Quantitative Finance, 2020
Vol. 20, No. 9, 1531–1551, https://doi.org/10.1080/14697688.2020.1736314

Stock-specific sentiment and return predictability

GUILLAUME COQUERET *

EMLYON Business School, 23 avenue Guy de Collongue 69130, Ecully, France

(*Received 3 May 2019; accepted 24 February 2020; published online 20 April 2020*)

This paper quantifies the impact of stock-specific news sentiment on future financial returns. Daily predictive regressions yield significant *t*-statistics for 7% at most of our sample of more than 1000 large stocks listed in the USA. While a few assets do run through pockets of predictability, the evidence suggests that the feedback effect is stronger in the reverse direction: returns are more likely to drive future sentiment than the other way around.

Keywords: News sentiment; Predictability; *p*-hacking

JEL Classification: G12, G17

1. Introduction

It is both an intuitive and a documented fact that news flows are drivers of asset returns. During the XX th century, these flows were dominated by scheduled and infrequent events, namely macroeconomic announcements and accounting disclosures in which earnings were notably scrutinised.† Over the recent decades, the continuous improvement in telecommunications and information technology has dramatically increased both the pace and variety of sources of news broadcasts. At the same time, social media and trading networks (e.g. eToro, ZuluTrade and Ayondo to name but a few) have changed the way investors craft their opinions and implement their trading strategies. Thus it has become a genuine challenge to synthesise the publicly available information on firms in indicators that track the propensity of a representative agent to buy or sell one stock in particular. This task is even more daunting at the aggregate level, e.g. for national equity indices.

Intrinsically, the crafting of such sentiment metrics, because they reflect attitudes and expectations towards the financial market, is valuable for investors. Consequently, data providers, such as Bloomberg or Thomson Reuters, now release such indicators, and other organisations also publish or sell similar indices (e.g. the American Association of Individual Investors, FinSentS, SentimenTrader, MarketPsych, RavenPack, etc.). Concurrently, the topic of sentiment has become mainstream in academia and the related literature is both vast and quickly expanding, as we will see in the next section. A macro view of published results suggests that measures of sentiment are often found to be effective conditioning variables or predictors in academic research, especially for aggregate variables. At first glance, this is easily interpreted: if sentiment is hypothesised to reflect a possibly partial, but shared, belief on financial assets, then it should indeed act as a driver of asset prices. Nonetheless, the overwhelming evidence supporting the impact of sentiment on financial markets can also stem from the publication bias towards significant results (see Thornton and Lee 2000, Lehrer *et al.* 2007, Mlinarić *et al.* 2017 and the references therein).‡ In an attempt to confirm and replicate the findings on the impact of sentiment on stock markets, we sought to perform one of the most common analyses in the field, namely, predictive regressions. However, in contrast to most existing contributions, we do not work at the aggregate national level (both for returns and sentiment), but at the asset level. Our conclusion is overwhelmingly negative and contrasts with the academic fervour surrounding sentiment.

One convenient feature of sentiment is that is can be evaluated at high frequency when analysing continuous news flows. Given the high coverage of firms (by news outlets, by financial analysts, or even in individuals' tweets), sentiment proxies can be computed on a daily basis at the stock

*Email: coqueret@em-lyon.com

† Early results on the impact of these events on stock prices can be found in Beaver (1968) and Mitchell and Mulherin (1994).

‡ A debate is ongoing in the scientific community regarding the quantification of so-called 'significant' results and the propensity of research to resort to *p*-hacking. This debate highlights the need to replicate major findings and encompasses a wide variety of fields: Medicine and Psychology (Kerr 1998, Jager and Leek 2013), Economics (Brodeur *et al.* 2016, Camerer *et al.* 2016, Ioannidis *et al.* 2017), Finance (Harvey *et al.* 2016, Harvey 2017, Chordia *et al.* 2017, Hou *et al.* 2018, Bartsch *et al.* 2018, Rytchkov and Zhong 2018), Statistics (Kuffner and Walker 2019) and multidisciplinary fields (Ioannidis 2005, Head *et al.* 2015, Benjamin *et al.* 2018 and, very recently, Amrhein *et al.* 2019).

level, at least for large corporations. This granularity in the data has only been available for a decade or so. In their early study on sentiment, Brown and Cliff (2004) confess: '*Our choice of market aggregates is primarily driven by data limitations*'. Nowadays, the storage and computational capacities are such that these limitations have faded. This makes sentiment proxies good candidates for independent variables in daily predictive regressions at the stock level. Indeed, daily or intraday sentiment scores stand in contrast with classical stock-specific attributes, such as dividend yields or book values, which are typically updated every quarter.†

Given a broad (in the cross-section) and deep (chronological) set of prices and sentiment scores, it is computationally simple to automate batches of predictive regressions and analyse their outcomes. Heuristically, sentiment seems like a reasonable signal for the purpose of stock return forecasting. Unfortunately, our results do not support this intuition. One could argue that our work is antithetical to that of Novy-Marx (2014): we find no forecasting power for a plausible predictor, while Novy-Marx (2014) reports significant predictability with unconventional variables (e.g. the temperature of the Pacific Ocean). All in all, our conclusions echo the early work of Brown and Cliff (2004) and are likely to tone down, in some sense, the enthusiasm surrounding sentiment metrics and their impact on stock markets.

Our main contributions can be summarised as follows. First, to the best of our knowledge, our study on firm-level sentiment is the first large scope study of this kind, and it is based on third party data. Second, in spite of a large palette of tests and configurations, we find very little evidence of return predictability stemming from stock-specific sentiment. Only 7% of stocks have t-statistics above two in absolute value (5% level). When the threshold is set to 3, a much more conservative value (0.2% level), the proportion drops to less than 1% of stocks. These low figures do not improve considerably when altering predictors or return horizons, and the stocks that emerge as predictable are strongly model dependent. No one stock is consistently and robustly predictable across our numerous protocols. While this could be anticipated, it makes it hard to confidently single out stocks or forecasting models for investment purposes. In fact, a reverse interpretation of the news cycle suggests that returns can also drive sentiment and our Granger causality tests indicate that this hypothesis is the most likely.‡ Lastly, our large-scale study proved an ideal playground to implement recent advances in econometric tools such as pockets of predictability and Bayesianised p-values.

The remainder of the paper is structured as follows. Section 2 reviews published contributions related to sentiment in the financial literature. Section 3 presents the data and the construction of variables. Section 4 explores in-sample predictability relationships between sentiment metrics and future stock returns. Section 5 is dedicated to robustness checks and extensions. Finally, Section 6 concludes.

2. Related literature

First of all, the topic of investor sentiment has recently gained traction, both among practitioners and researchers. On the Financial Economics Network of the SSRN platform, approximately 100 related papers have been posted every year since 2015.§

Below, we review and attempt to classify a wide range of contributions in the field. For the sake of brevity, we cite published articles and choose to exclude working papers. Filtering out unpublished articles still leaves a large body of literature and we do not pretend that our survey is exhaustive.

There is no unambiguous definition of sentiment in financial economics. In their seminal article, Baker and Wurgler (2006) describe investor sentiment as an '*intrinsically elusive concept*' which is intimately linked to '*the propensity to speculate*'. Following the theoretical work of De Long *et al.* (1990), Baker and Wurgler (2007) and Zhou (2018) propose a convenient way to characterise sentiment: they define it as the factor that explains why asset prices diverge from their fundamental value. This generic definition leaves room for interpretation and the academic literature evaluates sentiment, in its broad sense, based on a wide range of sources:¶

- **Newspaper articles**. Sentiment can be extracted from news releases. Examples include *The New York Times* (Garcia 2013), the *Wall Street Journal* (Tetlock 2007, Tetlock *et al.* 2008, Dougal *et al.* 2012), the *Dow Jones News Service* (Chan 2003, Barber and Odean 2008, Tetlock *et al.* 2008), local newspapers (Engelberg and Parsons 2011), business outlets (Larsen and Thorsrud 2019) and wide-scope public news (Fang and Peress 2009, Griffin *et al.* 2011). Financial shows on TV have also been investigated in Busse and Green (2002). Combinations of market newsletters are used in Brown and Cliff (2005).
- **Internet sources**. They include: the content of internet posts or boards (Tumarkin and Whitelaw 2001, Antweiler and Frank 2004, Das and Chen 2007, Sabherwal *et al.* 2011, Chen *et al.* 2014, Leung and Ton 2015), the volume of Google queries (Da *et al.* 2011, 2015), Nasdaq webpages (Zhang *et al.* 2016), user-generated content, such as consumer reviews (Neal and Wheatley 1998) and micro-blogging websites such as Twitter (Bollen *et al.* 2011, Yang *et al.* 2015, Oliveira *et al.* 2017).

† Liquidity and option data are also available at daily frequencies, but they have already been explored by the literature. We refer for instance to Gallant *et al.* (1992) for the relationship between volume and price variations, to Baker and Stein (2004) for the connexion between bid-ask spreads and subsequent returns, and to Cremers and Weinbaum (2010), Xing *et al.* (2010), Bakshi *et al.* (2011) and Duan and Zhang (2013) for studies that link option data to future stock returns.

‡ Sentiment is rarely taken to be the dependent variable in economic analyses. One notable exception is Starr (2012) in which it is reported that news shocks impact sentiment, albeit at the aggregate level.

§ We counted those that mention the keyword '*sentiment*' in the title or in the abstract.

¶ We refer to Oliveira *et al.* (2017) and Zhou (2018) for complementary surveys. Note that given the amount of research on sentiment, it is hardly possible to be exhaustive when reviewing this topic.

- **The Baker and Wurgler index**. Baker and Wurgler (2006) introduce a measure of investor sentiment via a composite PCA-based index of six proxies that capture the macro-economic mood. Numerous articles work with this indicator (Yu and Yuan 2011, Chung *et al.* 2012, Mian and Sankaraguruswamy 2012, Hribar and McInnis 2012, Stambaugh *et al.* 2012, Cen *et al.* 2013, McLean and Zhao 2014, Antoniou *et al.* 2016, Chang *et al.* 2019, Kim *et al.* 2017, Shen *et al.* 2017, Chen *et al.* 2019) or modified versions thereof (Firth *et al.* 2014, Chang *et al.* 2015, Huang *et al.* 2015).
- **Other macroeconomic indicators**. Additional aggregate measures are also studied, such as customer confidence indices (Schmeling 2009, Antoniou *et al.* 2013, McLean and Zhao 2014), discounts on closed-end funds and net mutual fund redemptions (Lee *et al.* 1991, Neal and Wheatley 1998), and both the number of initial public offerings and monthly equity share in new issues (Da *et al.* 2014). Survey data is used in Solt and Statman (1988), Hengelbrock *et al.* (2013), Shefrin (2015) and Benhabib and Spiegel (2018). Finally, Brown and Cliff (2004) resort to a package of measures, encompassing the AAII survey, market newsletters, IPO statistics, liquidity, market performance, etc.
- **Extra-economic mood**. Aggregate markets are likely to react to psychological stimuli such as sunshine (Hirshleifer and Shumway 2003) or lack thereof (Kamstra *et al.* 2003), or to sport results (Edmans *et al.* 2007).
- **Special datasets**. Alternative data is increasingly popular among asset managers because it can provide informational advantage and cutting-edge insights.† Related analyses encompass proprietary datasets (Groß-Klußmann and Hautsch 2011, Sun *et al.* 2016), corporate financial disclosures (Jiang *et al.* 2019), returns on lottery-like stocks (Lutz 2016), aviation disaster data (Kaplanski and Levy 2010), Thomson Reuters proprietary sentiment (Yang *et al.* 2018), option data (Yang and Wu 2011) and retail trades (Kumar and Lee 2006).

Now, once sentiment is defined, it is usually exploited in three ways. The first one is conditioning. Researchers investigate whether patterns emerge in asset returns when sentiment is high or low. Examples include Brown and Cliff (2005), Baker and Wurgler (2006), Sabherwal *et al.* (2011), Yu and Yuan (2011), Mian and Sankaraguruswamy (2012), Stambaugh *et al.* (2012), Antoniou *et al.* (2013), Chang *et al.* (2015), Antoniou *et al.* (2016), Lutz (2016), Chang *et al.* (2019), Kim *et al.* (2017) and Shen *et al.* (2017).

The second purpose of sentiment is cross-sectional analysis. In this case, assets with favourable sentiment are assumed to outperform those with less favourable coverage. Positive results are found in Fang and Peress (2009) and Da *et al.* (2011), but no such evidence is uncovered in Das and Chen (2007). If the discrepancy between sorted portfolios is high enough, one can postulate the existence of a pricing factor related to sentiment.

Finally, the third major application is predictability: the sentiment proxies are used for forecasting purposes, very often in predictive regressions.‡ Lead-lag effects and Granger causality are discussed in Schmeling (2009) and Tirunillai and Tellis (2012). The impact on prices is studied in Busse and Green (2002), Chan (2003), Griffin *et al.* (2011), Groß-Klußmann and Hautsch (2011) and Hengelbrock *et al.* (2013) and the effect on volume of trading is evoked in Engelberg and Parsons (2011) and Zhang *et al.* (2016). Because one of our major interests lies in predictive regressions, we compile the results of 18 studies in table 1. In their study (section 4.1), Da *et al.* (2014) also perform lagged regressions, but their purpose is not to predict but rather to understand which factors drive return reversals. Hence, their study is not included in table 1.

One can underline that a prevalent pattern in this table is that at least one predictability relationship is found to be significant in each study, with the notable exception of Brown and Cliff (2004). In the macroeconomic literature, a few articles have sought to characterize the role of sentiment on aggregate fluctuations using elaborate econometric models. Among those, at least two studies (Barsky and Sims 2012, Fève and Guay 2018) report a very limited role for sentiment.

3. Data

3.1. Source and construction

The scope of the paper is restricted to corporations listed on US exchanges. The database was constructed in two stages. First, the daily opening and closing prices and sentiment metrics were downloaded from the Bloomberg platform.§ We obtained data for 4648 firms. In the second step, we filtered out those with insufficient coverage. The way we proceeded was the following: we built a coverage ratio equal to the number of days with well-defined sentiment data divided by the total number of days of data for the stock, namely

$$C_i = \frac{\#(\text{well-defined sentiment points for stock } i)}{\#(\text{days of data for stock } i)}. \qquad (1)$$

Then, we only kept those stocks with a sentiment coverage ratio above 20%, which is equivalent to having at least one data point per week, on average. The final dataset consists of

† Consequently, traditional data providers are facing competition on this segment due to newcomers such as Quandl and Neudata.

‡ A non-exhaustive list is: Neal and Wheatley (1998), Wang (2001), Antweiler and Frank (2004), Das and Chen (2007), Tetlock (2007), Tetlock *et al.* (2008), Schmeling (2009), Bollen *et al.* (2011), Dougal *et al.* (2012), Garcia (2013), Chen *et al.* (2014), Da *et al.* (2015), Huang *et al.* (2015), Lutz (2016), Zhang *et al.* (2016) and Oliveira *et al.* (2017).

§ The items are *PX_OPEN*, *PX_LAST* and *NEWS_SENTIMENT_DAILY_AVG*, respectively.

Table 1. *Summary of studies on sentiment predictability.* Both dependent and independent variables are briefly outlined. By lack of space, we do not report all results, only the most illustrative ones. We refer to the original articles for more details. We use the following abbreviations: ret. for returns, dep. for dependent, indiv. for individual, cond. for conditionally (right column) and pred. for predicability. SMB and HML stand for the classical Fama and French (1993) long-short portfolios, MS for Morgan Stanley, SA for Seeking Alpha, TRM for Thomson Reuters MarketPsych, NYT for New York Times and WSJ for Wall Street Journal. Finally, BW06 refers to the methodology of Baker and Wurgler (2006).

Article	Dep. Variable	Indep. Variable (Sentiment proxy)	Pred.
Neal and Wheatley (1998)	SMB and HML ret.	Discounts on closed-end funds	Yes
Neal and Wheatley (1998)	SMB and HML ret.	Net mutual fund redemptions	Yes
Neal and Wheatley (1998)	SMB and HML ret.	Odd-lot sales to purchase ratio	No
Wang (2001)	Agricultural future ret.	Large trader positions	Yes
Wang (2001)	Agricultural future ret.	Small trader positions	No
Antweiler and Frank (2004)	Aggregate ret. and volatility	Internet stock message boards	Yes
Brown and Cliff (2004)	Aggregate ret.	Wide range of indicators	No
Das and Chen (2007)	MS High-Tech index ret.	Internet stock message board	Yes
Das and Chen (2007)	Indiv. stock ret.	Internet stock message board	No
Tetlock (2007)	Dow Jones ret.	Negative WSJ reports	Yes
Tetlock *et al.* (2008)	Firm earnings	Negative words in news	Yes
Schmeling (2009)	Aggregate stock ret.	Consumer confidence	Yes
Bollen *et al.* (2011)	Dow Jones variations	Twitter mood	Yes
Chung *et al.* (2012)	Long-short portfolio returns	Baker and Wurgler (2006) index	Yes / Cond.
Dougal *et al.* (2012)	Dow Jones total ret.	Scheduling of WJS columnists	Yes
Garcia (2013)	Dow Jones total ret.	Positive/negative words in NYT	Cond.
Chen *et al.* (2014)	Indiv. stock abnormal ret.	Negative words in SA posts	Yes
Da *et al.* (2015)	Aggregate stock ret.	Internet search volume	Yes
Da *et al.* (2015)	Aggregate volatility (VIX)	Internet search volume	Yes
Huang *et al.* (2015)	S&P500 ret.	Partial Least Square version of BW06	Yes
Lutz (2016)	Indiv. & Agg. stock ret.	Return on lottery stocks	Cond.
Zhang *et al.* (2016)	Indiv. stock volatility	Negative terms in NASDAQ articles	Yes
Zhang *et al.* (2016)	Indiv. traded volume	Negative terms in NASDAQ articles	Yes
Zhang *et al.* (2016)	Indiv. stock ret.	Positive terms in NASDAQ articles	Yes
Sun *et al.* (2016)	S&P500 ret.	TRM Index	Yes
Oliveira *et al.* (2017)	S&P500 ret.	Twitter mood	Yes
Renault (2017)	S&P500 ETF ret.	StockTwits posts	Yes
Jiang *et al.* (2019)	Aggregate stock ret.	Tone of financial disclosures	Yes
Yang *et al.* (2018)	Intraday S&P500 ret.	Thomson Reuters news sentiment	Yes

1009 stocks. The list of all tickers is provided in table A1 in the Appendix. For more than 76% of these stocks, the starting point is January 2007, but for recent corporations, such as Facebook, the data may commence later, e.g. in May 2012. In a large majority of cases, the final point is November 2, 2017, but some firms were for instance acquired by other companies and their series end some time before that.† The distribution of first and last points for the sentiment metric is shown in the left graph of figure 1.

In terms of firm size, we provide the histogram of average market capitalisation in the right graph of figure 1. The largest one is Apple, with 400B$ and the smallest one is FBR & Co. with 267M$. Our sample is therefore not restricted to large companies, even though a sufficient news coverage (in terms of C_i) does require a minimal size. Overall, half of the firms in our sample have an average capitalisation *below* 7B$ and cannot hence be deemed *large*.

The sentiment score in Bloomberg is computed through the following steps. Based on a large and increasing amount of human-tagged news stories, a Machine Learning algorithm evaluates the company-related content of each new piece of information that is received in the Bloomberg databases. The algorithm assigns probabilities π_t^+, π_t^- and π_t^0 to each of the following three events: conditional on reading this information about a security, an investor having a long position in this security will be: bullish ($+1$), bearish (-1) or neutral (0). The raw sentiment metric is $\pi_t = \pi_t^+ - \pi_t^- \in [-1, 1]$. Given the continuous flow of information received, the real-time news sentiment is updated every minute and averaged over an 8-hour rolling window. The daily indicator we use is the daily average of the intraday sentiment indicator. In the sequel, we denote it with s_t^i, where the superscript i relates to the firms.

Sometimes, a company can spend 1 day or more with no news coverage, and hence the metric will not be defined.‡ In this case, we manually assign the previous available value to the corresponding missing data, meaning that sentiment during this period remained unchanged because no impactful piece of news has been published. The corresponding variable will be denoted with x_t^i. Note that for the sake of robustness, we will show that omitting the days during which sentiment is undefined has a very marginal impact on the results. Thus the bulk of our study will rely on x_t^i as the primary independent variable.

† Examples include Cameron International Corp., Health Net Inc., HomeAway Inc., and Youku Tudou Inc.

‡ For the largest companies, the flow of news is such that this is seldom a problem. For instance, between July 2015 and July 2017, the number of news stories analysed by Bloomberg for Apple Inc., JP Morgan Chase & Co. and the Boeing Company was equal to 861 243, 368 011 and 159 522, respectively. On average, this amounts to 1709, 730 and 317 stories per trading day.

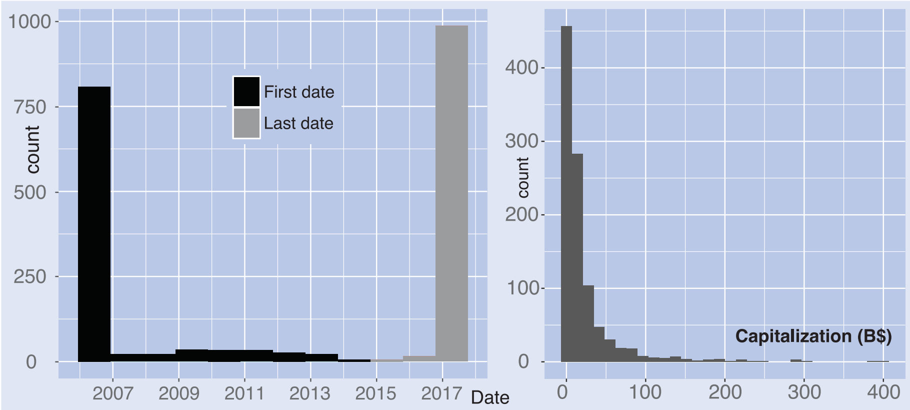

Figure 1. *Chronological range and average capitalisation*. On the left graph, we plot the distribution of first and last date for the time-series of the sentiment indicator. On the right graph, we show the cross-sectional distribution of average market capitalisation. The averaging is performed over all dates.

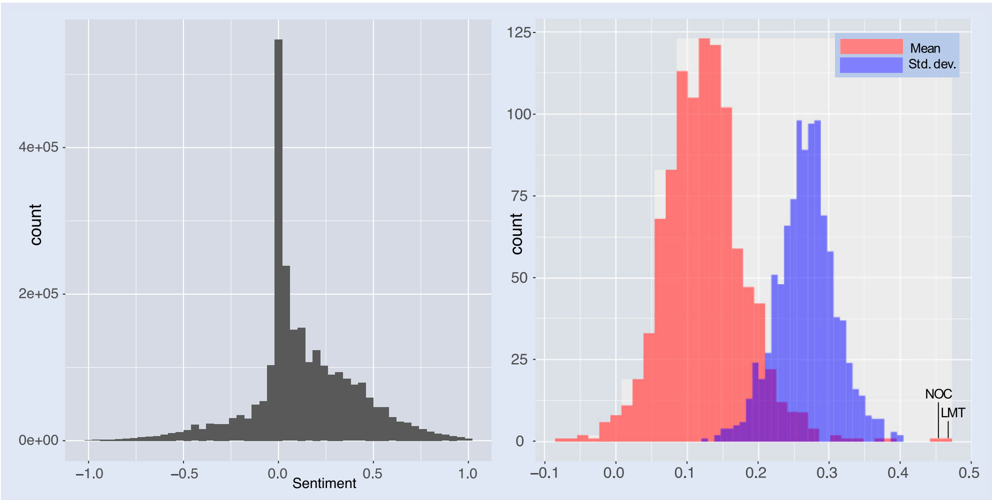

Figure 2. *Distribution of news sentiment*. On the left graph, we plot the histogram of all sentiment points x_t^i. On the right graph, we show the cross-sectional distribution of the averaged (across time) sentiment, as well as the corresponding standard deviation. The two firms with highest average sentiment are Northrop Grumman and Lockheed Martin. The width of bins is smaller for the 'Std. dev.' metric to ease readability.

Our data is sampled at the daily frequency. While many studies rely on monthly data to perform predictive regressions (e.g. Baker and Wurgler 2012), it is likely that nowadays the impact of news sentiment is incorporated into prices at a higher pace. For instance, Renault (2017) even reports significant predictability for aggregate intraday returns. Thus our baseline results will hold for daily regressions but we will also test the effect on longer horizon returns.

3.2. *Descriptive statistics*

In figure 2, we display the distribution of all sentiment values x_t^i (left graph) and the average values and standard deviations for each stock (right graph). The distribution of news sentiment appears skewed towards positive values. Two-thirds (67%) of the values are strictly positive. At the individual firm level, only 1.5% of stocks have an average sentiment below zero over the sample period.

Our baseline protocol relies on classical predictive regressions. The sentiment indicator s_t^i is released 10 minutes before the opening of the trading session, hence the most natural dependent variable is the excess return between the opening and closing of the trading day:

$$r_t^{i,OC} = \frac{P_t^{i,C}}{P_t^{i,O}} - 1 - r_t^f, \tag{2}$$

where $P_t^{i,O}$ is the opening price of stock i on day t and $P_t^{i,C}$ the corresponding closing price. For the sake of completeness, we will also consider a full day's excess return, from opening price to subsequent opening price:

$$r_{t+1}^{i,OO} = \frac{P_{t+1}^{i,O}}{P_t^{i,O}} - 1 - r_t^f. \tag{3}$$

This choice is rather unusual because returns are often computed from closing price to the next closing price. This choice is nonetheless dictated by the early release (in the morning) of the sentiment indicator.† In both equations, r_t^f is the daily rate of return of the US four weeks Treasury bill.

† Relatedly, following Berkman *et al.* (2012), Aboody *et al.* (2018) argue that individual investors often analyse news after work and

Table 2. *Descriptive statistics*: N stands for the number of observations and Q1 and Q3 for the first and third quartiles. The bar notation denotes the average over the time domain: $\bar{y}^i = (1/T)\sum_{t=1}^T y_t^i$.

	N	Min.	Q1	Median	Mean	Q3	Max.
s_t^i	999 704	−1.000	0.000	0.085	0.134	0.305	1.000
x_t^i	2 543 490	−1.000	0.000	0.075	0.127	0.295	1.000
$r_t^{i,OC}$	2 522 854	−0.724	−0.009	0.000	0.000	0.009	1.500
$r_t^{i,OO}$	2 475 910	−0.776	−0.011	0.000	0.001	0.011	2.338
$\bar{s}^i$	1009	−0.107	0.092	0.128	0.130	0.164	0.459
$\bar{x}^i$	1009	−0.082	0.088	0.121	0.124	0.156	0.459
$\bar{r}^{i,OC}$	1009	−0.004	0.000	0.000	0.000	0.000	0.006
$\bar{r}^{i,OO}$	1009	−0.003	0.000	0.000	0.000	0.000	0.003

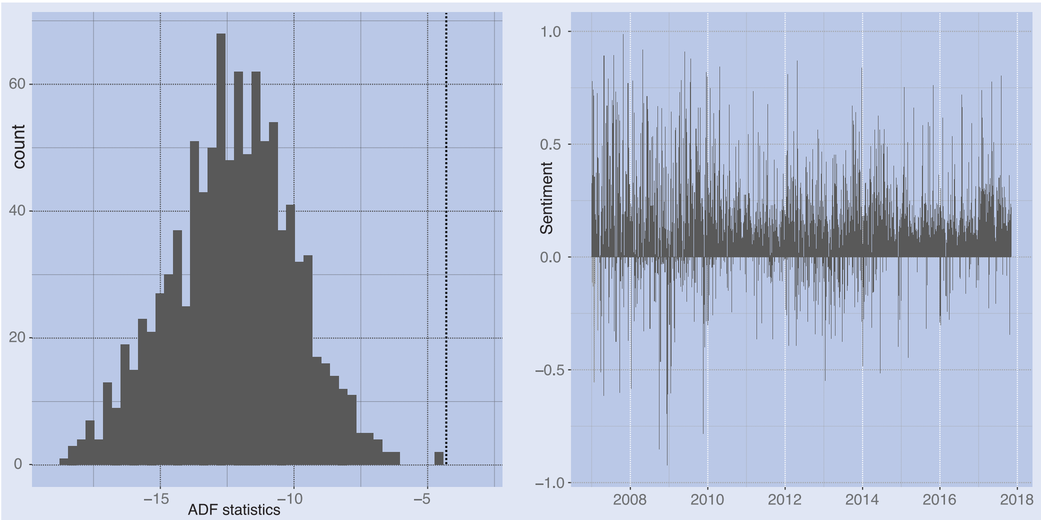

Figure 3. *Distribution of ADF statistics and sample sentiment plot*. On the left graph, we plot the histogram of the ADF statistics in the cross-section. The test is performed with constant and linear trend and the number of lags is equal to 6. The null hypothesis assumes the presence of a unit root. The vertical dotted line on the right corresponds to the 0.01 significance level (i.e. −3.96, see e.g. Hamilton 1994, Table B.6 Case 4). On the right graph, we plot the sentiment indicator for the AAPL (Apple Inc.) ticker.

In the first batch of regressions, we test both s_t^i and its completed version x_t^i as independent variables, but in the subsequent sections, we will focus on x_t^i solely because the results are not altered when switching from one to the other. Our baseline models are therefore fully specified by the following equations:

$$r_t^{i,OC} = \alpha^{i,OC} + \beta^{i,OC} y_t^i + \epsilon_t^{i,OC}, \quad y \in \{s, x\}, \tag{4}$$

$$r_{t+1}^{i,OO} = \alpha^{i,OO} + \beta^{i,OO} y_t^i + \epsilon_t^{i,OO}, \quad y \in \{s, x\}. \tag{5}$$

In line with many studies on predictability, we work with only one regressor.† Now that all variables are defined, we provide in table 2 their descriptive statistics, as well as those of their averages computed over the time domain.

Finally, and because the main objective of the paper is to run the predictive regressions (4) and (5), we must characterise the econometric properties of the regressors. Indeed, as is underlined in Stambaugh (1999), predictive regressions with autocorrelated predictors lead to biased OLS estimates.

Accordingly, we plot the statistics of the Augmented Dickey-Fuller (ADF) test with constant and linear trend on the left panel of figure 3. As is shown, the values are mostly far to the left of the 0.01 significance level, hence we can reasonably infer that stationarity is not much of an issue with these series. In order to illustrate this stationarity, we plot on the right graph of figure 3 a sample sentiment series, namely the one for the AAPL ticker (Apple Inc).

3.3. *Discussion on the choice of a third party signal*

In the literature reviewed in Section 2, the computation of sentiment often relies on complex derivations over macroeconomic indicators (e.g. Baker and Wurgler 2006 and the related articles) or on textual analysis (e.g. Das and Chen 2007, Zhang *et al.* 2016, Renault 2017 to cite but a few). Because we work at the company level, the first type does not apply, thus we must resort to the second type of methods.

The choice of Bloomberg is critical and dictated both by chronological depth and by access to news feeds. Bloomberg is one of the very rare data providers that have kept track of news sentiment prior to 2010 at the firm level. In addition, the ability of high-profile data providers to collect and handle immense volumes of data is unmatched. Proprietary datasets

are likely to place orders during the night. They find that overnight returns can serve as proxy of sentiment.

† The following articles do so: Welch and Goyal (2007), Dangl and Halling (2012), Johannes *et al.* (2014), Pettenuzzo *et al.* (2014) and Huang and Zhou (2017).

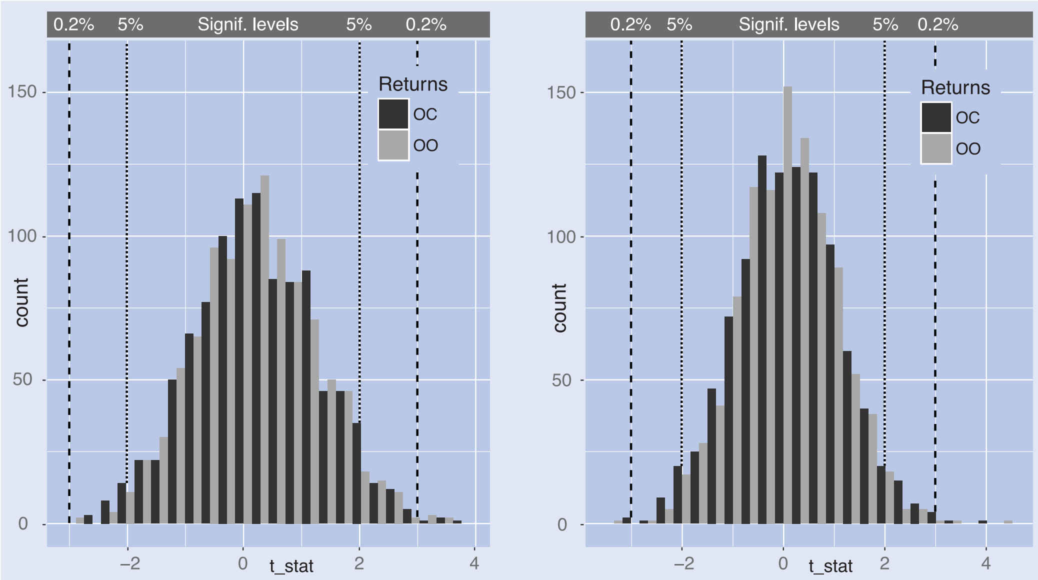

Figure 4. *Baseline results*. On the left graph, we plot the histogram of the t-stats of models (4) in black and (5) in grey when s_t^i is taken as the independent variable. The right graph plots the same statistic when x_t^i is taken to be the independent variable. The important thresholds (5% and 0.2% level) are highlighted with vertical dashed lines.

have much smaller scales. In their study, Zhang *et al.* (2016) work with 116 691 points of data, collected for 100 very large firms over 1255 days. This amounts to only *one* piece of news per company per day, on average. When screening data for large firms, Bloomberg scrutinises several hundreds (sometimes, thousands) pieces of news every day for each company. Such large volumes of data reduce the bias towards particular information sources and are likely to increases the pertinence of the daily sentiment indicator. In addition, the data is easily accessible, so that our results can be readily be replicated.

4. Predictability with sentiment

4.1. Baseline results

Given the large number of regressions (i.e. one per company) that we need to run, it is useful to resort to synthetic ways to present results. Because we aim to unveil the significance of the predictability relationship, we will essentially report the distribution of t-stats and omit the estimates.† There are alternative measures that characterise the performance and goodness-of-fit of regressions (e.g. the R^2), and we discuss this extension in Section 5.5. The present section is focused on in-sample results. Out-of-sample protocols are presented in the next section.

In figure 4, we plot the histograms of the t-stats corresponding to the beta estimates of models (4) and (5). The most important indicator is the proportion of stocks that reach the significant two-sided thresholds: two (at the 5% level) and three (at the 0.2% level).‡ We report these values below in Panel A of table 3. The figures are unambiguous: if the threshold is conservative (i.e. equal to 3), then less of 1% of the stocks show predicability relationships. If the threshold

Table 3. *Metrics for baseline results*: In Panel A, we report, for each batch of regressions, the proportion of t-stats that are above two or three in absolute value. In Panel B, we display the correlation of t-stats between each batch of regressions. Finally, in Panel C, we show the average proportion of significant t-statistics when regressing excess returns on 10 batches of i.i.d. random variables with normal distributions $\mathcal{N}(\mu = 0, \sigma = 1)$ and $\mathcal{N}(\mu = 0.12, \sigma = 0.28)$. The latter parameters correspond to an *average* sentiment trajectory, as depicted in the right graph of figure 2. The averages in Panel C are computed over the 10 batches.

PANEL A: Proportion of significant relationships				
Dependent variable	$r^{i,OC}$		$r^{i,OO}$	
Independent variable	s^i	x^i	s^i	x^i
Proportion of $\lvert t\rvert > 2$	0.072	0.064	0.048	0.035
Proportion of $\lvert t\rvert > 3$	0.004	0.004	0.006	0.004
PANEL B: Correlations of t-stats				
Dependent variable	$r^{i,OC}$		$r^{i,OO}$	
Independent variable	s^i	x^i	s^i	x^i
Dep./Indep. variable				
$r^{i,OC}/s^i$	1.000	0.608	0.808	0.489
$r^{i,OC}/x^i$		1.000	0.475	0.774
$r^{i,OO}/s^i$			1.000	0.611
$r^{i,OO}/x^i$				1.000
PANEL C: Random noise				
Dependent variable	$r^{i,OC}$		$r^{i,OO}$	
Independent variable	Random $\mathcal{N}(\mu = 0, \sigma = 1)$ noise			
Avg. Prop. of $\lvert t\rvert > 2$	0.042		0.044	
Avg. Prop. of $\lvert t\rvert > 3$	0.002		0.002	
Dependent variable	$r^{i,OC}$		$r^{i,OO}$	
Independent variable	Random $\mathcal{N}(\mu = 0.12, \sigma = 0.28)$ noise			
Avg. Prop. of $\lvert t\rvert > 2$	0.046		0.044	
Avg. Prop. of $\lvert t\rvert > 3$	0.002		0.003	

† This is not uncommon in the literature, see, e.g. Moskowitz *et al.* (2012).

‡ The 5% level is often associated with the more conservative 1% level. Following the recommendations of Harvey *et al.* (2016), we deliberately impose an even higher hurdle as our second choice of significance level.

is 2, then no more than 7% of firms see their future returns significantly driven by the sentiment indicator.

For the sake of naive comparison, we create a benchmark indicator which tracks the number of significant t-stats when regressing the excess returns against white noise. The noise is generated from two types of Gaussian distributions: the first is purely $\mathcal{N}(0, 1)$ distributed while the second one has a mean and variance that matches those of x_t^i. The corresponding proportions of significant t-statistics are gathered in Panel C of table 3. A simple rule of thumb is that resorting to random predictors yields 5% of significant t-statistics. This figure remains stable over several batches of simulations. We deduce that sentiment predictors increase this proportion only marginally.

A surprising feature of the four histograms is their apparent symmetry around zero. Heuristically, we should expect future returns to be positively impacted by the sentiment metric: high (*resp*. low) sentiment fuels positive (*resp*. negative) returns. Under this assumption, t-statistics should be scattered in the vicinity of some number substantially to the right of the origin. The fact that sentiment indices often negatively impact future returns across many firms is a second indication that news sentiment is unlikely to be a coherent and influential driver of asset prices.

In the subsequent sections, we investigate whether this overall lack of predictability holds when some specifications of the models are changed. Given the multiplicity of possibilities, we reduce the degrees of freedom by focusing solely on x_t^i as independent variable, thus excluding s_t^i. This is easily justified by the redundancy of the data in both variables which yields very similar outcomes (see Panel B of table 3). In the four batches of regressions above, the correlation between the t-stats obtained with s_t^i and those obtained with x_t^i is equal to 0.61 and this correlation is highly significant (the p-value of the corresponding linear model is smaller than 10^{-16}). We highlight that the correlations are even higher if we compare the dependent variables. This is confirmed by figure 4: the grey and black histograms are hardly distinguishable and their supports are very similar. As a consequence, we do not expect that switching from open–close returns to open–open returns will sharply alter the distribution of t-stats. Henceforth and unless specified otherwise, we will work with the open–close version of returns.

Before we turn to robustness checks, we perform a new set of regressions in which the dependent variables are returns computed over longer periods. The rationale behind this choice is simply that sentiment may take longer than simply 1 day to diffuse into prices. To this purpose, we introduce the following excess returns:

$$r_{t,t+k}^{i,OC} = \frac{P_{t+k}^{i,C}}{P_t^{i,O}} - 1 - r_{t,t+k}^{f}, \tag{6}$$

where $r_{t,t+k}^{f}$ is the return of the US T-Bill over the t to $t+k$ period. In a week without holidays, $k = 4$ corresponds to a weekly return: from Monday morning to Friday evening, for instance. Likewise, $k = 20$ is a good proxy for monthly returns. We now consider the predictive regression below

$$r_{t,t+k}^{i,OC} = \alpha_k^{i,OC} + \beta_k^{i,OC} x_t^i + \epsilon_{t,k}^{i,OC}, \tag{7}$$

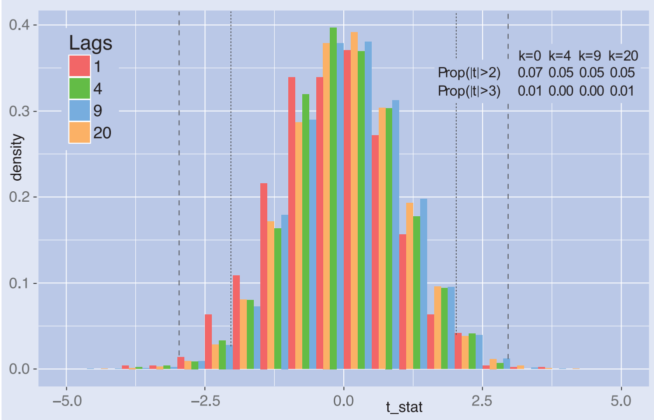

Figure 5. *Longer horizons*. We plot the distributions of t-stats related to the estimates of β_k^{OC} in (7), for $k \in \{0, 4, 9, 20\}$. To the right of the histogram, we compute the proportion of these t-stats that are larger than 2 or 3 in absolute value. These thresholds are represented with the dashed vertical lines.

and the distribution of t-stats related to $\beta_k^{i,OC}$ is shown in figure 5 for $k = 0, 4, 9, 20$.† The case $k = 0$ is simply the base case. To the right of the histogram, a short table displays the proportion of t-statistics larger than 2 or 3, in absolute value. The main pattern we observe is that choosing longer lags for returns does not significantly alter the shape of the distribution of the t-stats. The proportion of outstanding t-stats remains low: no more than 1% of them are larger than 3.

4.2. *Data snooping*

Our baseline results do not seem to support the assumption that sentiment, at the stock level, predicts future returns. This section is dedicated to the minimal (required) robustness checks: we investigate if the predictability relationships can be boosted by changes in the empirical protocol.

One of most the common p-hacking techniques is to rerun experiments on hand-picked subsets. In our framework, two dimensions naturally stand out: the cross-section of firms and the breadth of time-series. A legitimate argument in favour of filtering or sorting firms is that the ability of sentiment to predict a firm's future returns is contingent on the reliability of the sentiment thread (and hence on the coverage of this firm). We would expect predictability to be more salient if we considered firms with superior coverage. Similarly, the results might improve if we restrict the sample to recent dates because the quality of coverage improves in time.‡

† Technically, for each stock i, k regressions are run so that the returns in (7) do not overlap. Overlapping returns must be avoided to ensure that both the dependent variable and the residuals are not autocorrelated. For example, when $k = 20$, each firm will generate 20 t-stats and the blue histogram relates to all of these values. The k regressions correspond to the first k dates that are used to compute the first returns: for a starting point j, all returns are computed over $j + mk$ and $j + (m + 1)k$ for $m = 0, 1, \ldots$.

‡ A simple argument is that technology advances allow sentiment providers to process news more effectively. Notably, the sentiment metrics nowadays embed a growing variety of data sources (news articles, blog posts, social network content) and support (text, image, sound (e.g. podcast) and video). Furthermore, the increasing amount

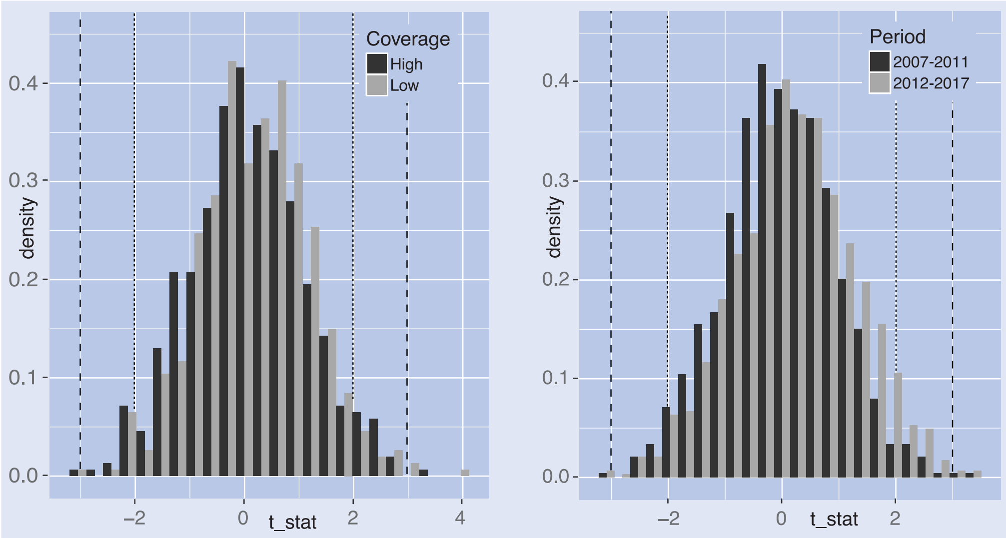

Figure 6. *Cross-sectional and chronological sorting*. We plot the distributions of t-stats related to the estimates of β^{OC} in (4). In the left graph, firms are sorted according to their coverage ratio (1). In the right graph, the regressions are run on two subsets of dates: those before January 2012, and those after.

The first dichotomy we test relates to coverage. Each stock can be characterised by its coverage ratio, as computed in (1). We thus sort firms into two groups: those with above-median coverage and those with below-median coverage.† We then run the predictive regressions (4) with $y = x$ (i.e. we use the completed version of the data). The distribution of the corresponding t-statistics is displayed in the left part of figure 6. The grouped histograms highlight the lack of difference between the two groups of firms. Among those with high (*resp*. low) coverage, 7.3% (*resp*. 5.6%) have a t-statistic above two in absolute value. In sum, it appears improbable that coverage be exploited to artificially improve significance in predictive regressions.

The second splitting variable is time. For each firm, we split the data into two: the dates (strictly) before January 2012 and those after this point. If there are more than 500 well-defined data points for returns and sentiment, then the predictive regression (4) is run (500 points amount to approximately two years). The distribution of the related t-statistics are shown on the right graph of figure 6. We notice a small shift between the two histograms: the t-stats related to the most recent period average to 0.21 while those computed before 2012 average to -0.01. This is encouraging because we expect the relationship to be indeed positive. Moreover, the proportion of statistics larger than two in absolute values increases from 3.9% (before 2012) to 7.3% (after 2012). This increase in predictability in the most recent years are possibly linked to the ever expanding scope of news coverage and to improvements in the processing of news flows. Nevertheless, in unreported results, we find that restricting the sample to the 2015–2017 period does not further improve these figures. Overall, it seems that the proportion of significant relationships can only be moderately boosted when restricting the study to the most recent periods.

The third variable that could impact our results is firm size. Indeed, Kumar and Lee (2006) report that small firms' returns are more sensitive to retail investor sentiment. Thus we split the sample into four groups of homogeneous market capitalization using the quartiles of average market capitalization across the whole sample. The small group comprises the firms with average market capitalization below 3.3B\$, the medium-small group has capitalization between 3.3B\$ and 8.2B\$, the medium-large one between 8.2B\$ and 22B\$ and the large one above 22B\$. The distribution of corresponding t-stats are shown in figure 7.

One takeaway from the plot is that the t-stats from the small firms (average value equal to 0.24) are more skewed to the right compared to the large firms (average t-stat equal to 0.08). This signals, on average, a more positive relationship between sentiment and future returns for small firms. Nevertheless, the discrepancy is not very pronounced and cannot be deemed significant.

Lastly, for the sake of completeness, we combined the two sorting variables and ran the regressions for firms with high coverage on the data spanning from 2012 to 2017. The proportion of t-statistics larger than 2 in absolute value was equal to 6.7%. If we consider the 250 firms with highest coverage over a period from 2015 onwards, the proportion reaches 7.7%. This modest figure summarises the findings of this section and underlines that consistent filters and subgroups do not generate incremental significance in the predictability relationship.

5. Robustness checks and further analyses

5.1. *Change in sentiment*

Whenever a relationship between two variables fails to hold, a classical adjustment is to perform tests on modifications of one (or both) of the variables. The analysis of variations instead of levels is arguably one of the most widespread

of human-tagged news leads to a refinement of the classification algorithms.
† Given that large firms tend to get more coverage, this sorting is a good proxy for a small company versus large company dichotomy.

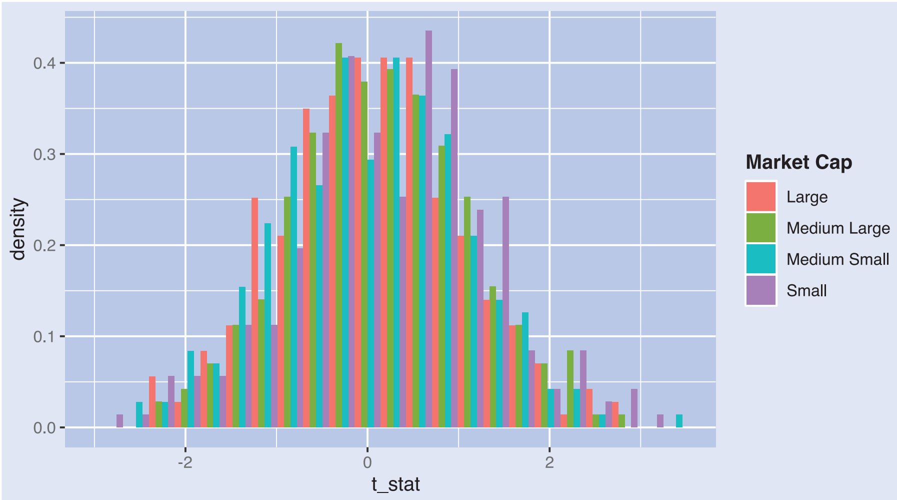

Figure 7. *The impact of firm size*. We plot the distributions of t-stats related to the estimates of β^{OC} in (4). The groups are the following: small (cap $<$ 3.3B\$), medium-small (3.3B\$ $<$ cap $<$ 8.2B\$), medium-large (8.2B\$ $<$ cap $<$ 22B\$) and the large (cap $>$ 22B\$).

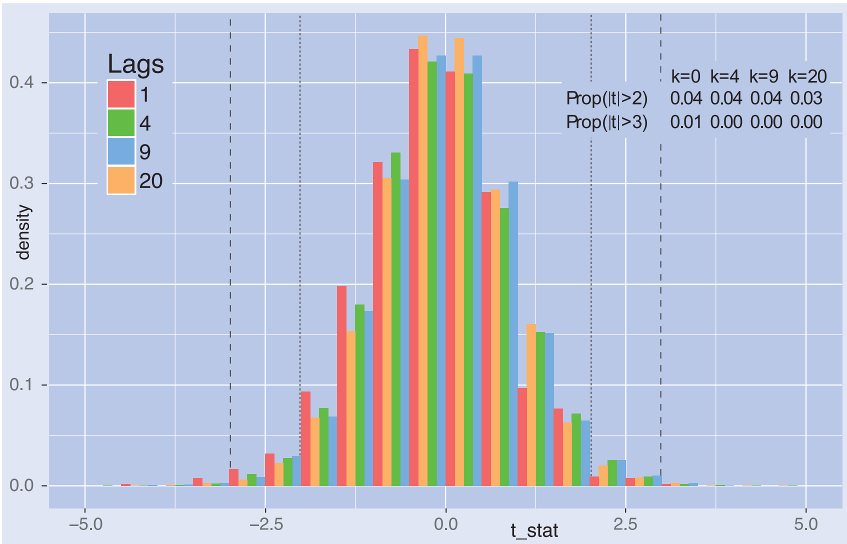

Figure 8. *Differences as predictors*. We plot the distributions of t-stats related to the estimates of β_k^{OC} in (8), for $k \in \{0, 4, 9, 20\}$. To the right of the histogram, we compute the proportion of these t-stats that are larger than 2 or 3 in absolute value. These thresholds are represented with the dashed vertical lines.

way of complementing the analysis and increasing the odds of uncovering outstanding results. Variations are furthermore appealing because they are more prone to display stationarity properties when working with time-series. In some case, this choice is justifiable since variations in the independent variable signal updates in the data that are potentially more valuable in explaining the dependent variable. This study is such one case because it can be debated that it is indeed the *changes* in sentiment (arising from the content of recent news) that will impact prices and not the level of sentiment itself. Accordingly, we introduce the corresponding model:

$$r_{t,t+k}^{i,OC} = \alpha_k^{i,OC} + \beta_k^{i,OC} \Delta x_t^i + \epsilon_{t,k}^{i,OC}, \qquad (8)$$

with $\Delta x_t^i = x_t^i - x_{t-1}^i$. Similarly to our baseline results, we test four horizons for the computation of the returns. The distribution of the corresponding t-statistics is shown in figure 8. The proportions of significant estimates remains very low, in fact, even lower than in the base case (figure 5).

5.2. *Macro-economic conditioning*

In the palette of robustness checks in Finance, conditioning protocols are commonplace. Their justification is theoretically rooted in the variation of agents' preferences throughout the business cycle: intuitively, investors are more risk-averse during recessions. Empirically, a handful of studies have validated the connection between economic conditions and the performance of prediction models. Henkel *et al.* (2011) dynamically assess the predictive power of the dividend yield and the rate of the 3 month US Treasury bill over the aggregate market returns in the G7 countries. They find that the adjusted R^2 is significantly higher during recessions, implying that returns are more predictable in the midst of economic contractions. Similarly, using combinations of forecasts, Rapach

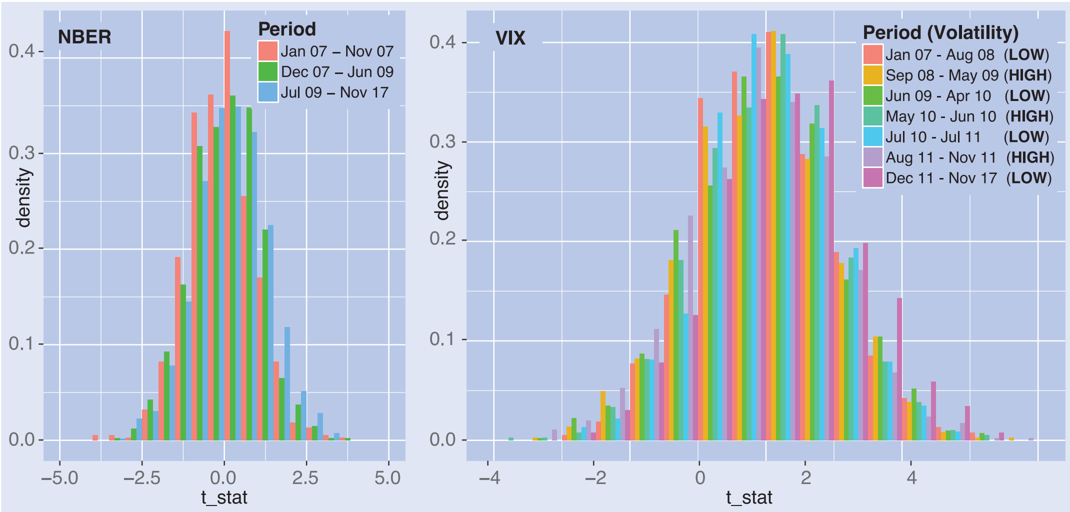

Figure 9. *Conditional results*. We plot the distributions of t-stats related to the estimates of β^{OC} in (4). The regressions are run on different periods. In the left graph, periods are determined by NBER cycles. In the right graphs, they correspond to high and low volatility periods for the aggregate market.

et al. (2010) find that predictability increases subsequently to NBER troughs and decreases after peaks.

In the spirit of these findings, we perform our baseline regressions on two sets of periods that reflect antithetic economic conditions. The first set of periods is related to NBER cycles. There are two splitting points in our sample: the peak of December 2007 and the trough of June 2009. We thus ran the model (4) on the three corresponding periods and the distribution of t-statistics is represented on the left graph of figure 9. Similarly to our previous results, all three histograms exhibit very light tails and the proportions of t-statistics larger than 2 in absolute value are all smaller than 7%. We nevertheless acknowledge a small shift from the red to the blue distribution, which means that on average the t-statistics related to the recent period are slightly larger, though not significantly. This echoes the findings of Section 4.2 and the right histogram of figure 6 in particular.

The second set of periods is driven by volatility. We use a smoothed version of the VIX index (via a moving average) to separate times of turbulence from times of relative calm on the equity markets. We find seven periods and run the regression on each one of them (the periods are graphically exhibited in the appendix, in figure A1). The distribution of the corresponding t-statistics is plotted in the right graph of figure 9. In line with our earlier conclusions, the histogram related to the latest sample, in pink, is clearly to the right of its counterparts, signalling a higher proportion of *positive* predictability relationships. Nonetheless, the proportion of significant results (with a threshold of 2 for the absolute t-statistics) is fairly stable across cycles: it oscillates between 4% and 7%. In any case, these modest values indicate that sentiment is a decent predictor of future returns only for a small proportion of firms.

5.3. *Controlling for market-wide sentiment*

All the sentiment series used in the analyses so far are stock-specific. Nonetheless, it may very well be the case that these metrics are correlated with a global, aggregated, version of sentiment. Hence, it can be worthy to remove the market-wide component of sentiment series before running the predictive regressions. The simplest way to achieve this is to resort to an initial regression:

$$x_t^i = a + bz_t + e_t^i, \tag{9}$$

where x_t^i is the raw stock-specific sentiment and z_t is the aggregate sentiment proxy. Thus the estimated errors (residuals) $\hat{e}_t^i = x_t^i - \hat{a} - \hat{b}z_t$ can be viewed as the raw sentiment in which the macro component has been removed. The new model then posits the relationship

$$r_{t+1}^i = \alpha^i + \beta^i \hat{e}_t^i + u_t^i \tag{10}$$

and seeks to explain future returns with the filtered sentiment $\hat{e}_t^i$. We use three proxies for market wide sentiment: the original BW series from Baker and Wurgler (2006),† the capitalization-weighted average of the sentiment series of our study‡ and the VIX. This last choice is motivated by the strong link between volatility and sentiment (see Wang *et al.* 2006, Baker and Wurgler 2007).

The distribution of t-stats stemming from regressions (10) is plotted in figure 10. The three histograms are very similar but do not stand out compared to our previous results. In all three cases, only 3–5% of the stocks are related to p-values smaller than 5%.

5.4. *Enhanced statistical tests*

It is known at least since Elliott and Stock (1994) that the econometric properties of regressors in predictive models can have an impact both on ordinary OLS estimates and on the related statistical tests. Several methods have been devised to overcome this issue and we point to Liu *et al.* (2019) and to

† Downloaded from http://people.stern.nyu.edu/jwurgler/data/Investor_Sentiment_Data_20190327_POST.xlsx
‡ Equally weighted series yield the same qualitative conclusions: the method used to aggregate sentiment series does not really matter.

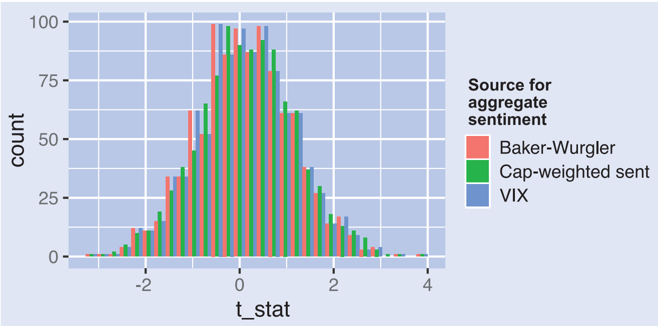

Figure 10. *Controlling for aggregate sentiment*. We plot the distributions of t-stats related to the estimates of β^i in (10). The dependent variable is the filtered sentiment estimated in equation (9).

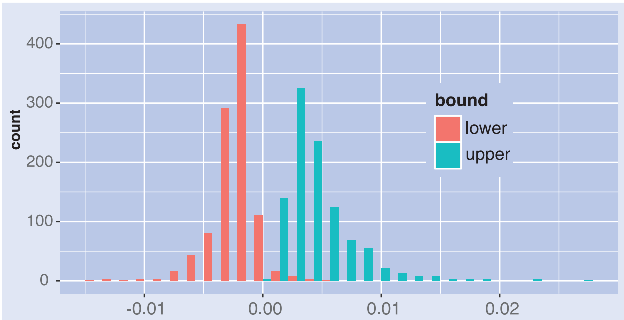

Figure 11. *Bounds for the Bonferroni Q-test*. We plot the distributions of lower and upper bounds of the confidence interval of estimates in the base case predictive regressions. The confidence level is 95% and the procedure is that of Campbell and Yogo (2006).

the references therein for an exhaustive account on this topic. Accordingly, we carry out the Bonferroni Q-test introduced in Campbell and Yogo (2006). The output is a confidence interval for the estimate of the β coefficients. The bounds of the interval are computed according to equations (14), (15) and (16) in Campbell and Yogo (2006).

As is done in Gao and Yang (2017), we compute the 95% confidence intervals for $\rho = 1$. We plot the distribution of lower and upper bounds in figure 11. The interesting pattern is that the bulk of lower bounds are to the left of zero and all upper bounds are to the right. This means that zero lies inside most of confidence intervals. In fact, out of the 1009 stocks, 62 only (6%) have confidence intervals that do not overlap with the origin. This figure is consistent with our findings so far and highlight that even robust tests fail to unveil a significant link between sentiment and future returns.

5.5. *Alternative performance metrics*

The results that have been reported so far focus only on t-statistics because our primary goal is to determine if the predictability relationship between sentiment and returns is strong. Most studies on predictability choose to report the R^2 (or its adjusted version) of the regression because it is the most straightforward indicator of goodness-of-fit.

In figure 12, we show the distributions of the R^2 of the regression model (7). A notable pattern emerges: longer horizons are associated to larger values of R^2. This is likely a corollary of the higher variance of the dependent variable.† Moreover, this is in line with many studies that find that predictability is more pronounced for large forecasting horizons.

The topic of testing the statistical significance of the R^2 is highly technical and we refer to Huang and Zhou (2017) for more details on the matter. Thus formal asset pricing-based tests at the stock level are out of the scope of this paper. Nonetheless, we will use a simple rule of thumb to coarsely assess the proportion of stocks with effective goodness-of-fit. For simplicity, let us consider weekly returns (lag $= 4$). In Ross (2009), the theoretical upper bound at the aggregate level (market returns) is 1.9%. The proportion of R^2 larger than this value in our results is equal to 0.4%. With a much less conservative specification based on some exogenous state variable (the consumption growth rate), Zhou (2010) recommends to apply a multiplicative discount of 0.1 or 0.15 on the bound of Ross (2009). In our regression outputs, 19.8% of the R^2 are larger than 1.9% $\times$ 0.15. Similar order of magnitudes are found for the other horizons (daily, monthly).

Finally, the right graph of figure 12 shows the strong positive connexion between the two predictability metrics: the absolute value of the t-statistic (x-axis) and the R^2 (y-axis). When looking at large absolute t-statistics, the relationship exhibits a quadratic-like pattern. Plainly, regressions associated with small absolute t-statistics are unlikely to yield outstanding R^2.

5.6. *Bayes to the rescue?*

In light of the debate on statistical significance spurred by Harvey (2017), we investigate the impact of prior odds on the likelihood that sentiment indeed predicts future returns. Harvey (2017) proposes a transformation which combines the t-statistic and prior odds into a Bayesianised p-value (Bpv). This value can then be interpreted as the probability that the null hypothesis is true. This last statement is typically false for the raw p-values. The formula is the following:

$$\mathrm{Bpv}(t, \mathrm{odds}) = \frac{e^{-|t|^2} \times \mathrm{odds}}{1 + e^{-|t|^2} \times \mathrm{odds}}. \tag{11}$$

The prior odds take the form x:1. When $x = 1$, then odds are even and it is assumed that the null has a 50% chance of being true. The prior probability that the null is true is equal to $x/(1 + x)$, hence if x is larger (*resp*. smaller) than 1, this probability is larger (*resp*. smaller) than 50%. In the present case, we are interested in small values of this probability because the predictability relationship can only hold if the null hypothesis is rejected.

In figure 13, we plot the cumulative distribution of the Bayesianised p-values for different values of prior odds. The

† Limiting values of R^2 are investigated in Ross (2009), Zhou (2010), Huang and Zhou (2017). In these articles, the upper bounds on R^2 are linked to the investment horizon through the risk-free rate and the stochastic discount factor. These bounds are indeed expected to increase with data latency. For instance, Huang and Zhou (2017) report higher values of R^2 for quarterly regressions compared to monthly regressions. Furthermore, Ross (2009) provides an upper bound of 0.25% for daily market returns and 1.9% for weekly market returns.

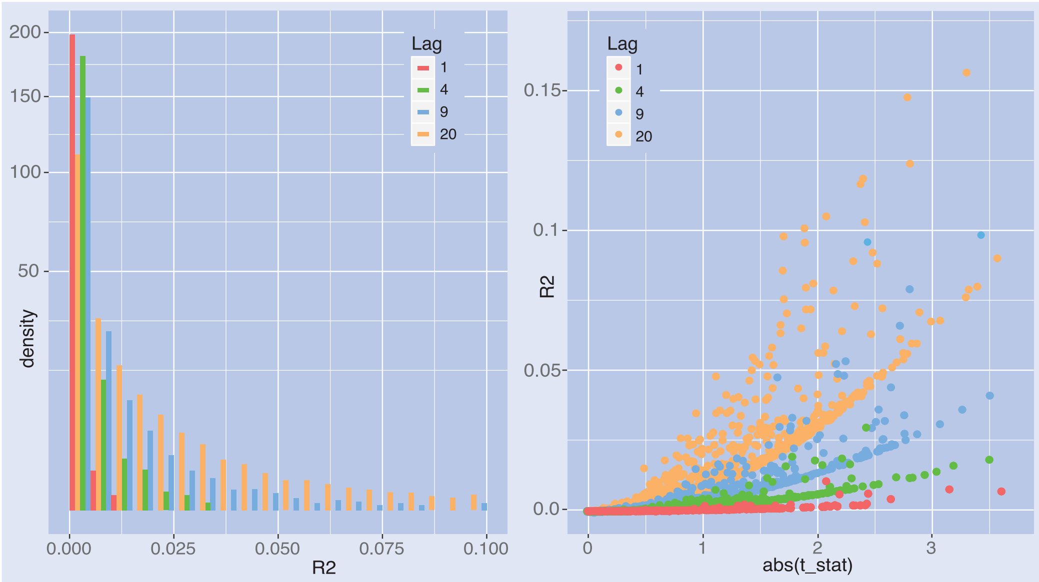

Figure 12. *R-squared*. On the left graph, we plot the distributions of R^2 stemming from the regression defined in (7). The y-axis has a highly nonlinear scale. On the right graph, the R^2 is shown as a function of the absolute value of the t-statistic related to the estimates $\hat{\beta}_k^{i,OC}$.

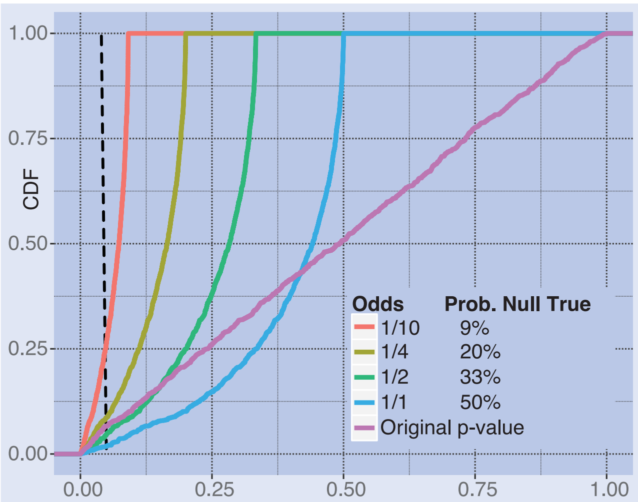

Figure 13. *Distribution of Bayesianised p-values.* We plot the distribution of Bpv defined in (11) with input t-statistics stemming from the baseline regressions (4). The dashed vertical line shows the 5% significance threshold.

original t-statistics are those from the baseline regression (4). We only consider the cases in which the odds are smaller than one, that is, when they are favourable to the assumption that sentiment does drive future returns. For comparison purposes, we also plot the distribution of the original p-values (in pink). This latter distribution is very close to uniform. Applying the Bayesian transform yields very different patterns for the probability of rejecting the null. It can be verified that the ratio of odds is the upper bound on the p-values.

We are interested in the amount of these values that are smaller than 5% (i.e. for which we can confidently reject the null hypothesis). Such proportion can be evaluated given the vertical dashed line. Unfortunately, under the most favourable odds for sentiment predictability (1/10:1), the proportion of probabilities that support the predictive relationship is equal to 26%. Hence, even with a very strong prior confidence in the ability of sentiment to predict future returns, the results to not support a substantial predictability relationship between news sentiment and stock returns.

5.7. *Pockets of predictability*

In a recent study, Farmer *et al.* (2018) show that the ability of classical predictors to forecast future aggregate returns is not constant through time, but rather clustered in windows of 20–400 days. They use local nonparametric estimators to show that both performance metrics (t-statistics and R^2) are highly time dependent. It is thus a logical extension to check whether this phenomenon holds true at the stock level with sentiment predictors. In line with Farmer *et al.* (2018), we run batches of regressions, on a stock-by-stock basis according to the protocol detailed below.

For computational purpose and without much loss of generality, we simplify the local non-parametric estimation procedure of Farmer *et al.* (2018). Because they show that local Taylor expansions yield no incremental insight, we stick to the local Nadaraya–Watson weighted method with the asymmetric Epanechnikov kernel:

$$K(x) = \frac{3}{2}(1 - x^2)\mathbf{1}_{\{x \in (0,1)\}}.$$

The reason for excluding the symmetric kernel is that we seek truly predictive relationship and not the sole value of the regression slope. Thus we must rely on past data only. The rolling windows have a size of 120 days, which amounts to roughly 6 months. We refer to chapter 12.3 in Campbell *et al.* (1997) for details on kernel regressions. For each stock, we obtain a continuum of estimates $\hat{\beta}_t^i$ and the corresponding standard errors $\sigma(\hat{\beta}_t^i)$ and R-squared $R_t^{2,i}$. As in Farmer *et al.* (2018), pockets are determined by a threshold c:

$$\mathcal{I}_t^i(c) = \mathbf{1}_{|\hat{\beta}_t^i/\sigma(\hat{\beta}_t^i)|>c}. \qquad (12)$$

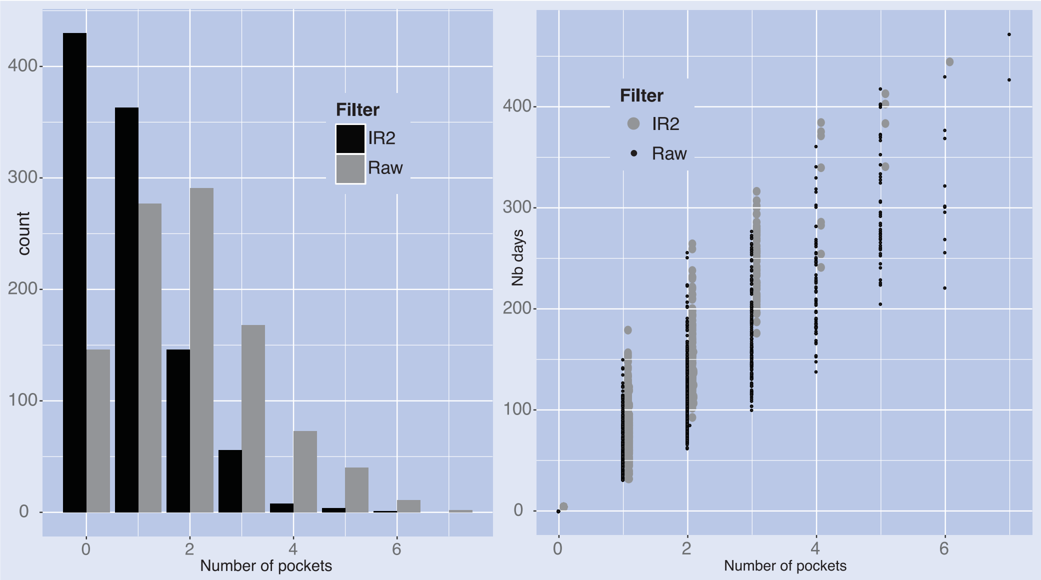

Figure 14. *Number of pockets and days of predictability.* On the left graph, we show the repartition of the number of pockets (1936 pockets have been identified in total and 881 of them pass the threshold of integrated R^2). On the right graph, each stock with at least one pocket is represented with one dot, and the y-axis is the corresponding total number of days of predictability computed in (13). The threshold for statistical significance in (12) is 2, which corresponds to a confidence level of 5%. The grey colour relates to the baseline results while the black colour corresponds to the pockets with an integrated R^2 above 3.

Table 4. *Most predictable stocks: correlations.* We compute the correlation between the proportion of t-statistics that are absolutely larger than 2 in Sections 4.1 (baseline), 5.1 (fine-tuned), 5.2 (macro-conditioning) and both the number of pockets of predictability and the total number of days of predictability.

	Fine-tuned	Macro-conditioning	Nb pockets	Length pockets
Baseline	0.081	0.090	0.028	0.031
Fine-tuned		−0.017	−0.042	−0.034
Macro-conditioning			0.286	0.322
Nb pockets				0.916

We choose to work with $c = 2$, which corresponds to a significance level of 5%.† For a given series of strictly positive $\mathcal{I}_t^i$ length of the jth pocket is computed as

$$L_j^i = (\tau_{j,start} - \tau_{j,stop} + 1) \times \mathbf{1}_{\{\tau_{j,start}-\tau_{j,stop}+1>30\}},$$

where $\tau_{j,start}$ and $\tau_{j,stop}$ are the chronological extremities of the pocket. We enforce a lower threshold to ensure the consistency of pocket because short pockets are more likely to be spurious. In their study, Farmer *et al.* (2018) find that the smallest pocket lasts 24 days and that average pocket duration ranges between 126 and 378 days. They also introduce a criterion for spurious pocket detection: the integrated R^2, defined as

$$IR^{2,i} = \sum_{t=\tau_{start}}^{\tau_{stop}} R_t^{2,i}.$$

Intuitively, a *true* pocket must either be short but have a strong predictive power (measured by the R^2) or long with possibly inferior predictive ability. This justifies the lower threshold of 30 days in the calculation of L_j^i. In their computations based on Monte Carlo simulations, Farmer *et al.* (2018) find that the 5% confidence level is reached whenever $IR^{2,i}$ is larger than 3 (see table 4 therein). Hence, in our results, we compare the baseline pockets (stemming from (12)) to those that are filtered using the criterion $IR^{2,i} > 3$. Finally, we also compute the total number of days of predictability for each stock:

$$L^i = \sum_j L_j^i. \tag{13}$$

In total, we ran 2.38 million regressions. We detail our results below. In figure 14, we characterise the pockets. On the left graph, we plot the distribution of the number of pockets, for each stock. Depending on this number, we display, on the right graph the total number of days within all corresponding pockets, for each firm. According to the value of the leftmost grey (*resp*. black) bar, there are 14% (*resp*. 43%) of the stocks that do not exhibit any predictability at all, at the 5% level. On the other hand, another 14% (7% after filtering) of stocks compile more than 200 days of predictability over the whole sample. Again, these figures illustrates the dispersion of predictability in the cross-section: the bulk of stocks can by no mean be forecast with news sentiment metrics, but some companies do experience several windows of predictability in the 11 years of our sample.

5.8. Which firms are the most predictable?

This section is dedicated to a macro view of the results obtained in Sections 4.1– 5.7. We compute the proportion of significant (5% level) t-statistics for each stock in Sections 4.1 (baseline), 5.1 (fine-tuned), 5.2 (macro-conditioning), and we consider the number of pockets and total number of days of

† With $c = 3$, i.e. a significance level of 0.2%, all pockets vanish.

Table 5. *Most predictable stocks.* We list the stocks which score the best on either one of the four criteria described in their respective panels.

Ticker	Prop. baseline	Prop. fine-tuned	Prop. Macro cond.	Nb pockets	Length pockets
PANEL A: High Proportion of significant baseline *t*-statistics					
JNS	0.618	0	0.200	4	250
LII	0.588	0	0	1	37
SNI	0.529	0.020	0	2	188
TYPE	0.853	0.073	0.200	3	218
PANEL B: High Proportion of significant fine-tuned *t*-statistics					
AMRS	0.059	0.361	0	2	223
FTR	0.176	0.333	0	2	68
VRX	0.118	0.569	0.100	2	159
PANEL C: High Proportion of significant macro-conditioned *t*-statistics					
HLF	0.118	0.049	0.300	5	328
LRCX	0.176	0.029	0.300	3	184
RAD	0.059	0.029	0.300	3	256
RMBS	0.176	0.022	0.300	4	282
SAM	0.029	0.051	0.300	3	270
SNY	0.118	0.020	0.300	6	433
PANEL D: Largest number of days of predictability					
LVLT	0	0	0.100	7	427
MLNX	0	0.022	0.200	6	430
MSI	0.118	0	0	7	472
SNY	0.118	0.020	0.300	6	433

predictability from Section 5.7. This amounts to five indicators across all stocks. In table 4, we display the correlation between these indicators.

By far, the largest correlation is between the two metrics originating from the analysis of predictability pockets; the link between the two quantities is self-evident. All other correlations are qualitatively low (0.32 at most, in absolute value), which means that each one of the four sets of protocols identifies different predictability patterns. The models related to fine-tuned predictors are even negatively related to the other ones. This highlights that slightly changing variables or empirical setups can lead to drastically altered results.

In table 5, we gather the companies for which the proportion of significant t-statistics are the highest, along with those for which the number of pockets of predictability is the largest. Only one stock, Sanofi SA (SNY) stands out in two different panels. Even when considering panels with bigger sizes, the overlap remains small. This again implies that the different forecasting protocols have little in common and that they cannot supply a list of assets that are consistently robust across predictors.

We close this meta-analysis by reversing the initial question: is it possible to find stocks for which t-statistics are all insignificant? Across our three regression-based studies (baseline, fine-tuned and macro-conditioned), only 83 stocks had sentiment estimates that are never associated with p-values lower than 5%. The fact that only 8% of our sample cannot be predicted with sentiment in our various attempts is an artefact of p-hacking. For any given firm, it is possible to craft a particular model that will indeed yield predictability (see, e.g. Novy-Marx 2014). Nevertheless, it is inconceivable to find one that will work well for a majority of stocks, even in sample, let alone out-of-sample.

5.9. *Synchronous and lagged regressions: who predicts who?*

Our baseline model seeks to infer future returns from the current level of sentiment. Because the indicator is released on the morning of day t and based on observations that occurred during day $t-1$, we finally run a last pool of regressions in which returns are lagged:

$$r^{i,OC}_{t-1,t} = \alpha^{i,OC}_k + \beta^{i,OC}_k x^i_t + \epsilon^{i,OC}_{t,k}, \qquad (14)$$

$$r^{i,OC}_{t-2,t-1} = \alpha^{i,OC}_k + \beta^{i,OC}_k x^i_t + \epsilon^{i,OC}_{t,k}. \qquad (15)$$

The first model (14) can be qualified as synchronous because the returns are computed over the period during which the news are collected for the estimation of the sentiment indicator. We add the '*lagged*' model (15) to test whether the predictability relationship is not reversed, i.e. if past returns drive future sentiment.† It is indeed likely that a firm's good or bad market performance for a given day is propagated in the news the following day. In figure 15, we plot the distribution of the corresponding t-statistics, along with that of the baseline model. The most striking pattern is the rightward shift of the two distributions, compared to the original one (in red). This signals that the link between sentiment and returns is more likely to be positive with synchronous and lagged return. Moreover, the proportion of significant t-statistics is also much higher (24% and 18% for synchronous and lagged returns, at the 5% level).

In short, these results demonstrate that the connexion between sentiment and current or past returns is undoubtedly more salient than that with future returns. This indicates

† Swapping the dependent and independent variable does not alter t-statistics.

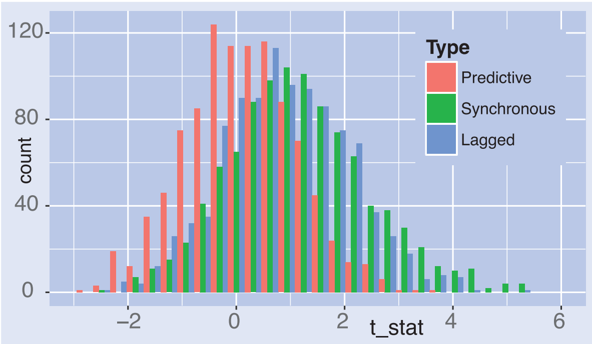

Figure 15. *Lagged regressions.* We plot the distribution of t-statistics for all three models (4) in red, (14) in green and (15) in blue.

that, at the daily frequency, sentiment indicators are mostly backward-looking and carry little predictive power.

We end this section with unreported robustness checks. We tried to control for aggregate market returns in the synchronous equation (14). In other words, we ran CAPM regressions with an additional sentiment component. At the 5% level, the proportion of outstanding sentiment-related t-statistics was equal to 6%. Finally, we ran Granger causality tests, with three lags. At the 5% level, sentiment was found to Granger-cause returns for 4.5% of stocks. The reverse relationship, in which returns Granger-cause sentiment, is much more salient: it is significant for 28% of stocks.

In sum, these results confirm those of Brown and Cliff (2004) obtained at the aggregate level: sentiment co-moves with returns, but fails to deliver short-term predictability. The highly oscillatory behaviour of sentiment moreover makes long-term predictability irrelevant.

5.10. *Twitter sentiment*

The bulk of our study relies on the sentiment metric computed by Bloomberg. Social media provides an alternative source of sentiment with respect to the stock market. Some services that analyse the content of tweets (e.g. Stocktwits) have been used as proxies of sentiment (Renault 2017, Audrino *et al.* 2018). Some authors have also processed raw data through network metrics, as in Yang *et al.* (2015) or via natural language processing tools (Bollen *et al.* 2011).

Just as for news flows, Bloomberg also processes Twitter feeds through Natural Language Processing algorithms and computes a sentiment score, at the firm level. Nonetheless, the time range is much smaller, as the sample only begins in 2015. This is why we exclude it from our main empirical framework. The correlation in all available pairs of news- and Twitter- scores is equal to 0.16, which indicates that both metrics are not redundant.

For the sake of completeness, we thus report the results of predictive regressions (4) and (5) when sentiment is this time proxied by Twitter content analysis. In figure 16, we plot the distribution of t-statistics obtained from the estimation. The proportion of t-statistics larger than 2 is equal to 8.8% (OC) and 7.8% (OO). When the threshold is 3, the proportions fall to 1.3% (OC) and 0.9%. These figures are in line with those

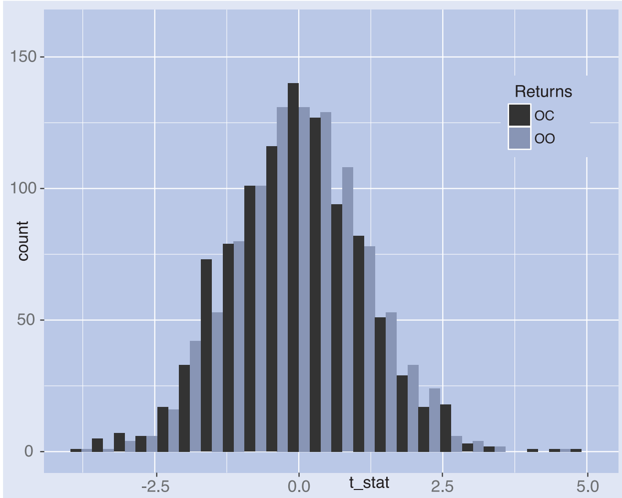

Figure 16. *Baseline results - Twitter predictor.* We plot the histogram of the t-stats of models (4) in black and (5) in grey when x_t^i is taken as the independent variable. The sample starts in January 2015 and ends in November 2017.

obtained with news sentiment as independent variable and confirm our previous findings with another set of predictors.

6. Conclusion

In the field of Financial Economics, researchers often seek pervasive factors that either structure the cross-section of stock returns or simply predict returns. Sentiment indicators that track the mood of investors towards indices and stocks have recently gained popularity and emerge as plausible drivers of market returns. This article evaluates the propensity of stock-specific sentiment to forecast or impact future returns and to drive their cross-section. In a nutshell, our results do not support the assumption that sentiment carries preeminent economic value for investors at the stock level.

First, one of the major takeaways of this study is that news-based sentiment is not a powerful predictor of future returns on a firm-by-firm basis. Only 7% of stocks benefit from a significant relationship at the 5% level and the proportion falls below 1% at the 0.2% confidence level. In fact, as the informational feedback loop could suggest, it is more likely that stock returns forecast sentiment than the reverse relationship.

In addition, sentiment-based predictability is very model dependent, meaning that the stocks that stand out in one empirical specification fail to reach significance in an alternative specification. This highlights the preponderance of regressors and our findings show that discovering reliable sentiment-based predictors that perform well across the cross-section of stocks is a complicated task. Nonetheless, our results indicate that sentiment-driven predictability has increased in the latest period, from 2012 to 2017. The refinement in the computation of the sentiment proxies is one possible explanation for this improvement.

In unreported results, we implemented portfolios based on raw sentiment scores and on predictions based on these scores.

Unsurprisingly, the results were disappointing. Monthly cross-sectional portfolios built on smoothed sentiment exhibit very similar returns which indicates that sentiment is not a priced factor.

We conclude with a final word on sampling frequency. Our study relies on daily data. Recent work at the aggregate level (Renault 2017, Yang *et al.* 2018) suggests that a low latency analysis would be better suited because it is likely that the informational value of sentiment is integrated into prices in a matter of seconds or minutes. Given the computational cost of a verification over the cross-section of stocks, we leave this extension for future work.

Acknowledgments

The author thanks two anonymous referees for their comments that have helped improve the quality of the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Guillaume Coqueret http://orcid.org/0000-0002-1596-4086

References

Aboody, D., Even-Tov, O., Lehavy, R. and Trueman, B., Overnight returns and firm-specific investor sentiment. *J. Financ. Quant. Anal.*, 2018, **53**(1), 1–21.

Amrhein, V., Greenland, S. and McShane, B., Scientists rise up against statistical significance. *Nature*, 2019, **567**(7748), 305–307.

Antoniou, C., Doukas, J.A. and Subrahmanyam, A., Cognitive dissonance, sentiment, and momentum. *J. Financ. Quant. Anal.*, 2013, **48**(1), 245–275.

Antoniou, C., Doukas, J.A. and Subrahmanyam, A., Investor sentiment, beta, and the cost of equity capital. *Manage. Sci.*, 2016, **62**, 347–367.

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

Audrino, F., Sigrist, F. and Ballinari, D., The impact of sentiment and attention measures on stock market volatility. SSRN Working Paper 3188941, 2018.

Baker, M. and Stein, J.C., Market liquidity as a sentiment indicator. *J. Financ. Mark.*, 2004, **7**(3), 271–299.

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

Baker, M. and Wurgler, J., Investor sentiment in the stock market. *J. Econ. Perspect.*, 2007, **21**, 129–151.

Baker, M. and Wurgler, J., Comovement and predictability relationships between bonds and the cross-section of stocks. *Rev. Asset Pricing Stud.*, 2012, **2**(1), 57–87.

Bakshi, G., Panayotov, G. and Skoulakis, G., Improving the predictability of real economic activity and asset returns with forward variances inferred from option portfolios. *J. Financ. Econ.*, 2011, **100**, 475–495.

Barber, B.M. and Odean, T., All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Rev. Financ. Stud.*, 2008, **21**(2), 785–818.

Barsky, R.B. and Sims, E.R., Information, animal spirits, and the meaning of innovations in consumer confidence. *Am. Econ. Rev.*, 2012, **102**, 1343–77.

Bartsch, V.-S., Dichtl, H., Drobetz, W. and Neuhierl, A., Data snooping in equity premium prediction. SSRN Working Paper, 2018.

Beaver, W.H., The information content of annual earnings announcements. *J. Account. Res.*, 1968, **6**, 67–92.

Benhabib, J. and Spiegel, M.M., Sentiments and economic activity: Evidence from us states. *Econ. J.*, 2018, **129**, 715–733.

Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E. -J., Berk, R., Bollen, K.A., Brembs, B., Brown, L., Camerer, C. and Cesarini, D.. Redefine statistical significance. *Nat. Human Behav.*, 2018, **2**(1), 6–10.

Berkman, H., Koch, P.D., Tuttle, L. and Zhang, Y.J., Paying attention: Overnight returns and the hidden cost of buying at the open. *J. Financ. Quant. Anal.*, 2012, **47**(4), 715–741.

Bollen, J., Mao, H. and Zeng, X., Twitter mood predicts the stock market. *J. Comput. Sci.*, 2011, **2**, 1–8.

Brodeur, A., Lé, M., Sangnier, M. and Zylberberg, Y., Star wars: The empirics strike back. *Am. Econ. J.: Appl. Econ.*, 2016, **8**(1), 1–32.

Brown, G.W. and Cliff, M.T., Investor sentiment and the near-term stock market. *J. Empir. Finance*, 2004, **11**(1), 1–27.

Brown, G.W. and Cliff, M.T., Investor sentiment and asset valuation. *J. Bus.*, 2005, **78**, 405–440.

Busse, J.A. and Green, T.C., Market efficiency in real time. *J. Financ. Econ.*, 2002, **65**, 415–437.

Camerer, C.F., Dreber, A., Forsell, E., Ho, T. -H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T. and Heikensten, E., Evaluating replicability of laboratory experiments in economics. *Science*, 2016, **351**(6280), 1433–1436.

Campbell, J.Y., Lo, A.W.-C. and MacKinlay, A.C., *The Econometrics of Financial Markets*, vol. 2, 1997 (Princeton University Press: Princeton, NJ).

Campbell, J.Y. and Yogo, M., Efficient tests of stock return predictability. *J. Financ. Econ.*, 2006, **81**, 27–60.

Cen, L., Lu, H. and Yang, L., Investor sentiment, disagreement, and the breadth–return relationship. *Manage. Sci.*, 2013, **59**, 1076–1091.

Chan, W.S., Stock price reaction to news and no-news: Drift and reversal after headlines. *J. Financ. Econ.*, 2003, **70**, 223–260.

Chang, C.-C., Hsieh, P.-F. and Wang, Y.-H., Sophistication, sentiment, and misreaction. *J. Financ. Quant. Anal.*, 2015, **50**(4), 903–928.

Chang, E.C., Lin, T. -C., Luo, Y. and Ren, J., Ex-day returns of stock distributions: An anchoring explanation. *Manage. Sci.*, 2019, **65**(3), 1076–1095.

Chen, H., De, P., Hu, Y.J. and Hwang, B.-H., Wisdom of crowds: The value of stock opinions transmitted through social media. *Rev. Financ. Stud.*, 2014, **27**(5), 1367–1403.

Chen, Y.-W., Chou, R.K. and Lin, C.-B., Investor sentiment, seo market timing, and stock price performance. *J. Empir. Finance*, 2019, **51**, 28–43.

Chordia, T., Goyal, A. and Saretto, A., p-hacking: Evidence from two million trading strategies. Working Paper, 2017.

Chung, S.-L., Hung, C.-H. and Yeh, C.-Y., When does investor sentiment predict stock returns? *J. Empir. Finance*, 2012, **19**(2), 217–240.

Cremers, M. and Weinbaum, D., Deviations from put-call parity and stock return predictability. *J. Financ. Quant. Anal.*, 2010, **45**(2), 335–367.

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

Da, Z., Engelberg, J. and Gao, P., The sum of all FEARS investor sentiment and asset prices. *Rev. Financ. Stud.*, 2015, **28**(1), 1–32.

Da, Z., Liu, Q. and Schaumburg, E., A closer look at the short-term return reversal. *Manage. Sci.*, 2014, **60**, 658–674.

Dangl, T. and Halling, M., Predictive regressions with time-varying coefficients. *J. Financ. Econ.*, 2012, **106**, 157–181.
Das, S.R. and Chen, M.Y., Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Manage. Sci.*, 2007, **53**, 1375–1388.
De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., Noise trader risk in financial markets. *J. Polit. Econ.*, 1990, **98**(4), 703–738.
Dougal, C., Engelberg, J., Garcia, D. and Parsons, C.A., Journalists and the stock market. *Rev. Financ. Stud.*, 2012, **25**(3), 639–679.
Duan, J.-C. and Zhang, W., Forward-looking market risk premium. *Manage. Sci.*, 2013, **60**, 521–538.
Edmans, A., Garcia, D. and Norli, Ø., Sports sentiment and stock returns. *J. Finance*, 2007, **62**, 1967–1998.
Elliott, G. and Stock, J.H., Inference in time series regression when the order of integration of a regressor is unknown. *Econ. Theory.*, 1994, **10**(3–4), 672–700.
Engelberg, J.E. and Parsons, C.A., The causal impact of media in financial markets. *J. Finance*, 2011, **66**, 67–97.
Fama, E.F. and French, K.R., Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.*, 1993, **33**, 3–56.
Fang, L. and Peress, J., Media coverage and the cross-section of stock returns. *J. Finance*, 2009, **64**, 2023–2052.
Farmer, L.E., Timmermann, A. and Schmidt, L.D., Pockets of predictability. Working Paper, 2018.
Fève, P. and Guay, A., Sentiments in svars. *Econ. J.*, 2018, **129**, 877–896.
Firth, M., Wang, K. and Wong, S.M., Corporate transparency and the impact of investor sentiment on stock prices. *Manage. Sci.*, 2014, **61**, 1630–1647.
Gallant, A.R., Rossi, P.E. and Tauchen, G., Stock prices and volume. *Rev. Financ. Stud.*, 1992, **5**(2), 199–242.
Gao, B. and Yang, C., Forecasting stock index futures returns with mixed-frequency sentiment. *Int. Rev. Econ. Finance*, 2017, **49**, 69–83.
Garcia, D., Sentiment during recessions. *J. Finance*, 2013, **68**, 1267–1300.
Griffin, J.M., Hirschey, N.H. and Kelly, P.J., How important is the financial media in global markets? *Rev. Financ. Stud.*, 2011, **24**(12), 3941–3992.
Groß-Klußmann, A. and Hautsch, N., When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *J. Empir. Finance*, 2011, **18**(2), 321–340.
Hamilton, J.D., *Time Series Analysis*, vol. 2, 1994 (Princeton University Press: Princeton, NJ).
Harvey, C.R., Presidential address: The scientific outlook in financial economics. *J. Finance*, 2017, **72**, 1399–1440.
Harvey, C.R., Liu, Y. and Zhu, H., … and the cross-section of expected returns. *Rev. Financ. Stud.*, 2016, **29**(1), 5–68.
Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. and Jennions, M.D., The extent and consequences of p-hacking in science. *PLoS Biol.*, 2015, **13**(3), e1002106.
Hengelbrock, J., Theissen, E. and Westheide, C., Market response to investor sentiment. *J. Bus. Finance. Account.*, 2013, **40**(7–8), 901–917.
Henkel, S.J., Martin, J.S. and Nardari, F., Time-varying short-horizon predictability. *J. Financ. Econ.*, 2011, **99**, 560–580.
Hirshleifer, D. and Shumway, T., Good day sunshine: Stock returns and the weather. *J. Finance*, 2003, **58**, 1009–1032.
Hou, K., Xue, C. and Zhang, L., Replicating anomalies. *Rev. Financ. Stud.*, 2018, Forthcoming.
Hribar, P. and McInnis, J., Investor sentiment and analysts' earnings forecast errors. *Manage. Sci.*, 2012, **58**, 293–307.
Huang, D., Jiang, F., Tu, J. and Zhou, G., Investor sentiment aligned: A powerful predictor of stock returns. *Rev. Financ. Stud.*, 2015, **28**(3), 791–837.
Huang, D. and Zhou, G., Upper bounds on return predictability. *J. Financ. Quant. Anal.*, 2017, **52**(2), 401–425.
Ioannidis, J.P., Why most published research findings are false. *PLoS Med.*, 2005, **2**, 696–701.
Ioannidis, J.P., Stanley, T. and Doucouliagos, H., The power of bias in economics research. *Econ. J.*, 2017, **127**, 236–265.
Jager, L.R. and Leek, J.T., An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 2013, **15**(1), 1–12.
Jiang, F., Lee, J.A., Martin, X. and Zhou, G., Manager sentiment and stock returns. *J. Financ. Econ.*, 2019, **132**(1), 126–149.
Johannes, M., Korteweg, A. and Polson, N., Sequential learning, predictability, and optimal portfolio returns. *J. Finance*, 2014, **69**, 611–644.
Kamstra, M.J., Kramer, L.A. and Levi, M.D., Winter blues: A sad stock market cycle. *Amer. Econ. Rev.*, 2003, **93**(1), 324–343.
Kaplanski, G. and Levy, H., Sentiment and stock prices: The case of aviation disasters. *J. Financ. Econ.*, 2010, **95**, 174–201.
Kerr, N.L., Harking: Hypothesizing after the results are known. *Pers. Soc. Psychol. Rev.*, 1998, **2**(3), 196–217.
Kim, J.S., Kim, D. -H. and Seo, S.W., Investor sentiment and return predictability of the option to stock volume ratio. *Financ. Manage.*, 2017, **46**(3), 767–796.
Kuffner, T.A. and Walker, S.G., Why are p-values controversial? *Am. Stat.*, 2019, **73**(1), 1–3.
Kumar, A. and Lee, C., Retail investor sentiment and return comovements. *J. Finance*, 2006, **61**, 2451–2486.
Larsen, V.H. and Thorsrud, L.A., The value of news for economics developments. *J. Econom.*, 2019, **210**(1), 203–218.
Lee, C., Shleifer, A. and Thaler, R.H., Investor sentiment and the closed-end fund puzzle. *J. Finance*, 1991, **46**, 75–109.
Lehrer, D., Leschke, J., Lhachimi, S., Vasiliu, A. and Weiffen, B., Negative results in social science. *Eur. Pol. Sci.*, 2007, **6**(1), 51–68.
Leung, H. and Ton, T., The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks. *J. Bank. Financ.*, 2015, **55**, 37–55.
Liu, X., Yang, B., Cai, Z. and Peng, L., A unified test for predictability of asset returns regardless of properties of predicting variables. *J. Econom.*, 2019, **208**, 141–159.
Lutz, C., The asymmetric effects of investor sentiment. *Macroecon. Dyn.*, 2016, **20**(6), 1477–1503.
McLean, R.D. and Zhao, M., The business cycle, investor sentiment, and costly external finance. *J. Finance.*, 2014, **69**, 1377–1409.
Mian, G.M. and Sankaraguruswamy, S., Investor sentiment and stock market response to earnings news. *Account. Rev.*, 2012, **87**(4), 1357–1384.
Mitchell, M.L. and Mulherin, J.H., The impact of public information on the stock market. *J. Finance*, 1994, **49**, 923–950.
Mlinarić, A., Horvat, M. and Šupak Smolčić, V., Dealing with the positive publication bias: Why you should really publish your negative results. *Biochemia Medica*, 2017, **27**, 1–6.
Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H., Time series momentum. *J. Financ. Econ.*, 2012, **104**, 228–250.
Neal, R. and Wheatley, S.M., Do measures of investor sentiment predict returns? *J. Financ. Quant. Anal.*, 1998, **33**(4), 523–547.
Novy-Marx, R., Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *J. Financ. Econ.*, 2014, **112**, 137–146.
Oliveira, N., Cortez, P. and Areal, N., The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert. Syst. Appl.*, 2017, **73**, 125–144.
Pettenuzzo, D., Timmermann, A. and Valkanov, R., Forecasting stock returns under economic constraints. *J. Financ. Econ.*, 2014, **114**, 517–553.
Rapach, D.E., Strauss, J.K. and Zhou, G., Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Rev. Financ. Stud.*, 2010, **23**(2), 821–862.
Renault, T., Intraday online investor sentiment and return patterns in the US stock market. *J. Bank. Financ.*, 2017, **84**, 25–40.
Ross, S.A., *Neoclassical Finance*, 2009 (Princeton University Press: Princeton, NJ).
Rytchkov, O. and Zhong, X., Information aggregation and p-hacking. SSRN Working Paper, 2018.

Sabherwal, S., Sarkar, S.K. and Zhang, Y., Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news. *J. Bus. Finance. Account.*, 2011, **38**(9–10), 1209–1237.

Schmeling, M., Investor sentiment and stock returns: Some international evidence. *J. Empir. Finance*, 2009, **16**(3), 394–408.

Shefrin, H., Investors' judgments, asset pricing factors and sentiment. *Eur. Financ. Manage.*, 2015, **21**(2), 205–227.

Shen, J., Yu, J. and Zhao, S., Investor sentiment and economic forces. *J. Monet. Econ.*, 2017, **86**, 1–21.

Solt, M.E. and Statman, M., How useful is the sentiment index? *Financ. Analysts J.*, 1988, **44**(5), 45–55.

Stambaugh, R.F., Predictive regressions. *J. Financ. Econ.*, 1999, **54**, 375–421.

Stambaugh, R.F., Yu, J. and Yuan, Y., The short of it: Investor sentiment and anomalies. *J. Financ. Econ.*, 2012, **104**, 288–302.

Starr, M.A., Consumption, sentiment, and economic news. *Econ. Inq.*, 2012, **50**(4), 1097–1111.

Sun, L., Najand, M. and Shen, J., Stock return predictability and investor sentiment: A high-frequency perspective. *J. Bank. Financ.*, 2016, **73**, 147–164.

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

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

Thornton, A. and Lee, P., Publication bias in meta-analysis: Its causes and consequences. *J. Clin. Epidemiol.*, 2000, **53**(2), 207–216.

Tirunillai, S. and Tellis, G.J., Does chatter really matter? Dynamics of user-generated content and stock performance. *Mark. Sci.*, 2012, **31**(2), 198–215.

Tumarkin, R. and Whitelaw, R.F., News or noise? Internet postings and stock prices. *Financ. Analysts J.*, 2001, **57**(3), 41–51.

Wang, C., Investor sentiment and return predictability in agricultural futures markets. *J. Futures Mark.*, 2001, **21**(10), 929–952.

Wang, Y.-H., Keswani, A. and Taylor, S.J., The relationships between sentiment, returns and volatility. *Int. J. Forecast.*, 2006, **22**(1), 109–123.

Welch, I. and Goyal, A., A comprehensive look at the empirical performance of equity premium prediction. *Rev. Financ. Stud.*, 2007, **21**(4), 1455–1508.

Xing, Y., Zhang, X. and Zhao, R., What does the individual option volatility smirk tell us about future equity returns? *J. Financ. Quant. Anal.*, 2010, **45**(3), 641–662.

Yang, A.S. and Wu, M. -L., Exploring the relationship between investor sentiment and price volatility. *Quant. Finance*, 2011, **11**(6), 955–965.

Yang, S.Y., Liu, A., Chen, J. and Hawkes, A., Applications of a multivariate hawkes process to joint modeling of sentiment and market return events. *Quant. Finance*, 2018, **18**(2), 295–310.

Yang, S.Y., Mo, S.Y.K. and Liu, A., Twitter financial community sentiment and its predictive relationship to stock market movement. *Quant. Finance*, 2015, **15**(10), 1637–1656.

Yu, J. and Yuan, Y., Investor sentiment and the mean–variance relation. *J. Financ. Econ.*, 2011, **100**, 367–381.

Zhang, J.L., Härdle, W.K., Chen, C.Y. and Bommes, E., Distillation of news flow into analysis of stock reactions. *J. Bus. Econ. Stat.*, 2016, **34**(4), 547–563.

Zhou, G., How much stock return predictability can we expect from an asset pricing model? *Econ. Lett.*, 2010, **108**(2), 184–186.

Zhou, G., Measuring investor sentiment. *Annu. Rev. Financ. Econ.*, 2018, **10**(1), 239–259.

Appendix

A. Tickers

Table A1. List of tickers

A	AA	AAP	AAPL	ABB	ABC	ABCO	ABT	ABX	ACAS	ACH	ACHC	ACIW
ACM	ACN	ACOR	ADBE	ADI	ADM	ADP	ADS	ADSK	AEG	AEGR	AEM	AEO
AEP	AES	AET	AFL	AGCO	AGIO	AGN	AGU	AIG	AIZ	AJG	AKAM	AKRX
AKS	AL	ALB	ALGT	ALJ	ALK	ALKS	ALL	ALNY	ALR	ALSN	ALU	ALV
ALXN	AMAG	AMAT	AMBA	AMCX	AME	AMGN	AMOV	AMP	AMRS	AMTD	AMX	AMZN
AN	ANET	ANF	ANGI	ANSS	ANTM	AON	APA	APC	APD	APOL	ARIA	ARMH
ARMK	ARRS	ARW	ASH	ASML	ASX	ATHN	ATI	ATVI	ATW	AU	AUO	AUY
AVH	AVP	AVT	AWAY	AWK	AXP	AXTA	AZN	AZO	BA	BAC	BAM	BAX
BBBY	BBD	BBDO	BBT	BBVA	BBY	BCE	BCR	BCS	BDX	BEAV	BEN	BG
BGCP	BHI	BHP	BID	BIDU	BIIB	BK	BKS	BLK	BLL	BLOX	BLUE	BMO
BMRN	BMY	BNS	BP	BR	BRCD	BRFS	BSFT	BSX	BT	BWA	BWLD	BWXT
BYD	C	CA	CACI	CAG	CAH	CAJ	CAKE	CAM	CAR	CAT	CAVM	CB
CBG	CBI	CBOE	CBS	CCE	CCJ	CCL	CCOI	CDE	CDNS	CE	CEA	CELG
CEO	CERN	CF	CFG	CFX	CGG	CHA	CHD	CHH	CHK	CHKP	CHL	CHTR
CHU	CI	CIE	CIEN	CIG	CL	CLF	CLGX	CLR	CLVS	CLX	CM	CMA
CMCSA	CMCSK	CME	CMG	CMI	CMS	CNC	CNI	CNQ	CNS	CNX	COF	COG
COL	COMM	COP	COST	CP	CPB	CPE	CPHD	CPN	CREE	CRH	CRM	CRUS
CRZO	CS	CSC	CSCO	CSIQ	CSOD	CSX	CTB	CTL	CTSH	CTXS	CUB	CUDA
CVC	CVE	CVS	CVX	CXO	CY	CYH	D	DAL	DATA	DB	DBD	DCM
DD	DDD	DE	DECK	DEO	DF	DFS	DG	DGI	DGX	DHI	DHR	DIS
DISCA	DISCB	DISCK	DISH	DK	DKS	DLB	DLPH	DLTR	DNB	DNKN	DNR	DO
DOV	DOW	DPS	DPZ	DRI	DTE	DUK	DVA	DVN	DWA	DXCM	E	EA
EAT	EBAY	EBR	EC	ECA	ECL	ED	EFX	EGN	EGO	EIX	EL	ELLI
EMC	EMN	EMR	ENB	ENDP	ENV	EOG	EPAM	EQT	ERIC	ERJ	ESRX	ESV
ETFC	ETN	ETR	EV	EVR	EW	EXC	EXPE	F	FANG	FAST	FB	FBHS
FBRC	FCN	FCX	FDS	FDX	FE	FET	FEYE	FFIV	FHN	FICO	FII	FIS
FISV	FITB	FL	FLDM	FLEX	FLIR	FLR	FLS	FLT	FMC	FMS	FNGN	FNSR
FORR	FOX	FOXA	FSLR	FTI	FTNT	FTR	G	GD	GDOT	GE	GFI	GG
GGB	GHDX	GIB	GILD	GIS	GLW	GM	GMCR	GME	GNC	GNW	GOL	GOLD
GOOG	GOOGL	GPN	GPS	GRA	GRFS	GRMN	GRUB	GS	GSK	GT	GWR	GWRE
GWW	H	HA	HAIN	HAL	HAR	HAS	HBAN	HBI	HBM	HD	HDB	HDP
HES	HFC	HIG	HII	HL	HLF	HLT	HMC	HMY	HNP	HNT	HOG	HOLX
HON	HP	HPQ	HRB	HRL	HRS	HSBC	HSIC	HSY	HTGC	HTWR	HTZ	HUM
HUN	HZNP	I	IAG	IART	IBKR	IBM	IBN	ICE	ICL	ICPT	IDTI	IFF
IHG	ILMN	IMPV	INCY	INFN	INFY	ING	INTC	INTU	INVN	IONS	IP	IPG
IPHI	IR	IRBT	IRWD	ISRG	IT	ITRI	ITT	ITUB	ITW	IVZ	IX	JAZZ
JBL	JBLU	JCI	JCP	JD	JEC	JKS	JLL	JNJ	JNPR	JNS	JOY	JPM
JUNO	JWN	K	KAR	KB	KBH	KBR	KCG	KEP	KEY	KGC	KITE	KLAC
KMB	KMT	KMX	KND	KNX	KO	KOS	KR	KSS	KSU	KT	KTOS	KYO
L	LAZ	LB	LDOS	LEA	LEG	LEN	LFC	LH	LII	LIVN	LL	LLL
LLTC	LLY	LM	LMT	LNC	LNKD	LOCK	LOGI	LOGM	LOW	LPI	LPL	LPLA
LRCX	LSCC	LUK	LULU	LUV	LVLT	LVS	LXK	LYB	LYG	LYV	M	MA
MAN	MAR	MAS	MAT	MBI	MBT	MC	MCD	MCHP	MCK	MCO	MDCO	MDLZ
MDP	MDR	MDRX	MDSO	MDT	MDVN	MENT	MET	MFC	MFG	MFRM	MGA	MGM
MITL	MJN	MKL	MKTO	MLM	MLNX	MMC	MMM	MMS	MNST	MNTA	MO	MOH
MON	MORN	MOS	MPC	MPWR	MRC	MRK	MRO	MRVL	MS	MSCC	MSCI	MSFT
MSI	MT	MTB	MTDR	MTSI	MTU	MTW	MU	MUR	MXIM	MXL	MYGN	MYL
N	NATI	NAV	NAVI	NBL	NBR	NCLH	NCR	NDAQ	NE	NEE	NEM	NFLX
NFX	NGG	NICE	NKE	NLSN	NMBL	NMR	NOC	NOK	NOV	NOW	NRG	NSC
NTAP	NTGR	NTRS	NUAN	NUE	NUVA	NVDA	NVO	NVS	NWL	NXPI	NXTM	NYT
OA	OAS	OC	OCN	ODP	OKE	OMC	ON	OPK	ORAN	ORCL	ORLY	OSK
OTEX	OXY	P	PANW	PAY	PAYX	PBA	PBI	PBR	PCAR	PCG	PCLN	PCRX
PE	PEG	PEGA	PEGI	PENN	PEP	PF	PFE	PFG	PG	PGR	PH	PHG
PHI	PHM	PII	PINC	PIR	PJC	PKG	PKI	PKX	PLCM	PM	PNC	PNR
PNRA	POL	POM	POST	POT	PPG	PPL	PQ	PRGO	PRU	PSO	PSX	PTCT
PTR	PUK	PVH	PX	PXD	QCOM	QEP	QGEN	QLIK	QLYS	R	RAD	RAI
RATE	RAX	RBS	RCI	RCL	RDC	RDN	REGI	REGN	RELX	REN	RENT	REXX
RF	RFP	RH	RHT	RICE	RIG	RIO	RJF	RKUS	RL	RMBS	RMD	RNG
RNR	ROK	ROST	RP	RPXC	RRC	RSG	RSPP	RTN	RY	RYAAY	SABR	SAM
SAN	SAP	SBGI	SBS	SBUX	SC	SCCO	SCHW	SCOR	SCTY	SD	SDRL	SEAS
SEE	SF	SFM	SGEN	SGMS	SGYP	SHAK	SHLD	SHPG	SHW	SID	SIG	SINA
SIRI	SIVB	SJM	SJR	SKM	SKX	SLAB	SLB	SLCA	SLF	SLM	SM	SMFG
SMG	SMI	SMTC	SN	SNCR	SNDK	SNE	SNI	SNN	SNP	SNPS	SNV	SNX

(*Continued*).

Table A1. Continued.

SNY	SO	SODA	SON	SPB	SPLK	SPLS	SPR	SPWR	SQNM	SRE	SRPT	SSL
SSNC	SSNI	SSYS	ST	STI	STJ	STLD	STM	STO	STR	STRZA	STT	STX
SU	SVU	SWK	SWKS	SWN	SYF	SYK	SYMC	SYNA	SYT	SYY	T	TAHO
TAP	TD	TDC	TDG	TEF	TEL	TEN	TERP	TEVA	TEX	TGT	THC	TI
TIF	TIME	TIVO	TJX	TKC	TM	TMH	TMK	TMO	TMUS	TOL	TOT	TPX
TRGP	TRI	TRIP	TRMB	TRP	TRU	TRUE	TRV	TS	TSCO	TSL	TSLA	TSM
TSN	TSRO	TSS	TTEK	TTM	TTWO	TU	TUP	TW	TWC	TWX	TXN	TXT
TYL	TYPE	UA	UAL	UHS	UIS	UMC	UN	UNH	UNM	UNP	UPS	URBN
URI	USB	USG	UTHR	UTX	V	VALE	VAR	VCRA	VEDL	VEEV	VFC	VG
VIA	VIAB	VIPS	VLO	VMC	VMW	VOD	VOYA	VR	VRNT	VRSK	VRTU	VRTX
VRX	VSAT	VVUS	VZ	W	WAT	WBA	WBC	WBK	WCG	WDAY	WDC	WEC
WEN	WFC	WFM	WFT	WHR	WIFI	WIN	WIT	WLL	WM	WMB	WMT	WNR
WPPGY	WPX	WSM	WU	WWAV	WWE	WYN	WYNN	X	XCO	XEC	XEL	XL
XLNX	XOM	XON	XRAY	XRX	XXIA	XYL	YGE	YNDX	YOKU	YPF	YUM	Z
ZAYO	ZBH	ZEN	ZG	ZION	ZLTQ	ZNH	ZTS					

A1. Volatility clustering

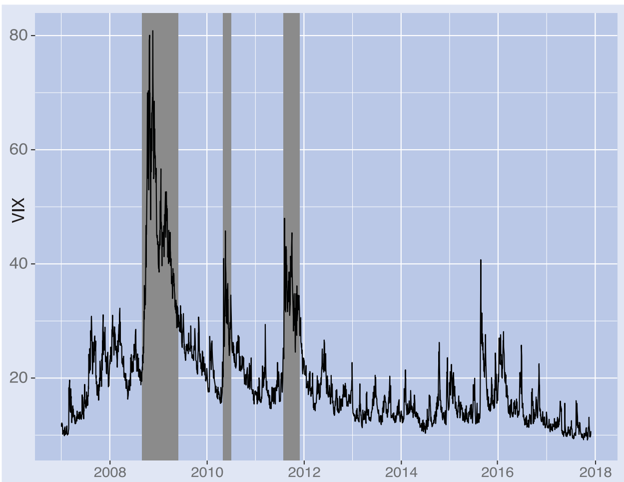

Figure A1. *VIX-based periods*. We plot the time-series of the VIX index. In grey, we show the time-spans of highly volatile market conditions.

A2. Sentiment smoothing

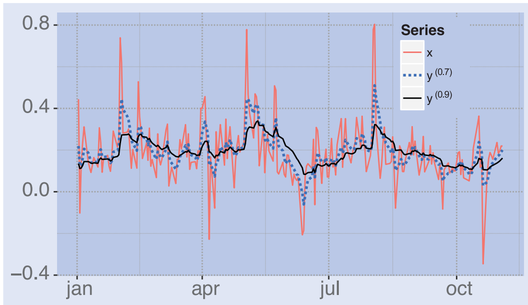

Figure A2. *Sentiment smoothing*. We plot the original sentiment indicator of Apple (AAPL ticker) for the first 11 months of 2017 (red curve). The smoothed versions ($\delta = 0.7$ and $\delta = 0.9$) are in dotted blue and black, respectively.