News Tone and Stock Return in Chinese Market

Huimin Ge Xiaoyan Zhang [*]

June 21, 2022

ABSTRACT

Using daily news tone data between 2017 and 2020, we examine whether news tones can predict stock returns in Chinese A-share market. We first document that the news tones significantly and positively predict the cross-sectional stock returns over next day and over the next 12-weeks. When we separate the news into online news and paper news, the online news exhibit strong predictive power for future returns, while the printed news only displays marginal predictive power. We hypothesize that the online news is more related to firm fundamentals, while the paper news is more linked to political aspects of firm information. Our results using earnings surprises and SOE subsamples provide supportive evidence for the hypothesis.

Keywords: Media, tone, text-analysis, cross-sectional prediction.
JEL Classification: G10, G11, G12, G14.

[*] Ge is from PBC School of Finance, Tsinghua University. Zhang is from PBC School of Finance, Tsinghua University. Contact information: Ge, gehm.17@pbcsf.tsinghua.edu.cn; Zhang, zhangxiaoyan@pbcsf.tsinghua.edu.cn. All remaining errors are ours.

News Tone and Stock Return in Chinese Market

Huimin Ge Xiaoyan Zhang

June 21, 2022

ABSTRACT

Using daily news tone data between 2017 and 2020, we examine whether news tones can predict stock returns in Chinese A-share market. We first document that the news tones significantly and positively predict the cross-sectional stock returns over next day and over the next 12-weeks. When we separate the news into online news and paper news, the online news exhibit strong predictive power for future returns, while the printed news only displays marginal predictive power. We hypothesize that the online news is more related to firm fundamentals, while the paper news is more linked to political aspects of firm information. Our results using earnings surprises and SOE subsamples provide supportive evidence for the hypothesis.

Keywords: Media, tone, text-analysis, cross-sectional prediction.
JEL Classification: G10, G11, G12, G14.

1. Introduction

Since investors' expectations of firm value are the primary determinants of their trading behavior and of asset prices, media news that provide information about firm value or reshape investors' beliefs may affect market activity and drive returns.[1] Meanwhile, market efficiency is directly related to the information environment, and is reflected in the speed of information dissemination. Several studies have investigated the predictability of stock returns in the U.S market based on news tone and find that different predictive pattern of news tone from different source.[2]

So what about the predictive power of news tone to the stock return in Chinese Market? Actually, the language, content and industry structure of financial news in China are quite different with the ones in the U.S. Meanwhile the participants of Chinese stock market also differ with the relatively larger trading proportion of the retail investors, which may cause the different reaction to financial news and different information dissemination. Thus, what is studied in U.S. might be no longer applicable in China. Since it's difficult to process Chinese text, limited evidence of predictive power of news tone is provided on the entire A share market even though the rapid growth of China's capital market impels a great demand and production for corporate financial news.

---

[1] There are active studies on the role of media news in the financial market (e.g., Chan 2003; Vega 2006; Tetlock 2007; Bushee et al. 2010; Griffin et al. 2011; Garcia 2013; Karabulut 2013).

[2] In addition to using the market news context to predict market returns (Tetlock 2007; Bollen et al. 2011; Karabulut 2013), researchers have analyzed firm news from television, the Internet, and newspapers using subsamples of stocks (e.g., 45 large stocks for Antweiler and Frank (2004) and stocks in the S&P 500 index for Tetlock et al. (2008)) on the U.S. stock market.

In this study, we examine the cross-sectional predictability of stock returns in Chinese stock market based on firms' news tones. We use the Financial News Database of Chinese Listed Companies (CFND) to investigate the predictability of stock returns by firms' news tone. The database provides comprehensive coverage of approximately 9.89 million online and 2.99 million printed articles from more than 400 websites and 600 newspapers. A firm's news tone is defined as the difference between the number of positive and negative articles on the firm, standardized by the total number of articles on the firm. The database also provides detailed information on the sources of news articles, such as online or newspaper prints, which enables examination of differences in the current information structures of various media.

Generally, we find that financial news tones positively predict cross-sectional next-day returns in Chinese stock market, with statistical and economic significance. For instance, a one-unit change in news tone leads to a 0.12% (30.49% annualized, with $t = 17.49$) change in daily return. Additionally, we document significant return prediction differences between news originating from the Internet and that in newspapers. While a one-unit change in online news tone leads to 0.13% change in daily return (with $t = 17.58$), the same change in newspaper news tone leads to a 0.02% change in daily return (with $t = 2.22$). That is, the online news tone can significantly predict the next-day return, while that of print news articles shows marginal predictive power for returns.

What information is contained in the news? To establish whether the news tone mainly captures fundamental information or merely investor sentiment, we first examine the long-horizon return predictability by news tone to determine whether there is a quick reversal pattern. If there is long term predictive power and no quick reversal, the information contained in news tone might be more related to fundamentals rather than temporary sentiments. The results show that the predictive power of the total news tone can last for up to twelve weeks. In addition, the persistence is significantly longer for small stocks and firms with low liquidity. We find that while online news predicts the long-horizon stock return and alpha, paper news shows marginal predictive power for returns on a shorter scale and no prediction for alpha.

Additionally, we directly link news to fundamental information. Previous literature mostly uses the standardized unexpected earnings (SUE) and cumulative abnormal returns (CARs) to capture the new information in fundamentals and market reaction to it around the earnings announcement day. We find that the total news tone prior to an earnings announcement day positively predicts both the earnings surprise and CAR, providing direct evidence that the news tone has information content regarding firm fundamentals. In addition, we find that the online news tone positively predicts the SUE and CAR to a lesser extent than the total news tone, and the paper news tone shows no predictive power for these proxies for fundamental information. It is interesting to find differences in information contained in online and paper news tones, and the results suggest to use total news tones to obtain the highest predictive power.

Given the differences between online and newspaper news tones' predictive power for future returns, we further investigate what factors lead to the differences. Previous studies show that online and paper news are distinct in writing style (Nielsen 2008), information content and coverage range, timeliness, readership, and perception of credibility (Flanagin and Metzger 2000). For instances, in terms of coverage, the media report more articles online without layout restrictions ; in terms of timeliness, the media report more timely online compared with the printed version; in terms of readership, the newspaper use more formal language and the news in it may be more difficulty to understand; and in terms of credibility, since the news in the newspaper are revised and proofread with many times, the printed news earn more credit.

Given the history of China's news industry, we hypothesize that the media generally report more fundamentals-related corporate news online, while newspapers, which typically serve as the government's mouthpiece, focus more on content such as regulatory penalties, social responsibility, and certain firm types, such state-owned enterprise (SOEs). Our earlier finding that the online news tone, but not paper news tone, can predict the SUE and CAR in all firms, is consistent with the hypothesis that online news contains more fundamental related information than paper news. To further disentangle the information content of different news media, we separate our samples into SOEs and non-SOEs with the hypothesis that the paper news contains more information for SOEs. Our findings reveal that the paper news tone significantly predicts SOEs' cross-sectional returns and the horizon can last for up to six weeks, yet it has no predictive power for non-SOEs' returns. In contrast, the online news tone can predict both SOEs' and non-SOEs' cross-sectional returns, the prediction horizon is longer for non-SOEs. This pattern is quite intuitive: Because most newspapers in China serve as the government's mouthpiece, they report more on information related to SOEs. In contrast, the Internet media, which are mostly market-oriented, focus more on non-SOEs.

Our study connects to several branches of the literature. First, our study is related to the literature that link news tones and future stock returns. For instance, Tetlock (2007), Bollen et al.

(2011) and Karabulut (2013) examines market level news, while Busse and Green (2002), Antweiler and Frank (2004), Tetlock et al. (2008), and Engelberg (2008) examine firm level news. Second, with heavy use of data from textual analysis, our paper is also linked to the text analysis literature (e.g., Antweiler and Frank 2005, Li 2010, Jegadeesh and Wu 2013, Manela and Moreira 2017). Third, given our focus on Chinese A-share stock returns and Chinese media news, our study is also connected with other papers on financial media in China, such as You et al.'s (2017) and Li et al. (2019).

Compared with previous studies, our study makes two significant contributions. First, we examine the cross-sectional prediction of stock returns by financial media news in Chinese market. Previous studies mostly focus on the U.S market or a subset of the Chinese market. We are among the first to establish a significant predictive relationship between news and returns on the entire A share market. Second, we examine the information content of news and differences between media types. We find that news may capture more fundamental-related information; compared with paper news that focuses more on SOEs, online news contains more fundamentals.

The paper is organized as follows: Section 2 introduces the data. Section 3 provides the main empirical results. Section 4 concludes.

2. Data and Descriptive Statistics

2.1 News data

Our news tone data are obtained from the Financial News Database of Chinese Listed Companies (CFND) in the Chinese Research Data Service Platform (CNRDS). The database covers news from more than 400 websites and 600 newspapers.[3] CFND uses a support vector machine (SVM) to measure news tones. The main approaches to tone extraction are the dictionary-based method and supervised machine learning. Because there is no standard Chinese financial dictionary, CFND use an SVM to extract the tone information. After the standard text data cleaning process, they randomly select news articles published before 2016 and manually label their tones. They then train the SVM and their classification model to automatically predict the tones of the articles. The precision and recall rate of the testing suggest that the approach used is quite reasonable. Since the manually labelled articles are randomly selected from those published before 2016, to avoid the looking-forward bias, we restrict our sample to the period January 2017 to December 2020, which yields 4,219,591 online and 1,229,996 in-print articles with unique news ids. For each news article with unique news id, we obtain the news title, report time (date for paper news and exact timestamp for online news), the number of each sentence, and news tone.

We aggregate the article-level news tones into firm-level news tone at daily frequency. The choice of the daily frequency is due to the full coverage of the news data and the timeliness of the information dissemination. Moreover, it has been verified that the Chinese stock market turnover is quite high. For each firm-day, we measure the firm-level tone as the difference between the number of positive and negative news, scaled by the total number of news articles, as follows:

---

[3] According to iResearch statistics in 2019, there were 30 financial online websites whose monthly number of users covered exceeded one million, and 20 of these are in this dataset. Moreover, among the 30 financial online websites with the highest rank scores on China Webmaster, 20 of them are in this dataset. Except the major financial websites, the database includes more than 400 other major, industry, or local websites, to include as much information as possible. The database contains 43 professional, mainstream financial newspapers with the most influence and that are most commonly used for research (e.g., You et al. 2017), as well as more than 500 other important newspapers and periodicals, including mainly important central, local morning, daily, evening, and other financial newspapers.

$$Tone_{i,t} = \frac{PosnewsNum_{i,t} - NegnewsNum_{i,t}}{PosnewsNum_{i,t} + NegnewsNum_{i,t} + NeunewsNum_{i,t}} \tag{1}$$

where $PosnewsNum_{i,t}$, $NegnewsNum_{i,t}$, and $NeunewsNum_{i,t}$ are the numbers of the positive, negative, and neutral news for firm i on day t. Based on the news sources, we further separate online news and printed news. That is, we calculate the firm-level media tones for the total (using all news), online, and paper news.

2.2 Return data and corresponding timelines

We obtain the return data including individual open and close stock prices from Wind Database. Since we establish the exact time stamp for each article for online news and report date for paper news, to ensure data availability for return prediction, we consider the close-to-close timeline in our main analysis. We regard the online news after the market closing time (15:00) on day t-1 and the news before the market closing time on day t as the news on day t. The report date for paper news remains unchanged. We then use the aggregated firm-level news tone to predict the close-to-close return. The close-to-close timeline for news reports and stock returns is displayed in Panel A of figure 1. The close-to-close return on day t+1 is calculated as follows:

$$R_{i,t+1}^{c} = \frac{P_{i,t+1}^{c} - P_{i,t}^{c}}{P_{i,t}^{c}} \tag{2}$$

For non-trading days, we aggregate the news tone on non-trading days with that on the first following trading day to predict the close to close return for the second following trading day. For the firms with no news, the tones, by definition, are assigned a value of zero to denote a neutral attitude of the media. To test whether the results are robust to different timelines or definition of neutral attitude, we additionally conduct our main analysis using an open-to-open timeline and no zero filling in the robustness check.

[Figure 1 about here]

2.3 Other financial data

We obtain other trading and financial statement information from Wind, and include in our analysis a set of financial variables that have been shown to have predictive power for asset returns. We use the product of the daily close price and the number of outstanding total A shares as the market capitalization to measure firm size. We calculate the earnings-price ratio as the ratio of a firm's net profit excluding non-operating revenue to the market capitalization in the latest financial report period and use it as a proxy for the firm's value. Liu et al. (2019) and Hou et al. (2020) show that the earnings-price ratio is a better proxy for firm value than the book-to-market ratio in Chinese market. We measure the market's trading activity using the turnover, which is measured as the average daily turnover in the previous one month. We include idiosyncratic volatility, which is calculated as the standard deviation of the residuals from the regression of stock returns on three Chinese factors, obtained from Liu, Stambaugh, and Yuan (2019), within the previous one month. All these variables vary over time for each stock at the daily level. For the time-series analysis, we use the value-weighed return for the entire A share market as the market return. Additionally, we use the civix published by Wind, calculated similarly to how CBOE calculates the VIX, as the market volatility index.

2.4 Summary statistics

Table 1 shows the summary statistics for the variables in the study. Panel A shows the media coverage in our sample period. There are 5,449,587 news articles with unique news ids during our sample period: 4,219,591 online and 1,229,996 in print. Note that news articles with different ids may have the same content since the reprint news is assigned with different id compared to the original one. The total and online news reports contain information on 3,947 firms, while the paper news reports cover 3,928 firms. Considering the number of listed firms on the A share market, most firms are covered by the financial media. While the average number of news reports per firm per year is 377, 292, and 88 for the total, online, and paper news, respectively, there are some "no news" days for some firms with less news coverage. Of the news reports, more than 40% are classified as positive news for the different media sources (42.8% for total, 41.6% for online, and 47.0% for paper news). This shows that the media tend to post positive information to investors. Compared with online news reports (with 26.9% neutral news and 31.5% negative news), paper news reports show a more positive attitude, with less negative news (16.9%) and more positive and neutral news (47.0% and 36.2%, respectively).

Panel B of Table 1 shows a description of the tones from the different media types and the correlation between them. By definition, the minimum and maximum of the key measures are -1 and 1, respectively. The average tone of the all firm-day observations for the total news data is 0.061, which shows that the media tend to provide investors with more positive than negative news. The average tone, standard deviation, and zero ratio for online news are 0.038, 0.374 and 0.791, respectively, while the respective figures for paper news are 0.042, 0.252, and 0.911. Comparing the news tones of the online and printed news, the latter's news tones have a larger mean, a smaller standard deviation, and a larger zero ratio, owing to the limited coverage resulting from the layout restrictions.

The correlations between the total and online news tones and between the total and printed news tones are 0.885 and 0.428, respectively. The correlation between the online and newspaper news tones is relatively small, with a value of 0.088. These observations indicate that the online and printed news tones are different, with little overlap, whereas the total news tone contains both the online and printed news tones in different proportions.

3. Empirical Results

In this section, we explore the predictive power of news tone for returns and its source. We first evaluate the return predictability by news tone based on all the information and different channels. To examine whether the information captured by the media is a reflection of investors' sentiment or firms' fundamentals, we assess the return predictability over a long horizon. Additionally, we directly relate the news content to the firms' fundamentals. Based on the difference in predictive power between online and paper news shown in the results, we check the driving force behind this difference.

3.1 Fama-MacBeth regression for daily predictions

Can the news tones provide helpful information about future returns? We use the Fama-MacBeth regressions to examine the predictive power of the news tones as follows.

$$Ret_{i,t+1} = b_{0,t+1} + b'_{1,t+1} Tone_{i,t} + b'_{2,t+1} Control_{i,t} + \epsilon_{i,t+1} \quad (3)$$

where the dependent variable $Ret_{i,t+1}$ is the close-to-close daily return for stock i on day $t+1$. The main independent variable is news tone of stock i on day t, $Tone_{i,t}$. From the first stage estimation, we obtain a time-series of the cross-sectional coefficients $b_{1,t+1}$. In the second stage, we compute the means b_1 and standard errors, and conduct inference using the time-series of these coefficients. Here, we use the news tone on day t to predict the next day's return. Thus, the standard errors of the time-series are adjusted using Newey–West (1987) with one lag. If news tone can predict future returns, we expect coefficient b_1 to be significantly positive. We first use only news tone for the prediction, and then control for the common variables that are predictive of future returns. We include the return on day t as a control variable. In addition, we include the natural logarithm of market capitalization, earnings-price ratio, turnover, and idiosyncratic volatility on day t, calculated from the previous month's trading data. The results are shown in Table 2.

Without the control variables, the coefficient on total tone is 0.12, with a t-statistic of 17.49. The positive and significant coefficient suggests that a news tone increase from 0 (neutral tone) to 1 (positive tone) leads to an increase in the stock return on the next day of 0.12%, on average. The spread between firms with total positive (1) and total negative (-1) tones would be 0.24%. When we add the control variables, the coefficient on total tone remains statistically and economically significant (0.07, t = 14.76). This shows that while the total news tone captures some information that is related to these well-known predictors, it contains unique information that could be used to predict future returns.

To compare the predictive powers of the online and paper news tones, we incorporate both in equation (3) and perform the Fama-MacBeth regression. The results show that the coefficients of the online and paper news tones are 0.13 (t = 17.58) and 0.02 (t = 2.22), respectively. This indicates that the spread between firms with total positive (1) and those with total negative (-1) online and paper news tones are 0.26% and 0.04.%, respectively. We find that the online news tones have predictive power for returns, whereas the paper news tones provide only marginal information. Moreover, when we add the common control variables, the coefficients of online and paper news tones become 0.07 (t = 14.44) and 0.01 (t = 1.45), respectively. This shows that when we control for the common return predictors, the online news tone retains unique information about future returns, whereas the information contained in the paper news tone is absorbed.

For the control variables, most of the coefficient signs are within expectation. While the earnings-price ratio positively predicts future returns, high turnover and idiosyncratic volatility show low future returns. However, the previous one-day returns significantly positively predict the next-day returns, which shows the one-day momentum in the Chinese market, the previous one-week returns significantly negatively predict the next-day returns and show a weekly reversal pattern. For firm size and the previous one-month return, the coefficients are not significant.

[Table 2 about here]

3.2 Portfolio sorting

We use portfolio sorting to examine the predictive power of news tone for cross-sectional stock returns in this section. We first sort the returns on day t+1 into tertiles based on the total news tones on day t, using a close-to-close timeline. We calculate the total news tone with all online and paper news following equation (1), and assign zeros if there is no news for each firm-day. Since there are many zeros in our sample, the firms are divided into three subgroups with negative, neutral, and positive tones, rather than based on the numerical values of the news tones. We then calculate both the value- and equal-weighted portfolio returns and alpha; alpha is calculated as the intercept of the regression of excess return in Chinese FF3 factors obtained from Liu, Stambaugh, and Yuan (2019). The results are shown in Panel A of Table 3.

There are 284 firms and 561 firms in the portfolios with negative and positive news, respectively, while approximately 2659 firms are not covered for each day, on average. We find that firms with positive news today tend to have higher next-day returns and alphas. The value-weighted, long-short portfolio sorted based on the total news tones has a 6.8-bp daily return spread (which is approximately 17.1% per year). The equal-weighted portfolio has the larger spread (21.3 bps). Meanwhile, the long-short portfolio return is almost the entire alpha (5.9 bps and 20.4 bps for the value- and equal-weighted portfolios, respectively) in all cases; since the portfolio is rebalanced daily, the strategy of sorting stock by news tone has little exposure to the three factors that represent long-term risk.

From unreported results, the average characteristic of the portfolio is that firms with neutral tones tend to have smaller capitalization and slightly larger turnovers. Intuitively, small firms tend to have less news coverage; additionally, by our definition, firms without news coverage on day t are assigned neutral tones. Because firms with positive tones tend to be larger than those with negative tones, they are more exposed to good news. Comparing the earnings-price ratios, we find that firms with more positive tones have higher earnings-price ratios, which partially reflects that news tone is related to firm performance. The idiosyncratic volatilities of the stocks are similar among the different groups.

Because we can clearly identify the media sources, we separately obtain the news tones from all the online and paper news. We the evaluate the predictive power of the news from the different media types for stock returns. The results are shown in Panels B and C of Table 2. For the online news, there are 282 firms and 450 firms in the portfolios with negative and positive news, respectively, while 2771 firms are not covered for each day, on average, which is on the same scale as that of the portfolio that was sorted by total news tone, since the zero ratios for the total and online news tones are similar. The value-weighted, long-short portfolio sorted by online news tone has a 0.05% daily return spread (14.62% per year, t =3.97) and a 0.06% alpha (t = 3.77). The equal-weighted, long-short portfolio sorted by online news tone has a larger daily return spread (0.23%, t = 21.17) and the same alpha with the same significance. For the paper news, the average number of firms in the portfolio with a neutral tone is slightly larger owing to the less coverage of paper news. The value- and equal-weighted return spreads for the long-short portfolio are 0.02% (t = 0.69) and 0.06% (t = 4.18), respectively.

The results show that the online news tones have predictive power for returns, while the paper news tones have limited predictive power for returns, with nonsignificant value-weighted and marginally significant equal-weighted portfolio results. Compared with those sorted by total news tone, both the magnitude and significance sorted by online news tone are smaller, which shows that the paper news could add marginal information about future returns to the online news. All the above results are consistent with the Fama-MacBeth regression in scale. Furthermore, when we analyze the characteristic pattern of the portfolio sorted by online and paper news tones, we find that while firms in the portfolio with neutral online and paper news tones remain the smallest in size owing to coverage, firms in the portfolio with positive paper news tones have lower turnover and idiosyncratic volatility.

[Table 3 about here]

We also examine whether the predictive power of the news tone is restricted or more prone to a specific firm type. We first sort the firms according to their characteristics on day t calculated from the previous one month data, and then sort them according to the news tones from the different sources and calculate the next-day return and alpha. The results are shown in Table 4. We first sort all stocks into tertiles based on market capitalization on day t and obtain small, medium, and large groups. Both the value-weighted returns and alphas for the long-short portfolios sorted by total and online news tones are larger in the small group (0.31%, t = 16.24; 0.31%, t = 15.96 for total news, and 0.31%, t = 15.09; 0.31%, t = 15.09 for the online news) than those in the large group (0.05%, t = 3.08; 0.04%, t = 2.75 for the total news and 0.04%, t = 2.47; 0.04%, t = 2.47 for the online news). For the portfolio sorted by paper news tone in the different size groups, the value-weighted return spreads and alphas are nonsignificant among large group and marginally significant among small and medium group.

We then categorize all the stocks into tertiles based on the Amihud (2003). The Amihud is the ratio of average absolute return to the trading volume in the previous one month. The value-weighted return spreads and alphas for the long-short portfolios sorted by total and online news tones are larger in the illiquidity groups (0.30%, t =12.61; 0.30%, t = 12.49 for total news and 0.34%, t =12.51; 0.34%, t = 12.45 for the online news) than in the liquidity groups (0.05%, t = 2.87; 0.04%, t = 2.54 for the total news and 0.04%, t = 2.25; 0.03%, t = 2.16 for the online news).

For the portfolios sorted by paper news tone in the different liquidity groups, the value-weighted return spreads and alphas are all nonsignificant.

All the results are intuitive: Small stock prices with less liquidity are more prone to impact from smaller changes in trading behavior and are thus slower in incorporating information. This explains why more significant results with larger magnitudes are obtained for the equal-weighted portfolios than for the value-weighted portfolios.

[Table 4 about here]

3.3 Long horizon predictions

Does the information captured by the media reflect investors' sentiment or firms' fundamentals? We first examine this question through long horizon prediction analysis. If the prediction direction reverses in the long run, then the news itself could capture more investor sentiment than fundamental information. In this subsection, we first extend the cross section prediction to a long horizon one. We then examine the cumulative CH-3 alpha for the long-short portfolio over the long horizon.

We use the news tone on day t to predict the k-week ahead return, as in the following formula:

$$Ret_{i,t+1+(k-1)w:t+kw} = c_{0,k} + c_{1,k}' Tone_{i,t} + c_{2,k}' Control_{i,t} + \epsilon_{i,t+1+(k-1)w:t+kw,} \quad (4)$$

where $Ret_{i,t+1+(k-1)w:t+kw}$ represents the k-week ahead weekly return rather than the cumulative return from day t. We assign 5 days to w, which is the number of trading days in a week. To account for serial correlation in the coefficients, the standard deviations of the time-series are adjusted using Newey-West (1987) with the corresponding lags. The control variables are the same as those in equation (3). We separately analyze the long horizon prediction for the total, online, and paper news tones. Panel A of Table 5 shows the results for the long-term prediction up to 12 weeks. If the news merely captured investor sentiment, we would observe the coefficient, $c_{1,k}$, decrease to zero within a short horizon and a reversal pattern in the long run under the assumption that the market would eventually discover the firms' fundamental values. We observe that the coefficient of news tone decreases from 0.104 (t = 8.164) to 0.052 (t = 4.538), respectively, for the weekly prediction. The return prediction for total news tone decreases with time but remains in the same direction; additionally, the prediction remains significant for up to the twelve weeks ahead. This shows that the media news generally has long term predictive power and captures more fundamental-related information. To save space, we do not report the coefficients of the control variables; however, their signs are within expectation, while the past return negatively predicts the 4-week ahead return, which shows the monthly reversal.

Comparing the online with the paper news, the coefficients for the online news tone are significant for up to twelve weeks (from 0.105, t = 8.208, to 0.055, t = 4.782), while the coefficients for the paper news tone decrease with time and become nonsignificant after ten weeks (from 0.060, t = 3.409, to 0.034, t = 2.029). Combined with the next-day prediction results, this shows that whereas the information contained in the online news is quickly absorbed by the market and may be more fundamental-related since it predicts the long-horizon future return, it takes one week for the paper news to be incorporated into the prices; the information effect then diminishes within a relatively short period (about half a month). Furthermore, for each prediction horizon, the magnitude and significance of the coefficient of the online news tone are larger than those of the paper news coefficient. We can thus infer that the online news contains more fundamental-related information than the paper news.

Furthermore, we examine the long-short portfolio's cumulative alpha in the long run to determine whether the public news tone can serve as a signal for a profitable strategy. We sort the stocks into tertiles based on the news tone and then hold the portfolios for up to 12 weeks. If the news truly reflected information about firms' fundamental values, then firms with more positive news tones would outperform those with negative news tones in the long run. Panel B of Table 5 shows the results for the long horizon long-short portfolio sorted by news tone on day t. We report the value-weighted, risk-adjusted return using the Chinese FF3 factors obtained from Liu, Stambaugh, and Yuan (2019). Given the overlapping data, we adjust the standard errors of the portfolio return time-series using Newey–West (1987), with the corresponding number of lags.

Over a one-week horizon, the value-weighted long-short portfolio alpha sorted based on total news tone increases from 0.07% (t = 4.36) to 0.10% (t = 2.49). When we increase the holding horizon, the alpha becomes nonsignificant after two weeks (0.14%, t = 2.17 for the two-week long-short portfolio) with the Newey-West adjustment. Compared with the k-week ahead Fama-MacBeth regression results, we find that for over one month, the cumulative return for the long-short portfolio would have much more exposure to the three factors that represent long-term risk. Thus, although the total news tone could predict the longer-horizon return, we still could not obtain the long-term, risk-adjusted return.

For the online news, the alpha for the long-short portfolio would become nonsignificant for up to one week (0.06, t = 1.67), which shows a shorter pattern than the one for the portfolio sorted by total news tone. The alpha for the long-short portfolio sorted by paper news tone remains nonsignificant irrespective of the duration of the holding period. Thus, after the risk adjustment, we could still obtain alpha from the online news, but no alpha from the paper news. All these observations corroborate the postulate that online news contains more fundamental-related information.

[Table 5 about here]

3.4 News and fundamentals

To directly relate news content to firm fundamentals, we examine the predictive power of news for earnings surprise. We first consider the predictive power of news tone for the SUE. The SUE is a commonly used measure to quantify surprise in the marketplace. The SUE score measures the deviation of the announced earnings from the previous mean. We use the average earnings over the previous eight quarters as the expected earnings, standardized by the standard deviation of the earnings over the same period. We then use the previous day's news information to predict the SUE. We perform the panel regression as follows:

$$SUE_{i,t} = \alpha_i + \mu_t + d_1 Tone_{i,t-1} + d_2{}' Controls_{i,t-1} + \epsilon_{i,t}, \tag{5}$$

where $Tone_{i,t-1}$ is the news tone one day before the earnings announcement day. We use the return on day t-1, previous one-week return, previous one-month return, size, and earnings-price ratio on day t-1 as the control variables. We first use the total tone to predict the SUE. If the news tone truly captured the fundamental information, we would expect the coefficient, d_1, to be significant and positive. The results are shown in Table 6. The coefficient of the total news tone is

0.61 (t = 13.04). We then use the online and printed news to predict the SUE. The coefficients of the online and printed news tones are 0.17 (t = 2.68) and 0.12 (t = 1.39), respectively.

We generally find that while the previous one-day news tones can significantly predict the SUE, the previous one-day printed news tones have no predictive power for the SUE. For the control variables, the previous one-day, one-week, and one-month returns positively predict the SUE. The coefficients of size and earnings-price ratio are positive and negative, respectively, which is as per expectation.

We also examine the predictive power of the news tone for the market reaction to earning announcements. We calculate the CAR using the earnings announcement day as the event day and then use the previous news information to predict the CARs. We use the [0,1] interval as the event window. Both the market and the three-factor models are used to estimate the abnormal return, and the results are robust to the CARs with different estimation models. In the main text, we assess the results from the three-factor model by using the [-280,-30] interval as the estimation window to obtain new coefficients for the three-factor model and the abnormal return.

The coefficients (and t-statistics) of the total, online, and printed news tones are, respectively, 0.30 (t = 8.73), 0.12 (t = 2.44), and 0. 07 (t = 1.12). Generally, the previous total and online news tones prior to the announcement day positively predict the CAR. The paper news tone has no predictive power for the CARs. These results support the claim that the media tend to report fundamental-related information online.

3.5 Online news vs. paper news

Our empirical results show that news can predict returns and contain fundamental information about firms. Moreover, we find that there are significant differences between online and newspaper news. What, then, is different between online and newspaper news? Given the history of China's news industry, we hypothesize that the media generally report more fundamental-related corporate news online, while the newspapers, which typically serve as the government's mouthpiece, focus more on content such as regulatory penalties, social responsibility, and certain types of firms, such SOEs. In the previous section, we found that while the online news tone could predict the SUE and CAR, the paper news tone could not. Thus, we establish the relative extents to which online and paper news focus on non-SOEs and SOEs.

We divide the companies into two subgroups according to the nature of their equity and use both the online and paper news tones to conduct return prediction over several periods, as in equation (4). The results are shown in Panel A of Table 7. We find that the online news tones positively predict SOEs' next-day returns (with coefficient 0.05, t = 5.37). For the long holding horizon, the online news tone predicts SOEs' future returns for up to 12 weeks. For the non-SOE subsample, the online news tone shows predictive power for both short and long horizons. The coefficient of the online news tone in predicting non-SOEs' next-day return is 0.08 (t = 9.03), which is on a larger scale than that for the SOE subsample. Additionally, the predictive power of the online news tone for non-SOEs' returns lasts for up to six weeks. This shows that while the online news tone has predictive power for both SOEs' and non-SOEs' returns, the predictive power for non-SOEs' returns is larger and persistent.

For the paper news tone, we find that it shows no predictive power for the total sample. However, when we split the sample into SOEs and non-SOEs, we find that the predictive power of the paper news tone for SOEs is significant and persistent. The coefficient of the paper news tone in predicting SOEs' next-day returns is 0.03 (t = 2.69), and decreases from 0.07 (t = 2.70) to 0.05 (t = 2.20) in predicting SOEs' k-week ahead return. For non-SOEs, the paper news tone shows no predictive power for both the short and long horizons.

The results are intuitive. Since the online media are mostly market-driven, they report fundamental information more for non-SOEs; however, online news still contains important information about SOEs. Newspapers, which are mostly controlled by the central or local state, serve as the government's mouthpiece, and focus more on SOEs. Thus, the paper news tone can predict SOEs' future returns rather than those for non-SOEs and the whole sample.

[Table 6 about here]

To investigate the difference in content between online and printed news, we decompose the online news tone into the two parts: one is related to the paper news tone and the orthogonal part. The detailed process is as follows:

$$Tone\_online_{i,t} = b_{0,t} + b_{1,t} Tone\_paper_{i,t} + \epsilon_{i,t} \tag{5}$$

We first perform the above regression to obtain the time-variant coefficient; we then calculate the following terms:

$$\widehat{Tone\_overlap_{\imath,t}} = \widehat{b_{1,t}}\, Tone\_paper_{i,t}$$

$$\widehat{Tone\_orthgonal_{\imath,t}} = \widehat{b_{0,t}} + \widehat{\epsilon_{\imath,t}} \tag{6}$$

The first part denotes the one related to the printed news in the online news and the second part denotes the orthogonal one. We then use the separate parts to sort the stocks and construct a value-weighted portfolio. The results are shown in Panel B of Table 7. We generally find that the value-weighted return and alpha spread for the long-short portfolio sorted by the part correlated to the printed news are nonsignificant (with mean 0.038 ($t = 1.34$) and alpha 0.013 ($t = 0.499$), while the return spread and alpha sorted by the orthogonal part are significant and of economic magnitude. The means are 0.051 ($t = 3.147$) and 0.047 ($t = 2.849$), and are similar to those for the portfolio sorted by pure online news tone. This supports the claim that the difference in predictive power between the online and paper news tones arises from the difference in the contents themselves.

[Table 7 about here]

4. Further Discussion

News tone can predict future stock returns. This predictive ability varies in different channels. In this section, we discuss several related issues to put news tone's predictive power into perspective. We first discuss whether news tone can predict the market's overall movement. We then consider the impact of stale news, which is an essential part of information in financial markets. Furthermore, we conduct robustness tests to ensure that our results do not depend on the timeliness of reports and calculation approach for news tone.

4.1. Predict market aggregate conditions

To examine whether the aggregate news tone at the market level can be regarded as a state variable to predict market return and volatility, we use market capitalization as the weight to aggregate firms' news tones into a time series index and then use it to predict the next-day market return and volatility. We find that both the online and printed news tones cannot predict the aggregate market return (-0.002, t =-0.230 and -0.012, t = -1.555), respectively, while only the total and online news tones can positively predict the market volatility (2.427, t = 2.799 and 2.296, t = 2.592, respectively) in Panel A of Table 8.

4.2 Stale news

Tetlock (2011) shows that investors overreact to stale information. Here, we examine whether the staleness of news affects the predictive power of news tone. The CFND provides a similarity measure for any two articles using the cosine distance of the vector space representation of the two articles. The articles with the earliest report time are assigned the original label, while for those that exhibit a high similarity with other articles but report in a later time, their original dummies are assigned zero. For each firm on day t, we calculate the ratio of the number of non-original news to the total number of news as a measure of the staleness of information. We then perform the Fama-MacBeth regression over different horizons, including the staleness and interaction term of news tone and staleness, as in the following equation:

$$Ret_{i,t+1+(k-1)w:t+kw} = e + e_{1,k}Tone_{i,t} + e_{2,k}Tone * Staleness + e_{3,k}Staleness$$
$$+c'_{4,k}Control_{i,t} + \epsilon_{i,t+1+(k-1)w:t+kw} \quad (7)$$

If investors overreact to old news, then the coefficient, $e_{3,k}$, of the staleness in predicting the k-week ahead returns should be negative; if the overreaction influences the predictive power of news tone, then the coefficient, $e_{4,k}$, of the interaction term should be negative. Panel B of Table 8 shows the results. For both the online and printed news tone, we find no significant influence of investors' overreaction.

4.3 Robustness: open-to-open prices

In the main research, we use the close-to-close timeline to perform the return predictions. As a robustness test, we use the open-to-open timeline instead. For online news, we have an exact timestamp. Before the market opens (9:00 am), the information reported in the morning (we use 8:00 am as the cutoff) is available to investors, who can trade on day t+1 based on the news reported on day t and day t+1 before 8 am, using the open price. Thus, we use the news reported between 8 am on day t and 8 am on day t+1 to predict the return on day t+1. The detailed timeline is shown in Panel B of figure 1.

For the newspaper data, if we assume that the news with a report time on day t is already accessible to investors on the morning of day t, then we should regard the news reported on day t as the day t-1 news. We use the news on day t to predict the open-to-open return on day t+1, which is calculated as follows:

$$R_{i,t+1}^{o} = \frac{P_{i,t+2}^{o} - P_{i,t+1}^{o}}{P_{i,t+1}^{o}} \tag{8}$$

The results are shown in Panel C of Table 8. We find that changing the timeline does not completely change our basic results. We still find the predictive power of the total news tone (0.06%, t = 2.97 for the value-weighted portfolio); additionally, while the online news (0.04%, t = 2.31 for the value-weighed portfolio) exhibits predictive power, the printed news does not (0.02%, t = 0.74 for the value-weighted portfolio). Compared with the previous results, the predictive power using the open-to-open timeline is weaker than that using the close-to-close timeline, which implies that investors focus more on news during market opening.

4.4. Robustness: no zero fillings

In the main analysis, we follow the literature[4] and use zeros for "no news" firm-days, assuming that "no report" means a neutral attitude from the media. Here, we discard the zeros and drop the "no news" firm-day observations from the analysis. The results are shown in Panel D of Table 8. We still sort the next-day returns by total news into three tertiles, with approximately 300 firms in each tertile. We then find that the return spread and alpha for the long-short portfolio remain economically and statistically significant, and on a similar scale to that of the main results (11.7 bps and 15.2 bps, respectively, for the value-weighted results, and 31.8 bps and 33.6 bps, respectively, for the equal-weighted results). The only difference lies in the firm characteristics. Because small firms tend to have no news coverage, in the main analysis, the portfolio with a neutral tone has the smallest average market capitalization; however, if we dropped the "no news" day observation, larger firms would tend to have neutral news reports while smaller firms, which are more regarded as the lottery, would have more positive tones.

5. Conclusion

In this study, we first examine the cross section of total news tone for stock returns in Chinese market. While many studies focus on subsamples on the U.S. market, we provide a Chinese market perspective with a high sensitivity for sentiment and turnover. The results show that news tone positively predicts the next-day cross-sectional returns. The return spread for the value-weighted long-short portfolio sorted by total news tone, on average, is 12 bps, which is approximately 30% per year. For the long horizon, the predictive power of news tone can last for up to six weeks. This

[4] Examples include Tetlock (2008), Hendershott et al. (2015), and Ke et al. (2019).

shows that news can capture fundamental-related information. We then directly link earnings surprise with the tones and find that the total news tone can positively predict the SUE and CAR.

Additionally, while the role of the media in financial markets is well-researched, studies that compare news from different channels and the corresponding market reactions are limited. The paper presses have been devoted to coping with the shocks from the emerging development of the Internet. However, no studies have examined whether the market reacts to online and paper news differently, and the driving force behind the difference. In this study, we compare online and printed news and the corresponding market reactions. We find that the online news tone can predict the cross-section of the excess stock return, and that this predictive power can extend to two weeks. The paper news tone has no predictive power at the daily frequency and only marginal predictive power for the long horizon.

Furthermore, we attempt to establish the potential drivers of the difference in predictive power between online and print news tones. We first focus on the information that is common to online and paper news. We use the earnings surprise prediction to assess whether the online news tone captures fundamental-related information. We find that while the online information can predict the earnings surprise, the paper news cannot. This corroborates the claim that the Internet news contains more information about firms' fundamental values. Additionally, we examine the predictive ability of online and paper news for SOE and non-SOE subsample returns, since the paper press tends to serve as the government's mouthpiece while the online media are mostly market driven. The results show that the paper news can predict SOEs' returns but not those of non-SOEs. While the online news can predict both SOEs' and non-SOEs' returns, the scales are larger and persistence longer for the non-SOE subsample. All these observations suggest that the information content in the two channels portrays firms differently.

References

Antweiler, W., & Frank, M.Z.(2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*. 59, 1259–1294.

Baker, S.R., Bloom,N., & Davis, S.J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*. 131,1593-1636.

Bushee, B.J., & Miller, G.S. (2012). Investor relations, firm visibility, and investor following. *Account Review*. 87, 867–897.

Chan W.S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Finance Economics*. 70, 223–260.

Engelberg, J. (2008). Costly Information Processing: Evidence from Earnings Announcements. *Working Paper*, Northwestern University.

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet Information Credibility. *Journalism & Mass Communication Quarterly*, 77(3), 515-540.

George, & Lisa M . (2008) The Internet and the Market for Daily Newspapers. *The B.E. Journal of Economic Analysis & Policy*, 8(1).

Griffin, J.M., Hirschey, & N.H., Kelly, P.J. (2011). How important is the financial media in global markets? *Review of Financial Studies*. 24, 3941–3992.

Jegadeesh, N., & Wu, D. (2013). Word Power: A New Approach for Content Analysis. *Journal of Financial Economics*. 110(3), 712-729.

Jiao, P., Veiga, A., & Walther, A. (2018) Social media, news media and the stock market. *Working Paper*.

Karabulut, Y. (2013). Can Facebook Predict Stock Market Activity? *Working Paper*, Goethe University.

Li, J., Chen, Y., Shen, Y., Wang, J., & Huang, Z.,(2019) Measuring China's stock market sentiment. *Working Paper*

Loughran, T., McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*. 66, 35–65.

Manela,A., & Moreira,A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*.123(1),137-162

Nielsen (2008). Writing Style for Print vs. Web. https://www.nngroup.com/articles/ writing-style-for-print-vs-web/

Seamans R , & Zhu F. (2014). Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers. *Management Science*, 60(2): 476-493.

Shah, D. V., Cho, J., & Eveland Jr, W. P., & Kwak, N. (2005). Information and expression in a digital age: Modeling Internet effects on civic participation. *Communication research*, 32(5), 531-565.

Stiefel A. (2014). The Effect of the Internet on Newspaper Readability. *Working Paper.*

Tetlock, P.C. (2007). Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance*. 62, 1139–1168.

Tetlock, P.C. , Saar-Tsechansky, M. , & Macskassy, S. (2008). More than words: quantifying language to measure firms' fundamentals. *Journal of Finance*, 63(3), 1437-1467.

Tetlock, P.C. (2011). All the news that's fit to reprint: do investors react to stale information? *Review of Financial Studies*. 24, 1481–1512.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics*. 82, 103–133.

You, J. , Zhang, B. , & Zhang, L.(2018). Who captures the power of the pen? *Review of Financial Studies*. 31(1),43-96.

Table 1. Summary Statistics

This table shows the summary statistics of our news data. The sample covers the period from the Jan. 1st 2017 to Dec. 31th 2020. The panel A shows the news coverage for different media sources. Each news in our sample with a unique id is tagged with one or more listed firms. We summary the number of the articles and the number of the firm covered. We also report the average number of the news per firm per year and the ratio of news with different tone categories from different media resources. Panel B show the summary statistics of our key measure. We first aggregate both online and printed news to get the tone from all news. Then we calculate the news tone from the online articles and paper articles separately according to the equation (1) and timeline in panel A of figure 1. The correlation of news tones from different information source is also displayed. All values are significant with of the 1% level of significance.

Panel A. News coverage

Type	No. of news	No. of firm covered	News per firm per year	% positive news	% negative news	% neutral news
All news	5,449,587	3947	377	0.428	0.282	0.290
Online news	4,219,591	3947	292	0.416	0.315	0.269
Paper news	1,229,996	3928	88	0.470	0.169	0.362

Panel B. News tones

	n	mean	Std. dev	10%	90%	Zero ratio	correlation	All news	Online news	Paper news
All news	3,412,021	0.061	0.392	0	0.667	0.759		1		
Online news	3,412,021	0.038	0.374	0	0.5	0.791		0.885	1	
Paper news	3,412,021	0.042	0.252	0	0	0.911		0.428	0.088	1

Table 2. Predicting the Next Day Return Using News Tone

This table shows the estimation results on whether the total news tone could predict the cross section of stock returns. The sample covers the period from January 2017 to December 2020 and our sample firms are all A share stocks listed in Shanghai and Shenzhen security exchange. We estimate Fama–MacBeth regressions, as specified in equation (3). The dependent variable is the raw return at day t+1. The natural logarithm of size, earning-price ratio, turnover, idiosyncratic volatility and return at the day t, return in the past week and month are controlled in the regression. LnSize is the natural logarithm of the total market capitalization of the stocks in ten billion. Earnings-price ratio is calculated as the ratio of the net profit excluding the non-operating revenue in the latest financial report period over the market capitalization at day t. Turnover is the average daily turnover in the last one month. Idiosyncratic volatility is calculated as the standard deviation of residuals from regression of stock returns on Chinese FF3 three factors within past one month.

	I	II	III	IV
All news tone	0.121	0.069		
	17.49	14.76		
Online news tone			0.129	0.074
			17.58	14.44
Paper news tone			0.018	0.008
		.	2.22	1.45
LnSize		-0.001		-0.001
		-0.08		-0.08
EP ratio		1.313		1.316
		6.63		6.63
Turnover		-0.551		-0.555
	.	-1.92		-1.94
Idio. Vol		-0.15		-0.15
		-6.49		-6.49
Past day return		0.052		0.052
		15.53		15.49
Past week return		-0.002		-0.002
		-1.42		-1.45
Past month return		0.000		0.000
		0.21		0.20
Adjusted R2	0.001	0.069	0.002	0.069

Table 3. Portfolio Sort By Total News Tone

This table shows daily return, alpha and characteristics of the portfolio sorted by news tone. Panel A shows the results sorted by the total news tone and Panel B and C show the results sorted by the news tone from the online and printed articles respectively. The sample covers the period from January 1st 2017 to December 31th 2020. The news tone and return are following close-to-close timeline shown in panel A of figure 2. The stocks are sorted into tertiles based on the news tone at day t and then the return is calculated at day t+1 in percentage. The alpha is calculated as the residual of regression of excess return on the Chinese FF3 three factors obtained from LSY (2019). Number of firms is the time-series average of the number of stocks within the portfolio. Size is the total market capitalization of the stocks in ten billion. Earnings price ratio is calculated as the annualized ratio of the earnings in the latest financial report period over the market capitalization at day t. Turnover is the annualized average daily turnover in the last one month. Idiosyncratic volatility is t calculated as the standard deviation of residuals from regression of stock returns on Chinese three factors within past one month.

Panel A. Portfolios sorted by all news

		VW returns				EW returns			
News tone	No. of firms	Ret	t-stat	CH3-alpha	t-stat	Ret	t-stat	CH3-alpha	t-stat
Negative	284	0.001	0.03	-0.038	-3.41	-0.113	-2.54	-0.099	-11.40
Neutral	2659	0.002	0.06	0.004	0.67	0.000	0.01	0.026	5.41
Positive	561	0.069	1.87	0.021	3.33	0.097	2.21	0.104	15.10
Positive-Negative		0.068	4.51	0.059	4.04	0.21	21.30	0.204	20.99

Panel B. Portfolios sorted by online news

		VW returns				EW returns			
News tone	No. of firms	Ret	t-stat	CH3-alpha	t-stat	Ret	t-stat	CH3-alpha	t-stat
Negative	282	0.014	0.35	-0.032	-2.98	-0.115	-2.58	-0.103	-11.23
Neutral	2771	0.007	0.18	0.006	1.23	0.002	0.05	0.027	5.72
Positive	450	0.072	1.91	0.023	3.10	0.113	2.57	0.120	14.54
Long-short		0.058	3.97	0.055	3.77	0.227	21.17	0.224	20.88

Panel C. Portfolios sorted by paper news

		VW returns				EW returns			
News tone	No. of firms	Ret	t-stat	CH3-alpha	t-stat	Ret	t-stat	CH3-alpha	t-stat
Negative	63	0.036	0.83	-0.009	-0.46	-0.019	-0.44	-0.019	-1.34
Neutral	3191	0.010	0.24	0.006	1.38	0.003	0.06	0.027	6.27
Positive	249	0.052	1.45	-0.009	-1.06	0.041	0.98	0.037	5.66
Long-short		0.016	0.69	0.000	0.02	0.060	4.18	0.056	3.89

Table 4. Double Sort Based on the Firm Characteristics and Tone News Tone.

This table shows the return and alpha of long-minus-short portfolio sorted by the different news tone within groups with different firm characteristics. Panel A shows the value-weighted portfolio return and alpha sorted by the different news tone among the different size group. Panel B shows the value-weighted portfolio return sorted by the different news tone among the different liquidity group. At day t+1, we first sort the stocks into three groups based on the firm characteristics and then within each tertile we sort the stocks into teriles further based on the news tone at day t+1. The size is the total market capitalization of the firm. The Amihud calculated as the average absolute return over the trading volume in the past one month is used as the proxy of illiquidity.

Panel A. Long short portfolio returns for different size groups

		Raw Returns			CH3 alpha		
		Small	Medium	Large	Small	Medium	Large
All news	mean	0.314	0.256	0.049	0.310	0.253	0.043
	t-stat	16.24	16.13	3.08	15.96	15.85	2.75
Online news	mean	0.314	0.265	0.038	0.314	0.265	0.038
	t-stat	15.09	14.81	2.47	15.09	14.81	2.47
Paper news	mean	0.087	0.078	0.016	0.077	0.077	0.001
	t-stat	2.42	2.49	0.65	2.15	2.44	0.06

Panel B. Long short portfolio returns for different liquidity groups

		Raw Returns			CH3 alpha		
		Low	Medium	High	Low	Medium	High
All news	mean	0.296	0.170	0.047	0.295	0.164	0.041
	t-stat	12.61	10.53	2.87	12.49	10.17	2.54
Online news	mean	0.342	0.181	0.035	0.343	0.176	0.034
	t-stat	12.51	10.05	2.25	12.45	9.79	2.16
Paper news	mean	0.044	0.074	0.016	0.037	0.074	0.002
	t-stat	0.95	2.44	0.64	0.79	2.42	0.09

Table 5. News Predictive Power for Future 12 weeks

This table displays results on whether news tone can predict individual stock returns at more distant horizons. The sample covers the period from January 2017 to December 2020 for all A share stocks listed in Shanghai and Shenzhen security exchange. We estimate Fama–MacBeth regressions as the equation (3). The dependent variable is the k-weeks ahead weekly return. To account for serial correlation in the coefficients, the standard deviations of the time-series are adjusted using Newey–West (1987) with corresponding lags. The control variables are same as the one day ahead return prediction in the equation (4). To save the space, we don't report the control variable.

Panel A. Fama-MacBeth regression over the next 12 weeks

News variable	All news tones		Online news tones		Paper news tones	
period	Coef.	T value	Coef.	T value	Coef.	T value
1 day	0.001	12.91	0.001	13.11	0.000	2.20
1 week	0.104	8.16	0.105	8.21	0.060	3.41
2 weeks	0.055	4.47	0.052	4.25	0.057	3.23
4 weeks	0.039	3.58	0.033	2.98	0.055	3.15
6 weeks	0.050	4.17	0.048	3.90	0.056	3.46
8 weeks	0.047	3.82	0.043	3.36	0.047	2.67
10 weeks	0.036	3.12	0.035	3.07	0.034	2.03
12 weeks	0.052	4.54	0.055	4.78	0.029	1.69

Panel B. Long short portfolio returns over the next 12 weeks (VW alpha)

News variable	All news tones		Online news tones		Paper news tones	
period	Ret	T value	Ret	T value	Ret	T value
1 day	0.066	4.36	0.055	3.68	0.018	0.77
1 week	0.104	2.49	0.062	1.67	0.067	0.95
2 weeks	0.142	2.17	0.080	1.45	0.113	0.92
4 weeks	0.175	1.43	0.096	1.02	0.273	1.22
6 weeks	0.240	1.20	0.136	1.01	0.298	0.87
8 weeks	0.297	1.07	0.169	0.86	0.340	0.71
10 weeks	0.309	0.80	0.090	0.33	0.425	0.70
12 weeks	0.250	0.54	0.099	0.29	0.277	0.39

Table 6. Information Content of News: Earnings Surprise

This table shows the prediction results of different news for the standardized unexpected earnings and cumulative abnormal return for earnings announcement. We use the past one day news tone from different media type to do SUE prediction as equation (4). For SUE, we use the average earnings in the past 8 quarters as the expected earnings and standardized by the standard deviation of the earnings in the past 8 quarters. Tone represents the different news tone at day t. We include the past one day, week and month return. LnSize is the natural logarithm of the total market capitalization of the stocks in ten billion. Earnings price ratio is calculated as the ratio of the earnings in the latest financial report period over the market capitalization at day t.

Earnings surprise news	SUE all	SUE online	SUE paper	CAR[0,1] all	CAR[0,1] online	CAR [0,1] paper
News tone	0.605	0.174	0.116	0.301	0.118	0.071
	13.04	2.68	1.39	8.73	2.44	1.12
LnSize	0.200	0.225	0.225	0.001	0.013	0.013
	8.41	9.46	9.42	0.03	0.72	0.72
EP ratio	-1.221	-0.735	-0.700	-1.611	-1.371	-1.347
	-1.70	-1.02	-0.98	-2.88	-2.45	-2.41
Past day return	8.165	9.147	9.417	-2.436	-2.008	-1.833
	9.38	10.48	10.85	-3.86	-3.18	-2.92
Past week return	3.136	3.386	3.414	-3.384	-3.268	-3.250
	7.97	8.59	8.66	-11.76	-11.35	-11.29
Past month return	0.394	0.487	0.494	-0.993	-0.954	-0.949
	1.92	2.37	2.40	-6.66	-6.39	-6.35
Intercept	-2.734	-2.992	-2.992	-0.201	-0.325	-0.327
	-8.48	-9.26	-9.23	-0.84	-1.37	-1.37
Adjusted R2	0.0191	0.0127	0.0125	0.0087	0.0062	0.0060

Table 7. Separate Online News and Paper News

This table tests the difference of the content of online and paper news. Panel A shows the prediction of the online and printed news tone for the stock return restricted to the subsample. We divide the sample into SOEs and non-SOEs and use the online and printed news tone as the key independent variables to do k-weeks ahead return prediction. Panel B shows the results of the decomposition exercise. We regress the online news on the paper news and get the parts of online news related and orthogonal to paper news. Then we use two parts to sort the next day return separately.

Panel A. SOEs vs. NSOEs

	Online news for SOEs		Online news for NSOEs		Paper news for SOEs		Paper news for NSOEs	
	Coef.	T value	Coef.	T value	Coef.	T value	Coef.	T value
1 day	0.052	5.37	0.077	9.03	0.026	2.69	0.006	0.62
1 week	0.075	3.18	0.093	5.02	0.073	2.70	0.040	1.76
2 weeks	0.034	1.61	0.049	2.60	0.056	2.10	0.040	1.65
4 weeks	-0.005	-0.26	0.044	2.59	0.053	2.06	0.032	1.39
6 weeks	0.023	1.14	0.060	3.45	0.054	2.20	0.031	1.27
8 weeks	0.032	1.45	0.036	1.84	0.045	1.64	0.003	0.13
10 weeks	0.050	2.42	0.029	1.67	0.010	0.39	0.039	1.63
12 weeks	0.000	0.02	0.053	2.79	0.012	0.49	0.041	1.68

Panel B. A decomposition exercise

	Part of online news related to paper news				Residual			
	Ret	T value	CH3-alpha	T value	Ret	T value	CH3-alpha	T value
Negative	-0.003	-0.06	-0.020	-0.94	-0.036	-0.66	0.002	0.11
Neutral	-0.034	-0.65	0.005	0.91	-0.033	-0.56	0.033	6.13
Positive	0.034	0.79	-0.008	-0.81	0.015	0.28	0.048	6.29
Long-short	0.038	1.34	0.013	0.50	0.051	3.15	0.047	2.85

Table 8. Additional Analysis

This table shows the additional tests. The sample covers the period from January 1st 2017 to December 31th 2020. Panel A and B shows the portfolio sorting results by the all news tone with the open-to-open timeline and without filling zeros for no news day respectively. The stocks are sorted into tertiles based on the news tone at day t and then the return is calculated at day t+1 in percentage. The alpha is calculated as the residual of regression of excess return on the Chinese FF3 three factors obtained from LSY (2019). The timeline of the news data and open-to-open return are shown in the panel B of figure 2. Panel C shows the aggregate prediction for the market conditions including the market return and volatility. We use the market cap as the weight to aggregate the news tone and all A share return and get the time series index and market return. We use the CIVIX as the market volatility. The market return or volatility at day t is controlled. The t statistics is Newey-west adjusted with one period lag. Panel D shows whether the investors overreact to the stale news and whether it affects the prediction of the news tone. We run the Fama-MacBeth regression as equation (4) which include the stale and the interaction term of the stale and news tone in the regression. To account for serial correlation in the coefficients, the standard deviations of the time-series are adjusted using Newey–West (1987) with corresponding lags. To save the space, we don't report the control variables which are same as before.

Panel A. Predict market conditions

Parameter	Mkt Ret	Mkt Ret	Mkt vol	Mkt vol
All news	-0.013		**2.427**	
	-1.56		2.80	
Online news		-0.002		**2.296**
		-0.23		2.59
Paper news		-0.012		-0.384
		-1.56		-0.47
Adjusted R2	0.0040	0.0040	0.0077	0.0043

Panel B. Staleness of news

	Online news tone		Online news tone *Stale		Stale		Paper news tone		Paper news tone*Stale		Stale	
period	Coef.	T value	Coef.	T value	Coef.	T value	Coef.	T value	Coef.	T value	Coef.	T value
1day	0.03	3.38	0.01	0.43	-0.02	-1.32	0.03	1.77	-0.01	-0.41	-0.01	-1.02
1week	0.05	2.00	0.03	0.89	**-0.08**	-2.25	0.11	2.94	-0.02	-0.38	-0.06	-1.91
2week	0.02	0.86	0.03	0.92	-0.06	-1.63	0.09	2.50	-0.04	-0.70	-0.04	-1.24
4week	0.01	0.33	-0.02	-0.41	-0.02	-0.61	0.02	0.68	-0.02	-0.45	0.00	-0.14
6week	0.05	2.08	0.02	0.48	0.05	1.39	-0.01	-0.36	0.07	1.33	0.01	0.45
8week	0.04	1.84	0.00	-0.06	0.03	0.82	0.03	0.92	0.03	0.49	-0.02	-0.51
10week	0.00	-0.03	0.01	0.38	0.01	0.22	0.01	0.27	0.04	0.64	-0.01	-0.45
12week	0.02	0.91	-0.07	-1.74	0.05	1.39	0.07	1.74	-0.02	-0.32	-0.01	-0.23

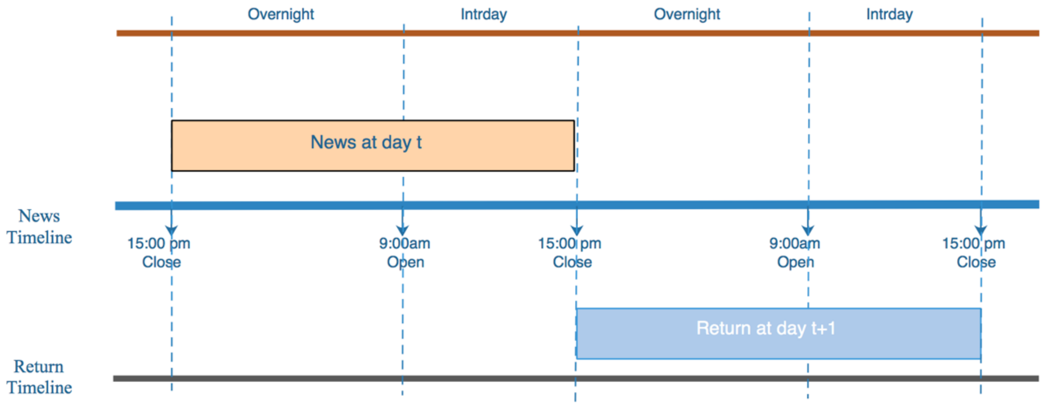
Panel C. Robustness check using open-to-open prices

						VW returns				EW returns			
News tone	No. of firms	Size	EP ratio	Turnover	Idio. Vol	Ret	t-stat	CH3-alpha	t-stat	Ret	t-stat	CH3-alpha	t-stat
Negative	278	31.810	0.437	6.552	0.651	-0.044	-0.91	-0.047	-1.45	-0.130	-2.24	-0.074	-2.36
Neutral	2516	9.760	0.393	6.552	0.589	-0.061	-1.14	-0.012	-0.42	-0.051	-0.84	0.020	0.70
Positive	579	44.390	0.443	5.796	0.620	0.011	0.24	-0.007	-0.23	-0.020	-0.36	0.030	0.99
Long-short						0.055	2.97	0.040	2.28	0.110	8.51	0.104	8.31

Panel D. Robustness check without filling zeros for no news day

						VW returns				EW returns			
News tone	No. of firms	Size	EP ratio	Turnover	Idio. Vol	Ret	t-stat	CH3-alpha	t-stat	Ret	t-stat	CH3-alpha	t-stat
Low	309	30.529	0.026	6.598	0.651	0.022	0.45	-0.014	-0.85	0.032	0.57	0.058	3.03
Medium	338	63.709	0.040	5.504	0.601	0.064	1.47	0.009	0.95	0.164	3.12	0.175	14.47
High	334	17.622	0.031	6.375	0.642	0.139	2.67	0.138	9.03	0.354	6.05	0.394	24.55
Long-short						0.117	4.35	0.152	6.29	0.318	14.44	0.336	15.34

Panel A. Close-to-close timeline



Panel B. Open-to-open timeline

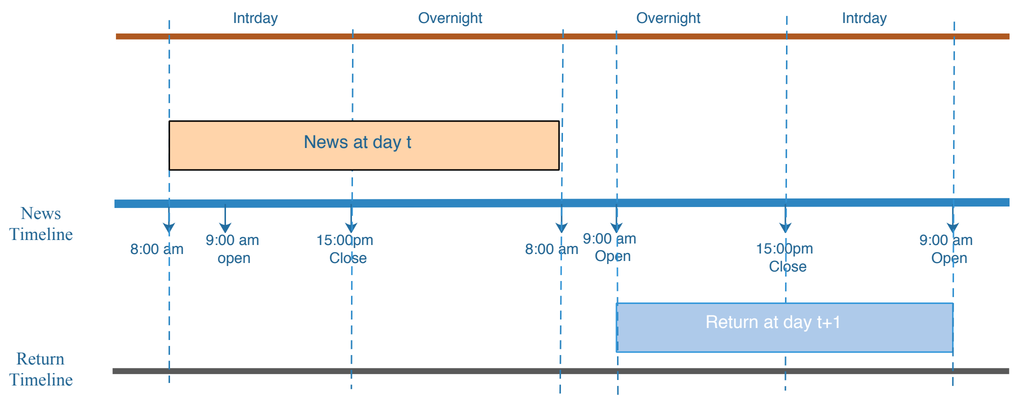


Figure 1. The Timeline of News and Return Prediction

This figure shows the timeline of news and return prediction. The Panel A shows the close-to-close timeline and Panel B shows the open-to-open timeline.