessarily limited to this particular historical period. Only the passage of time will permit an assessment of the extent to which these findings generalize to other time periods. There is evidence that we take as somewhat reassuring in this regard. Perhaps most importantly, the conventional risk factors have fairly conventional market effects during this period. In this sense the sample period does not appear to be anomalous, despite the sharp price movements in some stocks. When we split the sample into the boom period and the bust period, fairly consistent parameter values were obtained. While these facts are somewhat comforting, caution remains warranted as in any study of a new phenomenon in the real world.

Is it possible to base a trading strategy on the messages posted on Yahoo! Finance? Yes it is. A strategy that went short when a stock was very highly discussed, would have done well on average. However, the results show that it would also have been undertaking a great deal of risk. The high message posting stocks are very volatile, and they are also quite sensitive to established risk factors. The level of message posting also provides information about risk that matters to investors, but is not reflected in market, size, value, momentum, or liquidity factors.

References

[1] Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2002), "Parametric and Nonparametric Volatility Measurement," in L.P. Hansen and Y. Ait-Sahalia (eds.), Handbook of Financial Econometrics. Amsterdam: North-Holland, forthcoming.

[2] Antweiler, W. and M. Z. Frank, 2002, "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." Working paper, University of British Columbia.

[3] Campbell, J.Y., A.W. Lo, and A.C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton N. J.

[4] Das, S.R. and M. Chen, 2001, "Yahoo! for Amazon: Sentiment Parsing from Small Talk on the Web," Working Paper, Santa Clara University.

[5] Das, S.R. A. Martinez-Jerez, and P. Tufano, 2002, "e-Information," Working paper, Harvard University.

[6] DeLong, J., A. Shleifer, L. Summers, and R. Waldmann, 1990, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98, 703-738.

[7] Diether, K.B., C.J. Malloy, and A. Scherbina, 2002, "Difference of Opinion and the Cross-Section of Stock Returns," *Journal of Finance*, 57(5), 2133-2141.

[8] Duffie, D., N. Garleanu, and L.H. Pedersen, 2002, "Securities Lending, Shorting, and Pricing," *Journal of Financial Economics*, 66(2-3), November.

[9] Easley, D., S. Hvidkjaer, and M. O'Hara, 2002, "Is Information Risk a Determinant of Asset Returns?" *Journal of Finance*, 57(5), 2185-2221.

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

[11] Fama, E. and K. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 1, 55-84.

[12] Goyal, A. and P. Santa-Clara, forthcoming, “Idiosyncratic Risk Matters!” *Journal of Finance*.

[13] Jegadeesh, Narasimhan, and Sheridan Titman, 1993, “Returns to Buying Winners and Selling Losers: Implications for Market Efficiency,” *Journal of Finance*, 48, 65-91.

[14] Jegadeesh, N., and S. Titman, 2001, “Profitability of Momentum Strategies: An Evaluation of Alternative Explanations,” *Journal of Finance*, 56, 699-720.

[15] Lehmann, B.N., 1990, “Fads, Martingales, and Market Efficiency,” *Quarterly Journal of Economics*, 105, 1-28.

[16] Lo, A. and A.C. MacKinlay, 1990, “When are Contrarian Profits Due to Stock Market Overreaction?,” *Review of Financial Studies*, 3, 175-208.

[17] Miller, E. M., 1977, “Risk, Uncertainty, and Divergence of Opinion,” *Journal of Finance*, 32, 1151-1168.

[18] Newey, W. and K. West, 1987, “A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703-708.

[19] Pástor, L. and R.F. Stambaugh, 2002, “Liquidity Risk and Expected Stock Returns,” *Journal of Political Economy*, forthcoming.

[20] Tumarkin, R., and R.F. Whitelaw, 2001, “News or Noise? Internet Message Board Activity and Stock Prices,” *Financial Analysts Journal*, 57, 41-51.

[21] Wysocki, P.D., 1999, “Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards,” Working Paper, November, University of Michigan.

Appendix: Liquidity Measurement

In our analysis of market factors we replicate the liquidity measure introduced in Pástor and Stambaugh (2002). Liquidity is linked to order flow, which is captured by trading volume signed by the contemporaneous excess return of a given stock. Concretely, their liquidity measure is derived from the OLS estimates of $\gamma_{i,t}$ in the following estimating equation for the $d = 1, .., D$ daily observations of company i in month t:

$$r^e_{i,d+1,t} = \theta_{i,t} + \phi_{i,t} r_{i,d,t} + \gamma_{i,t}\text{sign}(r^e_{i,d,t}) \cdot v_{i,d,t} + \epsilon_{i,d+1,t} \tag{1}$$

Here, $r_{i,d,t}$ is the return on stock i on day d in month t, and $r^e_{i,d,t} \equiv r_{i,d,t} - r_{m,d,t}$ is the excess return relative to the market return $r_{m,d,t}$, which in turn is defined as the CRSP value-weighted market return. The $v_{i,d,t}$ variable is the dollar trading volume for stock i on day d in month t. Stocks with less than 15 observations during a given month were excluded from this regressions. Using N_t estimates $\hat{\gamma}_{i,t}$ for each month, the market-wide measure is given by $\hat{\gamma}_t = (1/N) \sum_{i=1}^{N_t} \hat{\gamma}_{i,t}$. Pástor and Stambaugh further employ the scaling m_t/m_1, where m_t is the total market capitalization at the end of month $t-1$. Next, they construct a measure of innovations in liquidity

$$\Delta\hat{\gamma}_t \equiv \left(\frac{m_t}{m_1}\right) \frac{1}{N_t} \sum_{i=1}^{N_t} (\hat{\gamma}_{i,t} - \hat{\gamma}_{i,t-1}) \tag{2}$$

and regress it on its lag as well as the lagged value of the scaled level series

$$\Delta\hat{\gamma}_t = a + b\Delta\hat{\gamma}_{t-1} + c\left(\frac{m_{t-1}}{m_1}\right)\hat{\gamma}_{t-1} + u_t \tag{3}$$

in order to produce serially uncorrelated residuals $\hat{u}_t$ that define the liquidity measure $\text{LIQ}_t \equiv \hat{u}_t$. We apply this procedure to obtain LIQ_t for all but the last month in our sample, which we lose due to autoregressive initialization.

Figures

Figure 1: Total Yahoo Posting Volume 1999-2001

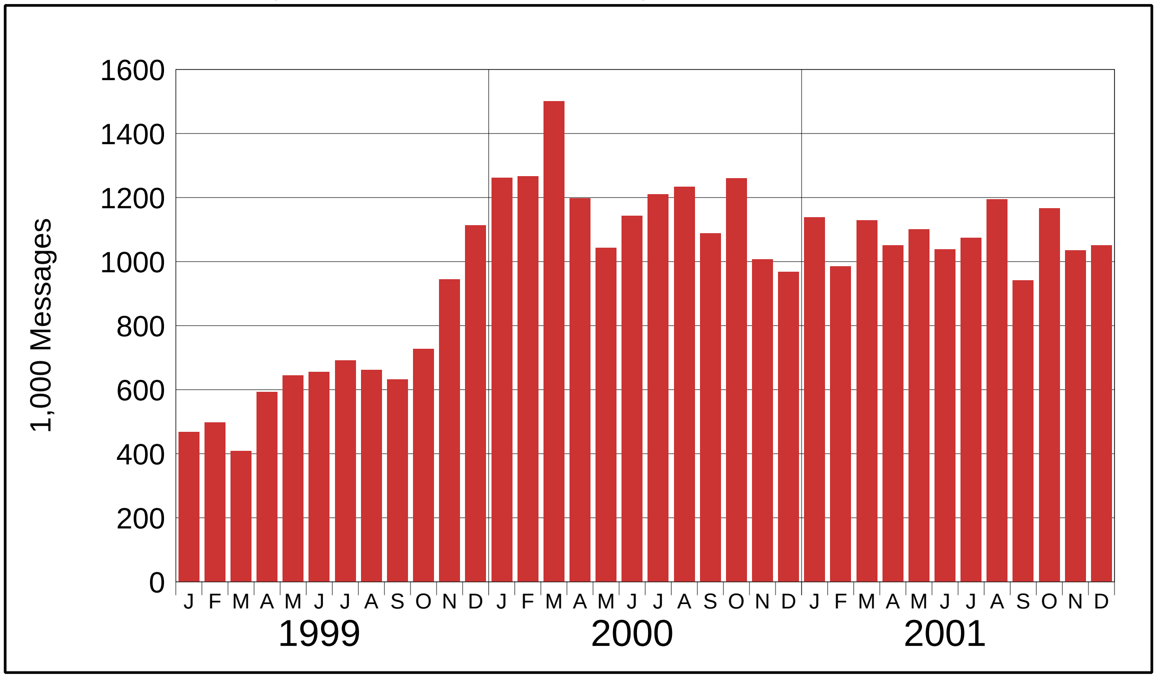


Figure 2: Rank Distribution of Posting Volume

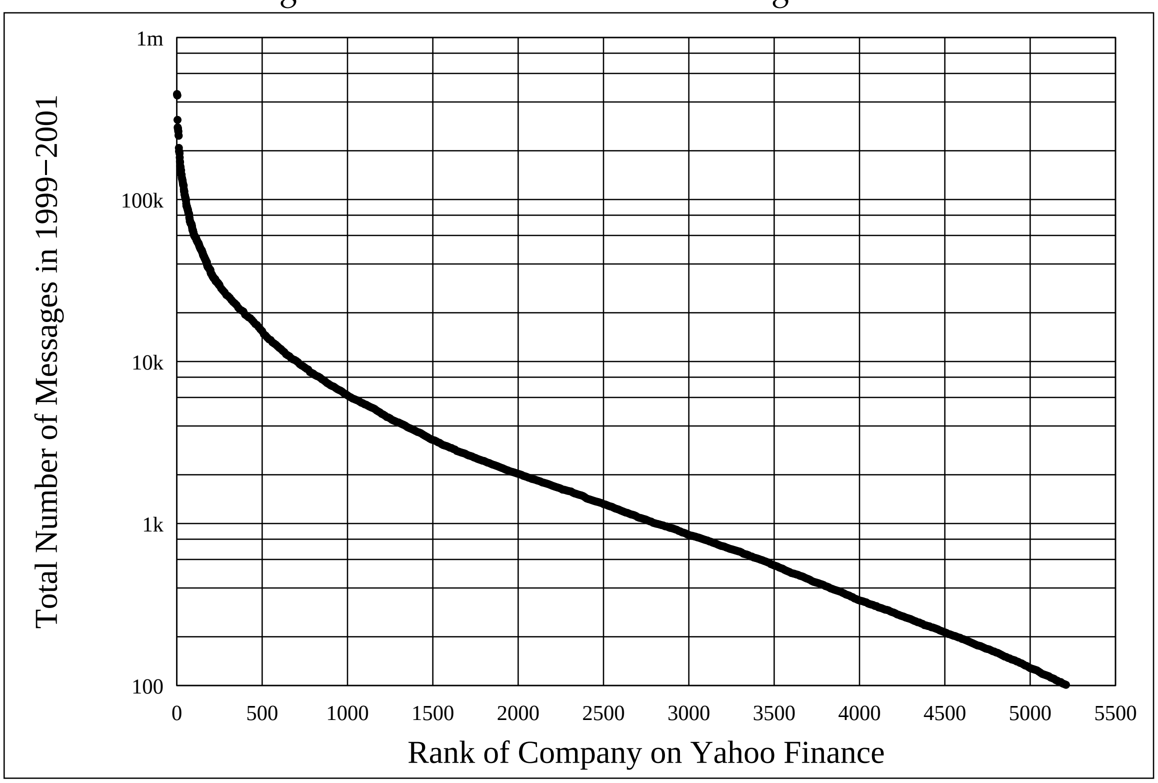
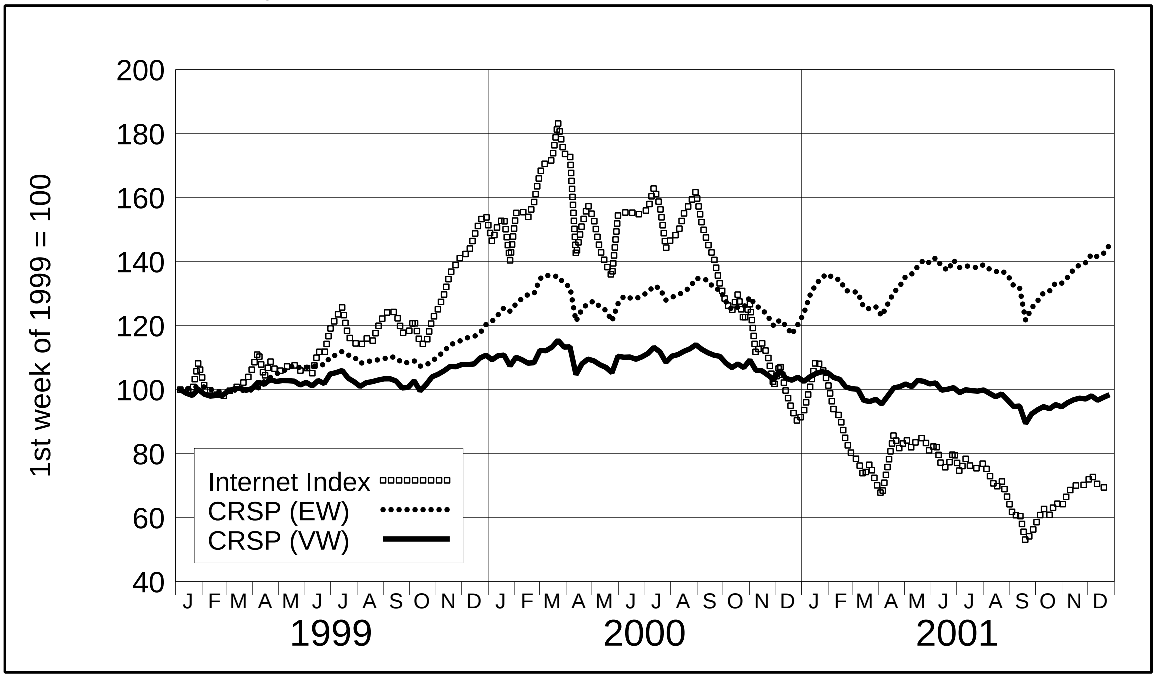


Figure 3: Price Movements of Benchmark Indices

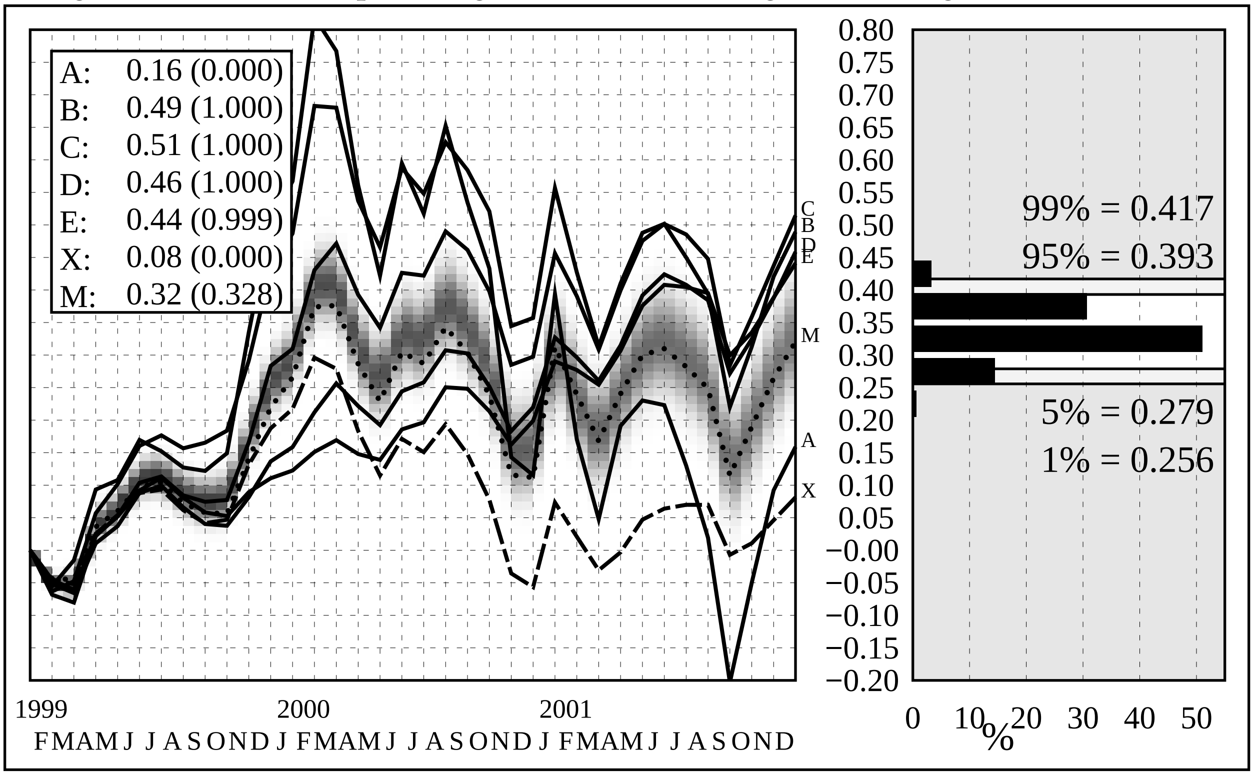


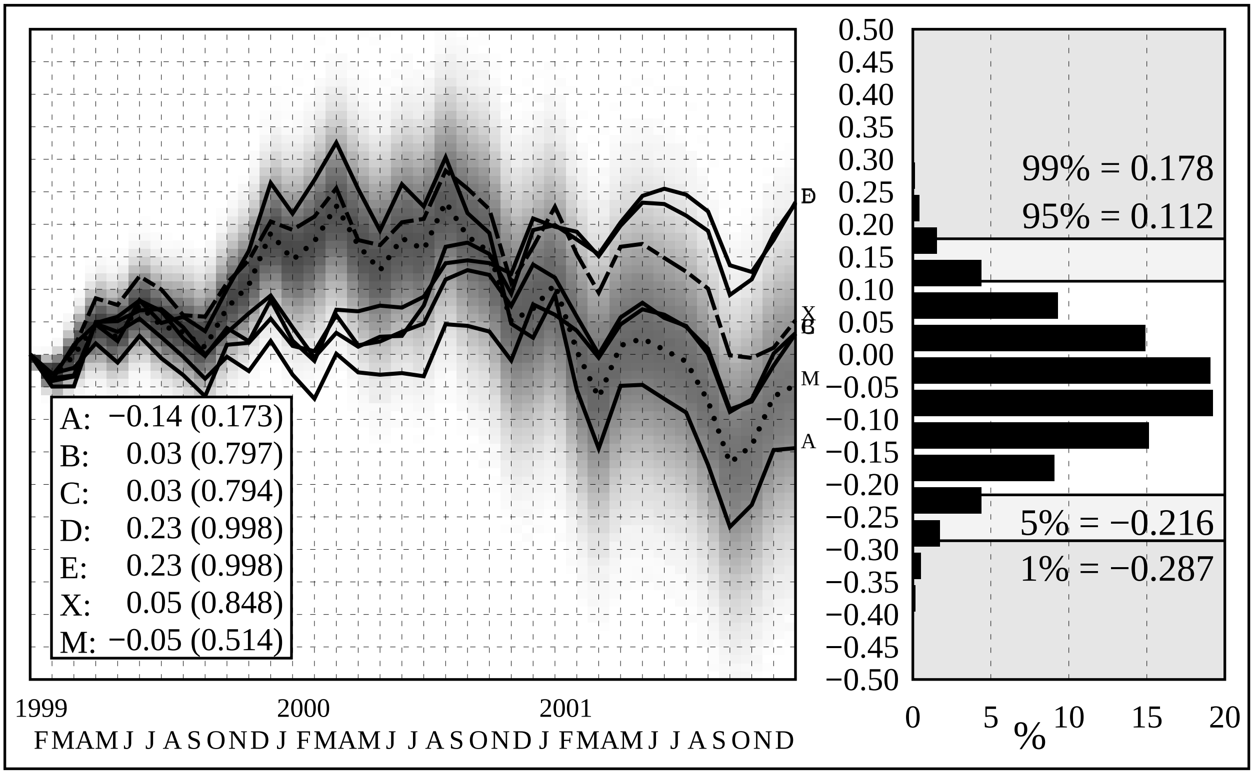
Note: 'Internet index' is the stock index with ticker symbol XLK. 'CRSP (EW)' and 'CRSP (VW)' are the equal-weighted and value-weighted stock indices provided by the Center for Research in Securities Pricing at the University of Chicago.

Figure 4 Notes

The top and bottom panels show the performance of the message board quintile portfolios (A-E) during the period February 1999 through December 2001. The additional X portfolio includes stocks that were not discussed on Yahoo. The dotted line with portfolio M corresponds to the overall market performance, which is taken to be the CRSP equal-weighted and value-weighted indices, respectively. The top panel shows the returns of the equal-weighted portfolios, while the bottom panel shows the returns of the value-weighted portfolios. Portfolios were determined through the Yahoo posting volume of the previous month. The main chart tracks the performance of each of the seven portfolios. The shading indicates the distribution of returns from the Monte Carlo simulation, and the histogram on the right shows the final distribution of returns from randomized portfolios of appropriate size. This Monte Carlo simulation involved 10,000 replications and was used to obtain P-values. The P-values corresponding to the 1%, 5%, 95% and 99% levels of confidence are indicated on the chart. The box insert in the main chart shows the final returns and corresponding P-values for the A-E and X portfolios.

Figure 4: Returns of equal-weighted and value-weighted Message Board Portfolios





Note: See previous page for detailed explanations.

Tables

Table 1: Double-Sorted Quantile Portfolio Returns

Panel A: Equal-Weighted Portfolios

Portfolio	Market Cap high	Market Cap medium	Market Cap low	all	Posting Volume
A (high)	-0.04 (0.000)	-0.15 (0.000)	0.56 (1.000)	0.16 (0.003)	924.5
B	0.12 (0.000)	0.52 (0.998)	0.71 (1.000)	0.49 (0.990)	79.32
C	0.13 (0.001)	0.57 (1.000)	0.77 (1.000)	0.52 (0.997)	23.81
D	0.30 (0.282)	0.37 (0.711)	0.67 (1.000)	0.46 (0.971)	7.19
E (low)	0.36 (0.681)	0.29 (0.272)	0.63 (1.000)	0.44 (0.945)	1.19
X (none)	0.14 (0.001)	-0.08 (0.000)	0.16 (0.003)	0.08 (0.000)	0.00
all	0.16 (0.004)	0.35 (0.572)	0.47 (0.979)	0.34 (0.515)	
Market Cap	7.37	0.21	0.04		

Panel B: Value-Weighted Portfolios

Portfolio	Market Cap high	Market Cap medium	Market Cap low	all	Posting Volume
A (high)	-0.15 (0.291)	-0.22 (0.170)	0.48 (0.999)	-0.14 (0.296)	924.5
B	0.01 (0.622)	0.44 (0.998)	0.58 (1.000)	0.03 (0.677)	79.32
C	-0.02 (0.571)	0.61 (1.000)	0.62 (1.000)	0.04 (0.685)	23.81
D	0.21 (0.932)	0.37 (0.992)	0.46 (0.998)	0.23 (0.945)	7.19
E (low)	0.22 (0.937)	0.34 (0.987)	0.40 (0.995)	0.24 (0.950)	1.19
X (none)	0.06 (0.725)	-0.08 (0.437)	0.07 (0.757)	0.05 (0.719)	0.00
all	-0.06 (0.468)	0.34 (0.987)	0.32 (0.983)	-0.05 (0.498)	
Market Cap	7.37	0.21	0.04		

Note: This table shows portfolio returns based on the combination of five quantiles of the distribution of Yahoo posting activity and three quantiles of market capitalizations. Returns are in logs. Each portfolio is rebalanced monthly based on market caps and posting volume of the previous month. The A-E quintile portfolios for a given month are defined by forming five equal-sized groups of all CRSP stocks (excluding penny stocks) by descending rank of Yahoo posting volume in the previous month. The X portfolio contains stocks that did not have a Yahoo message board. Stocks in the X portfolio were delisted due to merger, acquisition, or other types of exit. P-values based on the Monte-Carlo empirical distribution are shown in parentheses. The 'Posting Volume' column and 'Market Cap' row show average posting volume (messages per month per company) and average market capitalization (billion US Dollars). The time period is February 1999 through December 2001, with data for January 1999 used for initialization.

Table 2: Double-Sorted Quantile Portfolio Volatilities

Panel A: Equal-Weighted Portfolios

Portfolio	Market Cap high	Market Cap medium	Market Cap low	all	Posting Volume
A (high)	8.27 (1.000)	10.75 (1.000)	11.70 (1.000)	10.24 (1.000)	924.5
B	6.56 (0.000)	9.12 (1.000)	10.85 (1.000)	8.85 (1.000)	79.32
C	5.77 (0.000)	7.75 (0.673)	9.94 (1.000)	7.82 (0.982)	23.81
D	5.30 (0.000)	6.33 (0.000)	8.70 (1.000)	6.78 (0.000)	7.19
E (low)	5.34 (0.000)	5.50 (0.000)	7.83 (0.992)	6.22 (0.000)	1.19
X (none)	6.27 (0.000)	6.54 (0.000)	7.87 (0.999)	6.89 (0.000)	0.00
all	6.88 (0.000)	7.66 (0.050)	8.64 (1.000)	7.73 (0.502)	
Market Cap	7.37	0.21	0.04		

Panel B: Value-Weighted Portfolios

Portfolio	Market Cap high	Market Cap medium	Market Cap low	all	Posting Volume
A (high)	6.14 (0.952)	10.64 (1.000)	11.41 (1.000)	6.27 (0.997)	924.5
B	5.28 (0.000)	8.93 (1.000)	10.30 (1.000)	5.49 (0.000)	79.32
C	4.97 (0.000)	7.63 (1.000)	9.15 (1.000)	5.19 (0.000)	23.81
D	4.71 (0.000)	6.28 (0.998)	7.80 (1.000)	4.89 (0.000)	7.19
E (low)	5.18 (0.000)	5.53 (0.000)	6.86 (1.000)	5.24 (0.000)	1.19
X (none)	5.45 (0.000)	6.51 (1.000)	7.40 (1.000)	5.51 (0.000)	0.00
all	5.85 (0.136)	7.68 (1.000)	8.11 (1.000)	5.91 (0.301)	
Market Cap	7.37	0.21	0.04		

Note: This table shows realized portfolio volatilities (at annualized rates) based on the combination of five quantiles of the distribution of Yahoo posting activity and three quantiles of market capitalizations. Realized volatilities are calculated as monthly sums of absolute values of daily log returns. Each portfolio is rebalanced monthly based on market caps and posting volume of the previous month. The A-E quintile portfolios for a given month are defined by forming five equal-sized groups of all CRSP stocks (excluding penny stocks) by descending rank of Yahoo posting volume in the previous month. The X portfolio contains stocks that did not have a Yahoo message board. Stocks in the X portfolio were delisted due to merger, acquisition, or other types of exit. P-values based on the Monte-Carlo empirical distribution are shown in parentheses. The 'Posting Volume' column and 'Market Cap' row show average posting volume (messages per month per company) and average market capitalization (billion US Dollars). The time period is February 1999 through December 2001, with data for January 1999 used for initialization.

Table 3: Double-Sorted Quantile Portfolio Sharpe Ratios

Panel A: Equal-Weighted Portfolios

Portfolio	Market Cap high	Market Cap medium	Market Cap low	all	Posting Volume
A (high)	-.052 (0.000)	-.055 (0.000)	0.070 (0.363)	0.004 (0.002)	924.5
B	-.008 (0.000)	0.097 (0.763)	0.123 (0.956)	0.105 (0.845)	79.32
C	-.006 (0.000)	0.149 (0.996)	0.179 (1.000)	0.148 (0.996)	23.81
D	0.101 (0.800)	0.124 (0.959)	0.231 (1.000)	0.189 (1.000)	7.19
E (low)	0.146 (0.995)	0.105 (0.844)	0.291 (1.000)	0.217 (1.000)	1.19
X (none)	0.002 (0.001)	-.109 (0.000)	0.011 (0.004)	-.029 (0.000)	0.00
all	0.011 (0.004)	0.074 (0.421)	0.126 (0.965)	0.081 (0.529)	
Market Cap	7.37	0.21	0.04		

Panel B: Value-Weighted Portfolios

Portfolio	Market Cap high	Market Cap medium	Market Cap low	all	Posting Volume
A (high)	-.125 (0.326)	-.070 (0.589)	0.057 (0.959)	-.121 (0.346)	924.5
B	-.091 (0.487)	0.079 (0.978)	0.099 (0.988)	-.072 (0.578)	79.32
C	-.105 (0.418)	0.165 (0.999)	0.147 (0.998)	-.069 (0.590)	23.81
D	0.049 (0.951)	0.125 (0.995)	0.154 (0.998)	0.063 (0.966)	7.19
E (low)	0.051 (0.953)	0.126 (0.995)	0.168 (0.999)	0.066 (0.968)	1.19
X (none)	-.047 (0.692)	-.106 (0.415)	-.033 (0.746)	-.049 (0.681)	0.00
all	-.108 (0.407)	0.069 (0.971)	0.071 (0.973)	-.100 (0.443)	
Market Cap	7.37	0.21	0.04		

Note: This table shows portfolio Sharpe ratios (at annualized rates) based on the combination of five quantiles of the distribution of Yahoo posting activity and three quantiles of market capitalizations. Sharpe ratios are calculated as the portfolio return minus the riskfree rate, divided by the standard deviation of the portfolio return. Each portfolio is rebalanced monthly based on market caps and posting volume of the previous month. The A-E quintile portfolios for a given month are defined by forming five equal-sized groups of all CRSP stocks (excluding penny stocks) by descending rank of Yahoo posting volume in the previous month. The X portfolio contains stocks that did not have a Yahoo message board. Stocks in the X portfolio were delisted due to merger, acquisition, or other types of exit. P-values based on the Monte-Carlo empirical distribution are shown in parentheses. The 'Posting Volume' column and 'Market Cap' row show average posting volume (messages per month per company) and average market capitalization (billion US Dollars). The time period is February 1999 through December 2001, with data for January 1999 used for initialization.

Table 4: Properties of Quintile Portfolios Sorted on Message Board Posting Volume

Pf.	α	β	γ_1 (SMB)	γ_2 (HML)	γ_3 (UMD)	γ_4 (LIQ)
A (high)	3.441^c	1.259^c	1.062^c	-0.798^c	-0.395^c	-2.021
	(15.58)	(23.61)	(20.54)	(14.90)	(11.14)	(1.214)
B	2.153^c	1.117^c	0.924^c	-0.121^b	-0.335^c	-6.170^c
	(12.71)	(26.66)	(23.88)	(2.819)	(11.64)	(4.459)
C	1.528^c	1.036^c	0.760^c	0.216^c	-0.271^c	-3.741^b
	(10.65)	(26.99)	(22.32)	(5.331)	(11.89)	(3.224)
D	0.663^c	0.885^c	0.599^c	0.395^c	-0.233^c	-3.201^b
	(5.756)	(28.82)	(24.10)	(12.47)	(13.35)	(3.232)
E (low)	-0.098	0.657^c	0.491^c	0.403^c	-0.205^c	-4.140^c
	(1.015)	(24.78)	(24.03)	(14.80)	(13.79)	(5.065)
X (none)	0.339^b	0.473^c	0.601^c	-0.286^c	-0.385^c	-8.271^c
	(2.818)	(15.06)	(25.05)	(8.379)	(15.07)	(8.886)

Note: Portfolio quintiles are formed by ranking stocks discussed on Yahoo by their posting volume. The top quintile (A) contains the most actively-discussed stocks, and the bottom quintile (E) the least-discussed stocks. The X quantile contains stocks that were not discussed on Yahoo but are traded on the NYSE, AMEX, or NASDAQ. The factor estimates α, β, and γ_1–γ_4 are obtained from Generalized Method of Moments regressions for each portfolio. The GMM regressions use a Newey-West kernel with 1 lag. Monthly (1999/01-2001/12) data were used, with the first month lost to the initialization of the portfolio selection, and the last month lost due to the initialization of the liquidity measure. T-statistics of the estimates are shown in parentheses. Superscripts ^{a,b,c} indicate significance at the 95%, 99%, and 99.9% levels of significance.

Table 5: Alphas of Quintile Portfolios Sorted on Message Board Posting Volume

		Portfolio					
		A	B	C	D	E	X
Line	Model	(high)				(low)	(none)
1	CAPM alpha	3.616^c	2.711^c	2.223^c	1.335^c	0.502^c	0.207
		(16.87)	(16.61)	(15.82)	(12.09)	(5.508)	(1.886)
2	Fama-French alpha	3.123^c	1.934^c	1.374^c	0.534^c	-0.198^a	-0.138
		(14.51)	(11.76)	(9.830)	(4.726)	(2.106)	(1.252)
3	4-factor alpha	3.423^c	2.186^c	1.581^c	0.711^c	-0.043	0.319^b
		(15.71)	(13.01)	(11.11)	(6.218)	(0.445)	(2.683)
4	Pástor-Stambaugh	3.441^c	2.153^c	1.528^c	0.663^c	-0.098	0.339^b
		(15.58)	(12.71)	(10.65)	(5.756)	(1.015)	(2.818)

Note: Portfolio quintiles are formed by ranking stocks discussed on Yahoo by their posting volume. The top quintile (A) contains the most actively-discussed stocks, and the bottom quintile (E) the least-discussed stocks. The X quantile contains stocks that were not discussed on Yahoo but are traded on the NYSE, AMEX, or NASDAQ. Regressions are based on monthly (1999/01-2001/12) data for each portfolio, with the first month lost to initialization of the portfolio selection. The regression involving the liquidity measure also loses the last month due to initialization. A Generalized Method Moments estimator was used with a 1-lag Newey-West kernel. T-statistics of the estimates are shown in parentheses. Superscripts ^{a,b,c} indicate significance at the 95%, 99%, and 99.9% levels of significance.

Table 6: Market Factor Summary Statistics and Correlations with Message Board Factors

	Market	SMB	HML	UMD	LIQ	NMQ
			Panel A: Descriptive Statistics			
Mean	-0.281	0.877	0.688	1.145	-0.003	-0.807
Standard Deviation	5.321	6.591	6.315	8.556	0.112	11.164
Median	-0.316	0.855	0.840	1.935	0.011	-1.026
Minimum	-10.638	-16.690	-12.030	-25.100	-0.250	-23.892
Maximum	7.936	21.490	13.750	18.160	0.189	22.385
			Panel B: Correlations			
Noise-Minus-Quiet (NMQ)	0.782^{c}	0.654^{c}	-0.918^{c}	0.005	0.166	
	(7.319)	(5.045)	(13.49)	(0.029)	(0.969)	
Yahoo Volume Change (YYY)	0.389^{a}	0.245	-0.300	-0.120	0.067	0.443^{b}
	(2.461)	(1.474)	(1.833)	(0.705)	(0.385)	(2.841)

Note: T-statistics are shown in parentheses. Superscripts ^{a,b,c} indicate significance at the 95%, 99%, and 99.9% levels of significance. β is the CAPM beta, while γ_1 through γ_6 refer to the following other factors. HML (high minus low) and SMB (small minus big) are the two classical Fama-French market factors for size and book-to-market ratio, and UMD (up minus down) is the Jegadeesh-Titman momentum factor. LIQ is the Pástor-Stambaugh liquidity measure. NMQ (noise minus quiet) is our primary message board activity factor we have constructed, and YYY is a secondary factor that is constructed by first-differencing the total message board posting volume.

Table 7: Message Board Activity as a Market Factor

Line	Model	α	β	γ_1 (SMB)	γ_2 (HML)	γ_3 (UMD)	γ_4 (LIQ)	γ_5 (NMQ)
				Panel A: Conventional Models				
1	CAPM	1.327^c	1.127^c					
		(24.15)	(107.0)					
2	Fama-French	0.750^c	0.971^c	0.594^c	0.049^c			
		(13.50)	(71.28)	(48.21)	(3.317)			
3	4-factor	1.059^c	0.813^c	0.698^c	-0.051^c	-0.316^c		
		(18.37)	(54.84)	(55.86)	(3.296)	(30.07)		
4	Pástor-Stambaugh	1.031^c	0.813^c	0.704^c	-0.056^c	-0.319^c	-5.630^c	
		(17.74)	(54.61)	(54.88)	(3.581)	(30.16)	(12.16)	
				Panel B: Inclusion of NMQ Factor				
5	CAPM	1.360^c	0.256^c					0.532^c
		(24.42)	(13.35)					(48.04)
6	Fama-French	0.574^c	0.426^c	0.383^c	0.783^c			0.725^c
		(10.29)	(19.45)	(28.55)	(34.99)			(34.81)
7	4-factor	0.819^c	0.474^c	0.533^c	0.494^c	-0.242^c		0.503^c
		(14.29)	(22.37)	(41.96)	(22.40)	(23.73)		(26.82)
8	Pástor-Stambaugh	0.818^c	0.457^c	0.524^c	0.498^c	-0.241^c	-6.101^c	0.524^c
		(14.05)	(21.38)	(40.47)	(22.43)	(23.61)	(12.93)	(27.62)
				Panel C: Robustness checks (baseline model: line 8)				
9	YYY instead of NMQ	0.697^c	0.739^c	0.673^c	-0.053^c	-0.302^c	-5.023^c	0.070^c
		(11.63)	(47.09)	(51.67)	(3.359)	(28.71)	(10.78)	(17.27)
10	1999/01–2000/07	0.472^c	0.621^c	0.513^c	0.635^c	-0.156^c	0.202	0.490^c
		(6.420)	(24.56)	(30.84)	(17.31)	(7.725)	(0.276)	(20.65)
11	2000/08–2001/12	0.705^c	0.169^c	0.849^c	0.570^c	-0.164^c	-11.80^c	0.691^c
		(6.616)	(4.120)	(33.72)	(16.68)	(9.792)	(16.22)	(18.70)
12	Top Yahoo Stocks	4.219^c	0.310^a	0.400^c	0.148	-0.225^c	-3.289	1.428^c
		(10.34)	(1.963)	(4.273)	(1.006)	(3.477)	(1.103)	(10.30)

Note: This table reports the estimates of the NMQ (noise-minus-quiet) factor (γ_5) incorporated into Fama-French style regressions. HML (high minus low) and SMB (small minus big) are the two classical Fama-French market factors for size and book-to-market ratio. UMD (up minus down) is the Jegadeesh-Titman momentum factor, and LIQ is the Pástor-Stambaugh liquidity measure. The NMQ factor is constructed as the difference of returns between the A portfolio and the E portfolio in each month. Also reported are estimates corresponding to the YYY factor, which is based on the monthly percentage change in posting volume. Estimation period is 1999/01-2001/12, and all NYSE, AMEX, and NASDAQ traded stocks are included in the regressions. The first month of observations is lost due to initialization of the portfolio selection. The regression involving the liquidity measure also loses the last month due to initialization. A Generalized Method of Moments (GMM) estimator was used with a 1-lag Newey-West kernel. T-statistics of the estimates are shown in parentheses. Superscripts ^{a,b,c} indicate significance at the 95%, 99%, and 99.9% levels of significance.