The relevance of high-frequency news analytics for lower-frequency investment strategies*

David Happersberger[†a,b] Harald Lohre[a,b] Ingmar Nolte[a] Carsten Rother[b,c]

[a] *Centre for Financial Econometrics, Asset Markets and Macroeconomic Policy, Lancaster University Management School, United Kingdom*
[b] *Invesco Quantitative Strategies, Frankfurt am Main, Germany*
[c] *University of Hamburg, Germany*

This version: January 15, 2020

Abstract

This paper investigates whether lower-frequency investment strategies can be enhanced by the use of high-frequency news analytics. First, we study the cross-sectional characteristics of a broad set of indicators generated from news flow data. While long-short equity portfolios based on news sentiment indicators show promise in global and European stock universes, results for the US and Japan are rather modest. Second, we investigate the benefits of incorporating news-based equity factors into multi-factor investment strategies. Risk-based asset allocation strategies such as minimum-variance and risk parity strategies benefit from augmenting a portfolio of global equity factors by news sentiment-related equity factors. Also, active factor allocation strategies can be enhanced by utilizing the information embedded in news flow data. Factor timing using fundamental and technical time-series predictors generates statistically significant and economically relevant results. Similarly, a factor tilting strategy that exploits cross-sectional news-related information outperforms an equally weighted benchmark portfolio.

JEL classification: G11, G12, G17.

Keywords: News Analytics, News Sentiment, Portfolio Sorts, Factor Timing, Factor Tilting, Parametric Portfolio Policies.

*We thank Stefan Mittnik, and the participants of the 2018 CEQURA Conference on Advances in Financial and Insurance Risk Management in Munich for fruitful discussions and suggestions. Special thanks goes to Matthias Kerling for valuable guidance on data processing. This work was supported by funding from the Economic and Social Research Council (UK). Note that this paper expresses the authors' views that do not have to coincide with those of Invesco.

†Correspondence to: Centre for Financial Econometrics, Asset Markets and Macroeconomic Policy, Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, United Kingdom. Email: d.happersberger@lancaster.ac.uk.

1 Introduction

The proliferation of new alternative data sources opens for various new research avenues to enhance investment strategies, portfolio construction or risk forecasting. A highly promising route seeks to leverage news analytics that quantifies textual information from news wire articles and social media using natural language processing techniques. Interest in the relation between news and the stock market has been on the rise among both, academic scholars and industry practitioners. Tetlock (2007), Fang and Peress (2009), Heston and Sinha (2017) and Ke et al. (2019) are examples of this growing literature. While most studies concentrate on the short-term relationship between news and the cross-section of stock returns, there is only little evidence whether and how news analytics can be exploited for feasible investment strategies. We contribute to this strand of research by investigating the relevance of high frequency news analytics for lower-frequency investment strategies.

We use a unique global news data set to build a broad set of indicators to be tested in different low-frequency investment strategies. Specifically, we construct an international sample of real-time news releases at the firm-level between 2000 and 2017 collected by RavenPack.[1] RavenPack does not only provide the flow of news articles related to a firm but also quantifies the value-relevant information in each news article based on natural language processing algorithms. For example, a news article on a corruption scandal involving a firm's executives is associated with a negative score, while a news article regarding the successful development of a firm's new product is associated with a positive score.

Motivated by the literature on news analytics, we employ these firm-specific scores to derive news-based indicators which can be divided into the following four concepts: news volume, news sentiment, news trend and alternative news concepts. In particular, news volume, also referred to as media coverage or media attention, analyzes a firm's media presence (e.g., Barber and Odean, 2007; Fang and Peress, 2009). News sentiment was first studied by Tetlock (2007) and examines a news event's tone relating to a particular firm. News trend tries to detect time-series patterns in news sentiment (e.g., Leinweber and Sisk, 2011; Uhl et al., 2015). Alternative news concepts contain further, more complex ideas on how news analytics can be used to inform investment strategies. These include, for instance, the concept of news beta (Hafez, 2010) that measures the responsiveness of a firm's stock price to an overall news market sentiment or news significance that captures both mean and variance of news sentiment. To focus on pure news elements, we consider size- and industry-adjusted indicators. Applying various look-back windows we obtain a set of 36 indicators in total.

First, we examine the predictive content of the derived news-based indicators in the cross-section

[1]RavenPack is a leading news data provider and its database has been used in many studies, see e.g. Kolasinski et al. (2013); Dang et al. (2015); von Beschwitz et al. (2017).

of stock returns. To this end, we form equally weighted long-short portfolios according to the respective news indicators using a global universe of stocks. In contrast to Fang and Peress (2009), we do not find consistent evidence that firms with high media presence earn higher returns than firms with low media presence. Our finding might differ because we do not restrict our analysis to a few US newspapers and cover a much longer time horizon. Analyzing a significantly broader news data set including all types of news sources we expect our approach to lead to more robust results. In contrast, our findings support existing studies of Tetlock (2007) or Ke et al. (2019) who evidence that it is profitable to invest in long-short portfolios based on news sentiment indicators. Simple trading strategies earn significant returns and exhibit positive information coefficients, with Sharpe ratios above one. For the other concepts, we find news sentiment momentum and news significance to be particularly profitable. Notably, the performance of the news equity factors does not change when using a market-capitalization weighting scheme instead of equal-weighting. Performing spanning tests based on a standard set of equity factors (namely, value, quality, momentum, size and short-term reversal), we find the significant news indicators to still contribute in explaining the cross-section of stock returns, even though they exhibit a high correlation to the momentum factor.

Given that equity factors are found to exhibit region-specific effects (see e.g. Jacobs and Müller, 2020), we also conduct the cross-sectional analysis on a regional level. While the findings for Europe and a rest-of-the-world universe are even stronger than for the global universe, we do not evidence consistent significant cross-sectional stock return patterns for USA and Japan. Moreover, as long-term, factor-based investment management is usually based on equity factors that generate positive returns for longer horizons than one month, we additionally investigate long-term effects of the news-based equity factors. Most factors with significant one-month long-short portfolio returns exhibit a fast decay in subsequent months. Still, factors incorporating news sentiment over a longer horizon are more persistent and thus may be useful for long-term investment management.

With news sentiment equity factors earning significant returns and expanding the traditional equity factor investment opportunity set we next investigate whether news analytics are beneficial for multi-factor investment strategies. We first analyse whether risk-based factor allocation strategies can be enhanced by adding news-based factors to a representative set of global equity factors. Specifically, we consider an equally weighted portfolio, a minimum-variance portfolio and a risk parity portfolio. We document that all three risk-based allocation strategies benefit from augmenting the benchmark portfolio by news sentiment-related equity factors.

Given the time variation in equity factor returns a forecasting-based factor allocation strategy may add value over and above a passive factor allocation portfolio (see e.g., Asness, 2016; Arnott et al., 2016; Bender et al., 2018; Dichtl et al., 2019). We explore the benefits of active factor allocation when incorporating information from news flow data. To this end, we consider parametric portfolio policies that allow for timing factors conditioned on time series predictors and tilting factors based on cross-sectional factor characteristics. This approach avoids estimating the joint distribution of factor returns, but rather directly determines optimal factor allocation weights based on a set of information variables. A factor timing strategy relates factor returns to a variety of fundamental variables and technical indicators commonly used for predicting the equity risk premium. Based on the parametric portfolio policy framework of Brandt and Santa-Clara (2006), we assess the utility of information extracted from news flow data for factor timing strategies by comparing the resulting factor allocations to an equal-weighted benchmark. Using the same predictor set as Dichtl et al. (2019) we evaluate the time-series predictability of the fundamental variables and technical indicators for equity factors. We evidence that the statistical significance of the θ-coefficients defining the optimal portfolio weight of each factor in the factor allocation is limited. Nevertheless, factor timing is economically meaningful, as both factor timing strategies (including benchmark factors and adding news factors) outperforms the equal-weighted benchmark and therefore experience a positive information ratio. These gross figures have to be taken which a pinch of salt as the factor timing strategy requires a high turnover to follow the timing signals coming from fundamental variables and technical predictors. Accounting for transaction costs the performance drag is substantial leading to an underperformance compared to an equal-weighted benchmark and subsequently to a negative information ratio. Hence, the results amplifying the difficulty of effectively time factors after transaction costs.

For factor tilting we distill the set of news-based indicators on the level of equity factors to generate original equity factor characteristics. Utilizing the cross-sectional parametric policy framework of Brandt et al. (2009) we exploit the news-related factor characteristics to assess the predictive information embedded in the news flow data. Our empirical findings suggest that the benchmark portfolio of representative global equity factors benefits from utilizing news-based information. News sentiment-related factor characteristics show predictability in this portfolio utility context. Economically, we document higher risk-adjusted returns for the news-related tilting strategies compared to an equally weighted benchmark portfolio. When adding news-based equity factors to benchmark equity factor portfolio, predictability of the news-related factor characteristics weakens. Still, the economic relevance of the tilting strategies remain. All news sentiment-related tilting allocations exhibit positive information ratios, even after accounting for transaction costs.

We make two major contributions to the literature. First, we add to existing studies analyzing the cross-sectional effects of news flow data. While specific news phenomena have been examined for the US equity market by Tetlock (2007) or Fang and Peress (2009), among others, we study cross-sectional effects of various news indicators in a unified framework based on broad data set covering all types of news in global and regional universes, analyzing long-term effects as well. Second, to the best of our knowledge, this study is the first to assess the use of information embedded in news analytics for multi-factor investment strategies, including passive and forecasting-based factor allocation approaches.

The outline of the paper is as follows: Section 2 introduces the news analytics data and discusses the underlying ideas and the construction of the news-based indicators. Section 3 examines cross-sectional patterns in the derived news indicators, including regional and long-horizon analyses. In Section 4, we investigate the use of news-based indicators for multi-factor investment strategies, including factor timing and tilting. Section 5 concludes.

2 Condensing high-frequency news data into predictive indicators

2.1 News data

As main data source we utilize the news and sentiment data from RavenPack News Analytics. RavenPack systematically tracks, collects and analyzes real-time, firm-level business news from leading real-time news providers, including *Dow Jones Newswires*, the *Wall Street Journal*, *Barron's*, and other major publishers and web aggregators, including industry and business publications, regional and local newspapers, government and regulatory updates and trustworthy financial websites. In total, RavenPack features around 28,000 companies in over 130 countries (representing 98% of the investable global equity market) and covers news from a wide range of facts, opinions and corporate disclosures. The data are available from the year 2000, allowing for a backtest of over 18 years.

To transform unstructured news data items into structured granular data and indicators RavenPack Analytics 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. Second, RavenPack constructs a set of scores, rating different aspects of the relevant news items with respect to the respective entity based on natural language processing algorithms that effectively combine traditional linguistic analyses, financial expert consensus and market response methodologies. The following four major scores form the basis of the news indicators we will build:

- *Relevance (REL):* An integer score between 0 and 100, with higher values indicating greater relevance of the underlying news story for a given entity.

- *Event Relevance (EVR):* An integer score between 0 and 100 that reflects the relevance of the event in the story, with higher values indicating greater relevance.

- *Event Similarity Days (ESD):* An integer between 0 and 365 indicating the number of days since a similar event was detected over the last 365 days. The ESD thus allows to isolate the first news article in a chain of similar articles about a given news event.

- *Event Sentiment Score (ESS):* A granular score between -1.00 and $+1.00$ that represents the news sentiment for a given entity, where a negative (positive) score indicates negative (positive) sentiment and 0 indicates neutral sentiment. The ESS leverages RavenPack's event detection technology and produces a sentiment score every time an event is matched. In particular, the ESS is determined based on training sets in which experts with extensive experience and backgrounds in linguistics, finance and economics classify company-specific events and agree that these events generally convey a positive, neutral or negative sentiment.

2.2 Global equity data

To allow for a holistic investigation of the news analytics data, we assemble a representative and investable equity universe encompassing the constituents of global and regional equity indices from MSCI, FTSE, S&P, and STOXX. Company-specific data such as financial statement and price data are sourced from the Worldscope database. Having matched news and firm-level data, we consider a broad universe of 5,350 companies per month on average and 1,155,342 relevant news events in the sample period from January 2000 to December 2017. This translates to, on average, 94 news events per firm and month (cf. Table I).

[Table I about here]

Panel A of Table I gives further descriptive statistics of the number of news events per month and firm, reflecting a company's media presence which we call news volume in the following. We only consider relevant news events and therefore require a relevance score of at least 75. Initially, we do not restrict in terms of the event similarity days analytic since a repeated dissemination of the same or similar news events may be a useful indication of a company's media presence. As a consequence, we find a sample maximum of 57,528 relevant news events for one company within a month. Specifically, Facebook Inc.'s initial public offering in May 2012 was the biggest in technology history and therefore the major topic across all media channels.

The positive skewness and the huge maximum number of news indicate that news volume is largely driven by company size. Indeed, large companies account for the majority of news events: large companies have, on average, 208 news events per firm and month compared to 53 and 21 news events for medium-sized and small companies, respectively (see also Figure 1(a)). This fact is not only consistent with the literature on media and news indicating that large firms attract higher media attention but is also aligned with the intuition that large firms typically generate more news events (Ke et al., 2019, e.g.). To control for size effects, we will standardize the derived news indicators by market capitalization going forward (see details in Section 2.3).

[Figure 1 about here]

Figure 1(a) shows the evolution of news volume over the sample period. The number of news articles increases substantially from the beginning of the sample in 2000 to the year 2012, but stabilizes afterwards. In addition to RavenPack's changing media coverage, this time-series pattern is driven by both an increasing intensity of media coverage and a growing amount of firm activities. Figure 1(b) shows the evolution of the number of monthly news events per region. We differentiate between United States (USA), Japan (JAP), Europe (EUR), emerging markets (EM) and rest of the world (RES).[2] It is not surprising that US stocks exhibit the, by far, largest fraction of news events, followed by European stocks (cf. Table I). Figure 1(c) shows the number of daily news events over the years 2007 (upper part) and 2017 (lower part), conveying two different seasonal patterns: first, we observe a quarterly cycle that coincides with quarterly business reports (earnings announcements etc.).[3] Second, we observe a weekly cycle which is obviously due to a significantly reduced news dissemination on weekends. We control for both effects when constructing our indicators.

To explore the characteristics of the event sentiment score we examine Panel B of Table I. The number of ESS scores and firms is lower than the number of news events for two reasons: first, an event sentiment score is only assigned to a news event when it can be classified according to the RavenPack taxonomy. Second, we exclude news events with a neutral score and require the ESS to pass filters of 90 for relevance, event relevance and novelty to reduce noise (see Section 2.3 for more details on noise filtering). We observe that sentiment is slightly positive on average: the ESS has a mean of 0.17 and a median of 0.23, respectively. Panel (a) of Figure 2 shows the histogram of all event sentiment scores, when applying the described filters. We observe a slightly negative skewed and fat-tailed distribution. Panel 2(b) shows the evolution of the monthly ESS score averaged across firms, which is fairly stable with the exception of the time period of the global financial crisis in 2008.

[Figure 2 about here]

2.3 News-based indicators

In this section, we develop a broad set of indicators that aim to explain and predict (long-term) asset price variation utilizing information extracted from news flow data. The general use of news data for this purpose can be rationalized via the efficient markets hypothesis of Malkiel and Fama (1970), which can be seen as the theoretical basis for any return prediction analysis. Therein, market efficiency predicts that the expected return of a stock is dominated by unforecastable news, as this news is rapidly (in its starkest form, immediately) and fully incorporated in its price. The alternative hypothesis is that information in news flow data is not fully absorbed by market prices instantaneously, for reasons such as limits-to-arbitrage and limited attention (e.g. Baker and Wurgler, 2006; Tetlock,

[2]The rest of the world universe consists of the following developed countries: Australia, Canada, New Zealand, Israel and Hongkong. Emerging markets include those countries that are classified as emerging market by MSCI, FTSE, S&P, and STOXX. This classification is time-dependent. Emerging market countries are, for example, Brazil, Russia and India.

[3]As a robustness check, we perform an analysis excluding news events corresponding to earnings announcements when constructing the set of news indicators. Unreported results do not show significant differences to the results including earnings announcements data, suggesting that the analysis of news-based indicators is not solely driven by events concerning quarterly business reports.

2007; Ke et al., 2019). As a result, information contained in news flow data can be predictive of future asset prices. While this alternative hypothesis is by now considered uncontroversial for short horizons (e.g. daily or intradaily horizons), it is still not clear whether long-term investors can profit from information embedded in news flow data, facing investment horizons of one month or longer.

First, we filter the news data to reduce the noise in the signals. In particular, we only include firms with at least one news story. While it seems favorable to include as much information as possible (i.e. keep as many news events as possible), not all events are equally important. Therefore, we exclude news stories with neutral ESS and filter the data based on relevance, event relevance and event similarity days according to Hafez (2010), Kolasinski et al. (2013), Dang et al. (2015) and von Beschwitz et al. (2017): We only consider stories that are directly relevant to the mentioned company by only retaining data with a relevance score above 90. In a similar way, we only retain events with high relevance in a news story to avoid carrying unimportant news items, i.e. we require the event relevance score to be above 90. Furthermore, we only consider unique and novel news events. We hypothesize that the first instance of an event is most impactful and any subsequent repetition thereof can be expected to have a lesser impact. By retaining only events that have an event similarity days analytic above 90, we filter our data set down to only the most novel events within the last 90 days. As such, any analysis of the news events is less likely to be driven by the repetitive dissemination of the same or similar news events.[4]

In general, we proceed as follows when constructing a given news indicator: since our main analysis is conducted at a monthly frequency, we first aggregate the high-frequency news tick data to monthly indicators using indicator-specific functions. Second, we calculate each indicator for each firm in our investment universe using various look-back windows. As the required information differs among indicators, not all signals are based on the same number of firms. To mitigate concerns that our findings are salient to significant limits to arbitrage we require a minimum number of 300 firms in each month when deriving the signals.[5] Third, as industries tend to perform differently across the business cycle and may also be at different stages in their life cycle, it seems reasonable to assume that the information extracted from news flow data is likely to reflect the broad industry context, potentially confounded with cues about firm-specific performance. For this reason, we settle for a standardization based on industry classifications by subtracting from each score their industry averages and dividing by the industry-specific standard deviation. Fourth, since a firm's news volume and news sentiment are likely driven by company size, we cross-sectionally neutralize the indicators by their market capitalization. Appendix A gives further details on how we construct the individual

[4]We tested various filters around a value 90 but do not find significant differences in our results. Hence, we follow the studies from Hafez (2010); Kolasinski et al. (2013); Dang et al. (2015); von Beschwitz et al. (2017) that also use RavenPack news flow data. Notably, for some indicators we deviate from REL, EVR and ESD filters of 90 for indicator-specific reasons. For further information see the detailed indicator description in Appendix A.

[5]For these reasons, we refrain from analyzing signals with less than 300 firms.

news indicators.

The indicators that we derive from news flow data relate to various studies from the existing literature on news analytics and can be categorized into four broad concepts when building predictive signals.

2.3.1 News volume

News volume analyzes a firm's media presence measured by the number of news events within a specific time window. Existing studies suggest that a firm's media presence is related to its future stock price, however, the reported effects are ambiguous. The "attention grabbing effect" argues that investors are net buyers of stocks with high media presence (Chan, 2003; Barber and Odean, 2007; Da et al., 2011; Hillert et al., 2014). Associated returns of these attention-grabbing stocks are therefore (temporarily) higher than those of firms with low (or without) media presence. In contrast, the "neglect effect" advocates the slogan "no news is good news": Fang and Peress (2009) investigate the cross-sectional relation between media presence and expected stock returns and find that stocks with no media presence earn higher returns than stocks with high media presence even after controlling for well-known risk factors. We calculate a firm's average media presence within various time horizons (1, 3, 6 months) using different filter settings (REL>75 and REL>90, EVR>90, ESD>90).

2.3.2 News sentiment

News sentiment analyzes a news event's tone with respect to a particular firm. Positive sentiment corresponds to a news event that portrays positive surprises and opinions, resonating with generally good news or with an item that is better than expected. Numerous studies (e.g., Tetlock, 2007; Tetlock et al., 2008; Heston and Sinha, 2017; Wang et al., 2018) demonstrate that a firm's news sentiment contains information relevant to predicting its stock returns. For instance, Tetlock (2007) shows that high media pessimism, i.e. negative sentiment, forecasts falling stock market prices.[6] In this light, we construct various firm-specific sentiment indicators. We start with the simplest indicator by computing the monthly average of the event sentiment score over various look-back periods. Then, we construct a more robust version that compares the number of news events with positive event sentiment scores to the number of news events with negative event sentiment scores. This robust version is not dependent on the magnitude of the event sentiment score emerging from the proprietary model of the news data provider.[7] A further news sentiment indicator takes into account the temporal course within the time horizon (e.g. one month) by putting larger weight on

[6] For a detailed literature review on news sentiment see Uhl et al. (2015) or Coqueret (2018).

[7] Similar to our study, Wang et al. (2018) also base their study on news data from RavenPack analytics. To ensure the validity of the ESS provided by RavenPack they compute a simple sentiment score using common text processing techniques as a robustness check. Their findings show that both sentiment scores provide similar results.

more recent sentiment scores in the look-back window. Another empirical finding is that the market reaction to negative news is generally stronger than the reaction to positive news (Hafez et al., 2015). In this vein, we construct a firm-specific news sentiment indicator that gives different weights to positive and negative news. In particular, we employ a weighting scheme based on the prospect theory of Tversky and Kahneman (1992).

2.3.3 News trend

News trend relates to the dynamics in news sentiment rather than its average level. Analyzing associated time-series patterns, Leinweber and Sisk (2011) and Uhl et al. (2015) argue that longer-term news sentiment cycles exist and can be exploited for return predictions and investment strategies. The hypothesis is that a positive trend in a firm's news sentiment has a positive impact on its future returns. To extract noise and identify longer-term trends in the news-sentiment signal we follow Uhl et al. (2015) and use a frequency filter to construct a corresponding news sentiment momentum indicator. More simplistic approaches to determine time trends are (1) to compare the distribution of the ESS between two different points in time (similar to a simple t-statistic of a change in ESS) and (2) to regress the cumulative ESS on the time index.

2.3.4 Alternative news concepts

Alternative news concepts covers the indicators news beta, news dispersion and news significance. *News beta* measures the sensitivity of a firm's stock return to changes in market sentiment. To this end, we calculate an overall market news sentiment by averaging the ESS across firms for each month. The idea is that positive news beta stocks, on average, outperform the market while negative news beta stocks tend to underperform (Hafez, 2010). *News dispersion* looks at the intraday variation of the ESS, while *news significance* captures both mean and variation of the ESS within a specific time horizon.

3 News Analytics and the cross-section of stock returns

To examine the cross-sectional relevance of news analytics in a simple, non-parametric way we form long-short portfolios of stocks sorted by the derived news indicators (cf. Baker and Wurgler, 2006; Fang and Peress, 2009). Specifically, we divide the stock universe into monthly quintile portfolios based on the prevailing scores of the selected news indicator and compute the equally weighted average return of each portfolio during the following month.[8] If the information embedded in the

[8] We concentrate on an equal-weighting scheme when forming long-short portfolios as it is a simple and robust means of assessing the predictive power of the derived news indicators across the firm size spectrum, and is anecdotally

news indicator is already incorporated in stock prices, then the top quintile portfolio return should be similar to that of the bottom quintile portfolio. To test the pricing implications of news, we therefore form zero-investment trading strategies that are long stocks with the highest news scores and short stocks with the lowest news scores. Consequently, the ultimate long-short portfolio return emerges as the return difference between the top and bottom quintile portfolio returns.[9]

In this section, we first investigate the performance of news-based equity factors for the global stock universe. Second, we perform spanning tests to evaluate whether news factors contribute in explaining the cross-section of stock returns when also considering common equity factors such as value or momentum. Third, we conduct the cross-sectional analysis on a regional level, given that equity factors are found to exhibit region-specific effects. Fourth, we examine long-term effects of the news equity factors, because long-term, factor-based investment management is usually based on equity factors that generate positive returns for longer horizons than one month.

3.1 News-based equity factor evidence

Table II reports performance statistics of the monthly rebalanced long-short portfolio based on the set of news indicators applied to the global stock universe.[10] While the news data ranges from 2000 to 2017, the computation of indicators requires (at most) the last twelve months of data: hence, we start reporting monthly scores from 2001 to 2017.

[Table II about here]

It is interesting to note that most long-short portfolios based on news volume indicators deliver statistically insignificant returns over the sample period. The only exception is the news volume factor with low filter settings $VOL_{\mathrm{REL}>75,1}$, however, with a negative performance. Hence, our empirical findings neither support the "attention grabbing effect" of Barber and Odean (2007) nor the "neglect effect" of Fang and Peress (2009). The discrepancy to these studies may be explained by the fact that we do not restrict our study to a few US newspapers but analyze a significantly broader news data set including all types of news sources and cover a much longer sample period. Consequently, our approach inevitably leads to more robust results.

By contrast, we evidence that it is profitable to invest in long-short portfolios based on news sentiment indicators. Irrespective of the news sentiment indicator used, the ensuing return differential strategy is a statistically significant at the 1% level. Specifically, we find that a higher degree of

closer to the way that hedge funds use news text for portfolio construction (cf. Ke et al., 2019). Nevertheless, we also consider a market capitalization weighting scheme as robustness check.

[9] In the following, we also refer to the long-short portfolios as (equity) factors.

[10] We only report and discuss a representative set of news-based factors to save space.

sophistication in estimating news sentiment is rewarded. The ESS-based average sentiment factors earn higher monthly returns than the sentiment factors that only derive from the nature of a news event (positive/negative). For instance, $SENT_1$ has a 20 basis points (bps) pick-up in monthly return relative to $rSENT_{l=u=0,1}$. Still, performance can be further enhanced by weighting the individual news events. For example, the news sentiment factor that gives different weights to positive and negative news events ($wSENT_{pt,6}$) earns a monthly return of 2.78% at a 6-month time horizon (compared to 1.88% for $SENT$). The impact of the look-back window differs between weighted and non-weighted news sentiment factors. While the monthly returns for weighted factors increase with increasing look-back window (e.g. from 2.08% to 2.78% for $wSENT_{pt}$) the performance of non-weighted factors is fairly flat (e.g. from 1.98% to 1.88% for $SENT_6$). Notably, return benefits do not result from higher risk. In terms of Sharpe ratio, risk-adjusted returns range from 0.63 to 1.17, with the highest figure obtaining for the one-month news sentiment factors. Overall, using a much broader news data set, our findings are consistent with existing studies (e.g. Tetlock, 2007) that document that stocks with higher news sentiment earn higher returns than stocks with lower news sentiment.

Concerning news trend factors, we document less pronounced but still statistically significant results. In particular, the sentiment momentum factor ($SENTMOM$) has statistically significant return differential (t-statistic of 3.68) and a Sharpe ratio of 0.89. Moreover, the $aSENTMOM_6$ factor also earns a statistically significant return even though at a lower level (t-statistic of 1.95). Analyzing the conditional cross-sectional effects of the alternative news concept indicators provides different insights. While neither the news beta nor the news dispersion factor show statistically significant results, the news significance factor is more promising. We find statistically significant long-short returns in excess of 1%, which are more pronounced at longer horizons (3 and 6 months).

As a robustness check, we contrast the performance based on equal weights with that of market capitalization weights, allowing to gauge the relevance of our findings for actual portfolio implementation. Table III reports the results of the cap-weighted long-short portfolios, showing similar patterns to their equally weighted counterparts. Still, portfolios related to news sentiment have an overall good performance, yet significance is reduced. Hence, news flow data has stronger predictive power for future returns to small stocks, all else being equal. According to Ke et al. (2019), there are a number of potential economic explanations for this fact. First, small stocks receive less investor attention and thus respond more slowly to news. Second, the underlying fundamentals of small stocks are more uncertain and opaque and thus it requires more effort to process news into actionable price assessments. Third, small stocks are less liquid and therefore require a longer time for trading and thus for incorporating information into prices.

[Table III about here]

3.2 Mean-variance spanning

Factor-based investment managers usually do not restrict to invest in one equity factor only, but build on a complete set of factors to enjoy the benefits of factor diversification. Hence, it is crucial to evaluate whether the proposed news factors expand the investor's investment opportunity set. Figure 3 shows the return correlation matrix of the news factors including the standard set of equity factors, namely the Fama and French (1992, 2006) factors as well as the momentum and short-term reversal factors of Jegadeesh and Titman (1993).[11] By construction, most news factors are highly correlated within their concept category. We further find the momentum factor to be highly correlated with some of the news sentiment factors. This observation is reasonable as both factors pick up information from the current economic environment.

To statistically examine whether news-based equity factors are subsumed by traditional equity factors or do expand an investor's opportunity set, we employ the mean-variance spanning test of Kan and Zhou (2012). This examines whether adding assets to a set of benchmark assets improves the tangency or the global minimum-variance portfolio. It is based on a simple regression, which regresses the returns of the news factors, $r_{N,t}$, on the returns of a set of benchmark factors, $r_{b,t}$:

$$r_{N,t} = \alpha + \sum_{b=1}^{B} \beta_b r_{b,t} + \varepsilon_t. \tag{1}$$

If the news factors are fully explained by the set of benchmark factors, the estimated alpha, $\hat{\alpha}$, should be insignificant. To assess statistical significance, Kan and Zhou (2012) propose two sequential hypothesis tests. The first test examines the enhancement of the tangency portfolio: using the null $H_0^1 : \alpha = 0$. The second test investigates the additional benefit for the global minimum-variance portfolio: using the null $H_0^2 : \delta = 1 - \sum_{b=1}^{B} \beta_b = 0$. To this end, it imposes the restriction of $\alpha = 0$. Splitting up the hypotheses in this fashion allows to draw conclusions about the nature of the potential benefit of the news factors.

[Table IV about here]

Table IV reports the results of the spanning tests against a standard set of benchmark equity factors. We report regression statistics of Equation (1) as well as the test statistics of the step-down tests. We find most alphas for news sentiment, news trend and news significance factors to be significant at the 1% level suggesting that this set of news factors may contribute to explain the cross-section of stock returns. However, evaluating the R^2_{adj} we learn that the degree of added value decreases with the length of the respective factor's underlying time horizon: for instance, over 60% of the returns of the $SENT_6$ news factor can be explained by common equity factors (compared to

[11] See Table XI for a definition of the set of equity factors.

only 26% for $SENT_1$). In line with the correlation analysis, we find significant coefficients of the momentum factor for all news indicators, suggesting that price momentum effects are a crucial driver of the most promising news factors.

These findings are mainly confirmed by the step-down tests. The F1 test rejects the null hypothesis of $\alpha = 0$ for most news factors at the 1% significance level. Only SIG_1 and $aSENTMOM_6$ are not or at lower levels statistically significant. Likewise, the F2 test also rejects the null hypothesis of $\delta = 0$ for most news factors at the 1% significance level.

3.3 Robustness to different holding periods

Next, we investigate the persistence of the news indicators' predictive power, speaking to the ease with which these factors could be implemented in a portfolio. If the predictive power of the news indicators was to remain significant over several months, one could contemplate reducing the frequency and/or magnitudes of portfolio rebalances and, in turn, incur lower implementation costs. To this end, we test the performance of a strategy that represents an equally weighted average of the previous h monthly portfolios. The look-back period h is varied from one to twelve, meaning that a portfolio created twelve months ago could be used to harvest the next month's strategy returns. Figure 4 charts the associated cumulative returns for the news-based indicators. Table V reports to the corresponding statistics.

[Figure 4 about here]

The main findings are twofold: (1) Most factors with significant one-month long-short portfolio return exhibit a fast decay in the following months. The weighted sentiment factors with one month time horizon, however, rebound after 6 months. (2) Factors incorporating news sentiment at longer time horizon (e.g. $SENT_6$ and SIG_6) exhibit a rather stable and significant return pattern, indicating that these factors may be useful for long-term investment management.

[Table V about here]

3.4 Regional differences

Jacobs and Müller (2020) document regional differences when studying the pre- and post-publication return predictability of 241 cross-sectional anomalies in various international stock markets. They observe a surprisingly large discrepancy in the post-publication decline in long-short portfolio returns between the U.S. and international markets. In this vein, we divide the global stock universe into five regions—USA, Japan, Europe, rest of the world (RES) and emerging markets—and look for regional differences in the efficacy of the investigated news factors. Table VI reports the performance statistics of the long-short portfolio returns for the five regions.[12] News volume factors do not seem to be relevant in any of the five regions, similar to the global universe. The performance of news sentiment factors is mixed. For the USA only the one-month sentiment factors are slightly significant with Sharpe ratios of around 0.5, whereas we do not evidence any predictive power in Japan. For Europe and the rest of the world universe, the results are substantially better with significant returns and Sharpe ratios of around 1.2 on average. Similar to the global universe the best performing news sentiment factor are the time-weighted average sentiment factor with a Sharpe ratio of 1.37. Regarding the news trend concept the news sentiment momentum factor shows promising performance in all regions except Japan. In the USA, it exhibits a Sharpe ratio of 0.5, and it is even higher in Europe and RES (0.7 and 1, respectively). In the latter two regions the alternative news sentiment momentum factor is also significant in terms of long-short return. Alternative news concept factors do not deliver significant results for the USA and Japan, but for Europe and RES. Also, news significance factors are by and large significant in Europe and RES with best results at longer horizons. In summary, we evidence fairly weak results for the USA and Japan and strong results for Europe and the rest of the world universe. The findings for the USA may be rationalized by the fact that it is generally difficult to explain the cross-section of stock returns in the USA: U.S. markets seem to be simply more efficient than the other markets due to an extremely high analyst coverage, so that news are readily incorporated in stock prices (see McLean and Pontiff, 2016; Jacobs and Müller, 2020). The fact that average momentum returns have historically been low in the Japanese market (see Daniel et al., 2001; Hanauer, 2014) in conjunction with the finding that the momentum factor is highly correlated with news-based factors may explain the findings for Japan.

[Table VI about here]

We complement the results for the global developed countries by examining the relevance of the news-based factors for emerging markets. Overall, we find even stronger results than for the developed markets. The news volume factor with restrictive filters is (weakly) significant in return, increasing with the time horizon. News sentiment indicators show overall strong performance with high Sharpe ratios (up to a magnitude of 2.09). Moreover, we document significant results for news sentiment momentum factors. Regarding the alternative news concepts, we find news dispersion and news significance factors to perform with the latter exhibiting high Sharpe ratios.

4 News analytics and multi-factor investment strategies

As evidenced in Chapter 3.2, news-based equity factors may expand an investor's equity factor opportunity set. Building on these insights, we investigate in this section whether news analytics

[12] We exclude news factors with low coverage. In particular, we require an average of at least 100 firms per month.

may be beneficial for constructing multi-factor investment strategies.

To this end, we follow Dichtl et al. (2019) and construct a set of equity factors that includes not only the common factors used in Chapter 3.2 but also further equity factors widely used and well documented in academic research. The factors can be roughly assigned to the following four categories:[13]

- $Value$: cash flow yield (CFY), dividend yield (DY), book-to-market ratio (BTM), earnings yield (EY), and profitability ($PROF$)
- $Momentum$: 12-month price momentum ($MOM12$), short-term reversal (STR), and long-term reversal (LTR)
- $Quality$: asset turnover (AT), change in long-term debt ($DLTD$), change in shares outstanding (DSO), asset growth (AG), cash productivity (CP), profit margin (PMA), leverage (LEV), return on assets (ROA), sales-to-cash (STC), sales-to-inventory (STI), and accruals (ACC)
- $Size$: Size ($SIZE$)

Building on this benchmark set of factors we first examine whether risk-based multi-factor portfolios can be enhanced by adding news-based factors. In a second step, we investigate the benefits of utilizing news flow data for active factor allocation strategies. In particular, we use the parametric portfolio policies of Brandt and Santa-Clara (2006) and Brandt et al. (2009) to arrive at meaningful factor timing and tilting allocations along the lines of Dichtl et al. (2019).

4.1 Risk-based factor allocation

Taking an agnostic perspective regarding expected factor returns, risk-based factor allocations strategies are a common technique to construct well-diversified multi-factor portfolios. We examine how an equally weighted portfolio (1/N), a minimum-variance portfolio (MVP) and a risk parity portfolio (RP) adapts to the inclusion of news-based equity factors.[14]

Table VII provides the performance and risk statistics of the three strategies for the set of benchmark factors (Panel A) and the set of benchmark factors augmented by news-based equity factors (Panel B). We compute the first optimal portfolio weights over a 36-month window, which expands over time, so we obtain the first portfolio for January 2007 and the last for September 2017. We enforce full investment and non-negative factor weights. Overall, we document that all three

[13] See Dichtl et al. (2019) for a concise definition of each factor.

[14] The 1/N strategy rebalances monthly to an equally weighted allocation scheme. The minimum-variance portfolio is the mean-variance efficient portfolio that is expected to have the lowest possible portfolio variance. The risk parity strategy allocates capital so that the factors' risk budgets contribute equally to overall portfolio risk.

risk-based allocation strategies benefit from adding the $SENT_1$, $SENT_6$ and $wSENT_{pt,6}$ factors to the benchmark portfolio.[15]

[Table VII about here]

The 1/N strategy earns a 21 bps higher excess return at a decrease of 27 bps in volatility when including news-based factors. This results in a Sharpe ratio gain of 27 bps, but comes at the cost of a higher maximum drawdown (3.94% vs. 3.51%). For the MVP and the RP portfolio we observe generally lower excess returns and higher Sharpe ratios than for the 1/N portfolio (2.53% and 2.94% vs. 3.47% for the excess return; 2.13 and 2.25 vs. 1.60 for SR), but document similar gains in these measures when including news-based factors. The strengths of the MVP and RP portfolios reveal in downside risk hedging, translating to a significant reduction in maximum drawdown compared to the 1/N portfolio (0.85% and 1.14% vs. 3.94%). News-based factors help to further decrease this drawdown statistic (-9 bps for MVP and -13 bps for RP).

Due to the robustness and simplicity of the 1/N strategy (see DeMiguel et al., 2009), we benchmark the subsequent factor timing and tilting strategies using the 1/N strategy.

4.2 Factor timing

Utilizing time-series information embedded in a variety of fundamental variables and technical predictors, we want to improve performance over the equal-weighted benchmark. Employing a diversified factor set, one clearly see, that different factors tend to work better in different economic environments. Therefore, the identification of the state of the economic environment and using the predictive power embedded, should help to improve the risk-return profile for our factor allocation strategy compared to the equal-weighted benchmark.

Based on the parametric portfolio policy framework of Brandt and Santa-Clara (2006), we assess the utility of information extracted from news flow data for factor timing strategies by comparing factor timing portfolios with and without news-based factors using the predictor set as in Dichtl et al. (2019).

4.2.1 Methodology of Brandt and Santa-Clara (2006)

The parametric portfolio policy of Brandt and Santa-Clara (2006) directly translates any predictive power embedded in the predictor variables into optimal portfolio weights. Starting point is a dynamic maximization problem of a mean-variance investor with quadratic utility function and risk aversion

[15] The reported results are robust to the choice of news-based factors to be added to the set of benchmark factors, given that they are among the factors tested in the spanning tests in Table IV.

parameter γ, seeking to derive the optimal factor portfolio weights w_t:

$$\max_{w_t} E_t \left[w_t' r_{t+1} - \frac{\gamma}{2} w_t' r_{t+1} r_{t+1}' w_t \right] \tag{2}$$

where r_{t+1} is the vector of future excess returns of the N equity factors. The key idea of the authors is to parametrize the portfolio weights by assuming that the optimal portfolio strategy, w_t, is linear in the vector z_t of the K conditioning variables (of which the first element is simply a constant):

$$w_t = \theta z_t \tag{3}$$

where θ is an $N \times K$ matrix of parameters. Plugging the linear portfolio policy Equation (3) into Equation (2), Brandt and Santa-Clara (2006) showed that the original optimization problem is equivalent to running the following optimization:

$$\max_{\tilde{w}} E \left[\tilde{w}' \tilde{r}_{t+1} - \frac{\gamma}{2} \tilde{w}' \tilde{r}_{t+1} \tilde{r}_{t+1}' \tilde{w} \right] \tag{4}$$

where $\tilde{w} := vec(\theta)$ and $\tilde{r}_{t+1} := z_t \otimes r_{t+1}$.[16] Therefore, the original dynamic optimization problem can be restated as a static (unconditional) Markowitz optimization applied to an augmented set of equity factors that includes not only the original equity factors but also synthetic or "managed" ones. Each of these managed equity factors invests in a single equity factor according to the realization of one of the conditioning variables (see Brandt and Santa-Clara, 2006; Dichtl et al., 2019).

4.2.2 Predictor variables

The set of predictor or conditioning variables used in the parametric portfolio policy of Brandt and Santa-Clara (2006) includes both the fundamental variables of Welch and Goyal (2008) containing information about future states of the economy and factor-specific technical indicators and trading rules derived from past factor returns according to Neely et al. (2014). In particular, we employ the following variables:[17]

- *Fundamental variables*: dividend-to-price ratio (dp), dividend yield (dy), earnings-to-price ratio (ep), dividend payout ratio (de), stock variance ($svar$), book-to-market ratio (bm), net equity expansion ($ntis$), US T-bills (tbl), long-term yield (lty), long-term rate of return (ltr), term spread (tms), default yield spread (dfy), default return spread (dfr), and inflation ($infl$)

[16] Note that $vec(\cdot)$ is a linear transformation that converts the matrix into a column vector and $\otimes$ denotes the Kronecker product of two matrices.

[17] See Dichtl et al. (2019) for a detailed description of the predictor variables. Note that the fundamental variables are based on US fundamental data. As US data are, however, predictive for other developed countries' stock market returns (Rapach et al., 2013), applying these fundamental predictors in a global setting seems appropriate.

- *Technical variables*: Momentum (MOM_i for $i = 1, 3, 6, 9, 12$) and Moving average (MA_{s-l} for $s = 1, 2, 3$ and $l = 9, 12$)

To preserve their embedded information, we separately apply principal components analysis (PCA) to the fundamental variables and the technical indicators in the spirit of Neely et al. (2014), Ludvigson and Ng (2007, 2009), and Hammerschmid and Lohre (2018). This has the additional advantage of avoiding multicollinearity problems that arise due to high correlations within both groups of predictor variables. As a smaller number of predictors allows for a longer out-of-sample backtesting window, our main analysis is based on the first principal components (denoted as $FUN1$ and $TECH1$).

4.2.3 Empirical results

As described in the section above, the portfolio optimization estimates θ-coefficients to derive the optimal weight in each factor and therefore presents itself as an estimation framework. This allows us to compute the standard errors for the θ-coefficients and evaluate their significance. Following Brandt and Santa-Clara (2006) we calculate standard errors from the covariance matrix of $\tilde{w}$ as follows:

$$\frac{1}{\gamma^2}\frac{1}{T \ - \ N \times K}(\iota_T - \tilde{r}\tilde{w})'(\iota_T - \tilde{r}\tilde{w})(\tilde{r}'\tilde{r})^{-1} \tag{5}$$

where ι_T denotes a $T \times 1$ vector of 1s.

Table VIII shows the θ-coefficients as well as the standard errors. For the benchmark case we estimate 40 coefficients (20 factors $\times$ 2 conditioning variables), while we have 46 for the case including news factors (23 factors $\times$ 2 conditioning variables). Of the 40 θ-coefficients defining the optimal timing strategy in the benchmark factors case, only 10 are statistically significant at the 5% level. This number increases slightly to 12 when including the news factors. None of the θ-coefficients for the news factors show statistical significance.

[Table VIII about here]

To asses if a timing strategy is economically meaningful, we do a beauty contest of both factor sets and evaluate their performance profile over the sample period compared to the equal-weighted benchmark. Table IX shows the results when using fundamental and technical predictors to time factor weights compared to an equal-weighted benchmark. We compute the first optimal portfolio weights over a 72-month window, which expands over time, so we obtain the first portfolio for January 2008. For the risk-aversion parameter, γ, used in the quadratic utility function, we choose a value of 5, implying moderate risk aversion.

[Table IX about here]

Looking at the factor set including traditional academic factors only, the highest overweight during our sample period experiencing EY with an active weight of 10.38%, CFY (5.94%) and PROF (5.57%). On the other side, the least attractive factors over the sample periods are LEV with an active underweight of -6.79%, PMA (-5.46%) and DLTD (-5.12%). Based on the fundamental and technical predictors, the resulting factor allocation strategy experiences a gross return of 3.75%, outperforming the benchmark by 49 bps. As the re-weighting of the factors in the portfolio comes with a higher standard deviation of 3.36% (compared to the 2.45% volatility of the benchmark), risk-adjusted returns look slightly worse. The Sharpe ratio of the portfolio is 1.12, while those of the equal-weighted benchmark is at 1.33. As the timing strategy has a limited tracking error versus the benchmark, the outperformance results in a gross information ratio of 0.29. To maintain the optimal factor timing allocation a lot of turnover is needed. Hence, the resulting net return[18] of 1.12 underperforms the benchmark by 112 bps. This high two-way turnover of 836% p.a. leads to a negative information of - 0.66 after including transaction costs.

Including the news factors to the benchmark, the active weights look comparable. The biggest overweights are again in EY (11.02%), PROF (6.04%) and $SENT_1$ (5.79%). The biggest underweights compared to the equal-weighted benchmark are STC (-6.87%), ACC (-6.80%) and $wSENT_{pt,6}$ (-6.21%). Also, the performance characteristics are comparable between the two factor sets. The gross performance of 3.91% outperforming the benchmark by 44 bps, while coming at a higher risk. This leads the Sharpe ratio of the portfolio being lower than those of the benchmark. Comparing the two factor sets, the information ratio of 0.35 is a slight increase when adding news factors to the set of invested factors.

4.3 Factor tilting

A complementary way of equity factor investing exploits cross-sectional differences in factor characteristics by tilting the factor allocation according to those characteristics. Using the cross-sectional parametric policy framework developed by Brandt et al. (2009), we exploit cross-sectional factor characteristics based on the derived news indicators in addition to benchmark characteristics from Dichtl et al. (2019) to assess the relevance of the news analytics indicators. As before, we also compare factor tilting portfolios including news-based factors to portfolios without these factors.

[18]We account for three costs appearing in the management of such a dynamic factor allocation strategy: First, we rebalance the underlying factor portfolios to mimic the factor on a monthly basis. We subtract 75 bps for 100% turnover on the long side and additionally 40 bps on the short to account for the additional costs of shorting an asset. This is already reflected in the factor performance and all factor time-series are net of costs. Second, we assume that the factor portfolios are available as swaps, so we assume 96 bps per year for holding the swap. Third, we account for the rebalancing of the swap notional and assume 20 bps for turning over 100% of the notional.

4.3.1 Cross-sectional factor characteristics

To calculate news-based equity factor characteristics we follow Lee (2017) and use the idea of "factors within factors". That means, we first build quintile portfolios based on the chosen equity factor, such as value or momentum. We then compute the equally weighted average score of a news indicator across all stocks in each quintile portfolio. A factor's news characteristic is finally computed as the spread between the news score of the top and the news score of the bottom quintile portfolio.

In addition to a representative set of news-based characteristics we include the following factor characteristics that are well documented in the literature and used by Dichtl et al. (2019): factor valuation, factor spread, factor momentum, and factor crowding.[19]

4.3.2 Methodology of Brandt, Santa-Clara and Valkanov (2009)

We incorporate the standardized cross-sectional characteristics into the parametric portfolio policy of Brandt et al. (2009), which allows us to exploit the information content in a utility-based portfolio optimization. Specifically, we consider an investor seeking to maximize conditional expected utility over portfolio return $r_{p,t+1}$:

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} E_t\left[u(r_{p,t+1})\right] = E_t\left[u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right] \tag{6}$$

where $w_{i,t}$ denotes the portfolio weight for asset i and N_t denotes the number of assets at time t. Brandt et al. (2009) propose modeling the portfolio weight as a linear function of asset characteristics $x_{i,t}$:

$$w_{i,t} = f(x_{i,t};\phi) = w_{b,i,t} + \frac{1}{N_t}\phi'\hat{x}_{i,t}, \tag{7}$$

where $w_{b,i,t}$ denotes the benchmark weight, ϕ is the vector of coefficients to be estimated through utility maximization, and $\hat{x}_{i,t}$ denotes the standardized factor characteristics.

For a mean-variance utility function, the original problem can be restated[20] as

$$\max_{\phi} \phi'\hat{\mu}_c - \left(\frac{\gamma}{2}\phi'\hat{\Sigma}_c\phi + \gamma\phi'\hat{\sigma}_{bc}\right) \tag{8}$$

where $\hat{\Sigma}_c$ is the sample covariance matrix, $\hat{\mu}_c$ is the mean of the characteristic return vector, and $\hat{\sigma}_{bc}$ is the sample vector of covariances between the benchmark portfolio return and the characteristic-return vector. As all characteristics are standardized cross-sectionally at time t across all factors, deviations

[19] See Dichtl et al. (2019) for detailed description of these factor characteristics.

[20] For a detailed description see also DeMiguel et al. (2017).

from the benchmark are equivalent to a zero-investment portfolio (DeMiguel et al., 2017; Dichtl et al., 2019).

4.3.3 Empirical results

Table X shows estimation results and performance statistics for six news-related factor tilting allocations based on univariate parametric portfolio policies. Across the models for the benchmark equity factor portfolio (cf. Panel B), the only significant coefficients obtain for the tilting characteristics $SENT_1$ and $SENT_3$, suggesting a short-term sentiment effect among equity factors. Hence, factors with positive sentiment are overweighted relative to the equal-weighted benchmark while factors with negative sentiment are underweighted. The annualized returns of the corresponding parametric portfolio policy using $SENT_1$ and $SENT_3$ are 0.83 and 0.94 percentage points higher than the one for the equal-weighted benchmark, whereas the volatility is increased by 0.42 and 0.14 percentage points. These figures correspond to an information ratio of 0.50 and 0.56.

While statistically weak the news sentiment-related characteristics with longer horizon yet have positive information ratios as well: $SENT_6$, $wSENT_{td,6}$, $wSENT_{pt,6}$, and SIG_6 with information ratios of 0.56. 0.40, 0.40 and 0.58, respectively. Moreover, capturing news sentiment over longer horizon seem to be more profitable: The $SENT_6$ tilting portfolio has a higher Sharpe ratio than the $SENT_1$ tilting portfolio and than the equal-weighted benchmark (1.84 vs. 1.43 vs. 1.33). After accounting for transaction costs the $SENT_6$ strategy's return and Sharpe ratio are reduced to 3.08% and 1.32 compared to 2.24% and 0.92 for the equal-weighted benchmark. This reduction in (risk-adjusted) return is equivalent to an information ratio of 0.44 net transaction costs. Notably, news sentiment-related tilting allocations show similar performance statistics to allocations using common tilting characteristics such as factor crowding and factor spread and seem to be more profitable than those for factor momentum and factor valuation allocations.

[Table X about here]

While some news-related factor characteristics show predictability in this portfolio utility context for the benchmark equity factor portfolio, this turns when adding news-based equity factors to equity factor portfolio (cf. Panel C): none of the news sentiment-related factor characteristics exhibit significant coefficients if information from news flow data is directly incorporated in the equity factor portfolio. Yet, all news-related tilting allocations show positive information ratios, even after accounting for transaction costs. The economic relevance of news flow data is corroborated by overall higher (risk-adjusted) returns compared to the benchmark equity factor portfolio.

In a nutshell, our empirical evidence suggests that news sentiment information is valuable for constructing multi-factor allocation strategies. Thus, our findings are in line with Uhl et al. (2015) and Tetlock (2007) who document that news sentiment is useful for predicting future return movements.

5 Conclusion

This paper contributes to the literature on news analytics by investigating both its effects on the cross-section of stock returns and its ability to enhance multi-factor investment strategies. Studying the cross-sectional characteristics of a broad set of indicators generated from news flow data suggests that the insights gathered from firm-specific news sentiment analysis can find their way into implementable trading strategies in a manner that adds over and above common drivers of equity returns. Long-short portfolios based on news sentiment indicators seem to be particularly profitable in global and European stock universes, while results for the US and Japanese equity markets are rather moderate.

Assessing the information embedded in news flow data in risk-based and forecasting-based factor allocation strategies reveals interesting insights. An equally weighted portfolio as well as minimum-variance and risk parity strategies benefit from adding news sentiment-related equity factors to a portfolio of representative global equity factors. Building on these insights, we explore the benefits of active factor allocation when incorporating information from news flow data. Factor timing using fundamental and technical time-series predictors generates statistically significant and economically relevant results. Similarly, a factor tilting strategy that exploits cross-sectional news-related information outperforms an equally weighted benchmark portfolio. As both strategies require substantial turnover to follow the embedded information in the timing predictors or characteristics used, we experience a performance drag which is more pronounced for the factor timing than the factor tilting strategies.

References

Arnott, R., N. Beck, V. Kalesnik, and J. West (2016). How can 'smart beta' go horribly wrong? *Research Affiliates, February*.

Asness, C. S. (2016). The siren song of factor timing aka "smart beta timing" aka "style timing". *Journal of Portfolio Management 42*(5), 1–6.

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 61*(4), 1645–1680.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*(1), 3–18.

Barber, B. M. and T. Odean (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*(2), 785–818.

Bender, J., X. Sun, R. Thomas, and V. Zdorovtsov (2018). The promises and pitfalls of factor timing. *Journal of Portfolio Management 44*(4), 79–92.

Brandt, M. W. and P. Santa-Clara (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance 61*(5), 2187–2217.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies 22*(9), 3411–3447.

Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics 70*(2), 223–260.

Cohen, R. B., P. A. Gompers, and T. Vuolteenaho (2002). Who underreacts to cash-flow news? Evidence from trading between individuals and institutions. *Journal of Financial Economics 66*(2), 409–462.

Coqueret, G. (2018). The economic value of firm-specific news sentiment. *SSRN Working Paper*.

Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. *Journal of Finance 66*(5), 1461–1499.

Da, Z. and M. C. Warachka (2009). Cashflow risk, systematic earnings revisions, and the cross-section of stock returns. *Journal of Financial Economics 94*(3), 448–468.

Dang, T. L., F. Moshirian, and B. Zhang (2015). Commonality in news around the world. *Journal of Financial Economics 116*(1), 82–110.

Daniel, K., S. Titman, and K. J. Wei (2001). Explaining the cross-section of stock returns in japan: factors or characteristics? *The Journal of Finance 56*(2), 743–766.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies 22*(5), 1915–1953.

DeMiguel, V., A. Martin-Utrera, F. J. Nogales, and R. Uppal (2017). A portfolio perspective on the multitude of firm characteristics. *Available at SSRN: 2912819*.

Dichtl, H., W. Drobetz, H. Lohre, C. Rother, and P. Vosskamp (2019). Optimal timing and tilting of equity factors. *Financial Analysts Journal, forthcoming*.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (2006). Profitability, investment and average returns. *Journal of Financial Economics 82*(3), 491–518.

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

Fang, L. and J. Peress (2009). Media coverage and the cross-section of stock returns. *Journal of Finance 64*(5), 2023–2052.

Hafez, P. A. (2010). News beta - a new measure for risk & stock analysis. *RavenPack Research Paper*.

Hafez, P. A., J. A. Guerrero-Colon, and S. Duprey (2015). Thematic alpha streams improve equity portfolio performance. *RavenPack Research Paper*.

Hammerschmid, R. and H. Lohre (2018). Regime shifts and stock return predictability. *International Review of Economics & Finance, forthcoming*.

Hanauer, M. (2014). Is japan different? evidence on momentum and market dynamics. *International Review of Finance 14*(1), 141–160.

Haugen, R. A. and N. L. Baker (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics 41*(3), 401–439.

Heston, S. L. and N. R. Sinha (2017). News versus sentiment: Predicting stock returns from news stories. *Financial Analysts Journal 73*(3), 67–83.

Hillert, A., H. Jacobs, and S. Müller (2014). Media makes momentum. *Review of Financial Studies 27*(12), 3467–3501.

Hou, K., G. A. Karolyi, and B.-C. Kho (2011). What factors drive global stock returns? *Review of Financial Studies 24*(8), 2527–2574.

Jacobs, H. and S. Müller (2020). Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics 135*(1), 213–230.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Kan, R. and G. Zhou (2012). Tests of mean-variance spanning. *Annals of Economics and Finance 13*(1), 139–187.

Ke, Z. T., B. T. Kelly, and D. Xiu (2019). Predicting returns with text data. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-69).

Kolasinski, A. C., A. V. Reed, and M. C. Ringgenberg (2013). A multiple lender approach to understanding supply and search in the equity lending market. *Journal of Finance 68*(2), 559–595.

Lee, W. (2017). Factors timing factors. *Journal of Portfolio Management 43*(5), 66–71.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *Quarterly Journal of Economics 105*(1), 1–28.

Leinweber, D. and J. Sisk (2011). Event driven trading and the 'new news'. *Journal of Portfolio Management 38*(1).

Ludvigson, S. and S. Ng (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

Ludvigson, S. and S. Ng (2009). Macro factors in bond risk premia. *Review of Financial Studies 22*(12), 5027–5067.

Malkiel, B. G. and E. F. Fama (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance 25*(2), 383–417.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *Journal of Finance 71*(1), 5–32.

Neely, C. J., D. E. Rapach, J. Tu, and G. Zhou (2014). Forecasting the equity risk premium: The role of technical indicators. *Management Science 60*(7), 1772–1791.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics 108*(1), 1–28.

Rapach, D. E., J. K. Strauss, and G. Zhou (2013). International stock return predictability: What is the role of the united states? *Journal of Finance 68*(4), 1633–1662.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review 71*(3), 289–315.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*(3), 1139–1168.

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

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Uhl, M. W., M. Pedersen, and O. Malitius (2015). What's in the news? using news sentiment momentum for tactical asset allocation. *Journal of Portfolio Management 41*(2), 100.

von Beschwitz, B., D. B. Keim, and M. Massa (2017). First to "read" the news: News analytics and algorithmic trading. *INSEAD Working Paper*.

Wang, Y., B. Zhang, and X. Zhu (2018). The momentum of news. *SSRN Working Paper*.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.

Appendix A The set of news indicators

This section describes in detail how we construct indicators exploiting the news flow data from RavenPack News Analytics. All indicators are filtered using the relevance score (REL), the event relevance score (EVR) and the event similarity days score (ESD). Unless otherwise indicated, we require all scores to be above 90.

Let E_i be the i-th news event for a specific firm in a given time horizon, as classified by the RavenPack taxonomy. The publication time of a news event is denoted as $\tau(\cdot)$. Then, the news volume indicator at time t, VOL_t, is computed as the number of news events within time horizon h, i.e.

$$VOL_{t,h} = \sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}}, \tag{9}$$

where $I \subset \mathbb{N}$ is the number of all news events for a specific firm. In the empirical study, we calculate VOL using two filter settings: A less restrictive setting (REL > 75) to cover a firm's overall media presence and the standard setting (REL > 90, EVR > 90, ESD > 90) to restrict to the major events and thus only analyze a firm's meaningful media presence.

Let further $ESS(\cdot)$ be the event sentiment score of a news event. Then, the average firm-specific news sentiment indicator $SENT$ is given by

$$SENT_{t,h} = \frac{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}} ESS(E_i)}{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}}}. \tag{10}$$

The robust version of the news sentiment indicator, $rSENT$, is calculated as follows

$$rSENT_{t,h} = \frac{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t] \,|\, ESS(E_i) > u\}} - \sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t] \,|\, ESS(E_i) < l\}}}{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t] \,|\, ESS(E_i) > u,\, ESS(E_i) < l\}}}, \tag{11}$$

where l and u are lower and upper thresholds defining the range for the ESS. In the empirical analysis, we use two threshold settings: first, we differentiate between positive and negative news by setting $u = l = 0$. Second, we further exclude sentiment scores that are close to zero, i.e. $u = 0.1$ and $l = -0.1$.

To construct the weighted sentiment indicator $wSENT$, we denote the weight given to news event E_i by w_i. Consequently, this indicator is calculated as

$$wSENT_{t,h} = \frac{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}} w_i ESS(E_i)}{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}}}. \tag{12}$$

In the empirical study, we use two different weighting scheme: temporal decay and prospect theory. The former puts larger weight on sentiment scores closer to the end of time horizon h. This implies that the indicator is more reactive to recent news events and the corresponding news sentiment. The latter gives different weights to positive and negative news following evidence from prospect theory.

The news sentiment momentum indicator *SENTMOM* is constructed similar to the methodology of Uhl et al. (2015). Based on the SENT indicator, we first calculate crossing moving average time series of different time horizons (i.e. for $h = 1$ and $h = 12$ we get $SENT_{t,1} - SENT_{t,12}$) using a rolling window approach. Subsequently, we apply the cumulative sum (CUSUM) filter to this time series. See Uhl et al. (2015) for details on the CUSUM filter. Finally, the indicator series is normalized between -1 and 1.

Another way to calculate a trend indicator for news sentiment is to standardize a crossing moving average time series (e.g. for $h = 1$ and $h = 3$, see previous paragraph) by its sample standard error instead of applying the CUSUM filter. Specifically, the *aSENTMOM* indicator is computed as follows

$$aSENTMOM_{t,h} = \frac{SENT_{t,1} - SENT_{t-h}}{\sqrt{\sigma^2_{t,1}/VOL_{t,1} - \sigma^2_{t,h}/VOL_{t,h}}}, \tag{13}$$

where the sample variance $\sigma^2_{t,h}$ is given by

$$\sigma^2_{t,h} = \frac{\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}} \left(ESS(E_i) - SENT_{t,h}\right)^2}{\left(\sum_{i \in I} \mathbb{1}_{\{\tau(E_i) \in [t-h,t]\}}\right) - 1}. \tag{14}$$

The third news trend indicator, *REG*, is simply based on the t-statistic from regressing the cumulative sum of the ESS on the time index within time horizon h.

Among the alternative news concept indicators, *NEWSBETA* measures the responsiveness of a firm's stock return to an aggregate market news sentiment within a specific horizon. Specifically, the indicator value is calculated as the t-statistic from regressing a firm's stock return on a market capitalization-weighted average of the ESS across all firms in the universe.

The news significance indicator *SIG* measures the significance of the ESS (similar to a t-statistic) and thus captures mean and variation in the ESS. Specifically, it is given by

$$SIG_{t,h} = \frac{SENT_{t,h}}{\sqrt{\sigma^2_{t,h}/VOL_{t,h}}}. \tag{15}$$

The news dispersion indicator measures the variation in the ESS and is computed as

$$DISP_{t,h} = \frac{\sqrt{\sigma^2_{t,h}}}{SENT_{t,h}}. \tag{16}$$

All indicators except *SENTMOM* and the regression-based indicators are computed for $h = 1, 3, 6$, where h is measured in months. While *SENTMOM* uses multiple time horizons by definition, *REG* is calculated for $h = 6$, 12 and *NEWSBETA* for $h = 12$, 36, 60 due to sample size requirements for time-series regressions. In a final step, we standardize all indicators by company size and industry classification.

Table I
Descriptive statistics of news data

This table shows the descriptive statistics of news volume (Panel A) and the average event sentiment score (Panel B) per month and firm. For news volume, i.e. the number of news events per month, we require a relevance score above 75. For the ESS we require an (according to the RavenPack taxonomy) assigned and non-neutral ESS score as well as a relevance, event relevance and event similarity score above 90. For each panel, we show the overall statistics as well as statistics for the regions USA, Japan, Europe, rest of the world (RES) and emerging markets (EM) and for large, medium-sized and small firms. We show the following statistics: mean, median, minimum (Min), maximum (Max), variance (Var), standard deviation (Sd), skewness (Skew) and kurtosis (Kurt). Obs is the total number of observations and # Firms gives the average number of firms per month. The time period spans from January 2000 to December 2017.

	Mean	Median	Min	Max	Var	Sd	Skew	Kurt	Obs	# Firms
Panel A: News Events										
Overall	93.95	19	1	57,528	273,704	523.17	33.22	1,739	1,155,342	5349
USA	223.41	75	1	57,528	902,447	949.97	20.51	617	272,781	1263
Japan	41.53	10	1	24,704	49,800	223.16	32.97	2,398	106,144	491
Europe	85.11	23	1	41,395	143,970	379.43	27.62	1,383	280,823	1300
RES	55.63	13	1	12,207	35,348	188.01	17.51	542	158,896	736
EM	31.03	9	1	26,325	36,437	190.88	53.95	4,169	336,698	1559
Large	208.35	57	1	57,528	738,327	859.26	20.99	685	385,191	1783
Medium	52.83	19	1	22,643	48,775	220.85	51.52	3,538	385,038	1783
Small	20.62	7	1	18,684	13,717	117.12	92.88	11,454	385,113	1783
Panel B: ESS										
Overall	0.17	0.23	-1.0	1.0	0.15	0.39	-0.52	-0.30	851,220	3941
USA	0.16	0.18	-1.0	1.0	0.11	0.34	-0.34	-0.13	250,088	1158
Japan	0.18	0.27	-1.0	1.0	0.16	0.40	-0.60	-0.44	74,719	346
Europe	0.19	0.27	-1.0	1.0	0.15	0.39	-0.63	-0.11	199,378	923
RES	0.16	0.23	-1.0	1.0	0.18	0.42	-0.48	-0.51	104,366	483
EMM	0.17	0.27	-1.0	1.0	0.17	0.41	-0.55	-0.51	222,669	1031
Large	0.19	0.22	-1.0	1.0	0.11	0.33	-0.56	0.24	283,806	1314
Medium	0.16	0.22	-1.0	1.0	0.15	0.39	-0.45	-0.38	283,667	1313
Small	0.16	0.27	-1.0	1.0	0.18	0.43	-0.50	-0.67	283,747	1314

Table II
News equity factors: Global universe

This table shows performance statistics of equal-weighted long-short portfolios for a set of news indicators using the global stock universe. Annualized mean returns are calculated using the arithmetic average of simple returns. Standard deviation (Sd) and Sharpe ratio (SR) are annualized through multiplication by $\sqrt{12}$. Min and Max denote the lowest and highest monthly excess return in the sample period. MDD is the maximum drawdown. Mean return, Sd, Min, Max and MDD are given in percentage points. The last column gives the average number of firms per month. t-stat is the t-statistic for testing against the Null of a zero return effect. Mean returns and information coefficients are in boldface if significant at a 10% level or better. The time period is from January 2001 to December 2017.

Concept	Indicator	Return	t-stat	Sd	Min	Max	SR	MDD	Firms
News volume	$VOL_{\text{REL}>75,1}$	**-0.73**	-1.69	1.85	-1.97	3.39	-0.40	-13.92	3421
	VOL_1	-0.19	-0.64	1.24	-1.37	2.44	-0.15	-6.23	2772
	VOL_3	0.23	0.46	2.06	-2.35	4.89	0.11	-5.79	3576
	VOL_6	0.70	1.10	2.66	-2.64	6.18	0.26	-9.30	3774
News sentiment	$SENT_1$	**1.98**	4.96	1.70	-2.64	2.28	1.17	-2.83	2646
	$SENT_3$	**1.92**	3.70	2.19	-3.71	2.68	0.88	-5.69	3535
	$SENT_6$	**1.88**	3.01	2.61	-5.36	3.13	0.72	-7.83	3751
	$rSENT_{l=u=0,1}$	**1.78**	4.84	1.56	-2.71	2.02	1.14	-2.77	2646
	$rSENT_{l=u=0,3}$	**1.73**	3.67	1.99	-2.52	2.82	0.87	-3.42	3535
	$rSENT_{l=u=0,6}$	**1.70**	3.01	2.37	-4.51	3.11	0.72	-6.83	3751
	$wSENT_{td,1}$	**2.09**	4.82	1.84	-2.94	2.20	1.14	-3.26	2646
	$wSENT_{td,3}$	**2.03**	2.77	3.10	-6.15	3.11	0.66	-11.65	3535
	$wSENT_{td,6}$	**2.66**	3.28	3.41	-6.22	4.02	0.78	-12.34	3751
	$wSENT_{pt,1}$	**1.79**	3.96	1.91	-3.41	1.90	0.93	-3.91	2646
	$wSENT_{pt,3}$	**2.08**	2.68	3.28	-6.66	3.10	0.63	-13.25	3535
	$wSENT_{pt,6}$	**2.78**	3.24	3.60	-6.64	3.81	0.77	-12.79	3751
News trend	$SENTMOM$	**0.95**	3.68	1.06	-2.03	1.09	0.89	-3.34	2676
	$aSENTMOM_3$	0.30	0.98	1.31	-2.09	1.07	0.23	-4.84	2103
	$aSENTMOM_6$	**0.64**	1.95	1.37	-2.22	1.21	0.47	-3.94	2806
	REG_6	0.37	0.49	2.69	-3.29	3.36	0.14	-12.35	776
Alternative news concepts	$DISP_1$	0.89	1.31	2.90	-2.45	5.66	0.31	-6.11	2080
	$NEWSBETA_{60}$	1.43	1.58	2.77	-2.07	4.66	0.52	-3.53	2869
	SIG_1	**1.13**	2.37	2.02	-3.88	3.68	0.56	-5.20	2034
	SIG_3	**1.68**	3.31	2.15	-3.34	2.43	0.78	-6.41	3287
	SIG_6	**1.89**	3.06	2.59	-6.13	2.33	0.73	-9.11	3629

Table III
News equity factors: Cap-weighting

This table shows performance statistics of market capitalization-weighted long-short portfolios for a set of news indicators using the global stock universe. Annualized mean returns are calculated using the arithmetic average of simple returns. Standard deviation (Sd) and Sharpe ratio (SR) are annualized through multiplication by $\sqrt{12}$. Min and Max denote the lowest and highest monthly excess return in the sample period. MDD is the maximum drawdown. Mean return, Sd, Min, Max and MDD are given in percentage points. The last column gives the average number of firms per month. t-stat is the t-statistic for testing against the Null of a zero return effect. Mean returns and information coefficients are in boldface if significant at a 10% level or better. The time period is from January 2001 to December 2017.

Concept	Indicator	Return	t-stat	Sd	Min	Max	SR	MDD	Firms
News volume	$VOL_{\mathrm{REL}>75,1}$	-0.10	-0.36	1.17	-1.60	2.81	-0.08	-3.67	3421
	VOL_1	-0.02	-0.06	1.07	-1.15	2.41	-0.02	-3.68	2772
	VOL_3	0.26	0.64	1.69	-1.89	4.79	0.15	-3.99	3576
	VOL_6	0.42	0.82	2.15	-2.46	5.87	0.20	-6.70	3774
News sentiment	$SENT_1$	**1.89**	4.48	1.79	-2.56	2.14	1.06	-3.95	2646
	$SENT_3$	**2.11**	3.71	2.40	-3.75	2.35	0.88	-5.47	3535
	$SENT_6$	**2.04**	3.17	2.70	-4.78	2.99	0.76	-6.62	3751
	$rSENT_{l=u=0,1}$	**1.76**	4.17	1.80	-2.69	2.17	0.98	-4.10	2646
	$rSENT_{l=u=0,3}$	**2.05**	3.91	2.22	-2.50	2.51	0.93	-3.87	3535
	$rSENT_{l=u=0,6}$	**1.99**	3.24	2.57	-4.91	2.89	0.77	-7.57	3751
	$wSENT_{td,1}$	**1.89**	4.01	2.00	-3.05	2.49	0.95	-4.43	2646
	$wSENT_{td,3}$	**1.98**	2.54	3.30	-5.93	2.87	0.60	-12.18	3535
	$wSENT_{td,6}$	**2.98**	3.47	3.59	-5.70	4.20	0.83	-11.99	3751
	$wSENT_{pt,1}$	**1.71**	3.52	2.06	-3.37	2.11	0.83	-5.33	2646
	$wSENT_{pt,3}$	**2.05**	2.50	3.47	-6.52	2.65	0.59	-13.33	3535
	$wSENT_{pt,6}$	**3.17**	3.50	3.79	-6.43	3.92	0.83	-12.77	3751
News trend	$SENTMOM$	**1.03**	3.05	1.39	-2.41	1.48	0.74	-4.05	2676
	$aSENTMOM_3$	0.08	0.30	1.11	-1.50	1.77	0.07	-3.16	2103
	$aSENTMOM_6$	**0.74**	2.39	1.30	-1.34	2.31	0.57	-4.43	2806
	REG_6	0.33	0.48	2.47	-4.75	2.63	0.14	-9.43	776
Alternative news concepts	$DISP_1$	0.83	1.28	2.74	-2.40	6.44	0.30	-4.26	2080
	$NEWSBETA_{60}$	1.29	1.37	2.88	-2.45	4.93	0.45	-3.70	2869
	SIG_1	**0.76**	1.84	1.74	-4.04	1.78	0.43	-5.54	2034
	SIG_3	**1.38**	2.68	2.18	-2.89	2.31	0.63	-6.19	3287
	SIG_6	**1.57**	2.61	2.52	-5.28	2.10	0.62	-8.41	3629

Table IV
News equity factors: Spanning tests

This table shows results of spanning tests for the most promising equal-weighted news analytics factors in the global universe. As regressors we use the market return (represented by the MSCI World) and the equity factors value, quality, size, momentum (MOM) and short-tem reversal (STR) that are known to affect the cross-section of stock returns. We report the estimates for the intercept (alpha) and the equity factors. t-statistics are computed from Newey-West adjusted standard errors and are given in parentheses. In addition, the last two columns report the test statistics and corresponding p-values (in parentheses) of the Kan and Zhou (2012) step-down test. The null hypothesis is that news factors are spanned by the standard equity factors. F1 tests whether news factors improve the tangency portfolio, while F2 tests the ability of news factors to improve the minimum-variance portfolio. Coefficients and test statistics are in boldface if significant at a 10% level or better. The time period is from January 2001 to December 2017.

Concept	Indicator	Alpha	Market	Value	Quality	Size	Momentum	STR	R^2_{adj}	F1-Test	F2-Test
News sentiment	$SENT_1$	**0.002**	**-0.018**	0.003	**-0.077**	0.000	**0.039**	0.002	25.6%	**23.32**	**1,279.08**
		(4.28)	(-1.70)	(0.17)	(-2.52)	(-0.06)	(4.11)	(0.21)		(0.00)	(0.00)
	$SENT_3$	**0.001**	-0.003	-0.005	**-0.107**	0.007	**0.087**	**0.019**	57.9%	**19.46**	**1,112.73**
		(3.88)	(-0.39)	(-0.24)	(-2.93)	(0.63)	(10.46)	(2.29)		(0.00)	(0.00)
	$SENT_6$	**0.002**	-0.008	-0.032	**-0.121**	0.012	**0.107**	**0.022**	62.1%	**18.15**	**889.96**
		(3.83)	(-0.71)	(-1.22)	(-2.61)	(1.16)	(9.71)	(2.40)		(0.00)	(0.00)
	$rSENT_{l=u=0,1}$	**0.001**	**-0.015**	0.016	-0.046	-0.001	**0.036**	0.002	22.9%	**16.50**	**1,363.43**
		(3.96)	(-1.69)	(0.93)	(-1.58)	(-0.19)	(3.55)	(0.24)		(0.00)	(0.00)
	$wSENT_{td,1}$	**0.002**	**-0.018**	-0.006	**-0.079**	0.006	**0.051**	0.005	33.6%	**25.20**	**1,107.23**
		(4.46)	(-1.65)	(-0.34)	(-2.38)	(0.75)	(5.15)	(0.41)		(0.00)	(0.00)
	$wSENT_{pt,1}$	**0.001**	-0.014	-0.006	**-0.072**	0.006	**0.058**	0.006	37.3%	**15.86**	**1,123.29**
		(3.77)	(-1.29)	(-0.29)	(-1.96)	(0.77)	(5.55)	(0.52)		(0.00)	(0.00)
	$wSENT_{pt,3}$	**0.002**	-0.018	**-0.064**	**-0.130**	0.023	**0.142**	0.009	71.1%	**18.38**	**753.80**
		(3.77)	(-1.46)	(-2.83)	(-2.72)	(1.59)	(13.54)	(0.84)		(0.00)	(0.00)
	$wSENT_{pt,6}$	**0.002**	-0.019	**-0.056**	**-0.128**	**0.032**	**0.154**	0.003	70.1%	**24.15**	**546.39**
		(4.36)	(-1.56)	(-2.11)	(-2.40)	(2.07)	(11.72)	(0.22)		(0.00)	(0.00)
News trend	$SENTMOM$	**0.001**	-0.001	**-0.026**	-0.010	-0.001	**0.030**	0.003	28.2%	**13.60**	**3,112.57**
		(5.85)	(-0.18)	(-2.30)	(-0.78)	(-0.11)	(5.37)	(0.73)		(0.00)	(0.00)
	$aSENTMOM_6$	**0.001**	-0.011	-0.004	**-0.062**	-0.003	**0.042**	-0.006	36.4%	**4.62**	**2,344.13**
		(2.18)	(-1.58)	(-0.32)	(-2.23)	(-0.37)	(6.60)	(-0.72)		(0.03)	(0.00)
Alternative news concepts	SIG_1	0.001	-0.010	0.015	-0.038	0.001	**0.046**	0.000	24.0%	2.41	**1,044.36**
		(1.47)	(-0.92)	(0.68)	(-1.10)	(0.11)	(3.91)	(0.01)		(0.12)	(0.00)
	SIG_3	**0.001**	-0.007	-0.013	-0.046	**0.023**	**0.085**	0.010	58.7%	**10.09**	**1,143.65**
		(2.88)	(-0.88)	(-0.73)	(-1.44)	(2.39)	(9.22)	(0.94)		(0.00)	(0.00)
	SIG_6	**0.001**	-0.009	**-0.052**	-0.057	**0.024**	**0.110**	**0.019**	66.5%	**16.39**	**939.55**
		(4.15)	(-0.90)	(-2.38)	(-1.62)	(2.14)	(10.14)	(1.73)		(0.00)	(0.00)

Table V
News equity factors: Robustness to different holding periods

This table shows performance statistics of long-short portfolios based on the news indicators for the global stock universe and longer return horizons. Annualized mean returns are calculated using the arithmetic average of simple returns and are given in percentage points. We use different lags of the news indicator to return: 1, 3, 6, 9 and 12 months. t-stat is the t-statistic for testing against the Null of a zero effect. Mean returns are in boldface if significant at a 10% level or better. The time period is from January 2001 to December 2017.

Concept	Indicator	Ret.1M	tstat	Ret.3M	tstat	Ret.6M	t-stat	Ret.9M	t-stat	Ret.12M	t-stat
News volume	$VOL_{\mathrm{REL}>75,1}$	**-0.73**	-1.69	-0.11	-0.25	0.38	0.98	-0.51	-1.31	**0.89**	2.83
	VOL_1	-0.19	-0.64	0.37	1.26	**0.51**	1.71	-0.17	-0.61	**1.36**	4.49
	VOL_3	0.23	0.46	-0.12	-0.28	0.25	0.67	-0.15	-0.42	**1.22**	4.18
	VOL_6	0.70	1.10	0.03	0.05	0.32	0.64	**0.96**	2.27	**0.88**	2.43
News sentiment	$SENT_1$	**1.98**	4.96	0.08	0.22	0.42	1.29	**0.58**	2.06	0.19	0.73
	$SENT_3$	**1.92**	3.70	0.76	1.58	**0.87**	2.24	0.47	1.51	0.34	1.31
	$SENT_6$	**1.88**	3.01	**1.18**	2.20	**0.82**	1.91	**0.92**	2.69	0.37	1.24
	$rSENT_{l=u=0,1}$	**1.78**	4.84	0.39	1.19	0.50	1.58	**0.69**	2.57	0.09	0.39
	$rSENT_{l=u=0,3}$	**1.73**	3.67	**0.83**	1.95	**0.87**	2.18	0.52	1.64	0.27	1.00
	$rSENT_{l=u=0,6}$	**1.70**	3.01	**0.95**	1.86	0.49	1.17	**0.67**	2.04	0.42	1.39
	$wSENT_{td,1}$	**2.09**	4.82	-0.04	-0.10	0.33	0.92	**0.54**	1.80	0.16	0.54
	$wSENT_{td,3}$	**2.03**	2.77	**1.05**	1.78	**1.23**	2.39	0.54	1.26	0.24	0.69
	$wSENT_{td,6}$	**2.66**	3.28	**1.83**	2.69	**1.59**	2.86	**0.85**	1.78	**0.88**	2.18
	$wSENT_{pt,1}$	**1.79**	3.96	0.12	0.28	0.35	0.89	**0.72**	2.32	0.00	-0.01
	$wSENT_{pt,3}$	**2.08**	2.68	**1.13**	1.72	**1.58**	2.66	0.66	1.32	0.38	1.04
	$wSENT_{pt,6}$	**2.78**	3.24	**1.80**	2.39	**1.61**	2.59	**1.16**	2.21	**0.90**	2.06
News trend	$SENTMOM$	**0.95**	3.68	0.13	0.44	0.23	0.92	0.25	1.50	0.11	0.53
	$aSENTMOM_3$	0.30	0.98	0.31	1.13	0.31	1.28	0.22	0.92	0.07	0.31
	$aSENTMOM_6$	**0.64**	1.95	0.39	1.63	0.41	1.61	0.32	1.55	**0.34**	1.75
	REG_6	0.37	0.49	-1.21	-1.57	-0.24	-0.35	0.21	0.36	-0.54	-0.74
Alternative news concepts	$DISP_1$	0.89	1.31	0.39	0.58	-0.32	-0.46	**1.63**	3.07	-0.09	-0.17
	$NEWSBETA_{60}$	1.43	1.58	-0.23	-0.29	-0.58	-0.99	0.18	0.29	-0.69	-0.98
	SIG_1	**1.13**	2.37	**0.64**	1.73	**0.94**	2.29	**0.78**	2.36	**0.70**	2.25
	SIG_3	**1.68**	3.31	0.65	1.22	**0.84**	1.92	0.31	0.98	**0.54**	1.86
	SIG_6	**1.89**	3.06	**1.18**	2.38	**0.67**	1.65	**1.07**	3.52	**0.69**	2.32

Table VI
News equity factors: Regional universes

This table shows performance statistics of long-short portfolios based on the news indicators for the regional universes USA, Japan, Europe, rest of the world (RES) and emerging markets (EM) in addition to the global universe. Annualized mean returns are calculated using the arithmetic average of simple returns and are given in percentage points. t-stat is the t-statistic for testing against the Null of a zero effect. Mean returns are in boldface if significant at a 10% level or better. The time period is from January 2001 to December 2017.

		Global		USA		Japan		Europe		RES		EM	
Concept	Indicator	Return	t-stat	Return	t-stat	Return	t-stat	Return	t-stat	Return	t-stat	Return	t-stat
News volume	$VOL_{\mathrm{REL}>75,1}$	**-0.73**	-1.69	-0.57	-0.77	1.02	1.01	**-1.61**	-2.49	-0.83	-1.26	-0.16	-0.27
	VOL_1	-0.19	-0.64	-0.28	-0.45	1.65	1.36	-0.78	-1.49	0.28	0.41	**1.32**	1.79
	VOL_3	0.23	0.46	**1.87**	2.26	1.43	1.33	-0.77	-0.92	0.16	0.24	**1.53**	3.39
	VOL_6	0.70	1.10	**1.59**	1.76	0.83	0.62	-0.11	-0.11	1.00	1.29	**2.37**	4.86
News sentiment	$SENT_1$	**1.98**	4.96	**1.79**	2.27	-0.43	-0.32	**2.96**	4.94	**5.19**	5.63	**4.08**	4.95
	$SENT_3$	**1.92**	3.70	0.86	0.99	-0.25	-0.24	**3.54**	4.00	**4.98**	6.40	**4.85**	8.82
	$SENT_6$	**1.88**	3.01	1.01	0.99	-0.06	-0.06	**3.71**	4.17	**4.38**	4.72	**3.99**	8.08
	$rSENT_{l=u=0,1}$	**1.78**	4.84	**1.70**	2.55	-0.43	-0.39	**3.27**	5.66	**4.69**	5.31	**3.29**	4.21
	$rSENT_{l=u=0,3}$	**1.73**	3.67	0.93	1.26	0.11	0.11	**3.34**	4.05	**3.90**	5.21	**4.17**	8.66
	$rSENT_{l=u=0,6}$	**1.70**	3.01	0.78	0.83	0.29	0.32	**3.39**	3.81	**3.75**	4.03	**3.71**	8.39
	$wSENT_{td,1}$	**2.09**	4.82	**1.90**	2.30	0.40	0.31	**3.05**	5.04	**5.38**	5.63	**4.69**	5.53
	$wSENT_{td,3}$	**2.03**	2.77	1.10	0.86	-0.65	-0.61	**3.72**	4.17	**5.11**	5.99	**5.56**	9.56
	$wSENT_{td,6}$	**2.66**	3.28	**2.41**	1.76	0.75	0.69	**4.14**	3.89	**5.54**	5.66	**4.70**	8.69
	$wSENT_{pt,1}$	**1.79**	3.96	1.18	1.47	0.31	0.26	**3.44**	5.21	**4.91**	5.27	**4.61**	4.85
	$wSENT_{pt,3}$	**2.08**	2.68	1.23	0.93	0.15	0.15	**3.86**	4.01	**4.92**	5.46	**5.55**	9.81
	$wSENT_{pt,6}$	**2.78**	3.24	**2.44**	1.73	0.45	0.39	**4.31**	3.89	**5.55**	5.50	**5.15**	9.29
News trend	$SENTMOM$	**0.95**	3.68	**1.31**	2.03	-0.80	-0.72	**1.42**	2.84	**3.36**	4.04	**3.03**	4.03
	$aSENTMOM_3$	0.30	0.98	-0.21	-0.33	-0.21	-0.18	**1.13**	2.02	0.68	0.64	0.92	0.68
	$aSENTMOM_6$	**0.64**	1.95	-0.34	-0.62	-0.14	-0.15	**1.65**	2.52	**1.95**	2.70	**1.51**	3.26
	REG_6	0.37	0.49	0.65	0.64	-3.58	-0.44	**2.88**	1.95	-4.19	-0.55	4.06	1.05
Alternative news concepts	$DISP_1$	0.89	1.31	1.70	0.95	3.00	1.36	0.80	0.78	1.31	0.93	**-5.51**	-2.00
	$NEWSBETA_{60}$	1.43	1.58	2.56	1.20	2.42	1.33	**1.82**	1.87	0.66	0.41	0.55	0.76
	SIG_1	**1.13**	2.37	0.00	0.00	1.70	1.10	**2.11**	2.64	**4.45**	3.10	2.32	1.05
	SIG_3	**1.68**	3.31	0.41	0.47	0.16	0.16	**2.44**	2.66	**5.03**	6.59	**4.49**	6.72
	SIG_6	**1.89**	3.06	0.93	0.92	0.05	0.05	**3.73**	3.88	**4.87**	5.72	**4.64**	8.16

Table VII
Risk-based factor allocation

This table shows performance statistics of risk-based factor allocation strategies for the set of benchmark factors (Panel A) and the set of benchmark factors augmented by the news-based equity factors $SENT_1$, $SENT_6$ and $wSENT_{pt,6}$ (Panel B). Specifically, we examine an equally weighted portfolio (1/N), a minimum-variance portfolio (MVP) and a risk parity portfolio (RP). Annualized excess returns are calculated using the arithmetic average of simple returns. Standard deviation (Sd) and Sharpe ratio (SR) are annualized through multiplication by $\sqrt{12}$. Min and Max denote the lowest and highest monthly excess return in the sample period. MDD is the maximum drawdown. Excess return, Sd, Min, Max and MDD are given in percentage points. t-stat is the t-statistic for testing against the Null of a zero return effect. The performance statistics are based on the out-of-sample period from January 2007 to September 2017.

Strategy	Excess Return	Sd	Min	Max	SR	MDD	t-stat
Panel A: Benchmark factors							
1/N	3.26	2.45	-1.68	2.19	1.33	3.51	4.18
MVP	2.16	1.22	-0.50	1.37	1.77	0.94	5.55
RP	2.67	1.38	-0.79	1.74	1.94	1.27	6.07
Panel B: Benchmark + news factors							
1/N	3.47	2.18	-1.66	2.03	1.60	3.94	4.81
MVP	2.53	1.19	-0.49	1.55	2.13	0.85	6.42
RP	2.94	1.30	-0.66	1.56	2.25	1.14	6.79

Table VIII
News Factor Timing: Coefficients

This table shows the θ-coefficients for the fundamental ($FUN1$) and technical ($TECH1$) PCA factors that obtain in the parametric portfolio policy (PPP) for factor timing. We consider the PPP for the set of benchmark equity factors and the PPP for the set of benchmark factors augmented by the news-based equity factors $SENT_1$, $SENT_6$ and $wSENT_{pt,6}$. The coefficients are in bold-face if significant at the 5%-level. S.E. denotes the standard error of the coefficients. The sample period is from January 2002 to September 2017.

	Benchmark factors				Benchmark + news factors			
Predictor variable	$FUN1$	S.E.	$TECH1$	S.E.	$FUN1$	S.E.	$TECH1$	S.E.
$PROF$	-0.07	0.55	-1.68	1.47	-0.82	0.59	-1.56	1.47
CFY	0.98	0.94	0.74	1.26	1.96*	0.97	0.54	1.29
ACC	-1.99	1.09	-1.42	1.88	-1.73	1.11	-0.67	1.85
DY	-0.31	0.60	1.31	0.83	0.35	0.67	1.34	0.86
AT	-0.06	1.02	-2.10	1.16	-0.72	1.08	-0.66	1.17
BTM	-0.82	1.18	-0.88	1.58	-1.23	1.22	-0.09	1.63
$MOM12$	0.24	0.24	-0.66	0.43	-0.33	0.28	0.11	0.61
STR	-0.29	0.25	**-1.35**	0.53	-0.21	0.25	**-1.85**	0.54
LTR	0.03	0.55	**-1.36**	0.57	0.42	0.56	**-2.18**	0.62
$DLTD$	**-4.31**	1.41	-3.08	2.63	**-6.17**	1.56	-3.17	2.64
DSO	**-2.12**	0.82	1.02	1.62	-0.65	0.88	0.44	1.74
$SIZE$	**-0.80**	0.33	**-1.05**	0.46	**-0.98**	0.34	**-1.10**	0.47
AG	2.53	1.33	0.74	1.82	2.70	1.49	1.51	1.80
CP	**-2.62**	1.20	-1.56	2.06	**-4.51**	1.23	-1.61	2.10
PM	**-3.37**	1.04	1.61	1.28	**-4.30**	1.09	2.54	1.34
EY	**4.15**	1.02	**-3.83**	1.09	**4.80**	1.06	**-3.69**	1.16
LEV	-0.40	0.77	-0.11	0.91	-1.24	0.77	-0.49	0.92
ROA	-0.67	1.21	-2.60	1.44	**-2.91**	1.26	-2.11	1.53
STC	0.93	0.91	0.67	1.24	0.83	0.93	-0.27	1.23
STI	-0.92	0.69	-0.51	1.37	**-1.62**	0.68	-0.46	1.42
$SENT_1$	–	–	–	–	-0.91	0.93	-3.53	2.27
$SENT_6$	–	–	–	–	2.75	1.45	-1.12	2.38
$wSENT_{pt,6}$	–	–	–	–	1.59	1.21	-2.47	1.80

Table IX
News Factor Timing: Performance statistics

This table gives performance statistics of parametric portfolio policies (PPP) for factor timing. We use the first principal components of fundamental ($FUN1$) and technical ($TECH1$) predictor variables in the PPP. Panel A gives the PPP for the set of benchmark equity factors and Panel B gives the PPP for the set of benchmark factors augmented by the news-based equity factors $SENT_1$, $SENT_6$ and $wSENT_{pt,6}$. We include an equally weighted portfolio (1/N) as benchmark strategy for both sets. The performance statistics are based on the out-of-sample period from January 2007 to September 2017. Annualized excess returns are calculated using the arithmetic average of simple returns. Standard deviation (Sd) and Sharpe ratio (SR) are annualized through multiplication by $\sqrt{12}$. The information ratio (IR) uses arithmetic active returns of factor timing over the 1/N benchmark. Annualized turnover is stated as two-way turnover. All performance statistics are given in percentage points, except for Sharpe ratio.

	Excess Return			SR		IR		
Strategy	*gross*	*net*	Sd	*gross*	*net*	*gross*	*net*	Turnover
Panel A: Timing model with benchmark factors								
1/N	3.26	2.24	2.45	1.33	0.92	–	–	–
FUN1 + TECH1	3.75	1.12	3.36	1.12	0.33	0.29	-0.66	8.36
Panel B: Timing model with benchmark + news factors								
1/N	3.47	2.45	2.18	1.58	1.09	–	–	–
FUN1 + TECH1	3.91	0.96	3.08	1.27	0.31	0.35	-0.79	9.92

Table X
News factor tilting

The table gives estimation results and performance statistics of parametric portfolio policies (PPP) for factor tilting based on cross-sectional factor characteristics. We consider six news sentiment-related and four benchmark characteristics. Panel A gives the PPP for the set of benchmark equity factors and Panel B gives the PPP for the set of benchmark factors augmented by the news-based equity factors $SENT_1$, $SENT_6$ and $wSENT_{pt,6}$. We include an equally weighted portfolio (1/N) as benchmark strategy for both sets. The performance statistics are based on the out-of-sample period from January 2007 to September 2017. The second column gives the estimated coefficients of the PPP, marked in boldface if significant at a 5% level or better. Annualized returns are calculated using the arithmetic average of simple returns. Standard deviation and Sharpe ratio are annualized through multiplication by $\sqrt{12}$. The information ratio uses arithmetic active returns of factor timing over the 1/N benchmark. Annualized turnover is stated as two-way turnover. All performance statistics are given in percentage points, except for Sharpe ratio and t-statistic.

Characteristic	$\hat{\phi}$	Return p.a.		SD p.a.	Sharpe ratio		Maximum drawdown		t-statistic		Tracking error	Information ratio		Turnover p.a.
		gross	*net*		*gross*	*net*	*gross*	*net*	*gross*	*net*		*gross*	*net*	*two-way*
Panel A: Tilting model with benchmark factors														
1/N		3.26	2.24	2.45	1.33	0.92	3.51	4.01	4.18	2.87	–	–	–	
$SENT_1$	**5.03**	4.09	2.13	2.87	1.43	0.74	4.38	5.05	4.47	2.33	1.67	0.50	-0.07	5.02
$SENT_3$	**3.72**	4.20	2.80	2.59	1.62	1.08	5.32	5.92	5.09	3.39	1.69	0.56	0.33	2.21
$SENT_6$	2.85	4.31	3.08	2.34	1.84	1.32	4.67	5.20	5.78	4.14	1.77	0.56	0.44	1.39
$wSENT_{td,6}$	2.39	3.93	2.72	2.44	1.61	1.12	4.61	5.17	5.04	3.50	1.64	0.40	0.29	1.21
$wSENT_{pt,6}$	2.34	3.92	2.73	2.48	1.58	1.10	4.66	5.21	4.95	3.45	1.64	0.40	0.30	1.13
SIG_6	3.16	4.27	2.98	2.67	1.60	1.11	4.94	5.51	5.02	3.50	1.75	0.58	0.42	1.68
Crowding	**6.09**	4.87	3.34	3.29	1.48	1.02	3.85	4.51	4.64	3.19	1.83	0.88	0.60	2.89
Momentum	1.37	4.08	2.20	3.04	1.34	0.72	3.12	3.88	4.22	2.26	1.47	0.56	-0.03	4.61
Spread	**16.35**	4.01	2.73	3.07	1.30	0.88	3.55	4.07	4.09	2.75	1.83	0.41	0.26	1.59
Valuation	-1.51	3.72	2.50	2.41	1.54	1.04	4.12	4.67	4.84	3.26	1.72	0.27	0.15	1.30
Panel B: Tilting model with benchmark + news factors														
1/N	–	3.47	2.45	2.18	1.60	1.13	3.94	4.44	4.81	3.39	–	–	–	–
$SENT_1$	5.12	4.15	2.61	2.23	1.86	1.18	4.27	4.81	5.83	3.70	1.72	0.48	0.18	2.88
$SENT_3$	3.26	4.35	3.07	2.30	1.89	1.34	4.60	5.13	5.92	4.19	1.75	0.59	0.44	1.63
$SENT_6$	2.73	4.07	2.80	2.51	1.62	1.11	5.07	5.63	5.08	3.49	1.69	0.48	0.33	1.53
$wSENT_{td,6}$	2.49	4.29	3.07	2.36	1.82	1.30	4.67	5.21	5.70	4.07	1.72	0.56	0.44	1.31
$wSENT_{pt,6}$	2.46	4.26	3.05	2.36	1.80	1.29	4.70	5.24	5.65	4.04	1.70	0.55	0.44	1.26
SIG_6	3.17	4.40	3.16	2.39	1.84	1.32	4.74	5.27	5.78	4.15	1.75	0.62	0.49	1.41
Crowding	**6.53**	5.18	3.62	3.14	1.65	1.16	4.21	4.90	5.18	3.65	1.97	0.94	0.68	3.03
Momentum	1.12	3.93	2.04	2.65	1.49	0.77	2.84	3.55	4.66	2.40	1.39	0.44	-0.18	4.62
Spread	**15.62**	3.95	2.70	2.78	1.42	0.96	3.08	3.59	4.47	3.01	1.78	0.36	0.22	1.47
Valuation	-1.36	3.74	2.54	2.38	1.57	1.07	4.68	5.39	4.92	3.34	1.68	0.25	0.14	1.24

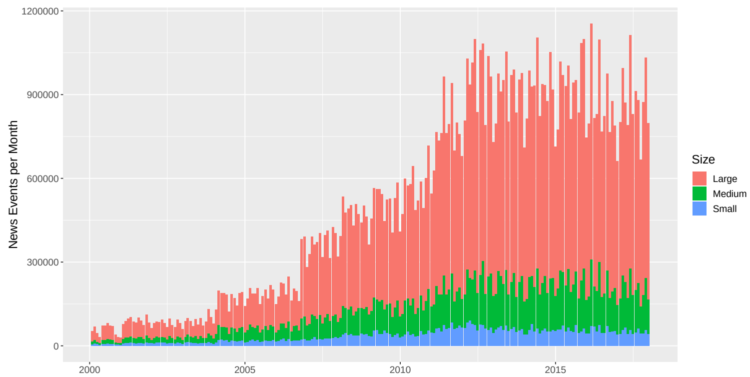
Table XI
Equity Factor Description

This table describes how we define common equity factors. The necessary data are sourced from the Worldscope database.

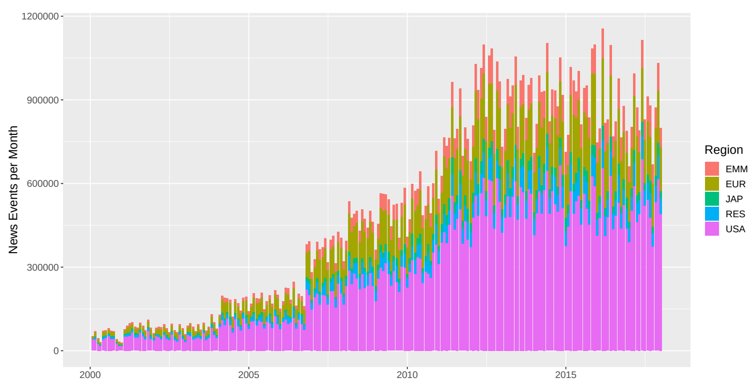
Factor	Description	Related studies
Value	We use cashflow yield as value factor. It captures the excess return of going long stocks with a high cashflow-to-price ratio and short those with a low cashflow-to-price ratio. Cashflows are measured as the sum of funds from operations, extraordinary items and funds from other operating activities	Sloan (1996); Da and Warachka (2009); Hou et al. (2011)
Quality	We use profitability as quality factor. This factor is long stocks with robust operating profitability and short stocks with weak profitability. Profitability is calculated as annual revenues less cost of goods sold and interest and other expenses, divided by book value for the last fiscal year-end.	Haugen and Baker (1996); Cohen et al. (2002); Fama and French (2006); Novy-Marx (2013); Fama and French (2016)
Momentum	We employ 12-month momentum that captures a medium-term continuation effect in returns by buying recent winners and selling recent losers. We control for the short-term reversal effect by excluding the most recent month $(t-1)$ at time t.	Jegadeesh (1990); Jegadeesh and Titman (1993)
Size	The size factors builds on the observation that stocks with a larger market capitalization tend to underperform stocks with smaller market capitalizations. The factor is going long stocks with the smallest market capitalization and short stocks with the highest market capitalizations.	Banz (1981); Fama and French (1992); Sloan (1996); Da and Warachka (2009); Hou et al. (2011)
Short-term reversal	This factor captures the short-term reversal effect in the cross-section of stock returns. The factor is long stocks with a weak previous month performance and short stocks with a high one.	Jegadeesh (1990); Lehmann (1990)

Figure 1. Characteristics of news volume

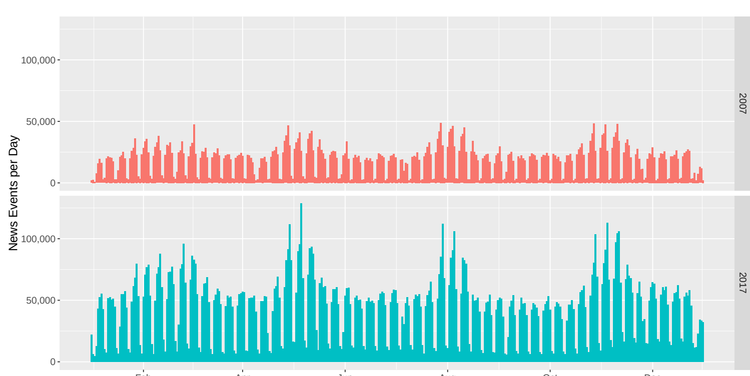
This figure illustrates various characteristics of news volume over the sample period from January 2001 to December 2017. Panel (a) shows monthly news events allocated to the following regions: United States (USA), Japan (JAP), Europe (EUR), emerging markets (EM) and rest of the world (RES). Panel (b) shows news volume per market capitalization (large, medium-sized and small companies). Panel (c) illustrates the yearly pattern of daily news events for the years 2007 and 2017.



(a) News volume per size



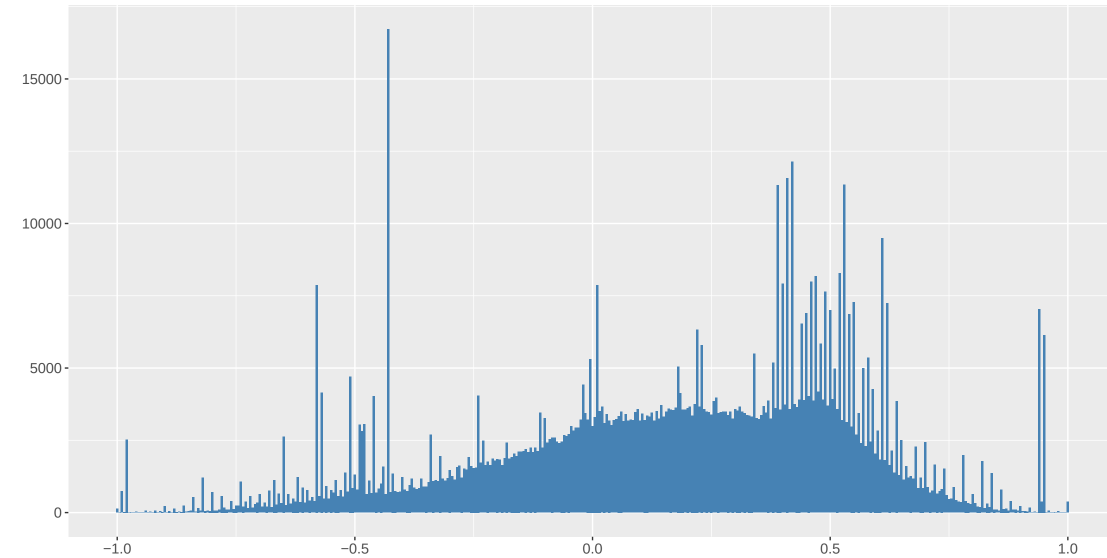
(b) News volume per region



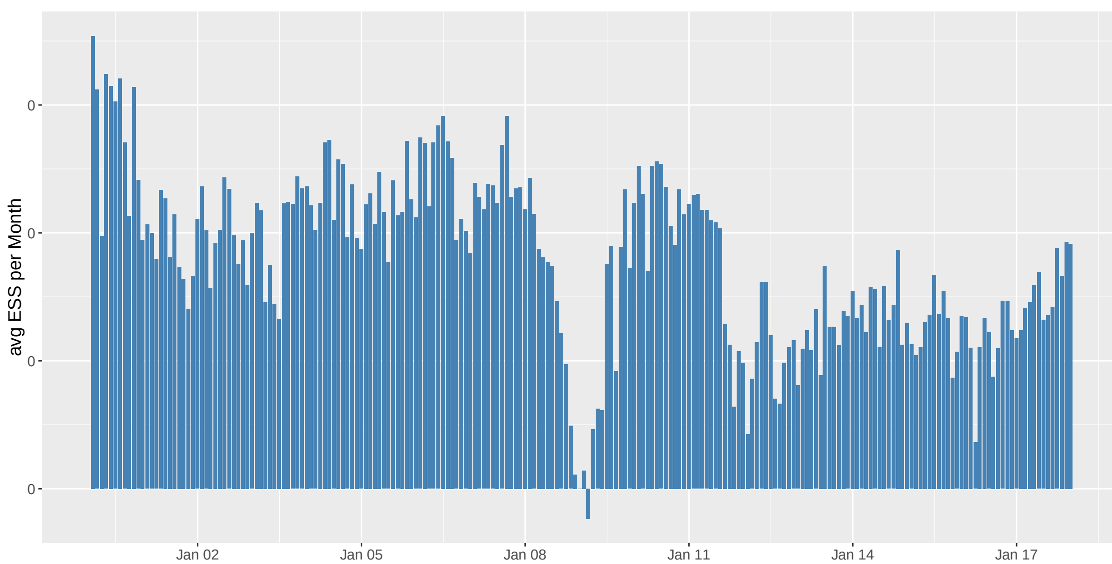
(c) Seasonality of news volume

Figure 2. Characteristics of news sentiment

Panel (a) shows the histogram of the ESS, whereas Panel (b) shows the monthly average event sentiment score across all firms. The sample period goes from January 2000 to December 2017.



(a) Histogram of ESS



(b) ESS over the sample period

Figure 3. Return correlation of news equity factors

This figure shows the correlation among news equity factors and traditional equity factors. Equity factors are derived from monthly return data for the global stock universe over the sample period from January 2001 to December 2017 and are grouped according to their concept category: news volume (A), news sentiment (B), news trend (c), alternative news concepts (D) and traditional equity factors (E).

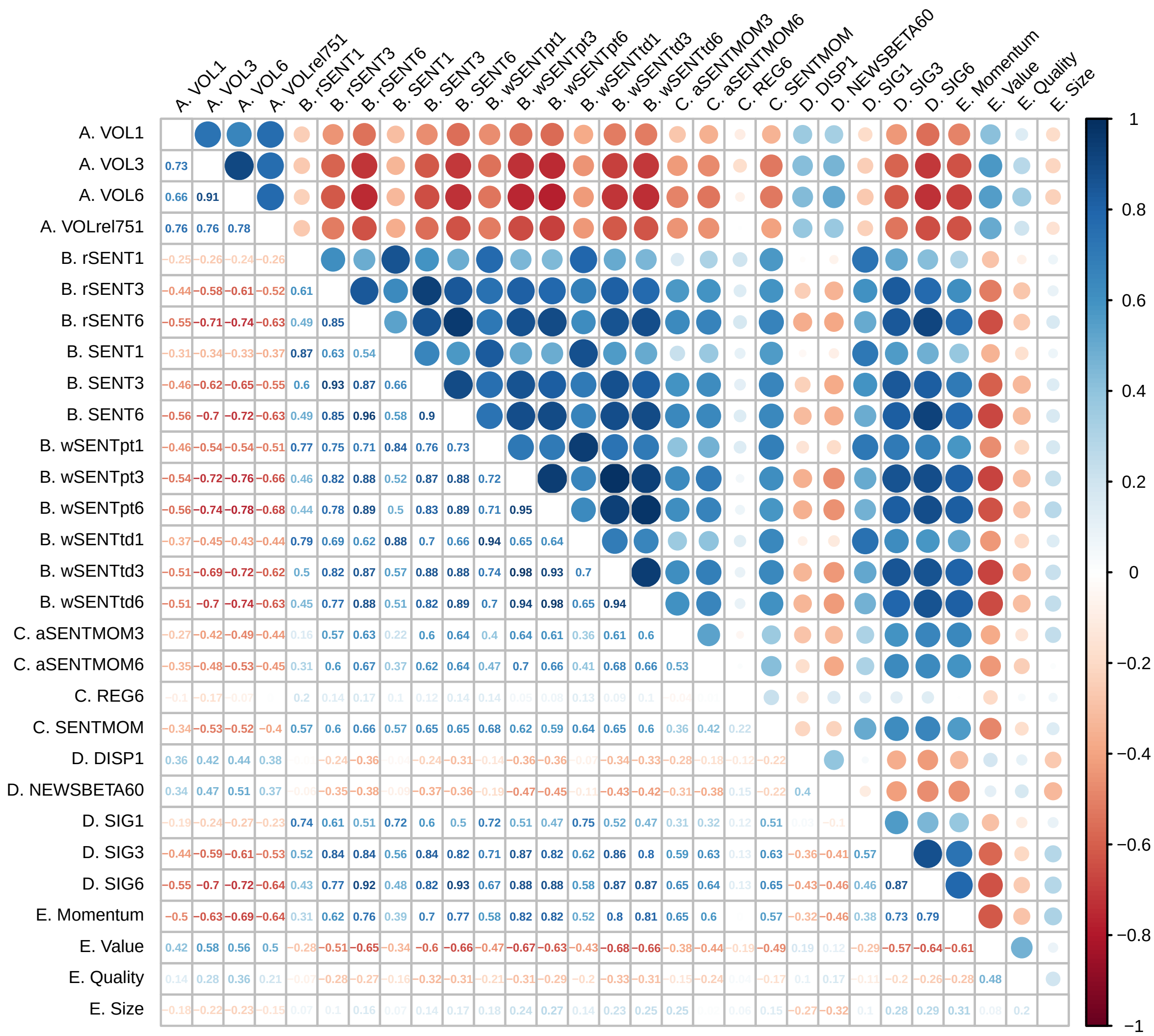
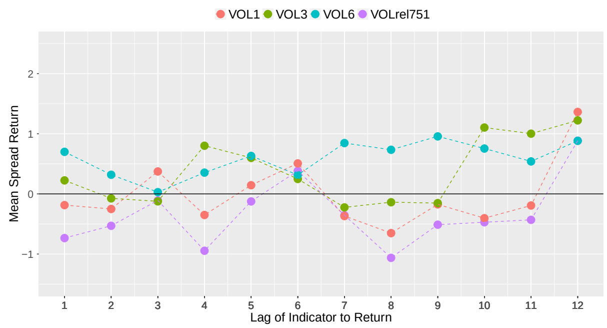
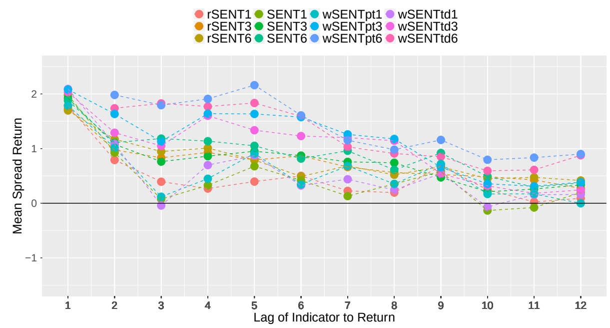


Figure 4. News equity factors: Long-horizon effects

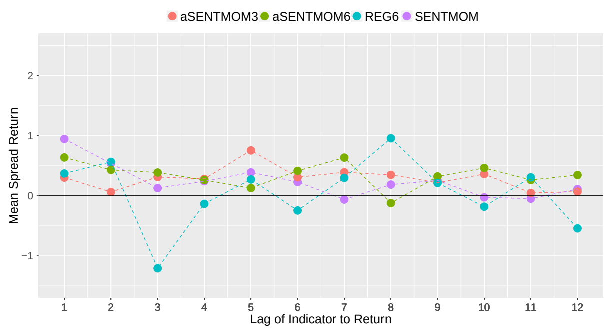
This figure shows the returns of cross-sectional long-short portfolios based on news volume (Panel A), news sentiment (Panel B), news trend (Panel C) and alternative news concepts (Panel D) indicators for the global stock universe from January 2001 to December 2017.



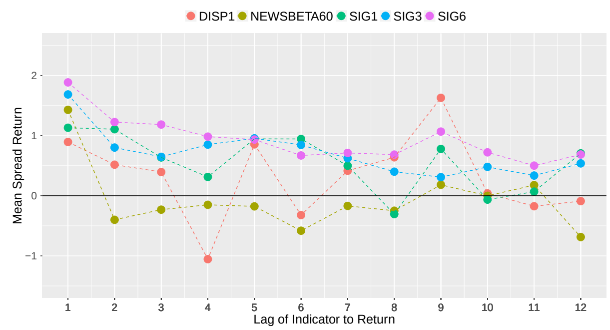
(a) News volume



(b) News Sentiment



(c) News Trend



(d) Alternative News Concepts