Attention: Implied Volatility Spreads and Stock Returns

Xuechen Gao
Assistant Professor of Finance
University of Central Arkansas
Email: xgao2@uca.edu
(Phone): 501-450-5338

Xuewu (Wesley) Wang
Associate Professor of Finance
School of Business, Quinnipiac University
275 Mount Carmel Avenue, Hamden, CT 06518
Email: xuewu.wang@quinnipiac.edu
(Phone): 617-935-3977

Zhipeng Yan
Associate Professor of Finance
Martin Tuchman School of Management, New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Email: zyan@njit.edu
(Phone): 973-596-3260

Abstract

Two competing hypotheses are proposed in the finance literature to explain the well-documented stock return predictability of the call-put implied volatility spreads: informed trading in the options market and mispricing. In this paper, we examine how the return predictability of volatility spreads changes as investor attention level varies. Using a new and direct measure of investor attention generated from the SEC's EDGAR log files, we find that as investor attention heightens, the volatility spread return predictability becomes more pronounced, providing favorable evidence for the informed trading hypothesis. A portfolio that longs stocks with the highest investor attention and the highest volatility spread and shorts stocks with the highest attention and the lowest volatility spread generates a Fama-French 5-factor monthly alpha of 2.43%.

JEL Classification: G10, G11, G14, G40

Keywords: Implied Volatility Spreads; the SEC's EDGAR Log Files; Investor Attention;

1. Introduction

It has been well established that the informed traders can choose to trade options first to capitalize on their private information (Easley et al. 1998) and that price discovery can take place in the options market (Chakravarty et al. 2004). Consistent with this notion, the existing literature has documented that economic variables constructed from the options market reliably predict subsequent stock returns.[1] One of such variables is the call-put implied volatility spread, defined as the implied volatility on the call options minus that on the put options, where calls and puts are matched on the underlying asset, strike price and expiration date (Bali and Hovakimian 2009, Cremers and Weinbaum 2010, Xing et al. 2010).

While the strong positive relation between implied volatility spreads and future stock returns survives different samples and time periods, what drives the predictive power remains controversial. Bali and Hovakimian (2009) argue that this predictability stems from informed trading in the options market, whereas Cremers and Weinbaum (2010) attribute it to mispricing, since they find the ability of volatility spreads to forecast future stock returns is strongest among relatively illiquid stocks that have liquid options. In addition, the predictability attenuates over time. This paper attempts to provide further evidence on the economic driver of the return predictability, and more importantly, examines the investment implication of such predictability.

Our approach is straightforward. We attempt to interpret the return predictability through the lens of investor attention and information acquisition. That is, we examine how the return predictive power changes as investor attention level varies. This allows us to differentiate the two competing hypotheses. If mispricing is the main driver for the return predictability, we expect weaker predictive power in the presence of heightened investor attention. After all, more attentive investors demanding more information help alleviate mispricing. In the extreme case, investor attention and the resulting information acquisition can be at such a high level that all value-relevant information is immediately produced and incorporated into asset prices, thus completely eliminating stock mispricing and return predictability. On the other hand, if informed trading is the main driver, we can expect stronger return predictability as investors become more attentive and their information acquisition activity intensifies. This is because as more investors pay more

[1] There is a vast amount of literature on informed trading in the options market. An incomplete list of papers includes Amin and Lee (1997), Easley et al. (1998), Cao et al. (2005), Pan and Poteshman (2006), Bali and Hovkimian (2009), Roll et al. (2010), Johnson and So (2012). We will survey this literature in Section 2.

attention, there is an increased likelihood of noise traders providing better camouflage for informed trading in the sense of Kyle (1985) and Admati and Pfleiderer (1988). As a result, informed traders may trade more aggressively, leading to stronger return predictability.

The power of these arguments hinges critically on how well we measure investors' attention level and their information acquisition activities. In this paper, we avail ourselves of a novel dataset that the Securities and Exchange Commission (SEC) has recently made available to the public: the SEC's EDGAR (Electronic Data Gathering, Analysis, and Retrieval) access log files. These access log files record all internet search traffic on its EDGAR system starting from February 2003. The log information includes each user's IP address as well as the firms and filings the internet user is requesting information about. Using investor activities on the EDGAR system, we can measure investor attention directly and unambiguously. While there are many alternative sources and measures for information acquisition activity, these log files have certain unique advantages for a number of reasons. [2] One prominent feature is that it has a long time series and a wide cross section, which enables us to conduct a series of asset pricing tests for constructed portfolios.

We first validate that stocks with higher volatility spreads in month t record higher returns in month t+1. For each month in our sample, we sort stocks into ten deciles based on the volatility spreads. We then calculate the average returns in the following month for each decile. We find that stock returns generally increase across the ten volatility spread deciles. More importantly, the return gap between the top and bottom deciles averages over 1.51 percent per month. Thus, we are able to verify that volatility spreads predict subsequent stock returns in the cross-section. We further show that this finding is robust to different return measures and the inclusion of conventional return predictors. An equal-weighting (value-weighting) portfolio that goes long on the stocks in the top volatility spread decile and short on the stocks in the bottom volatility spread decile generates a monthly Fama-French 5-factor alpha of 1.69 percent (1.51 percent) with a t-stat of 10.35 (6.91). It is thus profitable to utilize this predictability and construct investment portfolios solely based on the volatility spread.

Next, we examine how the profitability of the spread-only portfolio changes when stocks are further sorted based on investor attention as measured by the EDGAR access log files. To achieve this purpose, we conduct a double sorting procedure. Specifically, we sort stocks into ten deciles based on the volatility spreads. In addition, stocks are independently sorted into ten deciles based

[2] We will discuss these reasons in greater details in Section 3.

on investor attention in that month. Our focus is on the differences in the subsequent returns for stocks in the top and bottom volatility spread deciles across both the top and bottom attention deciles. For the bottom investor attention decile, returns on the top volatility spread decile are about 1.53 percent higher than the bottom volatility spread decile. In contrast, for the top investor attention decile, the return difference averages over 2.75 percent, illustrating that the return gap is much more conspicuous for the top attention decile than for the bottom attention decile. The striking difference in the return gap between the top and bottom attention deciles survives alternative sorting schemes as well as a rigorous multivariate regression test for statistical significance. Thus, our sample reveals much more pronounced return predictability for the top attention decile. This clearly constitutes favorable evidence towards the informed trading hypothesis and unfavorable evidence against the mispricing hypothesis.

Given the difference in the return gap, it would be interesting to examine whether professional money managers can capture a new source of alpha by focusing exclusively on the high attention stocks. Towards this end, we construct the spread-and-attention portfolios. Each month in our sample, we conduct the double sorting. The spread-and-high-attention portfolio is limited to only stocks with high investor attention and information acquisition activities, that is, stocks in the top attention decile. It goes long on the stocks in the top volatility spread decile and short on the stocks in the bottom volatility spread decile. We find that the spread-and-high-attention portfolio generates a Fama-French 5-factor monthly alpha of 3.36 (2.92) percent using equal (value) weighting. In comparison, the spread-and-low-attention portfolio that focuses on the low attention stocks only yields a Fama-French 5-factor monthly alpha of 1.70 (1.08) percent for equal (value) weighting.

One of the issues in using the SEC's EDGAR access log files is the reporting lag. Currently, the data is posted with a 6-month delay. Thus, the information used to construct portfolios is not available to investors when they need to build their portfolios. To resolve this issue, we turn to the investor attention measure 6 months ahead of the portfolio construction month. While this helps us get around the data availability issue, it begs the questions of the validity of this approach and the underlying economic rationale as to why this may be sensible and desirable. As pointed out by Li et al. (2018), investor attention as measured by the EDGAR log files tends to be persistent, especially given that institutional and sophisticated investors pay attention to and acquire information about certain stocks on a consistent basis. Thus, it can still be valuable to explore the profitability of the spread-and-attention portfolio by using the 6-month ahead investor attention measure. We indeed find that the spread-and-attention portfolio is still highly profitable even when we use the investor attention measure without any reporting lag. The equal-weighting (value-weighting) spread-and-high-attention portfolio yields a Fama-French 5-factor monthly alpha of 2.47 (2.43) percent with a t-stat of 3.05 (2.97).

This paper contributes to the existing literature along two dimensions. First, we add new evidence to the informed options trading literature by imposing investor attention on implied volatility spreads. Our findings shed light on the source of the return predictability. We argue that it is important to understand the source of the return predictability, especially for professional money managers who wish to design trading strategies to utilize this predictability. If the predictive power originates from informed trading, then such trading strategies can be sustainable. If mispricing drives the predictability, then such strategies may be much less appealing. The favorable evidence towards informed trading hypothesis implies that professional money managers do not have to concern themselves with the possibility that such predictability may disappear due to mispricing being corrected.

Our study also proposes a profitable trading strategy that practitioners can implement in the real time. The spread-and-high-attention portfolio generates a statistically and economically significant monthly alpha. The size of the monthly alpha seems to be large enough to accommodate reasonable transaction costs caused by monthly rebalancing of the portfolio. While we recognize the data limitation caused by the reporting lag of the access log files, we show that it is still profitable to use investor attention measure 6 months ahead of the portfolio formation month.

The rest of the paper is organized as follows. In Section 2, we review the literature on informed trading in the options market as well as investor attention. Section 3 discusses the data and methodology used in our empirical analysis. Section 4 contains our main empirical results. We conclude in Section 5.

2. Literature Review

In this section, we review two distinct strands of literature: the informed trading literature and the investor attention literature.

2.1. The informed options trading literature

In the Black-Scholes (1973) framework, options are redundant securities whose payoffs can be replicated by the underlying assets. Since options derive their value from the underlying assets, options trading should convey no new information to the market. However, in an incomplete market, options may no longer be redundant, and their trading can be informative of the underlying asset prices. Black (1975) points out that options provide higher leverage as compared to the underlying assets, which makes it desirable for investors who want to utilize leverage to magnify their returns. Consistent with this argument, the literature has documented ample evidence regarding the informational content of options trading. Amin and Lee (1997) document unusual activities in the options market immediately before earnings announcements. Options traders establish more long (short) positions before positive (negative) earnings news. Easley et al. (1998) build a theoretical model where sophisticated traders with private information elect to trade options first. Cao et al. (2005) show that takeover targets that experience the largest pre-announcement call-imbalance have highest announcement-day returns. Using a proprietary dataset, Pan and Poteshman (2006) document strong predictive power of their constructed put-call ratio on future stock returns. Roll et al. (2010) document higher options to stock volume ratio (O/S) around earnings announcements. Johnson and So (2012) show that the O/S ratio predicts future firm-specific earnings news, indicating that O/S reflects private information.

Using the options trading volume as the main measure for informed trading requires either high-frequency data on options trades and quotes or proprietary data. In view of this data limitation, a related strand of literature has turned to the implied volatility spread as a return predictor. Bali and Hovkimian (2009) show that implied volatility spread predicts future stock returns. Cremers and Weinbaum (2010) find that deviations from the put-call parity contain information about future stock returns beyond the short-sales constraints. Xing et al. (2010) examine the predictive power of a variation of the volatility spread, the volatility smirk in a cross section of stocks. Stocks with the steepest volatility smirks underperform stocks with the least volatility smirks by 10.9 percent on an annual basis after risk adjustments.

2.2. The investor attention literature

The notion of limited attention dates back to Kanehman (1973), who argues that human beings are subject to psychological biases and human brains have a capacity constraint when it comes to the cognitive-processing power. In contrast, the amount of information available to investors in the financial markets is enormous. Consequently, investors often fail to incorporate all relevant information due to limited attention.

The investor attention literature has grown substantially along two related themes: measures of investor attention and its economic implications. In a nutshell, the literature has developed various measures of investor attention and used them to investigate a multitude of interesting topics in asset pricing and market efficiency.

Measuring investor attention can be a daunting task since the exact determinants of investor attention and how investors allocate their limited attention remain largely unknown. Consequently, the literature has proposed a host of empirical proxies to measure investor attention. These empirical proxies include firm size; trading volume; information overload defined as the number of earnings announcements on a day (Hirshleifer, Lim, and Teoh 2009); event timing such as Fridays vs. Non-Fridays (DellaVigna and Pollet 2009) and trading hours vs. non-trading hours (Francis, Pagach, and Stephan 1992); Google's Search Volume Index (Da, Engelberg, and Gao 2011); search and browsing activities on Bloomberg terminals (Ben-Rephael Da, and Israelsen 2017); extreme returns (Kumar, Ruenzi, and Ungeheuer 2018); options trading volume (Wang 2017) etc.

Apparently, each investor attention measure has its own merits and limitations. For instance, larger firms typically receive more media and analyst coverage, hence they attract higher investor attention. However, firm size can also proxy for liquidity and information asymmetry. Thus, firm size is at best a noisy measure for investor attention. Trading volume is another widely used measure for investor attention. It has an intuitive appeal in that the more attentive investors are, the more likely they will trade, and the higher the trading volume. However, trading volume can also be driven by other considerations such as liquidity and divergence of opinions.

Using these proxies for investor attention, existing studies have gained many insights into how the financial markets respond to information-rich events such as earnings announcements (EAs). Francis, Pagach, and Stephan (1992) document a greater underreaction to EAs made during non-trading hours. Della Vigna and Pollet (2009) show that Friday EAs have more muted market response and stronger post-earnings announcement drift. Hirshleifer, Lim, and Teoh (2009) show that when investors are overloaded with too many EAs on a single day, the announcement-day reaction is much weaker, and the following drift is much stronger. Wang et al. (2018) (2018)

report stronger initial market response and weaker post-earnings announcement drift when the pre-earnings announcement option trading is more active.

Despite the existing studies, only a few studies have examined the implications of investor attention on portfolio construction. Similar to Li et al. (2018), we measure investor attention via user activities on the EDGAR website. Li et al. (2018) document that the market response for low-attention EAs is more muted, and the post-announcement drift is much more pronounced. They further construct the attention-based portfolio that exploits this differential market response and show that such a portfolio is highly profitable. Our paper shows that it is feasible and highly profitable to construct portfolios based on not only options volatility spreads but also investor attention.

3. Data & Methodology

We use a number of data sources to conduct our empirical analysis. When constructing our sample, we merge the two major databases: OptionMetrics and the SEC's EDGAR access log files. Supporting data and variables come from CRSP and Compustat. In what follows, we provide more detailed information on the sample construction process.

3.1. Options data

Options implied volatility data come from the Ivy DB OptionMetrics database, which has evolved into the industry standard for options-related research with data available since January 1996. In addition to the end-of-day summary data of options trading volume as well as the best bid and best offer prices for each optioned stock with contracts distinguished by options type (call or put), strike price, and expiration dates, OptionMetrics also provides standardized options price files (stdopd). These files contain prices, implied volatilities, and Greeks for at-the-forward-money call and put options with fixed days until expiration. Specifically, standardized option price files have options with 30, 60, 91, 182, 273, 365, 547, and 730 days until expiration. The strike prices are set to be equal to the forward price of the underlying stock with the forward delivery date matching the expiration date. For each stock on each day, OptionMetrics calculates an implied volatility surface using values interpolated from available options price data.

We follow Bali et al. (2016) and apply a number of filters when extracting the implied volatility data. The implied volatility, forward prices, and options prices provided by OptionMetrics must be greater than zero. The strike price is set equal to the forward price of the underlying stock.

Call and put options are matched on the strike price and maturity. In our empirical analysis, we limit ourselves to the standardized options with a maximum maturity of 120 days. Our robustness checks show that the results are virtually the same when other maturity cutoffs, such as 30 days, 60 days, or 180 days, are used.

3.2. Data on investor attention and information acquisition

As one of the main information repositories for original corporate filings, the SEC's EDGAR server experiences massive internet search traffic on a daily basis. Innumerable investors visit the EDGAR server to acquire filing information that interest them. Starting from February 2003, the SEC has been tracking such search traffic via the EDGAR access log files.[3] These log files contain detailed information about the users' IPs, corporations and filings, and the detailed time stamp (nearest to the second) etc. Recently, the SEC has released these log files to the public.[4] Since then, a fast-growing number of academic studies have utilized this dataset for research topics relating to investor attention and information acquisition (Lee et al. 2015; Drake et al. 2015, Loughran and McDonald 2016; Chen et al. 2017, Ryans 2018, Li et al 2018).

These studies contend that in spite of alternative sources that researchers can use to construct metrics for investor attention and information acquisition, the Edgar access log dataset have certain advantages for a number of reasons. First, the information collection process in these log files is fairly unique in that it allows researchers to remove accesses by computer programs or "robots" and retrieve number of downloads made by human beings. Second, the database is the original storage of corporate filings. It houses a complete set of corporate filings including annual reports (10-Ks), quarterly reports (10-Qs), 8-Ks, IPO prospectuses (Form S-1 and Form 424), insider trading files (Forms 3, 4 and 5), and SC 13D/13G etc. Search engines such as Google as well as the investor relations websites of many corporations such as ExxonMobil redirect investors to the EDGAR web site.[5] In these cases, it is possible that the log files pick up a

[3] Our sample starts from March 2003 since we require at least 15 days for any stocks in a month in order for us to calculate the average number of downloads by human beings. Our sample ends in December 2016 since the SEC has only posted data for the first half of 2017.

[4] A detailed description of the SEC Edgar access log dataset is available at: https://www.sec.gov/dera/data/edgar-log-file-data-set.html.

[5] Ryans (2017) provides some anecdotal evidence regarding investors' use of corporate websites for SEC filings. GE CFO noted that its 2013 annual report was downloaded only 800 times from GE's website during the entire year. In comparison, the EDGAR logs record 21,987 (4,325) downloads in the year (two months) following its filing for the same annual report. Thus, it is questionable whether investors predominantly use the company website for SEC filing retrievals.

significant fraction of filing retrievals. Third, these log files provide very rich information about the identity of the investors who request the filing information. Granular information on the masked Internet Protocol (IP) address, a timestamp, and the SEC accession number for every client request has been presented in easy-to-read comma-separated values (CSV) files. Efforts have been made to match the IP addresses to mutual fund managers to track down their portfolio choices and to match the SEC accession number to the original SEC filing and deduce the filing types (Chen et al. 2018). Fourth, these log files have a relatively long time series for a wide cross section of firms. In recent years, the number of observations is typically in millions or more for a day. All these reasons suggest that this dataset can be a very desirable choice for studies on investor attention and information acquisition and its implications on various financial markets.

When using this dataset, we need to remove all downloads by computer programs/robots so that we can construct measures for investor attention and information acquisition by human beings. Unfortunately, we have to make inferences about whether downloads are initiated by human readers since the IP addresses (the last octet is masked in a unique way) do not directly reveal whether it is a human being or a computer robot requesting information. To resolve this issue, the literature has proposed three alternative approaches: Drake, Roulstone, and Thornock (2015), Loughran and McDonald (2016), and Ryans (2018). These three approaches share one main feature in common in the screening methodology. The general idea is to calculate certain statistics for any user's download patterns and employ one or more tests to classify the user as either a human being or a computer robot. These approaches differ in their specific tests for computer robots.[6] Ryans (2018) compares these three approaches in great detail and constructs the number of downloads construed to be requested by human beings following each of the three approaches.[7] To measure investor attention in each month in our sample, we calculate the average number of downloads made by human beings in the month and label these three alternative measures as DRT_{pv}, LM_{pv}, and R_{pv} respectively in our empirical tests.

Table 1 provides the summary statistics for the main variables used in our empirical analysis. Note that we have over 320, 000 firm-month observations after matching OptionMetrics with

[6] Specifically, Drake, Roulstone, and Thornock (2015) remove requests from IP addresses accessing more than five filings in a given minute or more than 1,000 filings during a day. Loughran and McDonald (2017) simply define more than 50 requests in a single day from a particular IP address as a "robot." Ryans (2018) assumes humans download no more than 25 items or three different companies' items in a single minute; and humans download no more than 500 items in a single day.
[7] Professor James Ryans kindly posts the processed daily log files on his personal website at http://www.jamesryans.com/. We thank him for his generosity.

the access log files based on Central Index Keys (CIKs) and CUSIPs. This clearly speaks to the attractiveness of the log files. Also note that the implied volatility spreads are negative on average, which is consistent with what the literature has documented. More importantly, we notice that the number of downloads by human beings in a month averages around 813, 580, and 664 respectively based on the three robot-screening approaches. This strongly validates the argument that the access log files capture a significant fraction of investors' information acquisition activities.

****** Insert Table 1 about here ******

4. Empirical Results

4.1. Volatility spread on stock returns

Our empirical analysis starts with replicating the predictive power of implied volatility spreads on stock returns. Following the standard practice in the literature, we conduct a single sorting. Each month in our sample, we sort all sample stocks into ten deciles based on the volatility spreads. We then calculate the abnormal return of the stocks in the following month by subtracting the return of the similar size portfolio from the raw return (AR1), by subtracting the market return from the raw return (AR2), or by subtracting the return on a portfolio of similar size, book to market, and momentum from the raw return (AR3). We then average these abnormal returns across all the stocks in each decile in each month. We thus obtain a time series of the average abnormal returns for each of the ten decile portfolios. Table 2 reports the time series averages of these average abnormal returns. As we can see clearly, the average abnormal return generally increases across the ten volatility spread deciles. The bottom (top) volatility spread decile records an abnormal return AR3 of -81 (70) basis points on a monthly basis. The difference of 1.51 percent per month is statistically significant at 1 percent level. Using AR1 and AR2 reveals a similar pattern. Overall, results in Table 2 constitute strong evidence supporting the return predictability of implied volatility spreads.

****** Insert Table 2 about here ******

4.2. Regression analysis

We next subject the return results in Table 2 to a more rigorous regression test. We need to ensure that the return difference between the top and bottom volatility spread decile portfolios does not disappear once we control for other known return predictors. To achieve this purpose, we employ a Fama-MacBeth regression. Specifically, each month in our sample, we sort stocks into 10 deciles based on the volatility spreads. We then only retain stocks in the top and bottom deciles. We create a dummy variable *IsHgh* that takes the value of 1 if the stock is in the top volatility spread decile and 0 otherwise. Each month, we estimate the following regression equation:

$$Ret = \alpha_0 + \alpha_1 \cdot IsHgh + \alpha_2 \cdot Size + \alpha_3 \cdot Past1y + \alpha_4 \cdot Past1m + \alpha_5 \cdot BM + \varepsilon$$

The dependent variable, *Ret*, is the stock return in month t+1. Our focal variable is *IsHgh*. We include a set of return predictors that are well established in the literature: firm size (*Size*) measured as the natural logarithm of a firm's market capitalization in month t, past year stock return (*Past1y*) measured as the cumulative stock return in the past year (skipping the most recent month), past month stock return (*Past1m*), and book-to-market ratio (*BM*).

Table 3 reports the time series averages of the slope coefficient estimates before each independent variable. We notice that our focal variable *IsHgh* carries a positive slope coefficient of 1.1 percent that is highly statistically significant. Thus, the return gap between top and bottom decile volatility spread portfolios survives the inclusion of these return predictors. The magnitude of 1.1 percent is also largely consistent with the return difference result in Table 2.

****** Insert Table 3 about here ******

4.3. Spread-only portfolios

We further examine whether it is feasible to utilize the return predictability and reap economically significant profits. Towards this end, we form the monthly portfolios based on the volatility spreads. Each month, we sort all stocks into ten deciles. The spread-only portfolio goes long on the stocks in the top volatility spread decile and short on the stocks in the bottom volatility spread decile. The portfolio is rebalanced at the end of each month. We then calculate the returns to the equal-weighting and value-weighting portfolios and subject them to risk adjustments using standard asset pricing models.

Table 4 reports the monthly abnormal returns to these spread-only portfolios. It shows that these spread-only portfolios generate significant monthly alphas. The equal-weighting (value-weighting) spread-only portfolio yields a Fama-French 5-factor monthly alpha of 1.69 percent (1.51) percent with a t-stat of 10.35 (6.91). Using other factor models and the Daniel et al. (1997)

characteristic-adjustment procedure generates similar monthly alphas. The size of the monthly alphas seems to be more than enough to offset reasonable transaction costs due to the monthly rebalancing.

****** Insert Table 4 about here ******

4.4. Predictive power of volatility spreads in the presence of varying degrees of investor attention

Our analysis so far validates the existing literature on return predictability by volatility spreads. We now turn to investigate whether and how this predictability changes when we examine this predictability through the lens of investor attention. Such an exercise will not only allow us to ascertain the source of the return predictability, but also enlighten practitioners who may want to design implementable trading strategies.

We start with a standard double-sorting procedure. Each month in our sample, stocks are sorted into ten deciles based on the volatility spreads. Stocks are also independently sorted into ten deciles based on the investor attention measure DRT_{pv}.[8] We calculate the abnormal return to each stock in each following month by subtracting the return of the similar size portfolios from the raw return (AR1), by subtracting the market return from the raw return (AR2), or by subtracting the return on a portfolio of similar size, book to market ratio and momentum from the raw return (AR3). For both the top and bottom investor attention deciles, we average these abnormal returns for each volatility spread decile. Our focus is on the return difference on the top and bottom spread deciles.

Table 5 reports the return gap between the top and bottom volatility spread deciles for both the high-attention and low-attention groups. A few observations are obtained. First, the return gap is statistically and economically significant across the three abnormal return specifications. Take AR3 as an example. The top spread decile stocks register an average abnormal return of 83 (104) basis points across the bottom (top) investor attention deciles. In comparison, the bottom spread decile stocks log an average abnormal return of -70 (-171) basis points across the bottom (top) investor attention deciles. Thus, the return predictability survives the sorting by investor attention. More importantly, we notice that the return gap is more pronounced for the top attention decile than the bottom attention decile. The difference in the return gap across the top

[8] The results are quantitatively the same using LM_{pv} or R_{pv}. These results are available upon request.

and bottom attention decile is at least 1.22 percent per month. This clearly speaks to the greater return predictability in the presence of heightened investor attention and information acquisition. Our interpretation is that as more investors acquire more information, the increased likelihood of more noise trading in the sense of Kyle (1985) and Admati and Pfleiderer (1998) provides better camouflages for informed traders, who initially do not want to trade too aggressively since doing so would reveal their private information too quickly.

****** Insert Table 5 about here ******

4.5. Regression analysis

Similar to Section 4.2, we employ a Fama-MacBeth regression to ensure that the return predictability is robust to the inclusion of established return predictors. Each month in our sample, we sort stocks into ten deciles based on the volatility spreads. Independently, we sort stocks into ten deciles based on the investor attention in that month, DRT_{pv}. We then only retain stocks in the top and bottom deciles of volatility spreads and investor attention. We create two dummy variables: *VsHgh* and *AttnHgh*. *VsHgh* takes the value of 1 if a stock is in the top volatility spread decile and 0 otherwise. *AttnHgh* takes the value of 1 if a stock is in the top investor attention decile and 0 otherwise. Each month, the following regression equation is then estimated:

$$Ret = \beta_0 + \beta_1 \cdot VsHgh + \beta_2 \cdot AttnHgh + \beta_3 \cdot VsAttn + \beta_4 \cdot Size + \beta_5 \cdot Past1y + \beta_6 \cdot Past1m + \beta_7 \cdot BM + \varepsilon$$

Our focal variable is the *VsAttn*, the interaction term between *VsHgh* and *AttnHgh*. Given what we have documented in Table 5, we expect a positive slope coefficient between both *VsHgh* and *VsAttn*. We include the same set of return predictors as control variables: *Size*, *Past1y*, *Past1m*, and *BM*.

Table 6 reports the time series averages of the slope coefficient estimates before each independent variable along with the statistical significance. Our result shows that *VsHgh* still carries a positive slope coefficient estimate that is highly significant. Interestingly, a negative but insignificant slope coefficient estimate shows up before *AttnHgh*. Thus, using investor attention alone does not seem to predict future stock returns. However, the interaction term *VsAttn* has a parameter estimate of 2.83% with a p-value of 0.001. Thus, stocks in the top volatility spread and top attention deciles have much higher subsequent returns than stocks in the top volatility spread but bottom attention deciles.

****** Insert Table 6 about here ******

4.6. Spread-and-Attention portfolios

Given the dramatic patterns in the return difference in top and bottom attention deciles, a natural question to ask is: can professional money managers exploit this differential return pattern and construct profitable trading strategies? To address this question, we augment the spread-only portfolios with the power of investor attention and construct the spread-and-attention portfolios. We show that such spread-and-attention portfolios can generate sizeable abnormal returns.

To form the portfolios, we again conduct the two-way independent sorting on implied volatility spreads and investor attention in each month. The spread-and-high-attention portfolio is limited to the top investor attention decile. It goes long on the stocks in the top volatility spread decile and short on the stocks in the bottom volatility spread decile. This portfolio is rebalanced at the end of each month. We construct both the equal-weighting and value-weighting portfolios in this manner. We further calculate the monthly returns to these portfolios and examine their monthly alphas using standard asset pricing models.

For the purpose of comparison, we also construct the spread-and-low-attention portfolios and examine their monthly alphas. Table 7 presents the monthly abnormal returns to the spread-and-attention portfolios. Our results reveal that the spread-and-high-attention portfolio generates significantly higher abnormal returns. For instance, the Fama-French 5-factor alpha stands at 3.36 (2.92) percent for the equal-weighting (value-weighting) spread-and-high-attention portfolio with a t-stat of 3.95 (3.99). In contrast, the spread-and-low-attention portfolio yields much lower but still significant alphas. The monthly abnormal returns are also economically significant, as they are large enough to cover 1% transaction costs suggested by Novy-Marx and Velikov (2016) for investment strategies with monthly rebalancing of portfolios. Using other asset pricing models as well as the Daniel et al. (1997) risk adjustment leads to similar alphas.

****** Insert Table 7 about here ******

As we point out earlier, the SEC's EDGAR log files currently have a reporting lag of 6 months. Thus, the investor attention measure needed to form stocks into portfolios in each month is not available to portfolio managers. To circumvent this data availability issue, we turn to Li et al. (2018) and propose to use the investor attention measure 6 months ahead. Li et al. (2018) shows that the investor attention measure from the log files displays strong persistence around quarterly earnings announcements. One plausible interpretation is that institutional and sophisticated investors pay attention to certain stocks on a consistent basis. These investors possibly have maintained and established their portfolio positions over a relatively long horizon. Thus, they consistently acquire information about these portfolio stocks.

Table 8 presents the monthly alphas to the spread-and-attention portfolios by using the 6-month-ahead investor attention. We notice that such spread-and-high-attention portfolio still generates significant monthly abnormal returns. This certainly reinforces the predictive power of volatility spreads. In addition, the difference in the monthly alphas between the spread-and-high-attention and spread-and-low attention portfolios well exceeds 1 percent per month. This suggests that using the 6-month-ahead investor attention can still provide significant benefits to professional money managers.

****** Insert Table 8 about here ******

5. Conclusion

In this paper, we make a simple effort to further ascertain the source of the predictive power of implied volatility spreads on subsequent stock returns. Our novel idea of superimposing investor attention and information acquisition on return predictability allows us to gain more insights into the drivers of the return predictability. Our empirical results support the informed trading hypothesis rather than the mispricing hypothesis. Thus, professional money managers can expect to develop sustainable trading strategies to exploit this predictability. Our simple spread-and-attention portfolios generate sizeable and significant monthly abnormal returns. The difference in the monthly alphas between the spread-and-high-attention and spread-and-low-attention portfolios indicates that portfolio managers should focus on stocks with heightened investor attention when designing and implementing such trading strategies.

REFERENCES

Admati Anat, Paul Pfleiderer, 1988, A theory of intraday patterns: volume and price variability, Review of Financial Studies Vol. 1, No. 1, 3 - 40

Amin Kaushik and Charles Lee, 1997, Option trading, price discovery, and earnings news dissemination, Contemporary Accounting Research, Vol. 14, Issue 2, 153-192

Bali Turan, Armen Hovakimian, 2009, Volatility spreads and expected stock returns, Management Science, Vol. 55, No. 11, 1797 - 1812

Bali Turan, Robert Engle, Scott Murray, 2016, Empirical asset pricing: the cross section of stock returns, John Wiley & Sons Inc., Hoboken, New Jersey

Barber, Brad M., and Terrance Odean. 2008. "All that glitters: the effect of attention and news on buying behavior of individual and institutional investors." Review of Financial Studies, vol. 21, no. 2: 785–818

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It depends on where you search: institutional investor attention and underreaction to news, Review of Financial Studies, vol. 30, no. 9: 3009–3047

Black Fischer, 1975, Fact and fantasy in the use of options, Financial Analysts Journal 31, pp36–41, 61–72 (July/August 1975)

Cao Charles, Zhiwu Chen, and John Griffin, 2005, Informational content of option volume prior to takeover, Journal of Business 78, 1073-1109

Chakravarty Sugato, Huseyin Gulen, and Stewart Mayhew, 2004, Informed Trading in Stock and Option Markets, Journal of Finance, Vol. 59, No. 3, 1235-1258

Chemmanur, Thomas J., and An Yan. 2009, Advertising, attention, and stock returns, Working Paper, Boston College and Fordham University

Cremers Martijin, David Weinbaum, 2010, Deviations from put-call parity and stock return predictability, Journal of Financial and Quantitative Analysis, Vol. 45, No. 2, 335 – 367

DellaVigna, Stefano, and Joshua M. Pollet, 2009, Investor inattention and Friday earnings announcements, Journal of Finance, vol. 64, no. 2: 709–749

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011, In search of attention, Journal of Finance, vol. 66, no. 5: 1461–1498

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock. 2012. "Investor information demand: Evidence from Google searches around earnings announcements." *Journal of Accounting Research* vol. 50, no. 4: 1001–1040.

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock. 2015. "The determinants and consequences of information acquisition via EDGAR." *Contemporary Accounting Research* vol. 32, no. 3: 1128–1161

Easley David, Maureen O'Hara, and P.S. Srinivas, 1998, Option volume and stock prices, evidence on where informed traders trade, Journal of Finance, Vol. 53, No. 2, 431-465

Francis, Jennifer, Donald Pagach, and Jens Stephan, 1992, The stock-market response to earnings announcements released during trading versus non-trading periods, Journal of Accounting Research, vol. 30, No. 2, 165–184

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh. 2009, Driven to distraction: extraneous events and underreaction to earnings news, Journal of Finance, vol. 64, no. 5: 2289–2325

Hou, Kewei, Wei Xiong, and Lin Peng. 2009. A tale of two anomalies: the implication of investor attention for price and earnings momentum, Working paper, Ohio State University

Johnson Travis, and Eric So, 2012, The option to stock volume ratio and future returns, Journal of Financial Economics, 106(2): 262-286

Kahneman, Daniel, 1973, Attention and Effort, Englewood Cliffs, NJ: Prentice-Hall

Kumar Alok, Ruenzi Stefan, Ungeheuer Michael, 2018, Daily winners and losers, working paper, University of Miami

Kyle Albert, 1985, Continuous auctions and insider trading, Econometrica, Vol. 53, No. 6, 1315-1335

Li Ruihai, Wang Xuewu, Yan Zhipeng, Zhao Yan, 2018, Sophisticated investor attention and market reaction to earnings announcements: evidence from the SEC's EDGAR log files, Journal of Behavioral Finance forthcoming

Loughran, Tim, and Bill McDonald, 2017, The use of EDGAR filings by investors, *Journal of Behavioral Finance* vol. 18, no. 2: 231–248.

Madsen, Joshua, and Marina Niessner, 2014, Is investor attention for sale? The role of advertising in financial markets, Working Paper, Yale University

Ofek Eli, Matthew Richardson, Robert Whitelaw, 2004, Limited arbitrage and short sales restrictions: evidence from the options markets, Journal of Financial Economics, Vol. 74, No. 2, 305 – 342

Roll Richard, Educardo Schwartz, and Avanidhar Subrahmanyam, 2010, O/S: the relative trading activity in options and stock, Journal of Financial Economics, 96(1), 1-17

Ryans, James, 2018. "Using the EDGAR Log File Data Set." Working Paper, London Business School

Wang, Xuewu, 2017, Investor Attention Strategy, Journal of Behavioral Finance, Vol. 18, no. 4: 390-399

Wang, Xuewu, Zhipeng Yan, Qunzi Zhang, and Xuechen Gao, 2018, Investor attention and stock market under-reaction to earnings announcements: Evidence from the options market, Journal of Futures Markets, Vol. 38, No. 4, 478-492

Xing Yuhang, Xiaoyan Zhang, Rui Zhao, 2010, What does the individual option volatility smirk tell us about future equity returns, Journal of Financial and Quantitative Analysis, Vol. 45, No. 3, 641 – 662

Table 1: Summary Statistics of Main Variables

This table presents the summary statistics of the main variables used in the empirical analysis. We match the IV DB OptionMetrics database with the SEC's EDGAR access log files to construct the sample observations. *Vs* is the option implied volatility spread defined as the call option implied volatility minus the put option implied volatility, where call options and put options are matched on the underlying asset, strike price, and expiration date. A simple average is taken in the presence of multiple matches. BM is the book to market ratio. Size the natural logarithm of the market capitalization. Past1y is the stock return over the past 12 months (skipping the most recent month). Past1m is the stock return over the past month. Three alternative measures of investor attention, DRT_{pv}, LM_{pv}, and R_{pv}, are employed based on the computer robot screening schemes outlined in Drake, Roulstone, and Thornock (2015), Loughran and McDonald (2017), and Ryans (2017), respectively. For each of the measures, the average monthly human downloads is reported.

Variable	No. of Obs	Mean	Std Dev	P5	P25	Median	P75	P95
vs	323,544	-0.006	0.1	-0.109	-0.02	-0.002	0.013	0.08
bm	323,094	1.921	20.69	0.068	0.284	0.505	0.843	2.31
size	323,544	14.244	1.56	11.874	13.152	14.156	15.201	17
past1y	323,544	0.161	0.67	-0.515	-0.144	0.086	0.33	1
past1m	323,544	0.012	0.13	-0.179	-0.052	0.009	0.069	0.2
DRT_{pv}_sum	323,544	812.654	2704.47	92	232	465	882	2444
LM_{pv}_sum	323,544	580.394	2564.57	71	175	331	608	1613
R_{pv}_sum	323,544	664.106	2209.99	84	210	405	741	1893

Table 2: Average Abnormal Returns: Single Sorting by Implied Volatility Spread

This table presents the abnormal stock returns for stocks sorted by the implied volatility spread. Each month t from March 2003 to December 2016, stocks are sorted into ten deciles based on their implied volatility spread as defined in Table 1. Monthly abnormal returns for month t+1 are calculated by subtracting the return of the similar size portfolio from the raw stock return (AR1), by subtracting the market return from the raw stock return (AR2), or by subtracting the return on a portfolio of similar size, book to market ratio, and momentum from the raw stock return (AR3). Each month we calculate the cross sectional averages of these monthly abnormal returns of the ten deciles to obtain average abnormal returns, AR1, AR2, and AR3, respectively. This table reports the time-series average of AR1, AR2, and AR3 for each of the ten deciles. The last row reports the difference in the abnormal returns between the top and bottom deciles. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively.

Decile	**AR1**	**AR2**	**AR3**
Bottom	-0.0069	-0.0055	-0.0081
2	0.0003	0.0013	-0.0008
3	0.0018	0.0024	0.0006
4	0.0003	0.0008	-0.0005
5	0.0024	0.0032	0.0014
6	0.0015	0.0027	0.0002
7	0.0039	0.0054	0.0027
8	0.0035	0.0052	0.0020
9	0.0047	0.0064	0.0032
Top	0.0087	0.0104	0.0070
Top - Bottom	0.0156^{***}	0.0159^{***}	0.0151^{***}

Table 3: Predicting Stock Returns Using Implied Volatility Spread

This table presents the Fama-MacBeth regression results of predicting the stock returns using implied volatility spread as defined in Table 1. Each month from March 2003 to December 2016, stocks are sorted into ten deciles based on the implied volatility spread. Only stocks in the top and bottom deciles are retained in the regression analysis. The following cross-sectional regression is then estimated:

$$Ret = \alpha_0 + \alpha_1 \cdot IsHgh + \alpha_2 \cdot Size + \alpha_3 \cdot Past1y + \alpha_4 \cdot Past1m + \alpha_5 \cdot BM + \varepsilon$$

Ret is the stock return in month t+1. *IsHgh* is a dummy variable that takes the value of 1 if a stock is in the top volatility spread decile and 0 otherwise. *Size* is the natural logarithm of a firm's market capitalization measured in month t. *Past1y* is the stock return in the past year (skipping the most recent month). *Past1m* is the stock return in month t-1. *BM* is the book to market ratio. This table reports the time series average of the slope coefficient estimates before each independent variable and their corresponding P-values. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively.

Variable	**Estimate**	**P-value**
Intercept	0.0104	0.4694
IsHgh	0.0110***	<.0001
Size	-0.0006	0.5028
Past1y	0.0048	0.6737
Past1m	-0.0074	0.3354
BM	0.0005***	0.0052

Table 4: Abnormal Returns to Spread-Only Portfolios

This table presents the abnormal returns to portfolios formed by implied volatility spread sorting. Each month t from March 2003 to December 2016, stocks are sorted into ten deciles by the implied volatility spread. Only stocks in the top and bottom deciles are retained to form the long-short portfolio. This portfolio longs the stocks in the top volatility spread decile and shorts the stocks in the bottom volatility spread decile. The portfolio is rebalanced at the end of each month. This table presents the monthly abnormal returns to the equal-weighting and value-weighting portfolios. T-statistics are shown in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level respectively.

	Equal-weighting	Value-weighting
CAPM Alpha	0.0167***	0.0145***
	(10.66)	(6.99)
Fama - French Alpha	0.0167***	0.0144***
	(10.61)	(6.95)
Carhart Alpha	0.0165***	0.0143***
	(10.69)	(6.89)
5-Factor alpha	0.0169***	0.0151***
	(10.35)	(6.91)
DGTW Char adj.	0.0149***	0.0106***
	(9.73)	(4.97)

Table 5: Average Abnormal Returns: Double Sorting by Volatility Spread and Investor Attention

This table presents the abnormal stock returns for stocks sorted by the implied volatility spread and investor attention. Each month t from March 2003 to December 2016, stocks are sorted into ten deciles based on their implied volatility spread as defined in Table 1. Stocks are also independently sorted into ten deciles based on the investor attention measure as defined in Table 1. Monthly abnormal returns for month t+1 are calculated by subtracting the return of the similar size portfolios from the raw stock return, by subtracting the market return from the raw stock return, or by subtracting the return on a portfolio of similar size, book to market ratio, and momentum from the raw stock return. Cross sectional averages of these monthly abnormal returns are calculated to obtain average abnormal returns, AR1, AR2, and AR3, respectively. This table reports the time-series average of AR1, AR2, and AR3 respectively for the top and bottom deciles of the two sorting variables. The last row reports the difference in the abnormal returns between the top and bottom deciles. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively.

Panel A: Average Abnormal Return AR1		
	Attn Decile	
Spread Decile	Bottom	Top
Bottom	-0.0064	-0.0173
Top	0.0096	0.0122
Difference	0.0160***	0.0295***
Panel B: Average Abnormal Return AR2		
Bottom	-0.0052	-0.0157
Top	0.0106	0.0134
Difference	0.0158***	0.0290***
Panel C: Average Abnormal Return AR3		
Bottom	-0.0070	-0.0171
Top	0.0083	0.0104
Difference	0.0153***	0.0275***

Table 6: Predicting Stock Returns Using Volatility Spread and Investor Attention

This table presents the Fama-MacBeth regression results of predicting the stock returns using implied volatility spread and investor attention as defined in Table 1. Each month from March 2003 to December 2016, stocks are sorted into ten deciles based on the implied volatility spread. Stocks are also independently sorted into ten deciles based on investor attention. Only stocks in the top and bottom deciles are retained in the regression analysis. The following cross-sectional regression is then estimated:

$$Ret = \beta_0 + \beta_1 \cdot VsHgh + \beta_2 \cdot AttnHgh + \beta_3 \cdot VsAttn + \beta_4 \cdot Size + \beta_5 \cdot Past1y + \beta_6 \cdot Past1m + \beta_7 \cdot BM + \varepsilon$$

Ret is the stock return in month t+1. *VsHgh* is a dummy variable that takes the value of 1 if a stock is in the top volatility spread decile and 0 otherwise. *AttnHgh* is a dummy variable that takes the value of 1 if a stock is in the top investor attention decile and 0 otherwise. *VsAttn* is the interaction term between *VsHgh* and *AttnHgh*. *Size* is the natural logarithm of a firm's market capitalization measured in month t. *Past1y* is the stock return in the past year (skipping the most recent month). *Past1m* is the stock return in month t-1. *BM* is the book to market ratio. This table reports the time series average of the slope coefficient estimates before each independent variable and their corresponding P-values. ***, **, and * denote statistical significance at 1%, 5%, and 10% respectively.

Variable	Estimate	P-value
Intercept	0.0248	0.281
VsHgh	0.0083***	0.004
AttnHgh	-0.0081	0.267
VsAttn	0.0283***	0.001
Size	-0.0015	0.349
Past1y	-0.0004	0.916
Past1m	0.0016	0.899
BM	0.0007	0.600

Table 7: Abnormal Returns to Spread-and-Attention Portfolios

This table presents the abnormal returns to portfolios formed by both implied volatility spread and investor attention sorting. Each month t from March 2003 to December 2016, stocks are sorted into ten deciles by the implied volatility spread. Stocks are also independently sorted into ten deciles by investor attention in month t. Only stocks in the top and bottom deciles are retained to form the long-short portfolio. For both the top and bottom investor attention deciles, this portfolio longs the stocks in the top volatility spread decile and shorts the stocks in the bottom volatility spread decile. The portfolio is rebalanced at the end of each month. This table presents the monthly abnormal returns to these spread-and-attention portfolios using both equal weighting and value weighting. T-statistics are shown in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level respectively.

	Equal-weighting		Value-weighting	
	Bottom Attn Decile	Top Attn Decile	Bottom Attn Decile	Top Attn Decile
CAPM Alpha	0.0176***	0.0337***	0.0120***	0.0265***
	(5.99)	(4.12)	(3.73)	(3.74)
Fama - French Alpha	0.0177***	0.0338***	0.0121***	0.0265***
	(5.99)	(4.12)	(3.73)	(3.72)
Carhart Alpha	0.0175***	0.0338***	0.0118***	0.0268***
	(5.94)	(4.21)	(3.66)	(3.75)
5-Factor alpha	0.0170***	0.0336***	0.0108***	0.0292***
	(5.44)	(3.95)	(3.19)	(3.99)
DGTW Char adj.	0.0156***	0.0266***	0.0143***	0.0130
	(3.76)	(3.19)	(3.28)	(1.63)

Table 8: Abnormal Returns to Spread-and-Attention Portfolios: Considering Reporting Lag

This table presents the abnormal returns to portfolios formed by both implied volatility spread and investor attention sorting. Each month t from March 2003 to December 2016, stocks are sorted into ten deciles by the implied volatility spread. Stocks are also independently sorted into ten deciles by investor attention in month t-6. Only stocks in the top and bottom deciles are retained to form the long-short portfolio. For both the top and bottom investor attention deciles, this portfolio longs the stocks in the top volatility spread decile and shorts the stocks in the bottom volatility spread decile. The portfolio is rebalanced at the end of each month. This table presents the monthly abnormal returns to these spread-and-attention portfolios using both equal weighting and value weighting. T-statistics are shown in parentheses. ***, **, and * represent statistical significance at 1%, 5%, and 10% level respectively.

	Equal-weighting		Value-weighting	
	Bottom Attn Decile	Top Attn Decile	Bottom Attn Decile	Top Attn Decile
CAPM Alpha	0.0139***	0.0250***	0.0091***	0.0223***
	(4.83)	(3.19)	(2.75)	(2.82)
Fama - French Alpha	0.0138***	0.0256***	0.0094***	0.0225***
	(4.77)	(3.26)	(2.82)	(2.82)
Carhart Alpha	0.0138***	0.0259***	0.0093***	0.0224***
	(4.77)	(3.29)	(2.79)	(2.80)
5-Factor alpha	0.0140***	0.0247***	0.0094***	0.0243***
	(4.74)	(3.05)	(2.78)	(2.97)
DGTW Char adj.	0.0156***	0.0258***	0.0119***	0.0113
	(4.16)	(3.04)	(3.46)	(1.29)