Investor sentiment and the distribution of stock returns

Maria Osipenko[*], Rui Ren[†]

March 1, 2023

Abstract

We study how the distribution of stock returns is influenced through investor sentiment by applying expectile regression to abnormal NASDAQ returns over the period April 2015 - August 2019. Thereby, we differentiate between different aspects of investor opinion: investor perception, attention and associated sentiment, and use firm and time specific proxies for each of the aspects. We consider purely financial indicators, as book to market ratio, size, shares turnover, together with scheduled and unscheduled events, as earnings announcements and news releases, in an interplay with sentiment (or polarity) of news content.

We find, that the majority of the concerned effects depend on the expectile level τ, which itself can be understood as the 'state of the world' or 'extremeness level' of individual abnormal returns. While not always significant for the mean regression, the proxies of the investment opinion show significant effects on the conditional distribution of asset returns beyond its mean.

Our expectile regression approach allows us to naturally generalize the previous studies of conditional mean regression to higher conditional moments and estimate investor opinion impact thereon. Significant impact of positive and negative sentiment on the shape of the conditional return distribution is tracted via expectile-based measures of dispertion and skewness. Moreover, using the relation of expectiles to expected shortfall, we tranfer our findings to the expected-shortfall-point of view and explore significance of the effects on conditional expected shortfall as well.

Introduction

Establishing the link between investor sentiment associated with certain information flow channels and stock return movements have captured attention of many researches. The results are still somewhat unaligned.

For instance, Das and Chen (2007) observe that the message volume (regarded as optimistic sentiment) is positively correlated with stock volatility and negatively correlated with stock returns. Luss and D'Aspremont (2015) show that text features from news articles predict the size of the movement but not the direction. Hagenau et al. (2013) and Nguyen et al. (2015) are able to significantly improve their return forecasts by using sentiment from news or social media. Shang et al. (2014) confirms a significant effect only for negative sentiment. Kim and Kim (2014) find no evidence for predictive power of sentiment neither on returns nor on volatility. Smales (2016) again relates returns and volatility changes to news sentiment. Recently, Audrino et al. (2020) and Fang et al. (2021) report an improvement in volatility forecasts when sentiment from news and social media is accounted for.

Thereby, there is no clear alignment on how to measure investor sentiment. Various ways has been undergone in order to assess investor sentiment in its broad sense. Overall the existing approaches to measuring investor sentiment can be divided into two categories. The first one uses financial proxies to measure sentiment, as e.g. in De Long et al. (1990), Barberis et al. (1998), Baker and Wurgler (2007). Candidate methods of measuring sentiment contain financial indicators as book-to-market ratios, earnings surprise, size (Friend and

---

[*]Hochschule für Wirtschaft und Recht Berlin; osipenko@hwr-berlin.de
[†]Humboldt-Universität zu Berlin

Lang (1988), Fama and French (1992), Kothari and Shanken (1997), Pontiff and Schall (1998), Maheu and McCurdy (2004)) and event indicators as news arrivals, earnings announcements, search inquiries. Berkman et al. (2009) also argues that varying degree in differences in opinion regarding some stocks may influence their price dynamics around earnings announcements. Among other proxies for the opinion gap they use share turnover in the time span around earnings announcements. The authors argument that stocks with high degree of opinion differences are overpriced and earn lower return on earnings announcement days due to opinion alignment.

The second approach is to infer investor sentiment directly from the relevant textual documents. Su et al. (2021) classifies the related approaches as dictionary-based and machine learning based ones. The first one relies on word lists with positive and negative words with the most famous dictionaries - Harvard (HV) and Loughran & McDonald (LM) word lists (e.g. Loughran and McDonald (2011), Jegadeesh and Wu (2013)). The machine learning approach uses statistical models to map the word content to a relevant variable as return (e.g. Antweiler and Frank (2004), Fang et al. (2021)).

In this paper, we use financial proxies along with dictionary-based textual sentiment of news content delivered by both HV and LM word lists. To structure this broad assessment of investor sentiment into smaller chunks, we adopt three investor opinion formation channels of Su et al. (2021): investor perception (risk perception and individual believes), captured by financial variables, event-related investor attention and investor sentiment in a narrow sense, specified as positive or negative surprises and the polarity of news content.

On the model side, the limited attention hypothesis (Corwin and Coughenour (2008) and Loh (2010)) and the opinion divergence hypothesis (Berkman et al. (2009) and Chatterjee et al. (2012)) shed light on the ways these particular aspects of investor opinion influence asset pricing. Füss et al. (2020) introduce aggregated investor sentiment in their asset pricing model. They show that the inclusion of the lexicon-based sentiment as a risk factor yields a better model performance.

In this paper we analyse the impact of the three opinion formation aspects on the distribution of abnormal daily returns. Thereby we encapsulate *investor attention* using the indicators of days around earnings announcements and news arrivals. The aggregated *investor perception* is captured by financial fundamentals as book-to-market ratios, size, shares turnover around earnings announcements and reported earnings per share. Earnings surprises and dictionary-based polarity assessment of firm-specific news content is used to proxy *investor sentiment* in its narrow sense. Our choice of the proxies is based on the empirical evidence found in the previous literature provided below.

Besides the different approaches on the sentiment side, there is no consensus on which parameters of the conditional return distribution are affected. Significant effects were shown for conditional return expectation (e.g. in Hagenau et al. (2013) and Nguyen et al. (2015)) and for conditional variance (e.g. in Audrino et al. (2020) and Fang et al. (2021)). Cao et al. (2002) argue, that not only second but also third moments of return distribution are effected by information arrival due to inert "sidelined" investors. Engelberg et al. (2018) go beyond and close up the tails of the return distribution by considering return anomalies. They find a sign-changing asymmetric effect on information days. Recently, He (2022) confirms significant asymmetric sentiment effects on low and high quantiles of return distribution.

In this paper, we estimate the effects of investor sentiment *on the whole distribution* of abnormal returns by breaking it down to a fine grid of tail indices. That way we are neither constrained to a particular distributional assumption nor to certain conditional moments of the return distribution.

A natural choice of such tail indices are quantiles (Koenker and Bassett (1978)) of the return distribution, $q_\alpha(r_{i,t})$, for a range of α-values in $(0,1)$ capable of representing the whole return distribution without explicit distributional assumptions. By means of quantile regression for a set of $\alpha \in (0,1)$ distributional effects of investor opinions can be estimated. Ni et al. (2015), Ma et al. (2018), Al-Nasseri et al. (2021), and He (2022) use quantile regression to address the distributional effects of sentiment. The authors find different effects for different quantile levels and point out that, overall, extreme markets are largely driven by investor sentiment.

Instead of using α-quantiles for representing the conditional return distribution, an alternative measure τ-expectiles introduced in Newey and Powell (1987) can be employed. The τ-expectile is the real number,

such that the τ-proportion of the expected distance to it lies below and the $(1-\tau)$-proportion lies above it. As quantiles generalize the concept of the median, expectiles generalize the concept of the mean. Therefore, the expectile regression for a range of τ-values in $(0,1)$ naturally generalizes the mean regression and is capable of capturing the effects on the whole distribution of returns in our case. This fact makes our analysis comparable with the previous findings based on the mean regression for $\tau = 0.5$ on one hand and capable of capturing distributional effects for $\tau \in (0,1) \setminus \{0.5\}$ on the other.

In summary, our expectile regression approach allows to study heterogeneous effects over the whole distribution of returns breaking it down to different expectile levels (similar to quantile regression). As opposed to quantile regression, it covers the mean regression as a special case of $\tau = 0.5$ and facilitates simple and efficient estimation and inference on the obtained parameters (see Waltrup et al. (2015) on the efficiency of expectles, Barry et al. (2022) for inference).

Based on our expectile regression results, we confirm significant asymmetric effects of information arrivals with positive and negative sentiment on the shape of the conditional return distribution using non-parametric expectile-based measures of disperion and skewness. Further advantage of our expectile-based approach lies in the risk management perspective. Well-known connection of expectiles to expected shortfall (Taylor (2008)) makes our investigation of the factors impacting the expectiles of the returns potentially useful for risk management insights. Using the estimated effects from our conditional expectile model, we obtain plug-in estimates for the respective effects on the conditional expected shortfall and evaluate their significance via bootstrap.

The contribution of our paper is threefold. First, by incorporating all three investor opinion channels in our model, we conduct a more comprehensive analysis of the respective effects, which makes our approach stand out from the previous related studies. Second, we estimate the effects of investor sentiment on the whole distribution of asset returns breaking it down to a fine grid of tail indices - expectiles - avoiding any distributional assumptions or model constrains to particular conditional moments. Third, based on the connection of expectiles and expected shortfall, we obtain the estimated effects of investor sentiment directly on the expected shortfall of returns, which is highly relevant for risk management and uncovered in the existing studies.

The rest of the paper is organized as follows. First, we introduce our data and the variables we use to proxy the investor opinion formation channels. Next, we describe our conditional expectile model and present our estimation results. A close up analysis of the effects on the resulting shape of the conditional return distribution and expected shortfall follows. Finally, we conclude and point to possible further research direction.

1 Proxies for investor opinion formation channels

We use data for NASDAQ stocks in the period April 2015 - August 2019 obtained from Bloomberg. Our dependent variable is the abnormal return of NASDAQ stocks in the mentioned period, computed as stock-specific return on the day of news arrival minus Center for Research in Security Prices (CRSP) value-weighted market return. The associated data on the size, the book-to-market (btm) ratios, the share turnover, and the associated sentiment features from news articles build up the set of explanatory variables.

The textual corpus in this paper contains news articles from NASDAQ News and Insights and is downloaded from the Blockchain Research Center. We select the top 100 companies regarding news frequency in the final data set. The stock ticker list is provided in the Appendix. All news articles published on weekends, holidays or after 4 pm on weekdays are merged with news articles on the next trading day. Table ref(tab:table00) in Appendix describes the textual corpus used to calculate dictionary based investor sentiment measure.

We combine this data with the records on earnings announcements from kaggle platform (accessed on 21.September 2022)

https://www.kaggle.com/datasets/tsaustin/us-historical-stock-prices-with-earnings-data containing earnings date, estimated earning per share (eps) and reported eps, and with the sector information from NASDAQ website https://www.nasdaq.com/market-activity/stocks/screener (accessed on 10. September 2022).

Based on the available data and on previous findings in the literature, we construct the proxies for the three channels of information flow: investor attention, investor sentiment and investor perception described in the following.

Investor attention

Chen et al. (2022), Andrei and Hasler (2020) and Dellavigna and Pollet (2009) among others highlight the impact of investor attention on financial markets. Dellavigna and Pollet (2009) refer to earnings announcements on different weekdays to capture the phenomena. Andrei and Hasler (2020) focus on optimal attention allocation to news in investor utility terms and point out that the optimal attention to news increases in uncertainty in their model. Drake et al. (2017) use search volume, analysts forecasts revisions, financial statements downloads and number of news releases to quantify the amount of attention in the cross-sectional comovement sense. The authors argue, that because of limited attention investors focus rather on categories of stocks (e.g. belonging to the same industry), than on individual tickers when it comes to choosing a stock to buy from a large cross-section of stocks. Yuan (2015) investigates the role of market-wide "attention-grabbing" events in capturing investor attention across time. The author points out that the rise in attention level causes investors to make trade decisions.

In the spirit of the studies mentioned above earnings announcements and news arrivals must be a valid indicator of investor attention allocation available on a daily and stock level.

In summary, as proxies for investor attention to particular stocks we employ:

- indicator of firm-specific news arrival at t $news_{i,t}$
- number of day-sector specific news at t normalized on the number of firms in the sector, $news\_sector_{i,t}$, and the overall number of news for the whole portfolio arrived at t $news\_day_{i,t}$
- indicator of earnings announcements days $ean_{i,t}$ as well as
- indicators of days before and after earnings announcements, $pre\_ean_{i,t}$ and $post\_ean_{i,t}$.

The last two together indicate a three-day period around the earnings announcements as in Berkman et al. (2009) and the time of news releases.

Investor perception

We infer investor perception using variables which are relevant for risk assessment. First, according to Friend and Lang (1988) the well-known size effect reflects a risk effect and enters therefore our proxy for risk perception. Second, Peasnell (1982) connects the value of stocks and the expectation of future abnormal earnings by:

$$v_{i,t} = bv_{it} + \sum_{h=1}^{\infty} \frac{1}{(1+r)^h} \mathbb{E}_t(r_{i,t+h}^a)$$

where $v_{i,t}$ is the market value and $bv_{i,t}$ is the book value of equity i, $r_{i,t}^a$ abnormal returns at t and r is the cost of capital. The resulting book-to-market (btm) ratio:

$$btm_{i,t} = \frac{bv_{it}}{v_{i,t}}$$

encompases therefore investor expectations. Here we take firm-specific btm ratios of the previous period, $btm_{i,t-1}$, as a proxy for firm performance expectations. As a further proxy of future performance expectations, we imploy firm-specific reported earnings per share, scaled to the interval $[-1, 1]$.

Next, according to Berkman et al. (2009), the share turnover around earnings announcements can be an indicator of opinion differences and consequently overpriced stocks, which leads to price correction upon information arrival as the authors argue. Therefore, we also control for this effect by including an interaction term between turnover and days earnings announcements indicator $days\_ean_{i,t}$ in our model. The later

indicates the announcement date itself, the day before and the day after the earnings announcements as in Berkman et al. (2009).

Finally Donnelly (2014) writes "The intuition is that the prices of growth stocks are much more sensitive to earnings expectations than those of value stocks." and "Accordingly, their prices are more sensitive than value stocks to any earnings surprise regardless of the sign." Therefore, we also include the interaction term between btm ratio < 1 and post earnings day $(btm_{i,t-1} < 1) \cdot post\_ean_{i,t}$ to control for this effect.

In summary, as proxies for investor perception we use firm specific:

- firm specific size
- btm ratio of the previous period
- reported earnings per share
- turnover around earnings announcements
- the interaction term between small btm ratio and the post earnings announcements day.

Investor sentiment

Investor sentiment in its narrow sense, understood as positive or negative polarity (e.g. in Frankel et al. (2022)), has been measured with several techniques in the associated literature.

On one hand, financial proxies were used to track investor sentiment (e.g. De Long et al. (1990), Barberis et al. (1998), Baker and Wurgler (2007)). In addition to already coverred measures under investor perception channel, we include earnings per share (as e.g. in Henry (2006)) and earnings surprises (as e.g. in Maheu and McCurdy (2004)) to capture investor sentiment on eanings announcement days.

One the other hand, investor sentiment can be directly inferred from textual documents, where a dictionary-based and supervised machine learning sentiment techniques stand out (see e.g. Jegadeesh and Wu (2013) for dictionary based methods and Fang et al. (2021) who successfully uses machine learning methods (random forest) to infer textual sentiment). The results of Fang et al. (2021) and Frankel et al. (2022) on using non-linear sentiment models suggest a possibly non-linear effect of sentiment, which we capture in our per-expectile model. For that reason, we use a simpler dictionary-based method and calculate NASDAQ news sentiment as the sum of the positive words minus the sum of the negative words divided by the sum of both positive and negative words. This measure of sentiment is also called net tone, and widely adopted in research and applications. Following prior research, we apply two dictionaries to disclosure sentiment: the Loughran & McDonald (LM) business-context dictionary and the Harvard (HV) psychosociological dictionary.

In summary, as proxies for investor sentiment we use:

- earnings surprise $surp_{i,t}$ computed as the relative difference of analysts forecast and reported earnings to reported earnings;
- dictionary-based polarity assessment of firm specific news content with Loughran & McDonald (LM) word list, $lm_{i,t}$;
- dictionary-based polarity assessment of firm specific news content with Harvard (HV) word list, $hv_{i,t}$;

We normalize the last two sentiment measures to the interval $[-1, 1]$.

Overall, the assignment of the proxies to an investor opinion channel is somewhat rudimentary, since thea are interconnected and blurred. Our classification is thus not the only one possible and should just give a structure to the analysed effects.

Exploratory data analysis provided in Appendix supports our hypothesis that the considered covariates impact location, scale and shape of the conditional return distribution. To capture the effects on the conditional return distribution we represent it through its conditional expectiles. A practical estimator of the heterogeneous effects over the return distribution can be obtained by expectile regression.

2 Expectile regression estimation for the panel of abnormal returns

As mentioned in the introduction, conditional expectile extends the associated mean regression and allows to assess the considered effects on the whole distribution of asset returns braking it down to a fine grid of expectiles. That way we are neither constrained to a particular distributional assumption nor to certain conditional moments of the return distribution. In this section we review the general set up of the conditional mean model for the panel data of stock returns and its generalization to conditional expectile regression (ER).

2.1 General model definition

Consider a conditional mean model for the panel data of stock return $(r_{i,t})$ of firm i at time t with $i = 1, \ldots, n$ and $t = 1, \ldots, T$:

$$r_{i,t} = \beta_0 + x_{i,t}^\top \beta + z_i^\top \gamma + u_i + \varepsilon_{i,t}$$

with response $r_{i,t}$ being the abnormal return of stock i at time t computed as specified in Section "Data". $x_{i,t}$ a vector of p time varying predictors, and z_i a vector of n time-invariant covariates for the stock i known at time t. u_i are firm specific random effects and $\varepsilon_{i,t}$ is an error term both with finite second moments and uncorrelated with $(x_{i,t}, z_i)^\top$ for all $i = 1, \ldots, n$ and $T = 1, \ldots, T$.

The model can be rewritten as:

$$r_{i,t} = \beta_0 + x_{i,t}^\top \beta + z_i^\top \gamma + v_{i,t}.$$

where $v_{i,t} = u_i + \varepsilon_{it}$.

The conditional expectile regression of Newey and Powell (1987) aims at finding the conditional expectile (given the covariates $x_{i,t}$ and z_i):

$$e_{\tau,t}(r_{i,t}) \equiv e_\tau(r_{i,t}|x_{i,t}, z_i) = x_{i,t}^\top \beta_\tau + z_i^\top \gamma_\tau, i = 1, \ldots, n, t = 1, \ldots T,$$

or in terms of the responses $r_{i,t}$:

$$r_{i,t} = x_{i,t}^\top \beta_\tau + z_i^\top \gamma_\tau + v_{\tau,i,t}$$

where $v_{\tau,i,t}$ is the error term with $e_\tau(v_{\tau,i,t}|x_{i,t}, z_i) = 0$ and $\tau \in (0,1)$ is the expectile level which specifies the desired "extremeness" level as Kneib (2013).

The related estimator minimizes the expectile loss function:

$$\arg \min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \rho_\tau \Big(r_{i,t} - x_{i,t}\beta(\tau) \Big)$$

with $\rho_\tau(y) = |\tau - \mathbf{1}(y < 0)|y^2$ is the expectile check function.

The resulting estimator has the form:

$$\hat{\beta}_\tau = \Big(\sum_{i=1}^n \sum_{t=1}^T \hat{w}_{i,t}(\tau) x_{i,t} x_{i,t}^\top \Big)^{-1} \Big(\sum_{i=1}^n \sum_{t=1}^T \hat{w}_{i,t}(\tau) x_{i,t} r_{i,t} \Big)$$

with $\hat{w}_{i,t}(\tau) = |\tau - \mathbf{1}(r_{i,t} \leq x_{i,t}\hat{\beta}(\tau))|$

and can be computed via iterative weighted asymmetric least squares as proposed by Newey and Powell (1987) on pooled data.

Diogo Barry et al. (2016) show asymptotic normality of the estimator under mild assumptions:

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \to \mathcal{N}(0, H^{-1}\Sigma H^{-1})$$

The variances-covariances matrix of the estimator can be estimated via:

$$\hat{H} = X^\top \hat{W} X/n, \text{where } \hat{W} = diag(\hat{w}_{1,1}(\tau), \ldots, \hat{w}_{n,T})$$

and

$$\hat{\Sigma} = X^\top \hat{W} \hat{\varepsilon} \hat{\varepsilon}^\top \hat{W} X/n$$

.

We compute robust standard error estimates for the ER-coefficients with R-Package sandwich (Zeileis et al. (2020)) allowing for heteroscedastic and serial errors.

2.2 Expectile regression with proxies of investor opinion

We use the conditional linear expectile model defined above for estimating the effects of investor opinion channels on NASDAQ the individual conditional expectiles of abnormal returns $e_{\tau,t}(r_{i,t})$. Our model specification is therefore:

$$
\begin{aligned}
e_{\tau,t}(r_{i,t}) =& \beta_{0,\tau} + \beta_{1,\tau} \cdot pre\_ean_{i,t} + \beta_{2,\tau} \cdot ean_{i,t} + \beta_{3,\tau} \cdot post\_ean + \beta_{4,\tau} \cdot news_{i,t} + \beta_{5,\tau} \cdot news\_sector_{i,t} \\
&+ \beta_{6,\tau} \cdot news\_day_{i,t} + \beta_{7,\tau} \cdot \log(size_{i,t}) + \beta_{8,\tau} \cdot btm_{i,t-1} + \beta_{9,\tau} \cdot days\_ean_{i,t} \cdot turn_{i,t-1} \\
&+ \beta_{10,\tau} \cdot (btm_{i,t-1} < 1) \cdot ean_{i,t} + \beta_{11,\tau} \cdot eps_{i,t} + \beta_{12,\tau} \cdot surp_{i,t} + \beta_{13,\tau} \cdot lm\_tone_{i,t} \\
&+ \beta_{14,\tau} \cdot hv\_tone_{i,t} + \gamma_{1,\tau}^\top \cdot sector_i + \gamma_{2,\tau}^\top \cdot firm_i
\end{aligned}
\tag{1}
$$

with

- $r_{i,t}$ the abnormal return of stock i at time t,
- $pre_e an_{i,t}$ indicator the day before an earnings announcement referred to company i,
- $ean_{i,t}$ indicator whether an earnings announcement referred to company i has been released at t,
- $post_e an_{i,t}$ indicator the day after an earnings announcement referred to company i,
- $news_{i,t}$ indicator whether a news report referred to company i has been released at t,
- $news\_sector_{i,t}$ number of news reports at t for the company i sector normalized to sector size,
- $news\_day_{i,t}$ number of news reports at t for all companies in portfolio normalized to portfolio size,
- $\log(size_{i,t})$ the logarithm of the size? of the company,
- $btm_{i,t-1}$ the book-to-market ratio of company i in $t-1$,
- $turn_{i,t-1}$ share turnover of company i in $t-1$,
- $days_e an_{i,t}$ indicator a three-day period around an earnings announcement referred to company i,
- $eps_{i,t}$ reported earnings per share for company i in t,
- $surp_{i,t}$ earnings surprise in % at t according to an earnings announcement and analysts forecast referred to company i,
- $lm\_tone_{i,t}$ and $hv\_tone_{i,t}$ are news polarity or sentiment measures obtained by using LM and HV dictionary respectively. These measures are computed as $\frac{\text{\# of positive words} - \text{\# of negative words}}{\text{\# of positive words} + \text{\# of negative words}}$ using the corresponding dictionary.

7

- $sector_i$ and $firm_i$ are the associated sector/firm dummy variables of the form $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$.

Within our model, we control for the effects of investor attention (β_1 to β_6), for the effects of investor perception (β_7 to β_{10}), as well as the effects of investor sentiment (β_{11} to β_{14}).

We fit the models as specified in Equation (1) for $\tau \in \{0.05, 0.1, \ldots, 0.95\}$ and show the resulting coefficients, their significance (based on robust standard errors) and the associated R^2 in Table 1.
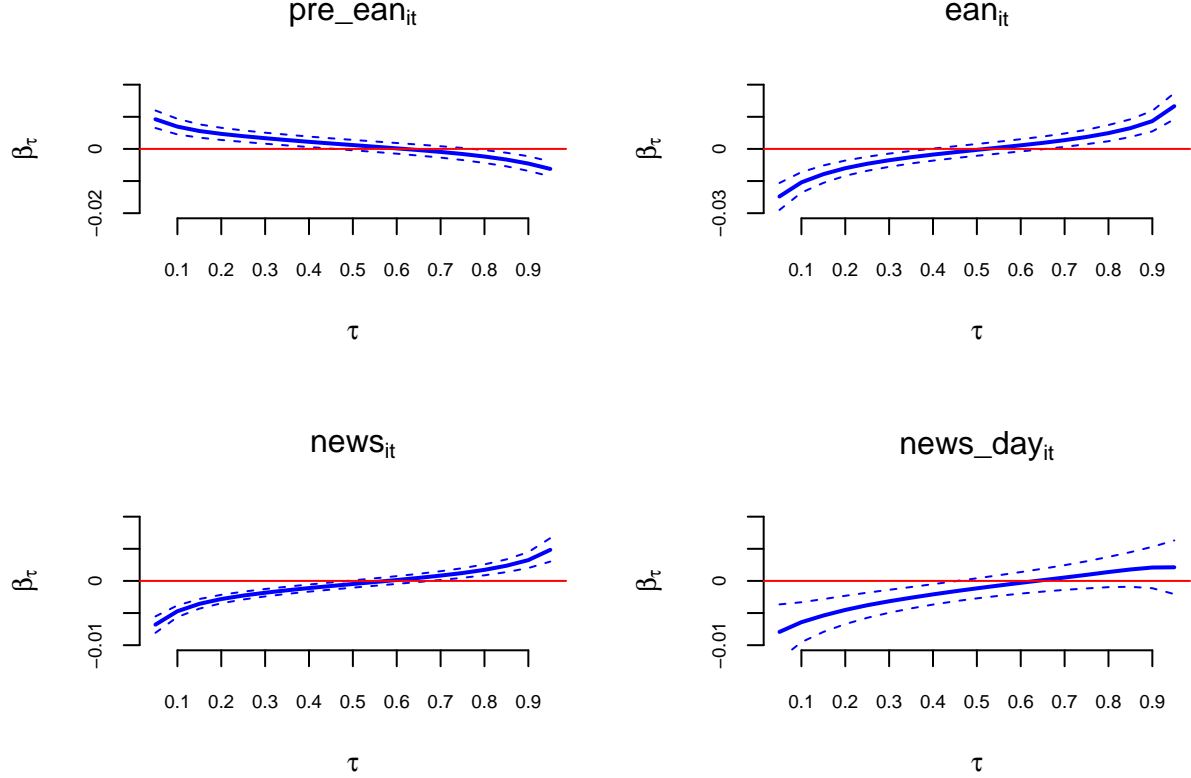

Figure 1: Coefficient profiles of selected investor attention related variables for different τ-levels

As reported in Table 1, many of the effects are insignificant for the conditional mean return but become significant for the conditional τ-expectiles with $\tau \neq 0.5$. That is, higher moments of the conditional distribution are affected by the considered investment opinion channel proxies. Also, the explained weighted variance proportion becomes higher in the tails of the conditional return distribution.

In the Figure 1, we plot the attention related effects with 95%-CI in Figure 1 across the τ-values. Note that the effects of $ean_{i,t}$ and $news_{i,t}$ are negative for low expectiles and become positive for higher τ-levels. Thus, these events tend to increase the dispersion in returns by "pushing" their expectiles apart. For the effect of $pre\_ean_{i,t}$ the opposite holds. However, the overall effect absorbs also the reversed contribution of $turn_{i,t} \cdot days\_ean_{i,t}$, such that both reduction and increase of dispersion are conceivable. The coefficient profile of $news\_day$ exhibits significant negative impact on lower expectiles producing a potential source of negative skewness in return distribution.

Figure 2 shows the effects related to investor perception with 95%-CI. Also here, higher values of btm and $turn \cdot days\_ean$ induce a higher expectile spread and consecutively higher dispersion. The negative influence of $btm < 1 \cdot post_e an$ and $log(size)$ is one-sided directed on either low or high expectiles.

In Figure 3 we plot the effects related to investor sentiment with 95%-CI. Overall, significant sentiment

Table 1: Coefficients and R^2 of the ERRE models with investor opinion proxies and different τs. The effects of investor attention, investor perception and investor sentiment are highlighted in blue, green and orange respectively.

	$\hat{\beta}_{0.01}$	$\hat{\beta}_{0.05}$	$\hat{\beta}_{0.1}$	$\hat{\beta}_{0.2}$	$\hat{\beta}_{0.3}$	$\hat{\beta}_{0.4}$	$\hat{\beta}_{0.5}$
Constant	-0.1802***	-0.09073***	-0.05876***	-0.0269*	-0.00722	0.00874	0.0238*
$pre\_ean_{i,t}$	0.01512***	0.00924***	0.00694***	0.00469***	0.00334***	0.00221***	0.0012
$ean_{i,t}$	-0.03662***	-0.02219***	-0.01562***	-0.00899***	-0.00528***	-0.00265*	-0.00044
$post\_ean_{i,t}$	-0.01184	-0.00442	-0.00052	0.00254	0.0039	0.00472	0.00556*
$news_{i,t}$	-0.01309***	-0.0068***	-0.00472***	-0.00282***	-0.00184***	-0.00111***	-0.00048*
$news\_sector_{i,t}$	0.00066	0.00111	0.00119	0.00112	0.00106*	0.00106*	0.00105*
$news\_day_{j,t}$	-0.00983*	-0.00794***	-0.00642***	-0.00452***	-0.00317***	-0.00209***	-0.00115
$\log(size_{i,t})$	0.00534*	0.00246*	0.00149*	0.00054	-0.00006	-0.00055	-0.00102*
$btm_{i,t-1}$	-0.0016*	-0.00124***	-0.00086***	-0.00042*	-0.00009	0.00019	0.00047*
$turn_{i,t-1}\cdot days\_ean_{i,t}$	-0.00973***	-0.00559***	-0.00397***	-0.00248***	-0.00166***	-0.00103*	-0.00049
$(btm_{i,t-1}\leq 1)\cdot post\_ean_{i,t}$	-0.03967*	-0.02863***	-0.02314***	-0.01663***	-0.01221***	-0.00872*	-0.00581*
$supr_{i,t}$	0.02436***	0.02096***	0.01921***	0.01653***	0.01501***	0.01399***	0.01337***
$eps_{i,t}$	0.0039	0.00299***	0.00233***	0.00175***	0.00148***	0.00133***	0.00121***
$lm\_tone_{i,t}$	0.00607***	0.00267***	0.00194***	0.00153***	0.00148***	0.00139***	0.00141***
$hv\_tone_{i,t}$	0.01437***	0.00805***	0.00589***	0.00392***	0.00292***	0.0022***	0.00161***
R^2	0.34275	0.20106	0.12776	0.06055	0.02666	0.00933	0.0364

|	$\hat{\beta}_{0.5}$	$\hat{\beta}_{0.6}$	$\hat{\beta}_{0.7}$	$\hat{\beta}_{0.8}$	$\hat{\beta}_{0.9}$	$\hat{\beta}_{0.95}$	$\hat{\beta}_{0.99}$
Constant	0.0238*	0.03992***	0.05966***	0.08768***	0.13773***	0.19665***	0.38881***
$pre\_ean_{i,t}$	0.0012	0.00018	-0.00094	-0.00236*	-0.00455***	-0.00623***	-0.00927***
$ean_{i,t}$	-0.00044	0.00169	0.00409*	0.00738***	0.01299***	0.01995***	0.03885***
$post\_ean_{i,t}$	0.00556*	0.00666*	0.00835*	0.01117*	0.01679*	0.0235*	0.03431***
$news_{i,t}$	-0.00048*	0.00014	0.00083*	0.00172***	0.00324***	0.00484***	0.00933***
$news\_sector_{i,t}$	0.00105*	0.00106*	0.00109*	0.00117*	0.00139	0.00172	0.00037
$news\_day_{j,t}$	-0.00115	-0.0003	0.00053	0.00138	0.00208	0.00212	0.00192
$\log(size_{i,t})$	-0.00102*	-0.00154***	-0.00218***	-0.00312***	-0.00481***	-0.00687***	-0.01378***
$btm_{i,t-1}$	0.00047*	0.00075***	0.00106***	0.00143***	0.00192***	0.00227***	0.00244***
$turn_{i,t-1}\cdot days\_ean_{i,t}$	-0.00049	0.00005	0.00065	0.00145*	0.00286***	0.00405***	0.00654***
$(btm_{i,t-1}\leq 1)\cdot post\_ean_{i,t}$	-0.00581*	-0.0032	-0.00059	0.00227	0.00613	0.0099	0.02214*
$supr_{i,t}$	0.01337***	0.01305***	0.01304***	0.01382*	0.01791*	0.02242	0.03353
$eps_{i,t}$	0.00121***	0.00114*	0.00108*	0.00097	0.0007	0.0001	-0.00163
$lm\_tone_{i,t}$	0.00141***	0.00146***	0.00156***	0.00174***	0.00203***	0.00218***	0.00119
$hv\_tone_{i,t}$	0.00161***	0.00107*	0.00051	-0.00012	-0.00102	-0.00158	-0.0023
R^2	0.00833	0.02487	0.05887	0.13014	0.20764	0.37761	

R^2 0.00364
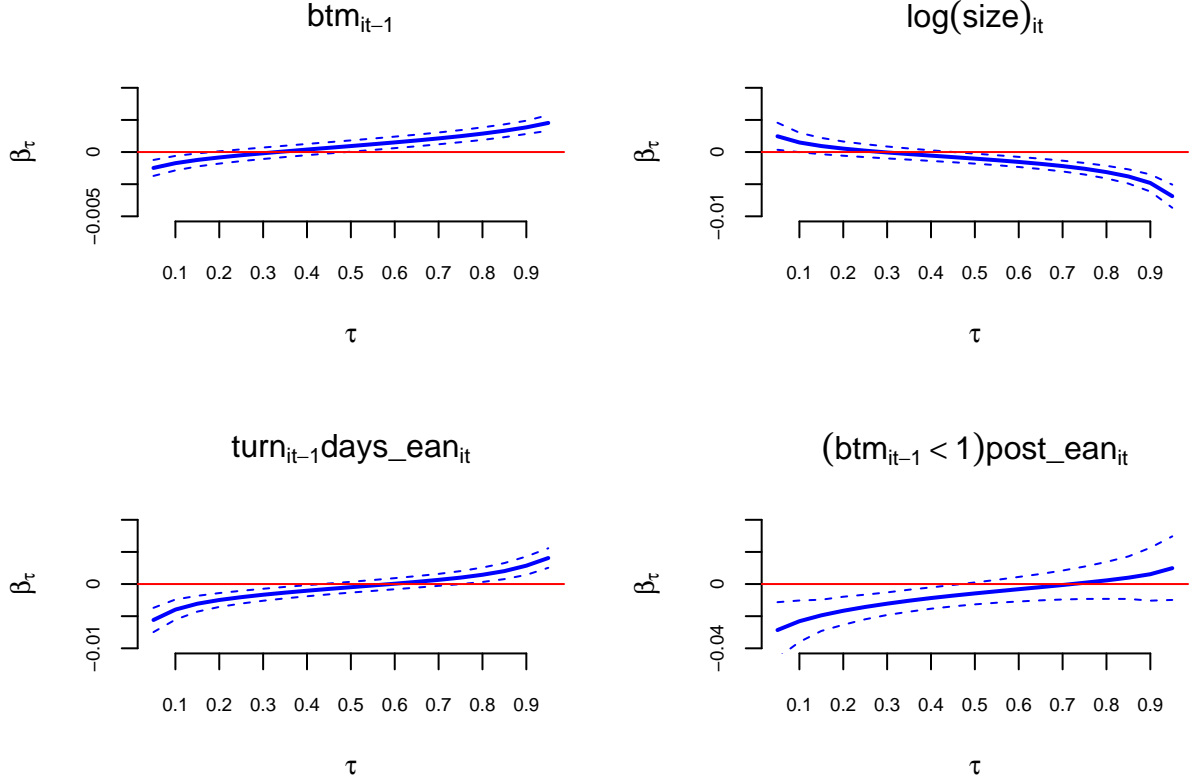
Figure 2: Coefficient profiles of selected investor perception related variables for different τ-levels

effects are all positive, as a result, the sign of the sentiment tone determines the expected direction of the conditional return reaction. Whereas, *lm_tone* tend to shift the whole distribution, *surp*, *hv_tone* and *eps* show significant impact primary on lower expectiles.

The effect of news releases is positive for higher expectiles and is sharpened by the sign of the associated sentiment tone. Simultaneously, we observe a strongly negative effect of news arrivals for low return expectiles regardless of the news tone. That is, also positive news expand the gap between return expectiles and increases thus the dispersion of associated abnormal returns. This effect can be connected to the 'sidedness' of Sarkar and Schwarz (2009) where the disproportionaly higher seller or buyer quote at news arrivals.

In summary, some of the effects such as *pre_ean*, *ean*, *news*, *btm*, and *turn · days_ean* impact both higher and lower expectiles with opposite effects (negative versus positive). As a result, they have a "widening" effect on the distribution as expectiles below and above the mean are pushed apart. *lm_tone* shifts all expectiles either to the right or to the left depending on the sign of the tone. Other considered effects are one-sided and change either the right or the left tail of the return distribution. We explore the resulting changes in shape of the conditional return distribution regarding its dispersion and skewness as well as tail measures in the next section.
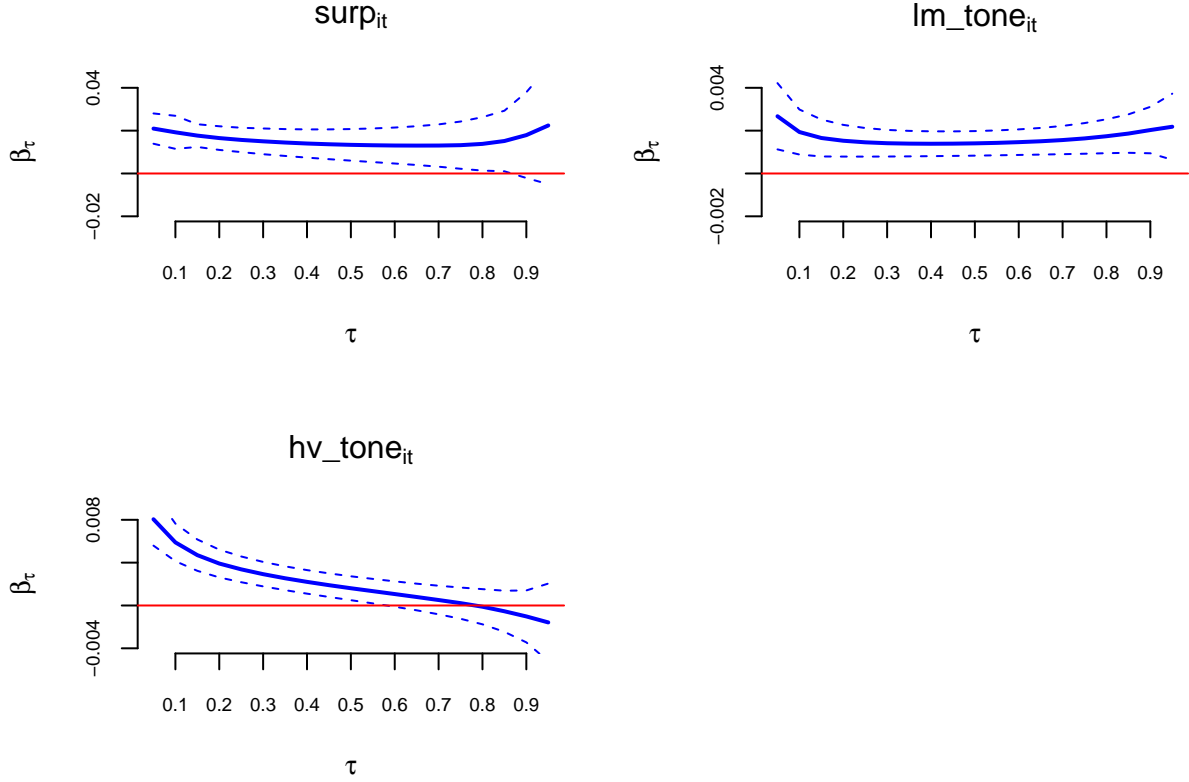
Figure 3: Coefficient profiles of the investor sentiment related variables for different τ-levels

3 Sentiment effects on the shape of the return distribution

In this section, we quantify the effects of the considered events as news releases and earnings announcements on the shape of the return distribution. To capture the later, we use recently proposed measures of dispertion and skewness based on expectiles (Eberl (2022)) as well as the connection of expectiles and expected shortfall established by Taylor (2008).

We use boostrap to check for significance of our results. The associated confidence intervals are obtained by cross-sectional and temporal bootstrap as defined in Kapetanios (2008) using 1000 bootstrap draws. To preserve possible serial dependence, we use block bootstrap for the time-series dimension with block length of 60.

3.1 Expectile-based dispertion and skewness measures

The estimated the conditional expectile models facilitate comparison of the resulting impact on the distribution via expectile-based measures. One useful concept - *interexpectile difference* (or range) was introduced in financial context by Bellini et al. (2018) to measure variability of a risk-neutral distribution. Their interexpectile difference for a random variable X with a finite expectation $\mathbb{E}(X)$ is defined as:

$$\Delta_\tau = e_{1-\tau}(X) - e_\tau(X)$$

for $\tau \in (0, 0.5)$.

The authors show that the proposed measure is consistent with the convex order. Recently, Eberl (2022) demonstrate that empirical interexpectile range is more efficient than the classical standard deviation for skewed and long-tailed distributions. Moreover, a class of expectile-based measures of skewness was proposed in Eberl and Klar (2022). For comparison here, we use their $s_{2\tau}$ (denoted further as s_τ) defined as:

$$s_\tau = \frac{1}{1-2\tau} \cdot \frac{e_{1-\tau}(X) + e_\tau(X) - 2\mathbb{E}(X)}{e_{1-\tau}(X) - e_\tau(X)}$$

for $\tau \in (0, 0.5)$. This skewness measure is normalized to $(-1, 1)$ and property $s_2(\tau) \geq 0$ ($s_2(\tau) \leq 0$) holds whenever $\tau \in (0, 0.5)$ and the distribution is right-skewed (left-skewed).

We take $\tau = 0.1$ for our analysis which was also used in Eberl and Klar (2022) and consider the straight forward adjustment for our conditional expectile regression setup of the model in (1). In our case, the estimated conditional interexpectile range becomes for $\tau = 0.1$ (for $\tau = 0.25$ analogous):

$$\hat{\Delta}_{0.1,i,t} = \hat{e}_{0.9,t}(r_{i,t}) - \hat{e}_{0.1,t}(r_{i,t}) = x_{i,t}^\top(\hat{\beta}_{0.9} - \hat{\beta}_{0.1}) + z_i^\top(\hat{\gamma}_{0.9} - \hat{\gamma}_{0.1})$$

where $x_{i,t}$ is a vector of the time-varying firm characteristics from (1) and z_i is a vector of firm characteristics which do not change over time. Similar, estimated conditional expectile-based skewness measure $\hat{s}_{\tau,i,t}$ is for $\tau = 0.1$:

$$\hat{s}_{0.1,i,t} = \frac{1}{1 - 2 \cdot 0.1} \cdot \frac{\hat{e}_{0.9,t}(r_{i,t}) + \hat{e}_{0.1,t}(r_{i,t}) - 2\hat{e}_{0.5,t}(r_{i,t})}{\hat{e}_{0.9,t}(r_{i,t}) - \hat{e}_{0.1,t}(r_{i,t})}$$

For each firm and each period, we compute $\hat{\Delta}_{0.1,i,t}$. Then, we average the resulting values considering time indices t, at which different events occur, separately. Thereby, the polarity (positive or negative) of an event is assessed via earnings suprises in case of *ean* and *post_ean*, and via dictionary based textual sentiment (average of *lm_tone* and *hv_tone*) for *pre_ean* and *news*. The resulting means are denoted as $\overline{\Delta}_{0.1}^{no\ event}$ and $\overline{\Delta}_{0.1}^{event}$. The associated significant changes in the interexpectile ranges for particular events, $\overline{\Delta}_{0.1}^{event} - \overline{\Delta}_{0.1}^{no\ event}$, as well as increase factors, $\overline{\Delta}_{0.1}^{event} / \overline{\Delta}_{0.1}^{no\ event}$, are provided in Table 2. As already mensioned, significance is obtained via cross-sectional and temporal bootstrap confidence intervalls.

Table 2: Mean interexpectile ranges $\overline{\Delta}_{0.1}^{event}$ and mean expectile-based skewness $\overline{s}_{0.1}^{event}$ on event days. The increases in dispersion compared to the case of no event ($\overline{\Delta}_{0.1}^{event} - \overline{\Delta}_{0.1}^{noevent}$) are given in brackets. Symbols '***','**','*' indicate significance on 0.1%, 1%, 5% level respectively by bootstrap confidence intervalls.

event	$\overline{\Delta}_{0.1}^{event} (\overline{\Delta}_{0.1}^{event} - \overline{\Delta}_{0.1}^{noevent})$	$\overline{\Delta}_{0.1}^{event} / \overline{\Delta}_{0.1}^{noevent}$	$\overline{s}_{0.1}^{event}$
noevent	0.0216 (-)	1.00	0.0036
pre_ean_positive	0.0246(0.0026)	1.12	-0.0084
pre_ean_negative	0.0315(0.0092**)	1.43	-0.0503
news_positive	0.0254 (0.0038***)	1.18	0.0239*
news_negative	0.0274 (0.0058***)	1.27	-0.0427*
ean_positive	0.0588 (0.0372***)	2.72	-0.0018
ean_negative	0.0659 (0.0443***)	3.05	-0.0255
post_ean_positive	0.0808 (0.0592***)	3.74	0.0048
post_ean_negative	0.0861 (0.0645***)	3.99	0.0084

We observe a substantial increase in dispersion on days around earnings announcements[1]. This effect possibly traces back according to Berkman et al. (2009) to opinion difference and opinion alignment after earnings announcements. Our findings confirm therefore the results of previous studies detecting increase in volatility on such event days. Also the asymmetric effect on dispertion shows up in higher increase of the interexpectile ranges in cases where the sentiment is negative and aligned with recently provided evidence on attention-based

---

[1]A significant increase of the dispersion on the days before earnings announcements shows up only for negative sentiment and iresults from the interplay of the effects of sentiment variables, *pre_ean* and *turn* $*$ *days_ean*.

trading in Barber et al. (2022). The effect of positive or negative news on the changes in interexpectile range ais much more subtle. News releases enhance the increase in dispersion of over 20% on average.

In addition, the resulting per-event average expectile-based skewness, $\overline{s}_{0.1}^{event}$, is also presented in Table 2. It reveals, that news releases induce significant positive and negative skewness of the return distribution when textual news polarity is positive or negative respectively, which generalizes the previous findings of Ekholm and Pasternack (2005) for the case of negative news releases.

4 The impact of the investor opinion channels on the conditional expected shortfall

A frequently used risk measure is the α-expected shortfall, denoted as ES_α. It is connected to the quantile q_α of the target distribution. For the lower tail of the later this risk measure is defined as $ES_\alpha = \mathbb{E}(Y|Y < q_\alpha)$.

Jones (1994) shows that there is a one-to-one correspondence between quantiles and expectiles of a distribution. Based thereof, Taylor (2008) proposes to use expectiles to estimate ES_α with $\alpha \le 0.5$ via:

$$ES_\alpha = \left(1 + \frac{\tau}{(1 - 2\tau)\alpha}\right) e_\tau - \frac{\tau}{(1 - 2\tau)\alpha} e_{0.5}, \tag{2}$$

where $\alpha \in (0, 0.5]$ denotes the quantile level, $\tau \in (0, 1)$ is the associated expectile level, and e_τ is the respective expectile. As Taylor (2008), the same relation holds for the conditional expected shortfall with respect to the conditional expectile.

Yao and Tong (1996) show the existence of a bijective function $\tau : (0, 1) \to (0, 1)$ which connects the quantile level α to the expectile level τ such that $q_\alpha = e_{\tau(\alpha)}$.

Thus, if the relation $\hat{\tau}(\alpha)$ or its estimate $\hat{\tau}(\alpha)$ is available, a plug-in estimator for the expected conditional shortfall based on the estimated conditional expectile analog to the autoregressive version in Taylor (2008) can be obtained:

$$
\begin{aligned}
ES_\alpha(r_{i,t}|x_{i,t}, z_i) &= \left(1 + \frac{\hat{\tau}(\alpha)}{(1 - 2\hat{\tau}(\alpha))\alpha}\right) e_{\hat{\tau}(\alpha)}(r_{i,t}|x_{i,t}, z_i) - \frac{\hat{\tau}(\alpha)}{(1 - 2\hat{\tau}(\alpha))\alpha} e_{0.5}(r_{i,t}|x_{i,t}, z_i) \\
&= \tilde{\beta}_{0,\alpha} + \tilde{\beta}_\alpha^\top \cdot x_{i,t} + \tilde{\gamma}_\alpha^\top \cdot z_i,
\end{aligned} \tag{3}
$$

with

$$\tilde{\beta}_\alpha = \hat{\beta}_{\hat{\tau}(\alpha)} - \frac{\hat{\tau}(\alpha)}{(1 - 2\hat{\tau}(\alpha))\alpha} \cdot \left(\hat{\beta}_{0.5} - \hat{\beta}_{\hat{\tau}(\alpha)}\right). \tag{4}$$

and $\tilde{\beta}_0$ and $\tilde{\gamma}$ are defined analogous.

To obtain the actual estimates, we first estimate the conditional expectile model as in (1) for a fine grid of $\tau = 0.001, 0.0025, 0.005, 0.01, 0.02, \dots, 0.1$. Following Taylor (2008), we construct an empirical cumulative distribution function by counting the observations below the estimated expectile profile for each of the τ-levels, such that for each τ we have the corresponding quantile index α. We then interpolate linearly between the resulting points to obtain $\hat{\tau}(\alpha)$ particular for $\alpha \in \{0.01, 0.05, 0.1\}$. The estimated correspondence is plotted in the left plot of Figure 4.

For instance, in our model specification the lower quantile levels $\alpha \in \{0.01, 0.05, 0.1\}$ correspond to the expectile levels $\tau \in \{0.0027, 0.0214, 0.0533\}$ respectively, see also Figure 4.
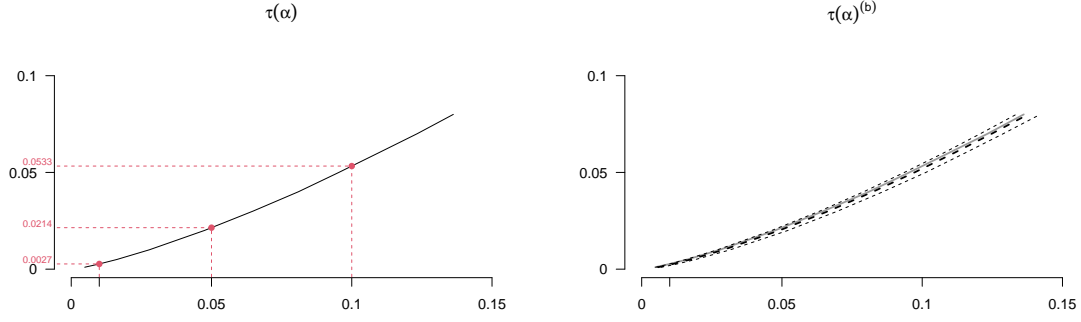
13

Figure 4: The estimated relation between the expectile level τ and the quantile level α for the NASDAQ return distribution. The left plot shows estimated $\tau(\alpha)$ with expectile levels shown for quantile levels $\alpha \in \{0.01, 0.05, 0.1\}$. The right plot shows estimated $\tau(\alpha)$ using the original NASDAQ return data (solid line) with boostrap based mean and 95% confidence intervals (dashed lines).

For the target $\alpha \in \{0.01, 0.05, 0.1\}$ we compute the resulting coefficients $\tilde{\beta}_\alpha$ which stand for the investor opinion effects on the conditional expected shortfall.

The bootstrapping is carried out as follows. For each bootstrap draw $b = 1, \ldots, B$, we, first, estimate the ER model for a fine grid of τ-values in the left tail $0.001, 0.0025, 0.005, 0.01, 0.02, \ldots, 0.1$. Then using the fitted conditional expectiles, we estimate the correspondence $\hat{\tau}^{(b)}(\alpha)$ by assigning to each τ-value the proportion of observations α in the bootstrap sample lying below the expectile profile $\{e_{tau}(r_{it}^{(b)})\}_{i=1,t=1}^{N,T}$. By interpolating linearly between the points, we determine the needed τ-levels for the target α-levels of $0.01, 0.05, 0.1$. The estimated $\hat{\tau}(\alpha)$ with 95% bootstrap confidence intervals is presented in right plot of Figure 4.

Finally, we compute $\tilde{\beta}_\alpha^{(b)}$ for each bootstrap sample as specified in (4), that is: $\tilde{\beta}_\alpha^{(b)} = \hat{\beta}_{\hat{\tau}^{(b)}(\alpha)}^{(b)} - \frac{\hat{\tau}^{(b)}(\alpha)}{\left(1 - 2\hat{\tau}^{(b)}(\alpha)\right)\alpha} \cdot \left(\hat{\beta}_{0.5} - \hat{\beta}_{\hat{\tau}^{(b)}(\alpha)}\right)$. Based on the resulting bootstrap distribution of $\left\{\tilde{\beta}_\alpha^{(b)}\right\}_{b=1}^{B}$ we, then, construct the confidence intervals.

Table 3: The estimated coefficients $\tilde{\beta}_\alpha$ measuring the effects of investor opinion on the expected shortfall. '***', '**', and '*' indicate significance of 0.001, 0.01, and 0.05 respectively obtained based on boostrap confidence intervalls.

	$ES_{0.01}$	$ES_{0.05}$	$ES_{0.1}$
Constant	-0.2461	-0.1253	-0.0741
$pre\_ean_{i,t}$	0.0238**	0.017***	0.0129***
$ean_{i,t}$	-0.0557***	-0.0508***	-0.0388***
$post\_ean_{i,t}$	-0.0403	-0.0391	-0.03
$news\_i,t$	-0.0337***	-0.0185***	-0.0128***
$news\_sector_{i,t}$	-0.0149	-0.0013	0.0007
$news\_day_{i,t}$	0.0040	-0.0081	-0.0074*
$\log(size_{i,t})$	0.0104	0.0059	0.0038
$btm_{i,t-1}$	-0.0020	-0.0022	-0.0022
$turn_{i,t-1} \cdot days\_ean_{i,t}$	-0.0134*	-0.0095**	-0.0066***
$(btm_{i,t-1} \leq 1) \cdot post\_ean_{i,t}$	-0.0572*	-0.0397*	-0.0349
$supr_{i,t}$	-0.0213	0.0344	0.0249
$eps_{i,t}$	0.0001	0.0051	0.0038
$lm\_tone_{i,t}$	0.0162	0.0076	0.0047
$hv\_tone_{i,t}$	0.0405	0.0257**	0.0174**

For $\alpha = 0.01, 0.05$ and 0.1 we confirm significant effects of *pre_ean*, *ean*, *news*, turnover around earnings announcement days $turn \cdot days\_ean$ and the post earnings announcements effect of btm ratio smaller than one $(btm \leq 1) \cdot post\_ean$ for $\alpha = 0.01$. Neither *lm_tone* nor *ml_tone* show significant impact. However, for *hv_tone* we find a significant effect for $\alpha = 0.05$ and 0.1.

As in the case of conditional expectile model, earnings announcements and shares turnover around earnings announcements exhibit significant negative effects on the conditional expected shortfall. Also, the mentioned "buy on rumor, sell on news" effect (positive coefficient estimate for *pre_ean* and negative for *ean* and *news*) persists in the conditional expected shortfall which makes it risk management relevant.

Based on our bootstrap procedure, we confirm significant effects of *pre_ean*, *ean*, and *news* on ES_α with $\alpha = 0.01, 0.05, 0.1$. Besides that, turnover around earnings announcement days $turn \cdot days\_ean$ and text sentiment *hv_tone* show significant impact on $ES_{0.05}$ and $ES_{0.1}$.

As in the case of conditional expectile model, earnings announcements and shares turnover around earnings announcements exhibit significant negative effects on the conditional expected shortfall. Significant effects of *pre_ean* , *ean* and *news* persist in the conditional expected shortfall consideration. Also, the overall significant asymmetric effect of news releases with positive and negative sentiment is confirmed in case of $ES_{0.05}$ and $ES_{0.1}$, which makes it risk management relevant.

5   Conclusion

We analyzed the distributional effects of the investor opinion on asset prices based on a panel of abnormal NASDAQ returns for the time period of April 2015 - August 2019. We adopted the information flow structure of Su et al. (2021) where investor opinion comprises such channels as investor attention, perception and sentiment. The proxies for each of the channels based on previous related studies acted as covariates in an expectile regression model introduced by Newey and Powell (1987). We use abnormal return as response to assess the distributional effects beyond the conditional mean regression and without imposing any assumptions on the response distribution.

The majority of the estimated effects depends on the expectile level or the "extremeness" τ in a nonlinear way and the associated estimates change sign when moving from the left to the right tail or vice versa of the conditional response distribution. A close up analysis of grabbing-attention events (new releases and days around earnings announcements) with positive and negative sentiment shows a significant asymmetric increase in dispersion, measured by interexpectile range, occuring on the event days. Particularly interesting appears the asymmetric skewness effect of news, measured through expectile skewness, since positive/ negative news induce significant left/right skewness of the distribution. One possible explanation for the change in skewness can be the "market sideness" of Sarkar and Schwarz (2009). Moreover, analogous significant effects are found in the conditional expected shortfall model, which we obtain by plugging in our conditional expectile formulation in the expected-shortfall-expectile relation derived by Taylor (2008). Our findings therefore have important implications for risk management assessment of the investor opinion formation channels.

A direction of possible further research would be to extend investor sentiment to topics based representation and assess the effect of different news topics on the conditional distribution and on the conditional expected shortfall of stock returns.

Appendix

List of the firm tickers in the dataset

The ticker list for the dataset is: 'AAPL', 'AMZN', 'TSLA', 'FB', 'BA', 'NFLX', 'DIS', 'EFX', 'BAC', 'INTC', 'F', 'GE', 'GM', 'MSFT', 'SBUX', 'AIR', 'AAL', 'IBM', 'JPM', 'CMG', 'WFC', 'C', 'TWTR', 'WMT', 'MCD', 'AMD', 'NVDA', 'JNJ', 'GS', 'BABA', 'CAT', 'MU', 'CSCO', 'XOM', 'CVX', 'BP', 'GOOGL', 'GPRO', 'COST', 'HD', 'NKE', 'KO', 'AXP', 'TGT', 'ATVI', 'CMCSA', 'DAL', 'LMT', 'T', 'ABBV', 'PFE', 'GILD', 'ADBE', 'CRM', 'VZ', 'AVGO', 'BX', 'LULU', 'BLK', 'UNH', 'KMI', 'BBY', 'PG', 'AGI', 'AMAT', 'MRK', 'M', 'BIDU', 'QCOM', 'FDX', 'AMGN', 'BMY', 'ORCL', 'BHP', 'MA', 'KR', 'MO', 'GME', 'PM', 'CHK', 'MMM', 'BBBY', 'COP', 'IRBT', 'MS', 'FCX', 'HAL', 'HPQ', 'UAL','JWN', 'CVS', 'V', 'EA', 'STZ', 'GLW', 'ADP', 'AZN', 'EBAY', 'ACN', 'PEP'.

News article data

In total, we have 30,424 news articles, and the average word count of the final news article sample is 1,114.

Table 4: Description of the final textual corpus

Year	News Articles (merged)	Average Word Count
2015	4,395	823
2016	5,391	840
2017	5,738	923
2018	8,465	1,352
2019	6,435	1,401
All years	30,424	1,114

Empirical data summary

We provide a standard summary of data for each variable in Table 5 and Table 6, a time period and sector wise expectile summaries of the return distributions in Table 7 and Table 8, and per-expectile summary of explanatory variables to give the reader an overview of the data structure in Table 9.

Table 5: Summary statistics for the variables: daily firm excess returns r, the logarithm of firm size $\log(size)$, firm btm ratio btm, and shares turnover $turn$.

	$r_{.,.}$	$\log(size_{.,.})$	$btm_{.,.}$	$turn_{.,.}$
mean	0.0001	24.7786	0.5322	1.8701
sd	0.0171	1.4178	1.5714	1.9573
skewness	0.3338	-0.9532	9.6404	3.0000

Table 6: Summary statistics for the variables: earning surprise $supr$ only for the days of earnings announcements, news sentiment based on LM dictionary denoted as lm_tone and on HV dictionary denoted as hv_tone for the days of news releases.

	$supr$	eps	lm_tone	hv_tone
mean	0.0737	1.3275	-0.0453	0.3844
sd	0.2255	1.5361	0.4586	0.2145
skewness	15.2487	3.0413	0.0584	-0.5822

In Table 7 empirical return expectiles $\hat{e}_\tau$ for $\tau = 0.05, 0.25, 0.5, 0.75$, and 0.95 and different sub periods of the considered time span are reported. We observe that the tail indices remain approximately the same over time. In Table 8 we report the empirical return expectiles $\hat{e}_\tau$ for $\tau = 0.05, 0.25, 0.5, 0.75$, and 0.95 sector wise.

Table 7: Empirical expectiles for the excess returns for the whole time span 2015-2019, for the first sub period 2015-2016 and for the second sub period 2017-2019.

	$\hat{e}_{0.05}$	$\hat{e}_{0.25}$	$\hat{e}_{0.5}$	$\hat{e}_{0.75}$	$\hat{e}_{0.95}$
2015-2019	-0.0190	-0.0058	0.0001	0.0060	0.0193
2015-2016	-0.0195	-0.0058	0.0002	0.0062	0.0207
2017-2019	-0.0186	-0.0058	0.0000	0.0059	0.0184

Table 8: Empirical expectiles for the excess returns for the whole time span 2015-2019 for each of the ten sectors

Sector	$\hat{e}_{0.05}$	$\hat{e}_{0.25}$	$\hat{e}_{0.5}$	$\hat{e}_{0.75}$	$\hat{e}_{0.95}$
Basic Materials	-0.0345	-0.0115	0.0001	0.0120	0.0373
Consumer Discretionary	-0.0197	-0.0061	0.0000	0.0061	0.0192
Consumer Staples	-0.0158	-0.0050	0.0000	0.0050	0.0151
Energy	-0.0245	-0.0079	-0.0005	0.0072	0.0255
Finance	-0.0122	-0.0041	0.0001	0.0044	0.0130
Health Care	-0.0143	-0.0048	-0.0001	0.0046	0.0138
Industrials	-0.0139	-0.0044	0.0003	0.0050	0.0146
Technology	-0.0196	-0.0057	0.0005	0.0068	0.0215
Telecommunications	-0.0115	-0.0038	0.0002	0.0041	0.0119
Utilities	-0.0193	-0.0069	-0.0007	0.0057	0.0190

Here we observe substantial variation of empirical expectiles between sectors. Based on this observation, we include sector and firm dummies to account for the posible individual variability and the variability across sectors.

In the following Table 9 per-expectile distributional summary is displayed. The summary statistics for the explanatory variables is conditioned on excess returns belonging to a specific empirical expectile interval: $[\min(r), \hat{e}_{0.05}), [\hat{e}_{0.05}, \hat{e}_{0.25}), [\hat{e}_{0.25}, \hat{e}_{0.5}), [\hat{e}_{0.5}, \hat{e}_{0.75}), [\hat{e}_{0.75}, \hat{e}_{0.95}),$ and $[\hat{e}_{0.95}, \max(r)]$. We report the mean and the standard deviation in parenthesis for the variables.

Table 9: Per-expectile means (standard deviations) of the variables shares turnover $turn$, logarithm of size $\log(size)$, btm ratios btm, indicator of news arrivals $news$, and indicator of earnings announcements ean.

	$turn$	$\log(size)$	btm	$news$	ean	eps
$r \leq \hat{e}_{0.05}$	3.187 (2.878)	23.937 (1.637)	0.679 (2.124)	0.350 (0.477)	0.047 (0.21)	1.213 (1.148)
$\hat{e}_{0.05} \leq r \leq \hat{e}_{0.25}$	1.913 (1.921)	24.747 (1.404)	0.559 (1.619)	0.258 (0.437)	0.011 (0.105)	1.347 (1.255)
$\hat{e}_{0.25} \leq r \leq \hat{e}_{0.5}$	1.459 (1.420)	25.040 (1.260)	0.474 (1.279)	0.241 (0.428)	0.009 (0.092)	1.607 (1.692)
$\hat{e}_{0.5} \leq r \leq \hat{e}_{0.75}$	1.479 (1.486)	25.040 (1.265)	0.457 (1.262)	0.244 (0.429)	0.009 (0.093)	1.414 (1.631)
$\hat{e}_{0.75} \leq r \leq \hat{e}_{0.95}$	1.871 (1.891)	24.767 (1.397)	0.552 (1.662)	0.266 (0.442)	0.015 (0.121)	1.545 (1.724)
$r \geq \hat{e}_{0.95}$	3.110 (2.784)	23.999 (1.626)	0.695 (2.183)	0.353 (0.478)	0.052 (0.222)	1.415 (1.514)

Interestingly, the extreme return states $r \leq \hat{e}_{0.05}$ and $r \geq \hat{e}_{0.95}$ are associated with substantially higher mean turnover and earnings announcements rate, slightly lower mean size and somewhat higher mean btm ratios and news arrivals rate.

Based on the data description above, one expects possibly different effects of the considered covariates on the location, scale and shape of the conditional return distribution. As in the summaries above, it is possible to represent the conditional return distribution through its conditional expectiles. A consistent estimator of the heterogeneous effects over the return distribution can be obtained by expectile regression.

References

Al-Nasseri, A., Menla Ali, F., and Tucker, A. (2021), "Investor sentiment and the dispersion of stock returns: Evidence based on the social network of investors," *International Review of Financial Analysis*, 78, 101910.

Andrei, D. and Hasler, M. (2020), "Dynamic Attention Behavior Under Return Predictability," *Management Science*, 66, 2906–2928.

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

Audrino, F., Sigrist, F., and Ballinari, D. (2020), "The impact of sentiment and attention measures on stock market volatility," *International Journal of Forecasting*, 36, 334–357.

Baker, M. and Wurgler, J. (2007), "Investor sentiment in the stock market," *Journal of economic perspectives*, 21, 129–152.

Barber, B. M., Huang, X., Odean, T., and Schwarz, C. (2022), "Attention-Induced Trading and Returns: Evidence from Robinhood Users," *The Journal of Finance*, 77, 3141–3190.

Barberis, N., Shleifer, A., and Vishny, R. (1998), "A model of investor sentiment," *Journal of financial economics*, 49, 307–343.

Barry, A., Oualkacha, K., and Charpentier, A. (2022), "A new GEE method to account for heteroscedasticity using asymmetric least-square regressions," *Journal of Applied Statistics*, 49, 3564–3590.

Bellini, F., Mercuri, L., and Rroji, E. (2018), "Implicit expectiles and measures of implied volatility," *Quantitative Finance*, 18, 1851–1864.

Berkman, H., Dimitrov, V., Jain, P. C., Koch, P. D., and Tice, S. (2009), "Sell on the news: Differences of opinion, short-sales constraints, and returns around earnings announcements," *Journal of Financial Economics*, 92, 376–399.

Cao, H. H., Coval, J. D., and Hirshleifer, D. (2002), "Sidelined Investors, Trading-Generated News, and Security Returns," *The Review of Financial Studies*, 15, 615–648.

Chatterjee, S., John, K., and Yan, A. (2012), "Takeovers and Divergence of Investor Opinion," *The Review of Financial Studies*, 25, 227–277.

Chen, J., Tang, G., Yao, J., and Zhou, G. (2022), "Investor Attention and Stock Returns," *Journal of Financial and Quantitative Analysis*, 57, 455–484.

Corwin, S. A. and Coughenour, J. F. (2008), "Limited Attention and the Allocation of Effort in Securities Trading," *The Journal of Finance*, 63, 3031–3067.

Das, S. R. and Chen, M. Y. (2007), "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science*, 53, 1375–1388.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990), "Noise trader risk in financial markets," *Journal of political Economy*, 98, 703–738.

Dellavigna, S. and Pollet, J. M. (2009), "Investor Inattention and Friday Earnings Announcements," *The Journal of Finance*, 64, 709–749.

Diogo Barry, A., Charpentier, A., and Oualkacha, K. (2016), "Quantile and Expectile Regression for random effects model," working paper or preprint.

Donnelly, R. (2014), "The book-to-market ratio, optimism and valuation," *Journal of Behavioral and Experimental Finance*, 4, 14–24.

Drake, M. S., Jennings, J., Roulstone, D. T., and Thornock, J. R. (2017), "The Comovement of Investor Attention," *Management Science*, 63, 2847–2867.

Eberl, A. and Klar, B. (2022), "Expectile-based measures of skewness," *Scandinavian Journal of Statistics*, 49, 373–399.

Eberl, A., K. B. (2022), "Stochastic orders and measures of skewness and dispersion based on expectiles." *Statistical papers*.

Ekholm, A. and Pasternack, D. (2005), "The negative news threshold—An explanation for negative skewness in stock returns," *The European Journal of Finance*, 11, 511–529.

Engelberg, J., McLean, R. D., and Potiff, J. (2018), "Anomalies and News," *The Journal of Finance*, 73, 1971–2001.

Fama, E. F. and French, K. R. (1992), "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

Fang, H., Chung, C.-P., Lu, Y.-C., Lee, Y.-H., and Wang, W.-H. (2021), "The impacts of investors' sentiments on stock returns using fintech approaches," *International Review of Financial Analysis*, 77, 101858.

Frankel, R., Jennings, J., and Lee, J. (2022), "Disclosure Sentiment: Machine Learning vs. Dictionary Methods," *Management Science*, 68, 5514–5532.

Friend, I. and Lang, L. H. (1988), "The size effect on stock returns: Is it simply a risk effect not adequately reflected by the usual measures?" *Journal of Banking & Finance*, 12, 13–30.

Füss, R., Guidolin, M., and Koeppel, C. (2020), "Sentiment Risk Premia in the Cross-Section of Global Equity," *SSRN Electronic Journal*.

Hagenau, M., Liebmann, M., and Neumann, D. (2013), "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems*, 55, 685–697.

He, Z. (2022), "Asymmetric impacts of individual investor sentiment on the time-varying risk-return relation in stock market," *International Review of Economics & Finance*, 78, 177–194.

Henry, E. (2006), "Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm," *Journal of Emerging Technologies in Accounting*, 3, 1–19.

Jegadeesh, N. and Wu, D. (2013), "Word power: A new approach for content analysis," *Journal of Financial Economics*, 110, 712–729.

Jones, M. (1994), "Expectiles and M-quantiles are quantiles," *Statistics & Probability Letters*, 20, 149–153.

Kapetanios, G. (2008), "A bootstrap procedure for panel data sets with many cross-sectional units," *The Econometrics Journal*, 11, 377–395.

Kim, S.-H. and Kim, D. (2014), "Investor sentiment from internet message postings and the predictability of stock returns," *Journal of Economic Behavior & Organization*, 107, 708–729, Empirical Behavioral Finance.

Kneib, T. (2013), "Beyond mean regression," *Statistical Modelling*, 13, 275–303.

Koenker, R. and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Kothari, S. and Shanken, J. (1997), "Book-to-market, dividend yield, and expected market returns: A time-series analysis," *Journal of Financial Economics*, 44, 169–203.

Loh, R. K. (2010), "Investor Inattention and the Underreaction to Stock Recommendations," *Financial Management*, 39, 1223–1251.

Loughran, T. and McDonald, B. (2011), "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance*, 66, 35–65.

Luss, R. and D'Aspremont, A. (2015), "Predicting abnormal returns from news using text classification," *Quantitative Finance*, 15, 999–1012.

Ma, C., Xiao, S., and Ma, Z. (2018), "Investor sentiment and the prediction of stock returns: a quantile regression approach," *Applied Economics*, 50, 5401–5415.

Maheu, J. M. and McCurdy, T. H. (2004), "News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns," *The Journal of Finance*, 59, 755–793.

Newey, W. K. and Powell, J. L. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847.

Nguyen, T. H., Shirai, K., and Velcin, J. (2015), "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, 42, 9603–9611.

Ni, Z.-X., Wang, D.-Z., and Xue, W.-J. (2015), "Investor sentiment and its nonlinear effect on stock returns—New evidence from the Chinese stock market based on panel quantile regression model," *Economic Modelling*, 50, 266–274.

Peasnell, K. (1982), "Some FormalL Connections Between Economic Values and Yields and Accounting Numbers," *Journal of Business Finance & Accounting*, 9, 361–381.

Pontiff, J. and Schall, L. D. (1998), "Book-to-market ratios as predictors of market returns," *Journal of Financial Economics*, 49, 141–160.

Sarkar, A. and Schwarz, R. A. (2009), "Market Sidedness: Insights into Motives for Trade Initiation," *The Journal of Finance*, 64, 375–423.

Shang, Z., Brooks, C., and McCloy, R. (2014), "Are investors guided by the news disclosed by companies or by journalists?" *Journal of Behavioral and Experimental Finance*, 1, 45–60.

Smales, L. A. (2016), "Time-varying relationship of news sentiment, implied volatility and stock returns," *Applied Economics*, 48, 4942–4960.

Su, Y., Qu, Y., and Kang, Y. (2021), "Online public opinion and asset prices: a literature review," *Data Science in Finance and Economics*, 1, 60–76.

Taylor, J. W. (2008), "Estimating Value at Risk and Expected Shortfall Using Expectiles," *Journal of Financial Econometrics*, 6, 231–252.

Waltrup, L. S., Sobotka, F., Kneib, T., and Kauermann, G. (2015), "Expectile and quantile regression—David and Goliath?" *Statistical Modelling*, 15, 433–456.

Yao, Q. and Tong, H. (1996), "Asymmetric least squares regression estimation: A nonparametric approach," *Journal of Nonparametric Statistics*, 6, 273–292.

Yuan, Y. (2015), "Market-wide attention, trading, and stock returns," *Journal of Financial Economics*, 116, 548–564.

Zeileis, A., Köll, S., and Graham, N. (2020), "Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R," *Journal of Statistical Software*, 95, 1–36.