News Coverage and Idiosyncratic Volatility of Stock Returns

Chu Zhang*

This version: August 2022

*Hong Kong University of Science and Technology (e-mail: czhang@ust.hk, phone: 852-2358-7684).
Zhiwen (Ryan) Zhuang provided excellent assistance. Financial support from the Research Grants Council of the Hong Kong SAR, China (General Research Fund no. HKUST16504818) is gratefully acknowledged. All remaining errors are mine.

News Coverage and Idiosyncratic Volatility of Stock Returns

Abstract

I examine the relationship between news coverage and idiosyncratic volatility of stock returns at the firm level. News sentiment score explains part of the idiosyncratic stock returns and news intensity is positively related to idiosyncratic volatility for firms with sufficient news coverage, adding to the set of fundamentals variables known to contribute to idiosyncratic volatility. The finding is consistent with the notion that higher idiosyncratic volatility is partially due to firm-specific information being impounded in stock returns, solving a puzzle of lacking such a relationship.

Key words: Idiosyncratic volatility of stock returns, news score, news intensity, firm-specific information, fundamentals.

1. Introduction

More than thirty years ago, Roll (1988) raises an issue of why only a small portion of stock return variations can be explained by the market return and industry returns in the US. During 1982-1987, the average R^2 (the proportion of squared returns, explained by the market and industry returns) of the market model is about 35% for monthly returns and 20% for daily returns on stocks traded on NYSE and AMEX, which tend to be larger stocks with higher R^2s than those traded on NASDAQ. Moreover, the source of unexplained, idiosyncratic variations in the stock returns is unknown, as whether or not there is firm-specific news traced to Wall Street Journal or Dow Jones Newswires does not make much difference for the market model performance. Roll (1988) attributes the unexplained idiosyncratic return variations to possible investor frenzy, when the efficient market hypothesis is still the dominant view held by most academics at the time. While little progress is made in the following decade to address Roll's (1988) inquiry on why the average R^2 is slow, the matter takes an abrupt turn by Morck, Yeung and Yu (2000) who document that, while the average R^2 of the market model in the US has become even smaller in late 1990s, it is typically low for developed markets and high for emerging markets. They argue that a low R^2 is a reflection that stock prices incorporate more firm-specific information, a good thing to have for a well-functioning market. This assertion seems to be undisputable given the fact of low average R^2s in developed markets and high average R^2s in emerging markets, but direct evidence has never been found. In addition, the question about the lack of trace of idiosyncratic volatility to firm news remains unanswered.

I reconsider these issues in this paper, focusing more on the second question by looking into news coverage more closely. The time is ripe because data resources on firm-specific news are much richer now than before. I use the data from RavenPack to extract daily news score and annual news intensity for every US firm listed on NYSE/AMEX/NADAQ from 2000 to 2019. From the RavenPack database, there are more than 5000 news platforms which distribute news about (listed) firms worldwide. With much increased coverage, will idiosyncratic returns be better explained, compared to that reported by Roll (1988)?

When Roll conducted his analysis using Dow Jones Newswires and Wall Street Journal, with constraints on the technology, the only information extracted is whether a firm on a given day is mentioned by news or not. There is not even information about whether the news is good or bad for the investors, let alone the magnitude of the potential news impact. Using advanced algorithms from modern technology, RavenPeck assigns each piece of news a score between 0 and 100 to indicate whether the news is relevant to firm value with 100 being the most relevant. It also assigns a sentiment score between 0 and 100 to indicate how good the news is for the firm investors, with 50 being neutral.[1] Since news pieces reported by different news platforms may refer to the same corporate event, RavenPack assigns a score to indicate the freshness of the news piece according to the time stamp the news is released. To the extent these scores reflect the truth, such a rich set of data can be very useful to conducting research on the relationship between stock returns and firm-specific information. The database has been fruitfully exploited in many studies.[2]

Despite the huge advantages of the news database, one obstacle for uncovering the relationship between news coverage and idiosyncratic return volatility is the fact that not all firms are equally covered by news platforms. There are strong positive correlations of news coverage with firm size and with trading volume of the stock. This is understandable, as large firms have large investor bases and news platforms are mostly commercially oriented with their profits tied to the readership. While large firms do have more events to be covered in news, the coverage appears to be disproportionate. As such, a simple regression of idiosyncratic return volatility on a news intensity measure unsurprisingly generates a negative slope coefficient. Even when firm size is controlled for, the picture is still unclear, as there are just too many firm-days for which there is no news coverage at all.

I therefore focus on firms with sufficient news coverage days at a minimum of 25 news days per year, roughly 10% of the trading days per year, to generate the main results. The firm-year

[1] The word *sentiment* used in this paper has no connotation of behavioral biases.

[2] An incomplete list includes Shroff, Verdi and Yu (2014), Dai, Parwada, and Zhang (2015), Dang, Moshirian and Zhang (2015) and Yang, Zhang and Zhang (2020).

sample with this news coverage threshold constraint, dubbed as the news sample, represents 59% of all possible firm-years. For the news sample, I construct news scores at daily frequency, taking into consideration of good-news orientation (i.e., good or bad news for investors), news relevance, and news freshness. For this sample of firms, a market model extended to include the industry return at daily frequency has a median R^2 of 25%. The news score can explain 4% of the remaining 75% of the idiosyncratic variance in total variance. When the other four of the Fama-French factors (Fama and French (2015)) are added to the extended market model, the median R^2 increases modestly to 30% and the news score can again explain 4% of the remaining 70% of the idiosyncratic variance in total variance. I also construct firm-year observations of idiosyncratic return volatility and a measure of news intensity for firm-years satisfying the news sample. The news intensity is found to be positively related to idiosyncratic volatility either at absolute level or relative to systematic volatility, after firm size is controlled for. The relationship is robust to the inclusion of other firm-specific variables known to explain idiosyncratic volatility.[3]

I also look into the sample firms that do not meet the news coverage threshold criteria (the no-news sample) and are therefore left out for the main analyses. The firms in the no-news sample tend to be smaller, less traded, more value than growth oriented, and have lower average earnings, but larger volatility of earnings. These features make the idiosyncratic volatility in the no-news sample greater than that of the news sample on average. Pooling the news sample and the no-news sample together obscures the relationship between news coverage and idiosyncratic volatility.

The paper makes the following contributions to the literature. Volatility of asset returns has always been in the center of asset pricing research. The theory that expected asset returns depend on their sensitivities to systematic factors has met various challenges, in the sense that the sensitivities to the factors that can explain much return variations do not explain much cross-sections of average returns (i.e. the failure of the CAPM), while factors such as those

[3]These firm characteristics include trading volume, past volatility of return-on-equity, and growth orientation, which have positive effect on idiosyncratic volatility of returns, and past mean return-on-equity which has negative effect on idiosyncratic volatility of returns.

constructed by Fama and French (2015) can explain some cross-sectional difference in expected returns, but little of return volatility (as the numbers above show). Therefore, understanding of idiosyncratic volatility becomes more important. Roll (1988), for example, concludes his analyses by attributing the unexplained idiosyncratic volatility to private information or investors' frenzy. In the literature of behavioral finance which becomes popular in the 1990s, noise trading is the buzzword that "explains" everything standard asset pricing models cannot explain. The finding by Morck, Yeung and Yu (2000) that idiosyncratic volatility is actually greater in mature markets than in emerging markets and that idiosyncratic volatility in the US is increasing over time points to a possible interpretation that large idiosyncratic volatility is indicative of large amount of firm-specific information being impounded in stock prices. This explanation appears to be plausible given the above-mentioned cross-market and time-series patterns. The only problem is that the evidence is hard to come by. The results presented in this paper offers a promising, though limited, support to the hypothesis that part of idiosyncratic volatility is justified and consistent with the notion that stock prices adjust to information relevant to firm values.

The rest of the paper is organized as follows. Section 2 gives a brief review of the literature on the patterns of idiosyncratic volatility of stock returns. Section 3 describes the data and sample used in this study. Section 4 shows the baseline results of factor models of daily returns with news scores for firms satisfying the 25-day-per-year threshold criteria. Section 5 examines the role of news intensity in explaining the absolute idiosyncratic volatility from factor models and the relative idiosyncratic volatility to total volatility. Section 6 characterize the patterns of idiosyncratic volatility for firm-years not satisfying the 25-day-per-year threshold criteria. Section 7 concludes.

2. Literature Review

The literature on stock return volatility is voluminous. This section gives a brief review of the literature on idiosyncratic volatility patterns in time series and cross sections, related to main focus of this paper.[4] Beside the early work by Roll (1988), there are two strands of the literature on these patterns. These two strands of inquiries are related in the sense that both involve idiosyncratic return volatility. But on the other hand, they are also different in the sense that the issues being addressed are distinct.

One strand of the literature begins with the finding reported by Campbell, Lettau, Malkiel and Xu (2001) that the average idiosyncratic volatility with respect to market returns and industry returns increased over time from 1960s to late 1990s, while the systematic risk remains basically unchanged over the same sample period. Schwert (2002) suggests that the unusually high return volatility during the late 1990s could be attributed to technology firms and hints that the increasing volatility is due to the increasing value of growth options associated with technology advances. Wei and Zhang (2006) show that, to a large extent, the increase in average idiosyncratic volatility is caused by the decrease in average earnings and increase in average earnings volatility, partially due to the addition of newly listed firms to the sample over time. Brown and Kapadia (2007) further examine the issue and find more evidence that newly listed companies contribute to increasing stock return volatilities. Cao, Simin, and Zhao (2008) use the market-to-book ratio of assets, and its variations as proxies for growth options and find that these growth option proxies contribute to explain the trend of the value-weighted average idiosyncratic volatility. Irvine and Pontiff (2009) extend the work by Wei and Zhang (2006) to consider the volatility of both earnings and sales, allegedly due to increased market competition, and they obtain similar results. Xu and Malkiel (2003) find that idiosyncratic volatility is positively related in cross sections to institutional holdings and the expected earnings growth made by analysts. Brandt, Brav, Graham, and Kumar (2010) question whether there is a sustained trend in average idiosyncratic volatility and document evidence of irrational trading behavior by retail investors during the late 1990s that allegedly caused a speculative bubble. Zhang (2010) compares the fundamentals-based theory with earnings volatility as the main explanatory variable and various non-fundamentals based explanations which work through trading volumes, either

[4]There is another literature pioneered by Ang, Hodrick, Xing and Zhang (2006, 2009) on the relationship between idiosyncratic volatility and stock returns. The literature is also large and interesting, but less related to the issues discussed in this paper. It is therefore, not reviewed here.

by institutional investors or by irrational retails investors. The comparison lends support to the fundamentals-based explanation, as the average idiosyncratic volatility ceases its upward trend after 2000 and fluctuates, while trading volume keeps going up. Bekaert, Hodrick, and Zhang (2012) also question whether there is a continued trend in idiosyncratic volatility in international markets. They focus more on econometric modeling issues than on the causes of the changing volatility, however.

The second strand of the literature begins with the work by Morck, Yeung and Yu (2000) which contains two results. One is that the average R^2 of the market model of the returns in the US has declined over time. This collaborates with the result of Campbell et al (2001) that average idiosyncratic volatility increases while market volatility remains unchanged over time. The difference is that Campbell et al (2001) focus on the average idiosyncratic volatility while Morck et al (2000) focus on the idiosyncratic volatility relative to the systematic volatility. The second result shows that the average market-model R^2 in mature markets of developed countries tend to be lower than that in emerging markets of developing countries. It is this observation that inspires the notion that a low average R^2 of the market model is indicative of market maturity. Morck et al (2000) posit that the idiosyncratic volatility is relatively high in mature markets because there is plenty of firm-specific information available which frequently gets impounded in stock prices and causes idiosyncratic volatility to be high, while the average systematic volatility is high in emerging markets because, due to lack of firm-specific information, investors react more responsively to macroeconomic information and infer values of firms without information from other related firms with some information, causing stock prices to comove. Morck et al (2000) also resort to weak property right protection for public investors in emerging markets as reasons for them to have a low average idiosyncratic volatility relative to systematic volatility. Jin and Myers (2006) make further hypotheses on institutional and cultural differences among countries of different stages of development and connect them to market model R^2s. Firm-level information opacity is a key variable in their study to explain the cross-country differences in the stock return R^2s. Follow-up studies have accumulate more evidence on this, which will not be reviewed here as it is not relevant to the current study. While cross-country studies generate much evidence consistent with the theory proposed by Morck et al (2000) and Jin and Myers (2006), some within-country studies provide opposite results. Piotroski and Roulstone (2004) and Chan and Hameed (2006), for example, find that firms are more covered by financial analysts actually have smaller idiosyncratic volatility. Dasgupta, Gan, and Gao (2010) argue that idiosyncratic volatility can be smaller for more transparent firms and provide evidence that idiosyncratic volatility typically becomes larger before seasoned equity offerings and cross-listings, but smaller afterwards. Teoh, Yang and Zhang (2009) document that low R^2 firms have smaller stock price responses to earnings and have more uncertainty about future earnings. In addition, low R^2 firms have worse information environment, as measured by earnings quality, earnings persistence, earnings predictability, and have higher probability of distress. In short, these studies show that the relationship between firm-specific information and idiosyncratic volatility has never been established empirically at the firm level within a market. Moreover, the direct link between corporate news and idiosyncratic volatility has never been established.

2.1. The Impact of News

There is a nascent literature on how news affect finance. Fang and Peress (2009) show that stock prices of firms covered and uncovered by media behave differently. Uncovered firms have higher average returns, and this is more evident for firms that have low institutional holding, low analyst following, smaller, and with more volatile stock prices. Dai, Parwada and Zhang (2015) show that media coverage reduces insiders' future trading profits by disseminating news on prior insiders' trades available from regulatory filings. These studies do not examine the direct relationship between corporate news and idiosyncratic volatility, however.

3. Data

The data used in this paper come from a variety of sources. Stock returns and market capitalization data are from CRSP and firms' accounting data are from Compustat. After merging the CRSP and Compustat data, the initial sample covers all common stocks listed on the NYSE, AMEX, and Nasdaq. Stocks included in this study are required to have a book equity greater than 1 million and non-missing information on market capitalization and return on equity. Returns and news variables are described here. Fundamentals and trading volume related variables will be described in a later section.

News data is from RavenPack News Analytics, which is a leading global news database used by practitioners in quantitative and algorithmic trading and by scholars in accounting and finance research. RavenPack collects and analyzes real-time, firm-level business news from leading news providers, such as Dow Jones Newswire, The Wall Street Journal, Barron's, and other major publishers and web aggregations, including industry and business publications, regional and local newspapers, government and regulatory updates, and trustworthy financial websites. Its coverage starts from 2000, so the sample is 2000 to 2019.

To measure the informational content of a news article, RavenPack implements two steps. First, it classifies news articles into news event categories according to the RavenPack taxonomy, and both the topic and a firm's role in the news article are tagged and categorized. All the news articles are grouped into 36 news categories. Among all the news stories, the top five news categories are earning (20.02%), products-services (10.84%), acquisitions-mergers (6.85%), analyst-ratings (6.71%), and technical-analysis (6.64%).

Second, RavenPack constructs a news sentiment score for each news piece based on professional algorithms, which were developed and evaluated by effectively combining traditional language analyses, financial expert consensuses, and market response methodologies. Specifically, the news sentiment score indicates whether and to what extent a news piece may have a positive, neutral, or negative effect on the firm value. This sentiment score, known as the Event Sentiment Score (ESS), is assigned to all relevant firms listed in the news report. It ranges from 0 to 100, with a value below (above) 50 indicating the negative (positive) sentiment of a given news piece. A score of 50 represents neutral sentiment. To facilitate our empirical analysis, the original ESS score minus 50 and divided by 50 is used. The adjusted sentiment score falls within the interval between minus one and one. To exclude as much repeated news as possible, the

Event Novelty Score (ENS) provided by RavenPack is used to generate a freshness indicator. The Event Novelty Score (ESS) ranges from 0 to 100, where 100 indicates the most fresh news. With ESS weighted by ENS, the cross-sectional pattern of news is less likely to be driven by the repetitive dissemination of the same or similar news articles. Likewise, to exclude as much irrelevant news as possible, relevance indicator (REL) provided by RavenPack is also used to adjust ESS. The relevance indicator (REL) ranges from 0 to 100, where 100 indicates the most relevant news. The news score (NS) used in the paper for a firm-day (i, d) is defined as follows.

$$\mathrm{NS}_{id} = \sum_{j \in d} \frac{\mathrm{REL}_{ij}}{100} * \left(\frac{\mathrm{ENS}_{ij}}{100}\right)^2 * \frac{\mathrm{ESS}_{ij} - 50}{50}, \tag{1}$$

where j indicates a piece of news related to firm i and the summation is taken within the day d. Only business day is considered here. Designation of days is made in order to match the time when stock exchanges are open, so that investors can react to the news. For a Tuesday to a Friday, a day is counted from the 4:00pm of the previous day to the 4;00pm on the current day. For a Monday, it is counted from 4:00pm on the last Friday to 4:00pm on the current day. In the definition, the freshness measure of the news is squared to further reduce the impact of stale news.

Among all the news pieces, one category is the unusual price changes. That is, a stock price experienced an unusual large change and a news piece reported it. While the stock price change may well be driven by information, it is obviously inappropriate to be included in the calculation of NS, as the purpose of this study is to see how much stock prices react to news, not the other way around. For this reason, news pieces within this category are all excluded in calculating NS.

News intensity (NI) is meant to indicate how strong a specific firm is covered by public news. It is defined as follows.

$$\mathrm{NI}_{id} = \sum_{j \in d} \frac{\mathrm{REL}_{ij}}{100} * \left(\frac{\mathrm{ENS}_{ij}}{100}\right)^2 * \left|\frac{\mathrm{ESS} - 50}{50}\right|. \tag{2}$$

Firm-year observations of news score NS_{iy} and news intensity NI_{iy} are calculated similarly as in (1) and (2) where y indicates a year and the summation is taken within the year y. For the same reason, news pieces within the category of unusual price changes are excluded in calculating NI.

It is evident that the relationship between idiosyncratic volatility and news coverage, if it exists, can only be discovered in a sample of firms with sufficient news for these firms. For that purpose, I decompose the entire sample of firm-years into the news sample and no-news sample. For a given year, a firm belongs to the news sample if and only if there are at least 25 days during the year in which there are news pieces about the firm. Under this definition, 59% of firms on average are included in the news sample. Note that the news sample and the no-news sample are not two fixed samples over time.

Table 1 reports summary statistics of stock returns and news related variables for the news sample. Panel A contains the panel mean, standard deviation, skewness, kurtosis and percentiles of daily returns on individual stocks of this sample during 2000-2019. It also reports the same statistics for the five factors of Fama and French (2015) at the daily frequency.[5] Compared with the market returns, the individual returns have much higher standard deviation, much more positive skewness, and much higher kurtosis.

Table 1 here

Panel B of Table 1 reports the panel mean, standard deviation, skewness, kurtosis and percentiles of the original ESS, ENS and REL measures assigned by RavenPack at the level of news pieces. All of the three measures are arranged between 0 and 100. The percentiles of REL show that more than 95% of the news pieces are relevant to the valuation of the firms. More than half of the news pieces are fresh news as the median value of ENS is 100. The median value of the sentiment score ESS is neutral, but the mean value of the sentiment tilts slightly to good news.

Panel C of Table 1 reports the summary statistics of news scores and news intensity when they are aggregated within a day. The distribution of the number of news pieces per day reveals that for more than half of the firm-days, there is no news at all, although on average there are 1.46 pieces of news per firm-day for the news sample. The resultant news scores and news intensity have huge positive skewness. To make them behave more appropriately in statistical

[5] They are calculated in the news sample, so the daily factor distributions are weighted by the number of firms in the news sample within the year.

sense, transformations are performed as follows. For news scores, since they can take either positive or negative values, Log transformation is defined as

$$\mathrm{LNS}_{id} = \mathrm{Sgn}(\mathrm{NS}_{id}) \log(1 + |\mathrm{NS}_{id}|), \tag{3}$$

where Sgn is the sign function, which is positive if NS_{id} is positive, negative if NS_{id} is negative, and zero if NS_{id} is zero. For the news intensity which is always non-negative,

$$\mathrm{LNI}_{id} = \log(1 + \mathrm{NI}_{id}). \tag{4}$$

The distributions of transformed news scores and news intensity are still quite right-skewed as seen from the table, but much reduced. The number of news piece, K_{id}, is very positively skewed as well. Near 25% of firm-days have just one piece of news. About 5% of firm-days have more than 7 pieces of news. These numbers, though not precise, provide a crude description of how news coverage is positively skewed. The news reports concentrate on small number of firms and on small number of days.

4. Idiosyncratic Volatility and News Score

For firm-years in the news sample, an extended market model is performed on the daily returns within the year. More specifically, the regression takes the form

$$r_{id} = b_0 + b_1\mathrm{MKT}_d + b_2\mathrm{MKT}_{d-1} + b_3\mathrm{IND}_{id} + b_4\mathrm{IND}_{i,d-1} + \varepsilon_{id}, \qquad d \in y, \tag{5}$$

where MKT is the market excess return, IND is the value weighted portfolio excess return of the industry firm i belongs to according to the Fama-French 49-industry classification, and ε is the generic term for the idiosyncratic component of the return with respect to the systematic factors of the model. One-day lagged market return and industry return are included to capture potential infrequent trading of the firm. The regression is run for daily returns for each firm-year. Idiosyncratic volatility IV_{iy} is calculated as the square root of the sum of squared residuals of the model: $\mathrm{IV}_{iy} = \sqrt{\sum_{d\in y} \hat{\varepsilon}_{id}^2}$, is the annualized sample standard deviation of the daily returns for stock i during year y.

Similarly, idiosyncratic volatility can be calculated from models with more systematic factors. The extended Fama-French five-factor model takes the form

$$\begin{aligned} r_{id} &= b_0 + b_1\mathrm{MKT}_d + b_2\mathrm{MKT}_{d-1} + b_3\mathrm{IND}_{id} + b_4\mathrm{IND}_{i,d-1} \\ &\quad + b_5\mathrm{SMB}_d + b_6\mathrm{HML}_d + b_7\mathrm{RMW}_d + b_8\mathrm{CMA}_d + \varepsilon_{id}, \qquad d \in y, \end{aligned} \tag{6}$$

where MKT, SMB, HML, RMW and CMA are the five factors used in Fama and French (2015).

Panel A of Table 2 reports the panel statistics of the estimated regression coefficients, their t-ratios, regression R^2, idiosyncratic volatility (IV), and the log of idiosyncratic volatility (LIV) from the extended market model. The same-day returns on the market portfolio and on the industry portfolio share the explanatory power fairly equally. The significance of the beta is higher for the industry portfolio than for the market portfolio. The one-day lagged returns on the market portfolio and on the industry portfolio explain only small portions of the individual returns. The median of R^2s is 0.25. The median of annualized idiosyncratic volatilities (IV) is and 32.82. The panel also reports the summary statistics of the log of idiosyncratic volatility (LIV) and the log of relative idiosyncratic volatility (LRIV), defined as the log of one minus R^2: LRIV = log $(1 - R^2)$. While IV is highly positively skewed, both LIV and LRIV are closer to be normally distributed with skewness close to zero and kurtosis slightly greater than that of the normal distribution.

Table 2 here

Panel B of Table 2 reports the summary statistics of the estimated regression coefficients, their t-ratios, regression R^2, idiosyncratic volatility (IV), and the log of idiosyncratic volatility (LIV) from the extended Fama-French five-factor model. The added four Fama-French factors divert certain explanatory power from the market returns and industry returns. The overall explanatory power with added factors does not increase much, however. The median of R^2s increases from 0.25 to 0.30. The median of annualized idiosyncratic volatilities (IV) reduces from 33.82 to 32.66.[6]

[6]While Fama-French factors do not exhaust all potential factors, other known factors created with similar

It is also worth discussing the methodology here. The choice of using daily data and estimating the extended market model and the extended Fama-French model within a year for each stock has its pros and cons. On one hand, daily returns contains more irregular observations which cause the model to be fit more poorly and generate larger idiosyncratic volatility than monthly returns. On the other hand, when the models are estimated within a year, the parameters are allowed to vary over years, generating much more flexibility than monthly returns in a period of, say, 60 months. Overall, there is no obvious bias associated with using daily returns.

To what extent the idiosyncratic component of returns can be attributed to news? To answer this question, the daily news score constructed with (1) is added to the extended market model and the extended Fama-French model. The purpose, however, is to capture the idiosyncratic component of returns, rather than the systematic component as the news score is mostly firm-specific in the US market.[7] Unlike firm characteristics used in the literature which are persistent and return-predictive, the news score is transitory at daily frequency and contemporary with the returns. To capture the effect of potential illiquid stock trading, the one-day delayed news score is also added to the regressions.

Panel A of Table 3 reports the results of adding news scores to the extended market model. Comparing with the results in Panel A of Table 2, it can be seen that the distributions of factor betas and their t-ratios remain virtually unchanged, testifying that news scores capture mostly idiosyncratic component of the returns. The majority of the coefficient of the same-day news score is significant, while most of the coefficient of the one-day lagged news score is insignificant. The median of R^2s increases from 0.25 to 0.29, similar to the effect of the four additional Fama-French factors. The median of annualized idiosyncratic volatilities (IV) net of the part captured by new scores is 33.04, similar to that IV from the extended Fama-French five-factor model.

$\boxed{\text{Table 3 here}}$

methodology on return-predictive firm characteristics share the same property that they tend to explain average return, rather than return variations. Even factors extracted using principal component analysis (or similar approaches) end up contributing small proportions of return variation after the first principal component. Details are available upon requests.

[7]Dang, Moshirian and Zhang (2015) show that news commonality is low in mature markets.

Panel B of Table 3 reports the results of adding news scores to the extended Fama-French five-factor model. Again, the distributions of factor betas and their t-ratios remain almost the same. The majority of the coefficients of the same-day news score is significant, while most of the coefficients of the one-day lagged news score is insignificant. The median of R^2s increases from 0.30 in the extended Fama-French five-factor model to 0.34. The median of annualized idiosyncratic volatilities (IV) net of the part captured by new scores is 31.84.

While the effect the news score on idiosyncratic component of returns is not huge, the return variation it captures is similar to that captured by the four Fama-French factors additional to the market return. This marks an important advance compared to the result of Roll (1988). Back to 1980s, Roll does not have the luxury of news databases which not only cover news providers broadly, but also provide news scores. By comparing firm-days that have news and have no news, Roll concludes that there is no essential difference between the two types, implying that news cannot explain idiosyncratic component of the returns at all. The results presented here show that news scores can indeed explain part of idiosyncratic component of the returns. Further reconciliation with Roll's result follows in later sections.

5. Idiosyncratic Volatility and News Intensity

This section investigates whether the news intensity can explain idiosyncratic volatility. The news intensity is meant to be a firm characteristics, similar to idiosyncratic volatility. The variable NI in (2) calculated within a year for each firm in the news sample is adopted for this purpose. The dependent variables are the log of idiosyncratic volatility generated from the extended market model (5) and the extended Fama-French five-factor model (6), corresponding to the literature following Campbell et al (2001) and the (log of) relative idiosyncratic volatility generated from the extended market model (5) and the extended Fama-French five-factor model (6), corresponding to the literature following Morck et al (2000).

The explanatory variables include the log of NI (LNI), the main interest, plus a few firm-specific variables used in the literature to explain idiosyncratic volatility. They are the log of market value of equity (LME) which is known to be negatively associated with idiosyncratic volatility, the sample mean of past twelve quarters of return-on-equity (MRROE) and the log of sample volatility of past twelve quarters of return-on-equity (LVROE) which are used in Wei and Zhang (2006) to explain the average idiosyncratic volatility trend, the log of dollar trading volume (LVOL) which is commonly thought to be related to volatility in general and is particularly related to the findings that trading by institutional investors and retail investors is responsible for the increased average idiosyncratic volatility during 1960-2000, and the log of market-to-book ratio of assets (LMBA) used by Cao et al (2008) to capture the growth option effect on the average idiosyncratic volatility trend. A few related variables are also considered. One of them is the log of idiosyncratic volatility of return-on-equity (LIVROE), constructed by first regression ROE of an individual firm on the aggregated market ROE and using the sample volatility of the residual ROE to form idiosyncratic volatility of ROE, suggested by Irvine and Pontiff (2009) who argue that idiosyncratic return volatility should be explained by *idiosyncratic* ROE volatility, among others. By the same token, the log of relative idiosyncratic volatility of stock returns should be explained by the log of relative idiosyncratic volatility of ROE (LRIVROE). For growth options, a seemingly more relevant variable is the log of market-to-book ratio of equity (LMBE).

Table 4 reports the estimated transition matrix based on the news intensity. A set of firms in a base year are classified into five equal numbered categories according to their news intensity. For those remain listed next year, the proportions of each category of firms being re-classified in various categories are calculated and form a transition matrix for that year. The time-series average of the transition matrices of all the base years during 2000-2018 is reported. Panel A is for all the firms. Panel B is for firms that never enter the no-news sample. Panel C is for firms that always stay in the news-sample. From the numbers in the table, it can be seen that the news intensity is quite persistent. It can be viewed as a firm characteristic, like other firm characteristics.

$\boxed{\text{Table 4 here}}$

Panel A of Table 5 lists summary statistics of the explanatory variables. A median value of LME equal to 6.88 means that the median equity size of the firms is 972.6 million US dollars. A median value of LVOL equal to 7.67 means that the median annual trading volume of the firms is 2,143.1 million US dollars. A median value of LMBA equal to 0.35 means that the median book-to-market assets ratio of the firms is 1.4. A median value of LMBE equal to 0.68 means that the median book-to-market equity ratio of the firms is 2.0. The difference between market-to-book assets ratio and market-to-book equity ratio means that the firms tend to have considerable amount of debt (measured in their book values in the calculation of market-to-book assets).

Table 5 here

Panel B of Table 5 reports the correlations among explanatory variables that will be used in the regressions of LIV and LRIV. It is worth pointing out that LNI is quite highly positively correlated with LME and LVOL: large firms and liquid firms (in terms of trading volume) tend to be covered by news. LME and LVOL are highly positively correlated themselves. MROE is modestly, positively correlated with LME: larger firms tend to make larger profits. MROE is modestly, negatively correlated with LVROE and LIVROE: firms that have low profits tend to have large variations in their profits. LVROE and LIVROE is extremely positively correlated. The correlation between LMBA and LMBE is positive, but modest.

The regressions of the log of idiosyncratic volatility (LIV) on the log of the news intensity (LNI) and other control variables take the form of

$$\begin{aligned}\mathrm{LIV}_{iy} &= b_0 + b_1\mathrm{LNI}_{iy} + b_2\mathrm{LME}_{iy} + b_3\mathrm{LVROE}_{iy} + b_4\mathrm{MROE}_{iy} \\ &\quad + b_5\mathrm{LVOL}_{iy} + b_6\mathrm{LMBA}_{iy} + \varepsilon_{iy}, \qquad (7)\end{aligned}$$

where LNI and LVOL are measured in the same year y as LIV, LME and LMBA are measured at the beginning of the year y, and MROE and LVROE are calculated using twelve quarters of quarterly ROE data before year y. In certain regressions, LMBA is replaced by LMBE. In certain regressions, LVROE is replaced by LIVROE. In one regression, both LVROE and LIVROE are included for a contest.

Panel A of Table 6 reports the panel regression results of log idiosyncratic volatility LIV from the extended market model (7) on combinations of explanatory variables including the log of the news intensity LNI for the news sample. First of all, when LIV is regressed on LNI only, the slope coefficient is significantly negative: without controlling other characteristics of the firms, more news coverage corresponds to lower idiosyncratic volatility. But once the log of firm equity size LME is controlled for, the coefficient of LNI becomes significantly positive. LME itself has a significantly negative coefficients in regressions of all possible combinations. Trading volume positively explains idiosyncratic volatility. With both LME and LVOL in the regression, the economic and statistical significance of LNI reduces. The past sample mean of ROE negatively explains idiosyncratic volatility, while the past sample volatility of ROE positively explains idiosyncratic volatility, as Wei and Zhang (2006) show for the sample 1976-2000 for their explanatory power for the idiosyncratic volatility trend. The log of idiosyncratic volatility of ROE, LIVROE, also positively explains LIV, but its explanatory power is slightly inferior to that of LVROE, despite its theoretical advantage claimed by Irvine and Pontiff (2009). In fact, when both LVROE and LIVROE are used in one regression, the sign of LIVROE changes. LMBA positively explains LIV, as shown by Cao et al (2008) in explaining the idiosyncratic volatility trend, and so does LMBE, but their importance ranks low compared with other variables, both evidenced by their t-statistics and their contribution in R^2.

Table 6 here

Panel B of Table 6 reports the panel regression results of log of idiosyncratic volatility from the extended Fama-French five factor model (7) on combinations of explanatory variables including the log of the news intensity for the news sample. Since the additional four Fama-French factors to the market factor do not explain much variations of the returns, the idiosyncratic volatilities of generated from the two models are closely related. Therefore, the results are not expected to differ qualitatively. Indeed, the results in Panel B are very similar to those in Panel A. The estimated coefficients in Panel B tend to be slightly smaller in magnitude, while the t-ratios tend to be slightly greater in absolute value.

The results presented in Table 5 are estimated with panel regression with year fixed effect and with firm-year clustering for calculating standard errors. Alternatively, the model can also be estimated with Fama-MacBeth cross-sectional regressions in which the model is estimated in the cross section within each year to generate coefficients for the year and the time-series averages of the annually estimated coefficients are then treated as the final coefficient estimates. The results from Fama-MacBeth cross-sectional regressions are qualitative the same. They are relegated to Appendix.

For the log of relative idiosyncratic volatility of the news sample, the regressions take the form of

$$\begin{aligned}\mathrm{LRIV}_{iy} &= b_0 + b_1\mathrm{LNI}_{iy} + b_2\mathrm{LME}_{iy} + b_3\mathrm{LIVROE}_{iy} + b_4\mathrm{MROE}_{iy} \\ &\quad + b_5\mathrm{LVOL}_{iy} + b_6\mathrm{LMBA}_{iy} + \varepsilon_{iy}, \qquad (8)\end{aligned}$$

where the main difference on the right-hand side is to use log of relative idiosyncratic volatility of ROE to match the log of relative idiosyncratic volatility of return on the left.

Panel A of Table 7 reports the panel regression results of the log relative idiosyncratic volatility from the extended market model (8) on combinations of explanatory variables including news intensity for the news sample. Like the regression for LIV, when LRIV is regressed on LNI only, the slope is negative. But, similarly, once the log of firm equity size LME is controlled for, the coefficient of LNI becomes significantly positive. LME itself has a significantly negative coefficients in regressions of all possible combinations. A major difference between the regressions of LRIV in Table 6 and LIV in Table 5 is that the coefficient of the log of trading volume to explain LRIV is significantly negative, hinting that trading volume explains (the systematic component of) the total volatility more than idiosyncratic volatility. The explanatory power of the log of the news intensity is not much affected by the log of the trading volume. The past sample mean of ROE still negatively explains the relative idiosyncratic volatility and the past sample relative volatility of ROE positively explains idiosyncratic volatility, with reduced explanatory power. Both LMBA and LMBE have stronger explanatory power for LRIV.

Table 7 here

Panel B of Table 7 reports the panel regression results of log relative idiosyncratic volatility from the extended Fama-French five factor model (8) on combinations of explanatory variables including the log of the news intensity for the news sample. The results are again very similar to those in Panel A. Again, the estimated coefficients in Panel B tend to be slightly smaller in magnitude, but the t-ratios also tend to be slightly smaller in absolute value.

Overall, the results presented in this section using the news intensity collaborate with those in the last section using the news score. News coverage contributes positively to absolute and relative idiosyncratic volatility, once firm size is controlled for. For the absolute idiosyncratic volatility, controlling trading volume is also important. The contribution of the news intensity to the idiosyncratic volatility is independent of other firm characteristics.

6. Idiosyncratic Volatility for the No-news Firm-years

For firm-years that have less than 25 news days per year, quantitative analyses of the effect of news coverage on idiosyncratic volatility are impractical. However, characterizing such firm-years, or especially firms, is still important for understanding idiosyncratic volatility. After all, news coverage is just one, though important one, channel for information about firms to be revealed to the public. For that reason, we examine the no-news sample and provide its characteristics.

Table 8 reports the summary statistics similar to those in Table 1. Panel A of Table 8 describes the firm-day returns of the individual stocks of the no-news sample and Fama-French factors during the no-news firm-years. It is easily seen that the individual returns are more volatile, more positively skewed, and above all, more fat-tailed. Figure 1 plots the histogram of the panel daily returns of the no-news sample along with that of the news sample.

Table 8 here

Figure 1 here

The Fama-French factor during the no-news firm-years are also more volatile than the news firm-years, but with smaller kurtosis. From Panel B, the relevance of news pieces in the no-news sample is slightly higher than that of the news sample and the freshness is even higher. Apparently, these firms are less followed by news agencies and, as a result, there are less staled news reports. The distribution of the event sentiment score for the no-news sample looks almost exactly the same as that of the news sample. Panel C reveals that, when news pieces are aggregated within a day, the sharp difference between the no-news sample with the news sample shows up. There are about 1% of the firm-days only with bad news and about 2% of the firm-days only with good news. On more than 95% of the firm-days there are no news at all.

Panel A of Table 9 reports the panel statistics of the estimated regression coefficients, their t-ratios, regression R^2, idiosyncratic volatility (IV), and the log of idiosyncratic volatility (LIV) from the extended market model for the no-news sample. The betas of the individual returns with respect to the same-day returns on the market portfolio and on the portfolio of finely defined industry tend to be smaller than those of the news sample, while the beta with respect to the lagged market return has a wide distribution. The median of R^2s is 0.06, much smaller than that of the news sample. The median of annualized idiosyncratic volatilities (IV) is 52.94, much greater than those of the news sample. Panel B of Table 9 reports the summary statistics of the estimated regression coefficients, their t-ratios, regression R^2, idiosyncratic volatility (IV), and the log of idiosyncratic volatility (LIV) from the extended Fama-French five-factor model for the no-news sample. As in the news sample, the additional four Fama-French factors do not add much explanatory power for return variations. The median of R^2s is 0.09, much smaller than those of the news sample. The additional four factors of the Fama-French factors reduce the median idiosyncratic volatility from 52.94 of the extended market model to 51.75 only.

Table 9 here

Panel A of Table 10 reports summary statistics of the explanatory variables for the firm-years of the no-news sample. A median value of LME equal to 4.73 means that the average median equity size of the firms is 113.3 million US dollars, much smaller than that of the news sample. A median value of LVOL equal to 4.29 means that the median annual trading volume of the firms is 73.0 million US dollars, also much smaller than that of the news sample. The median ROE is lower and the median volatility of ROE is higher than those of the news sample respectively. A median value of LMBA equal to 0.12 means that the median book-to-market assets ratio of the firms is 1.1. A mean median value of LMBE equal to 0.36 means that the median book-to-market equity ratio of the firms is 1.4. These numbers show that there are relatively more value firms in the no-news sample than the news sample.

Table 10 here

The regressions of LIV and LRIV on explanatory variables, such as those in Tables 6 and 7 for the news sample can also be conducted on the no-news sample, except that explanatory variables do not include the news intensity. Such regressions are performed and the results can be roughly described as follows. For LIV from the extended market model, the signs of all the coefficients are the same as those for the news sample. LME, LVOL and LMBA are less useful in explaining idiosyncratic volatility, while MROE and LVROE (LIVROE) are more useful. LIVROE appears to have greater t-ratios than LVROE, but when the two variables are used together, it is still LVROE that wins out. The results for LIV from the extended Fama-French five factor model for the no-news sample tell the same story. For LRIV from either the extended market model or the extended Fama-French five factor model, MROE lost its explanatory power. LRIVROE and LMBA (LMBE) are marginally significant, while LME and LVOL remain to be strong. The details are relegated to Appendix.

7. Conclusion

Roll (1988) examines NYSE/AMEX firms during 1982-1987. By comparing news days and no-news days of the same set of firms and finding no significant difference in idiosyncratic volatility between the two types of days, he concludes that idiosyncratic volatility is not explainable by corporate news. Using information about news coverage provided by RavenPack, I document that news coverage contributes to idiosyncratic volatility for firm-years that have sufficient news coverage. This is achieved by separating firm-years into two groups: one with at least 25 news days per year (the news sample) and the other less than that (the no-news sample). The firms in the no-news sample tend to be smaller, less traded, more value than growth oriented, with lower earnings and higher earnings volatility. These features determine that firms in this group have greater idiosyncratic volatility than the firms in the news sample. 41% of the firm-years falls into this group. As such, without separating the two groups, the relationship between idiosyncratic volatility and news coverage will be obscured.

Within the news sample, the news score at the daily frequency explains about 4% of the total return variations in either the extended market model or the extended Fama-French five-factor model. The news intensity at the firm-year level significantly explains both the log of absolute idiosyncratic volatility and the log of relative idiosyncratic volatility, once equity size is controlled for. Beside equity size, trading volume also affects the explanatory power of the news intensity as they are positively correlated. Other variables used in the literature to explain idiosyncratic volatility in earlier samples continue to be useful in the current sample of 2000-2019. These include earnings related variables, which are more useful for the absolute idiosyncratic volatility, and growth option variables, which are more useful for the relative idiosyncratic volatility.

In the extant literature, cross-sectional differences in idiosyncratic volatility are explained by firm characteristics, such as firm size, growth options, etc. The relationship of idiosyncratic volatility with these firm characteristics can have both rational asset pricing interpretations and behavioral interpretations. The contribution of the finding in this paper to the literature is that it is the first to lend support to the notion that higher idiosyncratic volatility can be partially due to more firm-specific information being impounded in the stock prices, a notion consistent with rational asset pricing. While many cross-country studies seem to align with the more general point that legal protection for investors and transparency of the market environment lead to lower market model R^2s, hence higher relative idiosyncratic volatility, direct evidence between information and idiosyncratic volatility within a country has never been documented. Therefore, the finding in this paper marks a breakthrough in this direction.

References

Ang, A., R. Hodrick, Y. Xing, X. Zhang, 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259-299.

Ang, A., R. Hodrick, Y. Xing, X. Zhang, 2009. High idiosyncratic volatility and low returns: international and further U.S. evidence. Journal of Financial Economics 91, 1-23.

Brandt, M. W., A. Brav; J. Graham, and A. Kumar, 2010. The idiosyncratic volatility puzzle: time trend or speculative episodes? Review of Financial Studies, 23, 863-899.

Brown, G., and N. Kapadia, 2007. Firm-specific risk and equity market development. Journal of Financial Economics, 84, 358-388.

Campbell, J. Y., M. Lettau, B. G. Malkiel and Y. Xu, 2001. Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. Journal of Finance, 56, 1-43.

Cao, C., T. Simin, and J. Zhao, 2008. Can growth options explain the trend in idiosyncratic risk? Review of Financial Studies, 21, 2599-2633.

Chan, K. and A. Hameed, 2006. Stock price synchronicity and analyst coverage in emerging markets. Journal of Financial Economics, 80, 115-147.

Dai, L., J. Parwada, and B. Zhang, 2015. The governance role of the media through news dissemination: Evidence from insider trading. Journal of Accounting Research, 53, 331-366.

Dang, T., F. Moshirian, and B. Zhang, 2015. Commonality in news around the world. Journal of Financial Economics,

Dasgupta S, J. Gan, and N. Gao, 2010. Transparency, price informativeness, and stock return synchronicity: theory and evidence. Journal of Financial and Quantitative Analysis, 45, 1189-1220.

Fama, E. F., and K. R. French, 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33, 3-56.

Fama, E. F., and K. R. French, 2015. A five-factor asset pricing model, Journal of Financial Economics, 116, 1-22.

Fang, L. and J. Peress, 2009. Media Coverage and the Cross-section of Stock Returns, Journal of Finance 64, 2023-2052.

Hameed, A., R. Morck, J. Shen, and B. Yeung, 2015, Information, analysts, and stock return comovement, Review of Financial Studies, 28, 3153-3187.

Irvine, P. J. and J. Pontiff, 2009. Idiosyncratic return volatility, cash flows, and product market competition. Review of Financial Studies, 22, 1149-1177.

Jin, L., and S. C. Myers, 2006. R2 around the world: new theory and new tests. Journal of Financial Economics, 79, 257-292.

Morck, R., B. Yeung, and W. Yu, 2000. The information content of stock markets: why do emerging markets have synchronous stock price movements? Journal of Financial Economics, 58, 215-260.

Piotroski J. and D. Roulstone, 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. Accounting Review 79, 1119-1151.

Shroff, N., Verdi, R.S., and Yu, G., 2014. Information environment and the investment decisions of multinational corporations. Accounting Review 89, 759-790.

Teoh, S.H., Yang, Y.G. and Zhang, Y. 2009. R-Square and Market Efficiency. Working paper.

Wei, S. X., and C. Zhang, 2006. Why did individual stocks become more volatile? Journal of Business, 79, 259-292.

Xu, Y. and B. G. Malkiel, 2003. Investigating the behavior of idiosyncratic volatility, Journal of Business, 76, 613-644.

Yang, Y.C., Zhang, B. and Zhang, C., 2020, Is information risk priced? evidence from abnormal idiosyncratic volatility, Journal of Financial Economics, 135, 528-554.

Zhang, C., 2010. A reexamination of the causes of time-varying stock return volatilities. Journal of Financial and Quantitative Analysis, 45, 663-684.

Table 1
Summary statistics of news and returns (news sample)

This table reports summary panel statistics of news measures and daily returns for firm-years that have at least 25 news days during the year, as well as the summary time-series statistics of factors. Panel A reports the mean, standard deviation, skewness, kurtosis and percentiles of daily returns (in percent) and Fama-French five-factors for the same firm-years. MKT is the market excess return. SMB is the small-minus-big return. HML is the high-minus-low book-to-market return. RMW is the robust-minus-weak profitability return, and CMA is the conservative-minus-aggressive investment return. Panel B reports the same statistics of news items at the level of news pieces. REL is the relevance to firm value. ENS is the freshness of the news. ESS is the event sentiment score. Panel C reports the same statistics of daily news score NS defined in (1) and its logarithm LNS, and news intensity NI defined in (2) and its logarithm LNI. K is the number of news pieces of a firm-day. The sample period is 2000-2019, with 2152.5 firms per year on average.

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
A. Daily returns									
r_{id}	0.05	3.73	10.08	1154.63	-4.59	-1.27	0.00	1.27	4.72
MKT_d	0.03	1.17	-0.15	12.63	-1.80	-0.43	0.07	0.56	1.64
MKT_{d-1}	0.03	1.17	-0.15	12.63	-1.80	-0.43	0.07	0.56	1.64
SMB_d	0.01	0.57	0.04	6.60	-0.88	-0.33	0.01	0.34	0.87
HML_d	0.00	0.65	0.57	13.07	-0.88	-0.29	-0.01	0.26	0.92
RMW_d	0.02	0.43	0.07	8.53	-0.62	-0.21	0.01	0.23	0.67
CMA_d	0.00	0.36	-0.45	16.31	-0.51	-0.18	-0.01	0.18	0.56
B. News pieces									
REL_{ij}	96.73	15.84	-4.64	22.51	100.00	100.00	100.00	100.00	100.00
ENS_{ij}	73.63	34.45	-0.97	2.47	1.00	56.00	100.00	100.00	100.00
ESS_{ij}	53.50	13.72	-0.13	3.64	28.00	49.00	50.00	63.00	78.00
C. Daily news									
NS_{id}	0.06	0.50	7.50	269.72	-0.24	0.00	0.00	0.00	0.56
LNS_{id}	0.04	0.25	2.01	19.93	-0.22	0.00	0.00	0.00	0.44
NI_{id}	0.18	0.64	10.37	254.49	0.00	0.00	0.00	0.02	0.97
LNI_{id}	0.11	0.28	3.60	19.15	0.00	0.00	0.00	0.02	0.68
K_{id}	1.46	6.24	25.77	1920.23	0.00	0.00	0.00	1.00	7.00

Table 2
The extended market model and Fama-French five-factor model (news sample)

This table reports the summary panel statistics of the coefficients and their t-statistics of the extended market model (5) and the extended Fama-French five-factor model (6) estimated using daily returns within a year for all firm-years that have at least 25 news days per year. It also reports the summary panel statistics of regression R^2, estimated idiosyncratic volatility, log idiosyncratic volatility and log relative idiosyncratic volatility. The sample period is 2000-2019, with 2152.5 firms per year on average.

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
A. The extended market model									
Const.	0.01	0.22	-0.03	27.33	-0.31	-0.07	0.01	0.10	0.31
t	0.09	0.95	-0.05	2.90	-1.48	-0.55	0.10	0.75	1.64
MKT_d	0.52	0.78	0.07	22.18	-0.52	0.04	0.45	0.92	1.81
t	1.64	2.55	0.02	5.31	-2.32	0.16	1.55	3.10	5.84
MKT_{d-1}	0.02	0.51	0.48	78.80	-0.66	-0.20	0.00	0.21	0.77
t	-0.02	1.19	-0.06	3.32	-1.98	-0.79	-0.01	0.76	1.91
IND_d	0.55	0.70	0.53	21.30	-0.47	0.15	0.53	0.96	1.64
t	3.70	6.37	13.17	410.21	-1.17	0.55	2.21	5.14	13.34
IND_{d-1}	0.03	0.43	1.37	128.10	-0.53	-0.13	0.02	0.18	0.63
t	0.09	1.12	0.01	3.32	-1.75	-0.64	0.10	0.82	1.94
R^2	0.29	0.21	0.65	2.59	0.02	0.11	0.25	0.43	0.68
IV	42.49	33.31	4.35	50.23	12.98	22.31	33.82	52.08	99.81
LIV	3.54	0.63	0.17	3.44	2.56	3.11	3.52	3.95	4.60
LRIV	1.27	1.42	0.48	3.39	-0.78	0.28	1.11	2.12	3.88
B. The extended Fama-French five-factor model									
Const.	0.02	0.22	0.48	30.68	-0.28	-0.06	0.02	0.10	0.33
t	0.16	0.93	-0.04	2.95	-1.37	-0.47	0.16	0.80	1.66
MKT_d	0.45	0.77	-0.86	55.00	-0.61	0.03	0.43	0.85	1.62
t	1.33	2.16	-0.01	5.65	-2.00	0.09	1.27	2.55	4.86
MKT_{d-1}	0.01	0.51	-0.10	110.16	-0.66	-0.20	0.00	0.20	0.75
t	-0.03	1.18	-0.08	3.32	-1.97	-0.80	-0.01	0.76	1.87
IND_{id}	0.48	0.70	0.92	32.69	-0.50	0.08	0.45	0.88	1.54
t	3.02	6.03	15.52	516.55	-1.14	0.27	1.64	4.10	11.60
$\text{IND}_{i,d-1}$	0.03	0.44	2.08	181.40	-0.52	-0.13	0.01	0.18	0.62
t	0.09	1.12	0.02	3.34	-1.76	-0.65	0.08	0.82	1.92
SMB_d	0.63	0.77	0.20	13.95	-0.40	0.10	0.58	1.09	1.91
t	1.89	2.34	0.35	4.01	-1.60	0.34	1.73	3.28	5.90
HML_d	0.01	0.83	-0.67	23.93	-1.26	-0.34	0.02	0.40	1.21
t	0.12	1.63	0.50	5.57	-2.38	-0.89	0.07	1.06	2.78
RMW_d	-0.21	1.03	-1.30	22.62	-2.03	-0.59	-0.06	0.33	1.06
t	-0.14	1.46	-0.14	3.85	-2.50	-1.06	-0.13	0.79	2.21
CMA_d	-0.02	1.17	1.88	115.66	-1.70	-0.53	0.00	0.48	1.61
t	-0.01	1.38	-0.10	4.37	-2.24	-0.87	-0.01	0.86	2.16
R^2	0.33	0.21	0.47	2.37	0.04	0.15	0.30	0.48	0.71
IV	41.25	32.81	4.40	51.22	12.39	21.47	32.66	50.44	97.69
LIV	3.51	0.64	0.17	3.42	2.52	3.07	3.49	3.92	4.58
LRIV	0.95	1.24	0.26	3.07	-0.92	0.08	0.83	1.74	3.18

Table 3
The extended market model and Fama-French five-factor model with news score (news sample)

This table reports the summary panel statistics of the coefficients and their t-statistics of the extended market model (5) and the extended Fama-French five-factor model (6) with news score, estimated using daily returns within a year for all firm-years that have at least 25 news days per year. It also reports the summary panel statistics of regression R^2, estimated idiosyncratic volatility, log idiosyncratic volatility and log relative idiosyncratic volatility. The sample period is 2000-2019, with 2152.5 firms per year on average.

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
A. The extended market model									
Const.	-0.03	0.22	-0.44	42.97	-0.36	-0.11	-0.02	0.06	0.26
t	-0.15	0.96	-0.06	3.11	-1.75	-0.78	-0.13	0.50	1.42
MKT_d	0.52	0.77	0.03	24.76	-0.51	0.05	0.46	0.92	1.80
t	1.68	2.55	0.04	5.20	-2.30	0.19	1.58	3.14	5.89
MKT_{d-1}	0.02	0.50	0.31	78.55	-0.65	-0.20	0.00	0.21	0.76
t	-0.02	1.19	-0.06	3.31	-1.97	-0.79	-0.01	0.77	1.91
IND_{id}	0.55	0.70	0.69	30.42	-0.47	0.15	0.53	0.95	1.62
t	3.72	6.34	13.01	402.96	-1.17	0.56	2.22	5.19	13.32
$\text{IND}_{i,d-1}$	0.03	0.43	1.84	135.85	-0.52	-0.13	0.02	0.18	0.62
t	0.09	1.12	0.02	3.28	-1.75	-0.64	0.10	0.82	1.93
NS_{id}	1.81	3.62	6.64	209.18	-1.27	0.19	1.04	2.61	7.09
t	1.99	2.39	0.46	3.94	-1.51	0.40	1.77	3.39	6.22
$\text{NS}_{i,d-1}$	0.02	1.89	-23.10	3569.35	-1.99	-0.47	0.00	0.50	2.13
t	0.01	1.21	0.05	3.69	-1.94	-0.77	-0.01	0.78	1.98
R^2	0.32	0.20	0.55	2.57	0.04	0.15	0.29	0.45	0.70
IV	41.51	32.41	4.26	48.70	12.81	21.81	33.04	50.77	97.72
LIV	3.52	0.63	0.18	3.43	2.55	3.08	3.50	3.93	4.58
LRIV	1.00	1.20	0.31	3.45	-0.82	0.18	0.90	1.72	3.17

Table 3 (cont'd)

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
B. The extended Fama-French five-factor model									
Const.	-0.01	0.22	-0.05	43.50	-0.33	-0.09	-0.01	0.07	0.28
t	-0.08	0.94	-0.05	3.16	-1.64	-0.70	-0.08	0.55	1.44
MKT_d	0.46	0.76	-1.02	67.50	-0.59	0.03	0.43	0.85	1.61
t	1.36	2.17	0.01	5.54	-1.96	0.11	1.29	2.59	4.92
MKT_{d-1}	0.01	0.50	-0.24	111.70	-0.65	-0.20	0.00	0.20	0.73
t	-0.03	1.18	-0.09	3.31	-1.97	-0.79	-0.02	0.76	1.87
IND_{id}	0.48	0.69	1.38	54.55	-0.50	0.08	0.45	0.87	1.53
t	3.03	6.00	15.35	507.61	-1.15	0.27	1.65	4.13	11.63
$\mathrm{IND}_{i,d-1}$	0.03	0.43	2.64	191.58	-0.51	-0.13	0.01	0.17	0.61
t	0.09	1.12	0.02	3.31	-1.75	-0.65	0.08	0.82	1.93
SMB_d	0.63	0.77	0.05	14.91	-0.39	0.10	0.58	1.08	1.90
t	1.92	2.35	0.35	3.95	-1.58	0.35	1.76	3.34	5.97
HML_d	0.01	0.83	-0.31	34.49	-1.25	-0.34	0.02	0.39	1.20
t	0.12	1.64	0.50	5.57	-2.39	-0.89	0.07	1.06	2.80
RMW_d	-0.21	1.01	-1.30	16.62	-2.01	-0.59	-0.06	0.33	1.04
t	-0.14	1.47	-0.13	3.85	-2.50	-1.07	-0.13	0.80	2.23
CMA_d	-0.02	1.15	1.99	128.56	-1.68	-0.52	0.00	0.48	1.59
t	-0.01	1.38	-0.10	4.35	-2.23	-0.86	0.00	0.86	2.18
NS_{id}	1.81	3.62	6.51	213.62	-1.27	0.19	1.04	2.61	7.09
t	2.02	2.41	0.46	3.93	-1.53	0.42	1.80	3.44	6.29
$\mathrm{NS}_{i,d-1}$	0.02	1.87	-23.57	3485.16	-1.98	-0.46	0.00	0.50	2.10
t	0.01	1.21	0.05	3.69	-1.95	-0.77	0.00	0.78	2.00
R^2	0.36	0.20	0.37	2.37	0.06	0.19	0.34	0.50	0.72
IV	40.27	31.91	4.32	49.80	12.21	21.00	31.84	49.14	95.74
LIV	3.48	0.63	0.19	3.42	2.50	3.04	3.46	3.89	4.56
LRIV	0.74	1.11	0.19	3.31	-0.96	-0.01	0.66	1.43	2.73

Table 4
Transition matrix of news intensity
This table reports estimated annual transition matrices for firms classified with equal numbers in five categories according to their news intensity (NI). Panel A is for all firms that existed in the base year and the next year, with 3314.9 firms per year on average. Panel B is for firms that never entered the no-news sample, with 2958.0 firms per year on average. Panel C is for firms that always stay in the news sample, with 1144.9 firms per year on average.

A. All firms					
Category	Lowest NI	2	3	4	Highest NI
Lowest NI	0.78	0.18	0.03	0.01	0.00
2	0.20	0.50	0.24	0.06	0.00
3	0.04	0.25	0.45	0.23	0.02
4	0.01	0.06	0.24	0.53	0.16
Highest NI	0.00	0.01	0.02	0.16	0.81
B. Firms that are never in the no-news sample					
Category	Lowest NI	2	3	4	Highest NI
Lowest NI	0.69	0.24	0.06	0.01	0.00
2	0.24	0.45	0.25	0.06	0.00
3	0.06	0.24	0.44	0.24	0.02
4	0.01	0.06	0.23	0.53	0.16
Highest NI	0.00	0.01	0.02	0.15	0.82
C. Firms that are always in the news-sample					
Category	Lowest NI	2	3	4	Highest NI
Lowest NI	0.57	0.30	0.10	0.03	0.00
2	0.20	0.42	0.30	0.07	0.00
3	0.05	0.21	0.46	0.26	0.02
4	0.01	0.03	0.16	0.61	0.20
Highest NI	0.00	0.00	0.01	0.12	0.88

Table 5
Summary statistics of firm-specific variables (news sample)
Panel A of the table reports the summary panel statistics of the firm-year log idiosyncratic volatility, log relative idiosyncratic volatility, and variables to be used in regressions to explain them, plus a few related variables. Panel B reports the correlation matrix. NI is news intensity, LNI is log news intensity, LME is log market equity, LVROE is log volatility of ROE, LIVROE is log idiosyncratic volatility of return-on-equity (ROE) calculated as the sample standard deviation of previous twelve quarterly ROEs, LRIVROE is log relative idiosyncratic volatility of return-on-equity, calculated similarly using the ROE residual from a market model of ROE. MROE is sample mean of previous twelve quarterly ROEs, LVOL is log trading volume, LMBA is log market-to-book assets ratio, and LMBE is log market-to-book equity ratio. The sample period is 2000-2019, with 2152.5 firms per year on average.

A. Summary statistics of variables

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
NI	47.15	65.18	8.05	122.07	9.83	17.27	28.94	53.91	134.80
LNI	3.50	0.79	0.63	3.46	2.38	2.91	3.40	4.01	4.91
LME	6.93	1.94	0.14	2.88	3.80	5.58	6.88	8.20	10.22
LVROE	2.63	2.35	1.24	5.98	-0.39	1.05	2.22	3.78	7.05
LIVROE	1.85	2.26	0.04	2.48	-1.75	0.23	1.76	3.49	5.68
LRIVROE	1.62	2.26	0.05	2.51	-1.98	0.00	1.54	3.24	5.45
MROE	0.42	1.47	-0.97	3.56	-2.69	-0.25	0.84	1.40	2.15
LVOL	7.48	2.31	-0.29	2.63	3.31	5.92	7.67	9.16	10.96
LMBA	0.48	0.54	1.12	4.43	-0.12	0.07	0.35	0.78	1.55
LMBE	0.76	0.81	0.66	5.23	-0.37	0.23	0.68	1.22	2.16

B. Correlation matrix: Pearson \ Spearman

	LNI	LME	LVROE	LIVROE	LRIVROE	MROE	LVOL	LMBA	LMBE
LNI		0.62	-0.04	-0.04	0.00	0.17	0.69	0.14	0.17
LME	0.66		-0.24	-0.24	-0.02	0.44	0.94	0.22	0.27
LVROE	-0.03	-0.24		0.99	-0.04	-0.28	-0.11	0.16	0.17
LIVROE	-0.03	-0.24	0.99		0.07	-0.28	-0.11	0.16	0.17
LRIVROE	0.00	-0.01	-0.04	0.06		0.01	-0.03	0.00	0.00
MROE	0.17	0.43	-0.42	-0.41	0.01		0.38	0.21	0.26
LVOL	0.70	0.93	-0.10	-0.11	-0.02	0.34		0.25	0.29
LMBA	0.12	0.19	0.19	0.19	0.00	0.00	0.23		0.93
LMBE	0.16	0.26	0.20	0.21	0.01	0.05	0.27	0.87	

Table 6
Explaining idiosyncratic volatility (news sample)

This table reports the panel regressions with year fixed effect and firm-year clustering of log idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on new intensity NI and other control variables for the news sample. LME is log market equity, LVROE is log volatility of ROE, LIVROE is log idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. R^2 (R^2_*) is the R^2 with (without) year fixed effect. The sample period is 2000-2019, with 2152.5 firms per year on average.

A. LIV from the extended market model											
LNI	-0.28		0.34	0.15		0.16	0.16	0.15	0.16	0.16	0.16
	(-24.1)		(16.0)	(11.0)		(11.7)	(11.4)	(10.8)	(11.5)	(11.2)	(11.8)
LME		-0.20	-0.30	-0.40	-0.38	-0.40	-0.40	-0.40	-0.40	-0.41	-0.40
		(-46.4)	(-37.1)	(-43.5)	(-35.6)	(-47.3)	(-43.2)	(-44.1)	(-48.3)	(-44.1)	(-47.5)
LVROE				0.05	0.05	0.04	0.04				0.08
				(15.5)	(14.4)	(13.4)	(12.9)				(3.8)
LIVROE								0.04	0.04	0.04	-0.04
								(14.7)	(13.2)	(12.8)	(-2.0)
MROE				-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05
				(-18.7)	(-21.6)	(-19.2)	(-19.2)	(-18.7)	(-19.0)	(-19.3)	(-19.3)
LVOL				0.16	0.18	0.16	0.16	0.17	0.16	0.16	0.16
				(15.3)	(19.0)	(16.7)	(15.9)	(15.5)	(17.1)	(16.2)	(16.7)
LMBA					0.10	0.11			0.11		0.11
					(6.5)	(7.5)			(7.6)		(7.9)
LMBE							0.04			0.04	
							(3.9)			(3.9)	
R^2	0.22	0.53	0.59	0.72	0.71	0.73	0.72	0.72	0.72	0.72	0.73
R^2_*	0.13	0.47	0.54	0.68	0.68	0.69	0.69	0.68	0.69	0.68	0.69
B. LIV from the extended Fama-French five-factor model											
LNI	-0.28		0.35	0.16		0.17	0.17	0.16	0.17	0.17	0.17
	(-23.2)		(16.7)	(11.7)		(12.4)	(12.1)	(11.5)	(12.2)	(11.9)	(12.6)
LME		-0.20	-0.31	-0.40	-0.38	-0.40	-0.41	-0.40	-0.41	-0.41	-0.40
		(-47.7)	(-37.5)	(-41.9)	(-33.8)	(-45.5)	(-41.8)	(-42.6)	(-46.5)	(-42.8)	(-45.5)
LVROE				0.05	0.05	0.04	0.04				0.08
				(15.1)	(14.7)	(13.3)	(12.8)				(4.0)
LIVROE								0.05	0.04	0.04	-0.04
								(14.3)	(12.9)	(12.6)	(-1.9)
MROE				-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05
				(-18.8)	(-21.3)	(-19.2)	(-19.3)	(-18.8)	(-19.0)	(-19.5)	(-19.3)
LVOL				0.16	0.18	0.16	0.16	0.17	0.16	0.16	0.16
				(14.6)	(18.0)	(15.9)	(15.2)	(14.8)	(16.2)	(15.5)	(15.9)
LMBA					0.11	0.12			0.12		0.12
					(6.8)	(8.0)			(8.0)		(8.3)
LMBE							0.04			0.04	
							(4.0)			(4.1)	
R^2	0.22	0.52	0.59	0.72	0.71	0.73	0.72	0.71	0.72	0.72	0.73
R^2_*	0.12	0.46	0.54	0.68	0.68	0.69	0.69	0.68	0.69	0.68	0.69

Table 7
Explaining relative idiosyncratic volatility (news sample)

This table reports the panel regressions with year fixed effect and firm-year clustering of log idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on new intensity NI and other control variables for the news sample. LME is log market equity, LRIVROE is log relative idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. R^2 (R^2_*) is the R^2 with (without) year fixed effect. The sample period is 2000-2019, with 2152.5 firms per year on average.

A. LRIV from the extended market model											
LNI	-0.88		0.43	0.43	0.42	0.53	0.51	0.51		0.53	0.52
	(-19.3)		(10.6)	(10.6)	(10.3)	(13.0)	(12.7)	(12.7)		(13.0)	(12.6)
LME		-0.51	-0.64	-0.64	-0.63	-0.50	-0.47	-0.47	-0.39	-0.45	-0.47
		(-48.5)	(-41.5)	(-41.4)	(-40.0)	(-16.5)	(-15.0)	(-15.1)	(-10.9)	(-15.0)	(-16.1)
LRIVROE				0.01	0.01	0.01		0.01	0.01	0.01	0.01
				(2.2)	(2.2)	(1.8)		(1.8)	(1.9)	(2.1)	(2.1)
MROE					-0.03		-0.04	-0.04	-0.06	-0.03	-0.03
					(-2.5)		(-4.0)	(-4.0)	(-4.7)	(-3.1)	(-3.5)
LVOL						-0.15	-0.16	-0.16	-0.11	-0.20	-0.18
						(-5.3)	(-5.7)	(-5.7)	(-3.5)	(-7.2)	(-6.6)
LMBA									0.32	0.34	
									(11.1)	(12.3)	
LMBE											0.18
											(12.1)
R^2	0.28	0.57	0.59	0.59	0.59	0.59	0.60	0.60	0.59	0.61	0.61
R^2_*	0.22	0.53	0.55	0.55	0.55	0.56	0.56	0.56	0.55	0.58	0.57
B. LRIV from the extended Fama-French five-factor model											
LNI	-0.74		0.41	0.41	0.40	0.49	0.47	0.47		0.49	0.48
	(-18.9)		(10.5)	(10.5)	(10.3)	(12.4)	(12.3)	(12.3)		(12.5)	(12.2)
LME		-0.44	-0.56	-0.56	-0.55	-0.45	-0.43	-0.43	-0.35	-0.42	-0.43
		(-43.3)	(-35.7)	(-35.6)	(-35.6)	(-16.0)	(-14.7)	(-14.7)	(-10.9)	(-14.8)	(-15.8)
LRIVROE				0.01	0.01	0.01		0.01	0.01	0.01	0.01
				(2.0)	(2.0)	(1.7)		(1.7)	(1.8)	(1.9)	(1.9)
MROE					-0.02		-0.03	-0.03	-0.04	-0.02	-0.02
					(-1.7)		(-2.9)	(-2.9)	(-3.6)	(-2.0)	(-2.3)
LVOL						-0.12	-0.13	-0.13	-0.08	-0.16	-0.14
						(-4.4)	(-4.6)	(-4.6)	(-2.6)	(-6.0)	(-5.5)
LMBA									0.29	0.31	
									(11.0)	(12.1)	
LMBE											0.17
											(11.8)
R^2	0.26	0.54	0.56	0.56	0.56	0.57	0.57	0.57	0.56	0.58	0.58
R^2_*	0.20	0.50	0.52	0.52	0.52	0.53	0.53	0.53	0.52	0.55	0.54

Table 8
Summary statistics of news and returns (no-news sample)

This table reports summary panel statistics of news measures and daily returns for firm-years that have less than 25 news days during the year, as well as the summary time-series statistics of factors. Panel A reports the mean, standard deviation, skewness, kurtosis and percentiles of daily returns (in percent) on individual stocks and Fama-French five-factors for the same firm-years. MKT is the market excess return. SMB is the small-minus-big return. HML is the high-minus-low book-to-market return. RMW is the robust-minus-weak profitability return, and CMA is the conservative-minus-aggressive investment return. Panel B reports the same statistics of news items at the level of news pieces. REL is the relevance to firm value. ENS is the freshness of the news. ESS is the event sentiment score. Panel C reports the same statistics of daily news score defined in (1) and news intensity defined in (2). K is the number of news pieces of a firm-day. The sample period is 2000-2019, with 1399.9 firms per year on average.

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
A. Daily returns									
r_{id}	0.08	5.80	16.94	3233.25	-6.91	-1.72	0.00	1.54	7.28
MKT_d	0.00	1.33	0.07	7.14	-2.14	-0.70	0.04	0.66	2.08
MKT_{d-1}	0.00	1.33	0.06	7.14	-2.14	-0.70	0.04	0.66	2.08
SMB_d	0.03	0.67	-0.51	6.96	-0.98	-0.33	0.05	0.41	1.04
HML_d	0.05	0.74	0.16	8.25	-1.09	-0.30	0.03	0.35	1.27
RMW_d	0.04	0.74	0.26	6.38	-1.08	-0.32	0.02	0.38	1.25
CMA_d	0.04	0.59	-0.89	12.49	-0.88	-0.21	0.02	0.31	1.00
B. New pieces									
REL_{ij}	99.98	1.11	-71.80	5156.12	100.00	100.00	100.00	100.00	100.00
ENS_{ij}	85.32	25.01	-1.71	4.95	24.00	75.00	100.00	100.00	100.00
ESS_{ij}	53.95	13.87	-0.12	3.97	33.00	50.00	50.00	64.00	76.00
C. Daily news									
	Mean	Std	Skew	Kurt	p1	p2	p97	p98	p99
NS_{id}	0.01	0.15	7.91	268.42	-0.04	0.00	0.00	0.21	0.50
LNS_{id}	0.01	0.09	4.51	95.54	-0.04	0.00	0.00	0.19	0.40
	Mean	Std	Skew	Kurt	p95	p96	p97	p98	p99
NI_{id}	0.02	0.17	13.11	266.65	0.00	0.00	0.20	0.38	0.75
LNI_{id}	0.02	0.10	8.43	87.94	0.00	0.00	0.18	0.32	0.56
K_{id}	0.15	1.05	16.95	876.39	0.00	1.00	1.00	2.00	4.00

Table 9
The extended market model and Fama-French five-factor model (no-news sample)

This table reports the summary panel statistics of the coefficients and their t-statistics of the extended market model (5) and the extended Fama-French five-factor model (6) estimated using daily returns within a year for all firm-years that have less than 25 news days per year. It also reports the summary panel statistics of regression R^2, estimated idiosyncratic volatility, log idiosyncratic volatility and log relative idiosyncratic volatility. The sample period is 2000-2019, with 1399.9 firms per year on average.

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
A. The extended market model									
Const.	0.06	0.54	13.00	914.57	-0.47	-0.07	0.05	0.19	0.60
t	0.25	0.91	-0.13	3.18	-1.29	-0.33	0.28	0.86	1.69
MKT_d	0.43	1.02	-3.10	180.24	-0.67	0.01	0.34	0.80	1.79
t	1.20	1.81	0.84	5.28	-1.25	0.02	0.97	2.13	4.49
MKT_{d-1}	0.05	0.94	-1.48	192.26	-0.92	-0.23	0.02	0.32	1.14
t	0.07	1.19	-0.08	3.53	-1.88	-0.69	0.07	0.84	2.00
IND_d	0.24	0.86	4.64	256.06	-0.70	-0.07	0.17	0.54	1.34
t	1.12	2.49	3.53	30.21	-1.43	-0.25	0.62	1.74	5.49
IND_{d-1}	0.04	0.76	1.14	246.66	-0.76	-0.16	0.03	0.24	0.88
t	0.15	1.10	0.04	3.26	-1.65	-0.57	0.15	0.87	1.97
R^2	0.12	0.15	1.77	5.92	0.01	0.02	0.06	0.17	0.45
IV	69.04	62.62	10.67	424.14	17.08	32.72	52.94	86.88	172.73
LIV	3.97	0.75	-0.40	4.62	2.84	3.49	3.97	4.46	5.15
LRIV	2.66	1.47	-0.06	2.60	0.22	1.56	2.76	3.74	4.94
B. The extended Fama-French five-factor model									
Const.	0.05	0.56	8.81	550.20	-0.46	-0.07	0.04	0.18	0.61
t	0.23	0.89	-0.12	3.19	-1.27	-0.34	0.26	0.83	1.65
MKT_d	0.42	1.27	-2.10	248.06	-0.92	-0.02	0.38	0.86	1.89
t	0.99	1.61	0.64	5.04	-1.28	-0.06	0.83	1.86	3.89
MKT_{d-1}	0.03	1.03	5.69	579.05	-0.97	-0.25	0.01	0.29	1.11
t	0.01	1.17	-0.12	3.51	-1.93	-0.73	0.02	0.77	1.89
IND_{id}	0.18	1.00	6.24	575.43	-0.87	-0.13	0.14	0.51	1.31
t	0.84	2.22	3.86	37.70	-1.50	-0.38	0.45	1.44	4.54
$IND_{i,d-1}$	0.04	0.84	-3.63	578.06	-0.77	-0.16	0.03	0.24	0.87
t	0.15	1.10	0.03	3.27	-1.65	-0.58	0.14	0.87	1.94
SMB_d	0.49	1.31	1.82	106.42	-0.91	-0.02	0.41	0.99	2.07
t	1.15	1.79	0.88	4.53	-1.28	-0.06	0.88	2.07	4.54
HML_d	0.07	1.65	2.18	244.60	-1.72	-0.40	0.07	0.55	1.78
t	0.14	1.24	0.13	4.69	-1.83	-0.64	0.13	0.92	2.14
RMW_d	-0.22	1.82	8.79	749.20	-2.40	-0.72	-0.09	0.36	1.56
t	-0.17	1.20	-0.06	3.95	-2.10	-0.94	-0.16	0.61	1.77
CMA_d	-0.02	2.22	0.27	172.73	-2.41	-0.64	-0.02	0.59	2.40
t	-0.04	1.16	-0.08	3.61	-1.95	-0.79	-0.03	0.72	1.82
R^2	0.16	0.16	1.49	4.73	0.02	0.04	0.09	0.23	0.50
IV	67.54	61.04	10.64	431.13	16.41	31.84	51.75	85.20	169.52
LIV	3.94	0.76	-0.42	4.67	2.80	3.46	3.95	4.45	5.13
LRIV	2.12	1.23	-0.30	2.48	-0.02	1.20	2.27	3.06	3.91

Table 10
Summary statistics of firm-specific variables (no-news sample)
Panel A of the table reports the summary panel statistics of the firm-year log idiosyncratic volatility, log relative idiosyncratic volatility, and variables to be used in regressions to explain them, plus a few related variables. Panel B reports the correlation matrix. LME is log market equity, LVROE is log volatility of ROE, LIVROE is log idiosyncratic volatility of return-on-equity (ROE) calculated as the sample standard deviation of previous twelve quarterly ROEs, LRIVROE is log relative idiosyncratic volatility of return-on-equity, calculated similarly using the ROE residual from a market model of ROE. MROE is sample mean of previous twelve quarterly ROEs, LVOL is log trading volume, LMBA is log market-to-book assets ratio, and LMBE is log market-to-book equity ratio. The sample period is 2000-2019, with 1399.9 firms per year on average.

A. Summary statistics of variables

	Mean	Std	Skew	Kurt	p5	p25	p50	p75	p95
LME	4.90	1.80	0.34	2.94	2.16	3.61	4.73	6.13	7.97
LVROE	2.62	2.36	1.25	6.05	-0.43	1.03	2.22	3.78	7.07
LIVROE	2.08	2.36	-0.16	2.43	-1.99	0.40	2.14	3.85	5.84
LRIVROE	1.84	2.37	-0.16	2.45	-2.23	0.16	1.90	3.61	5.64
MROE	0.10	1.51	-0.69	3.01	-2.86	-0.93	0.55	1.20	1.92
LVOL	4.55	2.42	0.42	2.47	1.13	2.58	4.29	6.34	8.78
LMBA	0.30	0.56	1.38	5.94	-0.35	-0.02	0.12	0.53	1.46
LMBE	0.43	0.88	0.57	5.13	-0.88	-0.10	0.36	0.89	2.00

B. Correlation matrix: Pearson \ Spearman

	LME	LVROE	LIVROE	LRIVROE	MROE	LVOL	LMBA	LMBE
LME		-0.22	-0.22	-0.01	0.35	0.88	0.38	0.40
LVROE	-0.21		0.99	-0.01	-0.43	-0.02	0.17	0.15
LIVROE	-0.21	0.99		0.10	-0.43	-0.03	0.17	0.15
LRIVROE	-0.01	-0.01	0.09		0.00	-0.02	0.01	0.00
MROE	0.33	-0.49	-0.49	0.00		0.22	0.08	0.11
LVOL	0.89	-0.01	-0.02	-0.02	0.18		0.42	0.42
LMBA	0.33	0.22	0.22	0.00	-0.11	0.38		0.95
LMBE	0.38	0.18	0.18	0.01	-0.07	0.40	0.89	

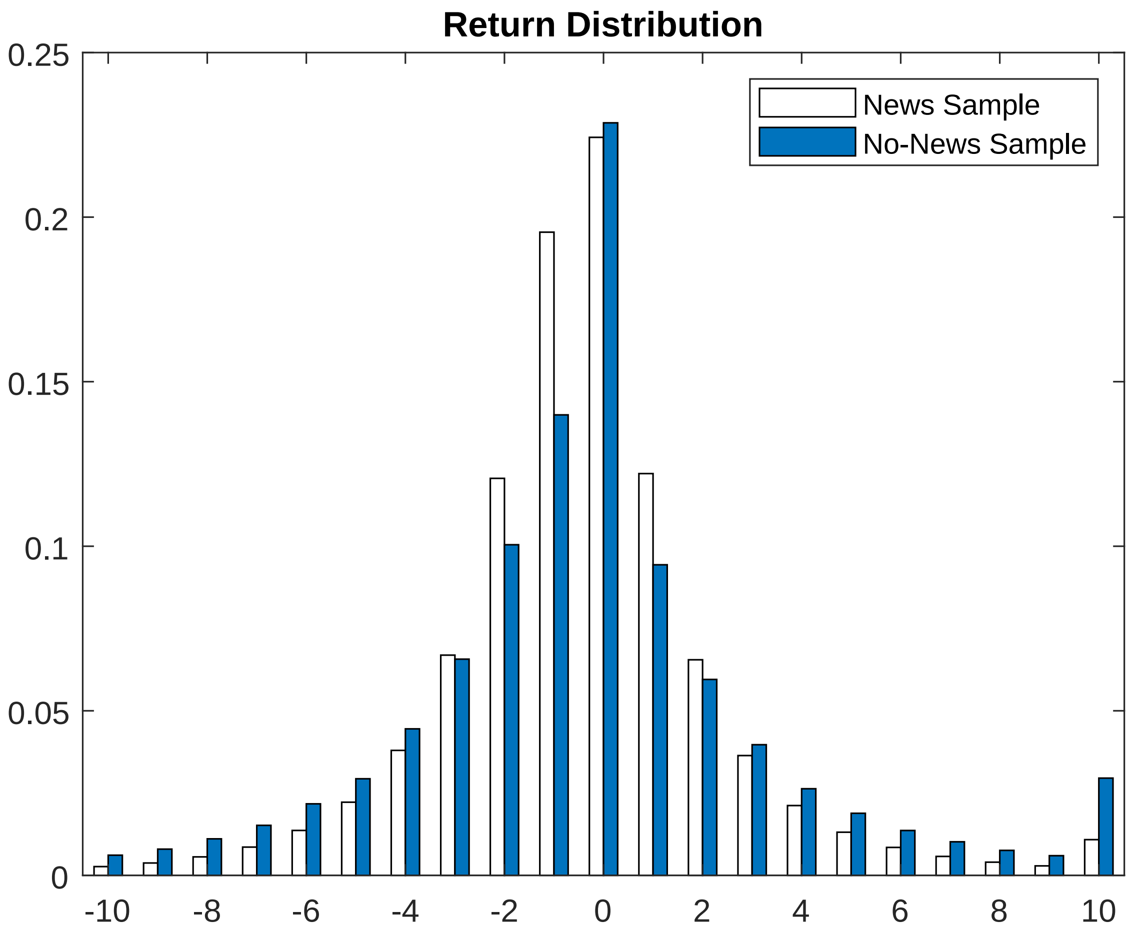

Figure 1. Histograms of daily returns: news sample vs no-news sample

This figure plots the panel histograms of daily returns (in percentage, winsorized at ±10%) of the news sample and no-news sample.

Table A1
Explaining idiosyncratic volatility (news sample)

This table reports the Fama-MacBeth regressions of log idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on news intensity NI and other control variables for the news sample with at least 25 news days per year for a firm. LME is log market equity, LVROE is log volatility of ROE, LIVROE is log idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. The sample period is 2000-2019, with 2152.5 firms per year on average.

A. LIV from the extended market model											
LNI	-0.27		0.34	0.15		0.16	0.16	0.15	0.16	0.16	0.16
	(-25.3)		(13.3)	(10.5)		(12.1)	(12.0)	(10.3)	(11.8)	(11.7)	(12.2)
LME		-0.20	-0.30	-0.41	-0.38	-0.41	-0.41	-0.41	-0.41	-0.41	-0.41
		(-33.6)	(-25.0)	(-58.8)	(-48.6)	(-58.7)	(-53.3)	(-59.4)	(-59.5)	(-54.2)	(-59.7)
LVROE				0.04	0.04	0.04	0.04				0.06
				(17.9)	(16.4)	(14.8)	(14.2)				(5.8)
LIVROE								0.04	0.04	0.04	-0.03
								(16.6)	(14.5)	(14.0)	(-2.6)
MROE				-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05
				(-21.2)	(-26.6)	(-23.8)	(-25.8)	(-20.8)	(-23.3)	(-25.4)	(-23.3)
LVOL				0.18	0.19	0.17	0.18	0.19	0.17	0.18	0.17
				(15.0)	(22.5)	(19.4)	(17.2)	(15.3)	(19.7)	(17.4)	(19.4)
LMBA						0.11			0.11		0.11
						(7.4)			(7.3)		(7.9)
LMBE							0.04			0.04	
							(4.2)			(4.2)	
B. LIV from the extended Fama-French five-factor model											
LNI	-0.26		0.35	0.16		0.17	0.17	0.16	0.17	0.17	0.17
	(-25.4)		(13.8)	(11.6)		(13.4)	(13.2)	(11.4)	(13.1)	(12.9)	(13.5)
LME		-0.20	-0.30	-0.41	0.00	-0.41	-0.41	-0.42	-0.41	-0.42	-0.41
		(-34.7)	(-25.6)	(-56.2)	(-45.4)	(-56.5)	(-51.7)	(-57.1)	(-57.5)	(-52.7)	(-57.2)
LVROE				0.05	0.05	0.04	0.04				0.06
				(16.5)	(16.4)	(14.0)	(13.5)				(5.6)
LIVROE								0.04	0.04	0.04	-0.02
								(15.4)	(13.6)	(13.3)	(-2.3)
MROE				-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05
				(-20.5)	(-26.1)	(-23.5)	(-25.7)	(-20.1)	(-23.0)	(-25.2)	(-23.1)
LVOL				0.18	0.19	0.17	0.18	0.18	0.17	0.18	0.17
				(14.9)	(21.8)	(18.8)	(16.8)	(15.1)	(19.1)	(17.0)	(18.7)
LMBA						0.11			0.11		0.11
						(7.7)			(7.7)		(8.3)
LMBE							0.04			0.04	
							(4.4)			(4.4)	

Table A2
Explaining relative idiosyncratic volatility (news sample)

This table reports the Fama-MacBeth regressions of log relative idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on news intensity NI and other control variables for the news sample with at least 25 news days per year for a firm. LME is log market equity, LRIVROE is log relative idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. The sample period is 2000-2019, with 2152.5 firms per year on average.

A. LRIV from the extended market model											
LNI	-0.84		0.41	0.41	0.40	0.47	0.46	0.46		0.47	0.46
	(-20.2)		(10.9)	(10.9)	(10.8)	(12.5)	(12.4)	(12.4)		(12.4)	(12.3)
LME		-0.50	-0.62	-0.62	-0.61	-0.49	-0.46	-0.46	-0.38	-0.45	-0.46
		(-50.4)	(-36.0)	(-35.9)	(-37.4)	(-17.2)	(-16.7)	(-16.8)	(-12.5)	(-16.6)	(-17.5)
LRIVROE				0.01	0.01	0.01		0.01	0.01	0.01	0.01
				(2.6)	(2.4)	(2.0)		(1.9)	(2.1)	(2.2)	(2.2)
MROE					-0.02		-0.03	-0.03	-0.05	-0.03	-0.03
					(-2.5)		(-4.2)	(-4.2)	(-5.2)	(-3.5)	(-3.9)
LVOL						-0.14	-0.14	-0.14	-0.11	-0.18	-0.17
						(-5.8)	(-6.3)	(-6.2)	(-4.2)	(-8.1)	(-7.4)
LMBA									0.31	0.33	
									(12.4)	(13.7)	
LMBE											0.18
											(14.5)
B. LRIV from the extended Fama-French five-factor model											
LNI	-0.71		0.39	0.39	0.38	0.43	0.42	0.42		0.44	0.43
	(-21.3)		(10.8)	(10.8)	(10.8)	(12.2)	(12.2)	(12.3)		(12.2)	(12.1)
LME		-0.43	-0.54	-0.54	-0.54	-0.44	-0.43	-0.43	-0.36	-0.41	-0.43
		(-53.4)	(-34.3)	(-34.2)	(-36.0)	(-17.2)	(-16.7)	(-16.9)	(-13.1)	(-17.0)	(-17.8)
LRIVROE				0.01	0.01	0.01		0.01	0.01	0.01	0.01
				(2.4)	(2.3)	(2.0)		(1.9)	(2.0)	(2.1)	(2.1)
MROE					-0.01		-0.02	-0.02	-0.03	-0.01	-0.01
					(-1.2)		(-2.3)	(-2.3)	(-3.6)	(-1.6)	(-2.0)
LVOL						-0.11	-0.11	-0.11	-0.08	-0.14	-0.13
						(-4.8)	(-5.1)	(-5.0)	(-3.0)	(-6.8)	(-6.1)
LMBA									0.28	0.30	
									(12.0)	(13.1)	
LMBE											0.16
											(13.6)

Table A3
Explaining relative idiosyncratic volatility (no-news sample)

This table reports the panel regressions with year fixed effect and firm-year clustering of log idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on news intensity NI and other control variables. LME is log market equity, LVROE is log volatility of ROE, LIVROE is log idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. R^2 (R^2_*) is the R^2 with (without) year fixed effect. The sample period is 2000-2019, with 1399.9 firms per year on average.

A. LIV from the extended market model								
LME	-0.19	-0.32	-0.33	-0.32	-0.32	-0.33	-0.33	-0.33
	(-19.6)	(-31.8)	(-32.6)	(-30.9)	(-32.1)	(-33.1)	(-31.1)	(-32.8)
LVROE		0.06	0.06	0.06				0.10
		(18.1)	(15.6)	(16.7)				(4.9)
LIVROE					0.06	0.06	0.06	-0.04
					(19.4)	(16.5)	(17.9)	(-2.2)
MROE		-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
		(-17.6)	(-14.6)	(-16.5)	(-18.0)	(-15.0)	(-16.8)	(-14.4)
LVOL		0.14	0.13	0.14	0.14	0.14	0.14	0.13
		(12.0)	(12.6)	(12.3)	(12.3)	(13.0)	(12.6)	(12.6)
LMBA			0.07			0.07		0.07
			(4.6)			(4.7)		(4.6)
LMBE				0.01			0.01	
				(1.0)			(1.0)	
R^2	0.43	0.65	0.65	0.65	0.65	0.65	0.65	0.65
R^2_*	0.34	0.60	0.60	0.60	0.59	0.60	0.59	0.60
B. LIV from the extended Fama-French five-factor model								
LME	-0.20	-0.33	-0.33	-0.33	-0.33	-0.33	-0.33	-0.33
	(-19.7)	(-31.7)	(-32.6)	(-31.2)	(-32.0)	(-33.0)	(-31.4)	(-32.9)
LVROE		0.06	0.06	0.06				0.10
		(18.7)	(16.1)	(17.2)				(5.0)
LIVROE					0.06	0.06	0.06	-0.04
					(20.0)	(17.0)	(18.4)	(-2.1)
MROE		-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
		(-17.3)	(-14.2)	(-16.1)	(-17.7)	(-14.5)	(-16.4)	(-13.9)
LVOL		0.14	0.13	0.14	0.14	0.13	0.14	0.13
		(12.0)	(12.5)	(12.3)	(12.3)	(12.9)	(12.6)	(12.5)
LMBA			0.07			0.07		0.07
			(4.8)			(4.9)		(4.8)
LMBE				0.01			0.01	
				(1.0)			(1.0)	
R^2	0.44	0.65	0.65	0.65	0.65	0.65	0.65	0.66
R^2_*	0.36	0.60	0.61	0.60	0.60	0.60	0.60	0.61

Table A4
Explaining relative idiosyncratic volatility (no-news sample)

This table reports the panel regressions with year fixed effect and firm-year clustering of log idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on news intensity NI and other control variables. LME is log market equity, LRIVROE is log relative idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. R^2 (R^2_*) is the R^2 with (without) year fixed effect. The sample period is 2000-2019, with 1399.9 firms per year on average.

A. LRIV from the extended market model								
LME	-0.64	-0.64	-0.65	-0.30	-0.29	-0.29	-0.30	-0.30
	(-28.8)	(-29.0)	(-31.0)	(-8.0)	(-8.3)	(-8.3)	(-8.1)	(-8.0)
LRIVROE		0.01	0.01	0.01		0.01	0.01	0.01
		(3.0)	(3.0)	(2.1)		(2.1)	(2.2)	(2.1)
MROE			0.04		-0.02	-0.02	-0.01	-0.01
			(3.3)		(-1.7)	(-1.7)	(-0.9)	(-1.4)
LVOL				-0.28	-0.29	-0.29	-0.30	-0.29
				(-11.8)	(-12.4)	(-12.4)	(-12.8)	(-12.6)
LMBA							0.10	
							(2.6)	
LMBE								0.04
								(1.8)
R^2	0.58	0.58	0.58	0.62	0.62	0.62	0.62	0.62
R^2_*	0.56	0.56	0.56	0.60	0.60	0.61	0.61	0.61
B. LRIV from the extended Fama-French five-factor model								
LME	-0.54	-0.54	-0.55	-0.27	-0.27	-0.27	-0.27	-0.27
	(-27.9)	(-28.0)	(-31.1)	(-7.5)	(-8.2)	(-8.3)	(-8.0)	(-8.0)
LRIVROE		0.01	0.01	0.01		0.01	0.01	0.01
		(2.6)	(2.7)	(1.9)		(1.9)	(2.0)	(1.9)
MROE			0.03		-0.01	-0.01	-0.00	-0.00
			(2.9)		(-0.8)	(-0.8)	(-0.0)	(-0.4)
LVOL				-0.23	-0.23	-0.23	-0.24	-0.23
				(-9.9)	(-10.5)	(-10.5)	(-11.1)	(-10.8)
LMBA							0.09	
							(2.7)	
LMBE								0.03
								(2.0)
R^2	0.60	0.60	0.60	0.63	0.63	0.63	0.64	0.64
R^2_*	0.58	0.58	0.58	0.62	0.62	0.62	0.62	0.62

Table A5
Explaining idiosyncratic volatility (no-news sample)

This table reports the Fama-MacBeth regressions of log idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on news intensity NI and other control variables for the no-news sample with less than 25 news days per year for a firm. LME is log market equity, LVROE is log volatility of ROE, LIVROE is log idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. The sample period is 2000-2019, with 1399.9 firms per year on average.

A. LIV from the extended market model								
LME	-0.20	-0.32	-0.33	-0.32	-0.33	-0.33	-0.32	-0.33
	(-31.0)	(-29.4)	(-28.5)	(-27.1)	(-30.0)	(-29.0)	(-27.6)	(-29.0)
LVROE		0.06	0.06	0.06				0.10
		(19.5)	(18.3)	(18.5)				(7.0)
LIVROE					0.06	0.06	0.06	-0.04
					(19.1)	(18.0)	(18.5)	(-3.1)
MROE		-0.06	-0.05	-0.06	-0.06	-0.05	-0.06	-0.05
		(-27.9)	(-23.4)	(-25.9)	(-26.9)	(-22.5)	(-24.7)	(-23.2)
LVOL		0.14	0.13	0.13	0.14	0.13	0.14	0.13
		(11.8)	(12.0)	(11.9)	(12.1)	(12.3)	(12.1)	(12.1)
LMBA			0.05			0.05		0.05
			(4.7)			(4.7)		(5.0)
LMBE				0.01			0.01	
				(0.6)			(0.7)	
B. LIV from the extended Fama-French five-factor model								
LME	-0.20	-0.33	-0.33	-0.33	-0.33	-0.33	-0.33	-0.33
	(-31.3)	(-29.4)	(-28.5)	(-27.3)	(-29.9)	(-29.1)	(-27.7)	(-29.0)
LVROE		0.06	0.06	0.06				0.10
		(19.6)	(18.4)	(18.5)				(7.1)
LIVROE					0.06	0.06	0.06	-0.04
					(19.0)	(18.0)	(18.4)	(-3.0)
MROE		-0.06	-0.05	-0.06	-0.06	-0.05	-0.06	-0.05
		(-26.0)	(-21.9)	(-24.5)	(-25.2)	(-21.2)	(-23.5)	(-21.8)
LVOL		0.13	0.13	0.13	0.14	0.13	0.13	0.13
		(11.6)	(11.8)	(11.6)	(11.9)	(12.1)	(11.9)	(11.8)
LMBA			0.05			0.06		0.06
			(4.8)			(4.8)		(5.1)
LMBE				0.00			0.01	
				(0.5)			(0.6)	

Table A6
Explaining relative idiosyncratic volatility (no-news sample)

This table reports the Fama-MacBeth regressions of log relative idiosyncratic volatility from the extended market model and the extended Fama-French five-factor model on news intensity NI and other control variables for the news sample with less than 25 news days per year for a firm. LME is log market equity, LRIVROE is log relative idiosyncratic volatility of ROE, MROE is mean of ROE, LVOL is log trading volume, LMBA is log market-to-book assets, and LMBE is log market-to-book equity. The sample period is 2000-2019, with 1399.9 firms per year on average.

A. LRIV from the extended market model								
LME	-0.64	-0.64	-0.65	-0.32	-0.31	-0.31	-0.32	-0.32
	(-34.1)	(-34.5)	(-35.1)	(-13.3)	(-13.0)	(-13.1)	(-12.8)	(-13.1)
LRIVROE		0.01	0.01	0.01		0.01	0.01	0.01
		(3.2)	(3.1)	(2.3)		(2.2)	(2.2)	(2.2)
MROE			0.03		-0.02	-0.02	-0.01	-0.01
			(2.8)		(-2.2)	(-2.2)	(-0.7)	(-1.4)
LVOL				-0.26	-0.27	-0.27	-0.28	-0.27
				(-11.5)	(-11.8)	(-11.9)	(-12.1)	(-11.9)
LMBA							0.15	
							(6.4)	
LMBE								0.05
								(3.9)
B. LRIV from the extended Fama-French five-factor model								
LME	-0.54	-0.54	-0.55	-0.29	-0.29	-0.29	-0.30	-0.30
	(-32.8)	(-33.2)	(-34.1)	(-13.5)	(-14.3)	(-14.3)	(-13.9)	(-14.1)
LRIVROE		0.01	0.01	0.00		0.00	0.00	0.00
		(2.5)	(2.5)	(1.7)		(1.7)	(1.8)	(1.8)
MROE			0.02		-0.01	-0.01	0.00	0.00
			(2.6)		(-1.4)	(-1.4)	(-0.0)	(-0.6)
LVOL				-0.21	-0.21	-0.21	-0.22	-0.21
				(-10.3)	(-10.9)	(-10.9)	(-11.2)	(-11.0)
LMBA							0.14	
							(6.4)	
LMBE								0.05
								(4.2)