Journal of Financial Stability 48 (2020) 100741

Contents lists available at ScienceDirect

Journal of Financial Stability

journal homepage: www.elsevier.com/locate/jfstabil

Wisdom of crowds before the 2007–2009 global financial crisis☆

Michael Chau[a], Chih-Yung Lin[b,*], Tse-Chun Lin[c]

[a] Faculty of Business and Economics, University of Hong Kong, Hong Kong
[b] Department of Information Management and Finance, National Chiao-Tung University, Taiwan
[c] Faculty of Business and Economics, University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 15 October 2018
Received in revised form 15 March 2020
Accepted 26 March 2020
Available online 6 April 2020

JEL classification:
G01
G14
G17
G21

Keywords:
Social media
Textual analysis
Financial crisis
Predictability
Seeking alpha

ABSTRACT

Our paper examines whether investor opinions expressed in social media predicted stock returns of financial firms during the 2007–2009 global financial crisis. We conduct a textual analysis of the articles published on the stock market insight website Seeking Alpha before the crisis and find that banks that were described in articles with a higher fraction of negative words experienced (1) sharper drops in stock prices, (2) larger increases in expected default probability, and (3) greater surges in nonperforming loans during the crisis. Our evidence suggests that wisdom of crowds provides valuable information on how banks weather a forthcoming crisis.

1. Introduction

The 2007–2009 global financial crisis is considered the second largest economic crisis in history after the Great Depression of the 1930s. Major financial institutions such as Fannie Mae, Freddie Mac, Bear Stearns, Lehman Brothers, Merrill Lynch, Citigroup, and American International Group (AIG) either failed or came close to doing so. American families' wealth decreased by $11 trillion in 2008, which equals the combined output of Germany, Japan, and the UK at that time.[1] Hence, whether the crisis could be foreseen or prevented has been extensively debated.

On the one hand, neither bank insiders nor financial economists anticipated the impending crisis. For example, Fahlenbrach and Stulz (2011) show that CEOs in the banking industry failed to reduce their own shareholdings, implying that the financial crisis was not expected by these key Wall Street players. Colander et al. (2009) argue that economists were not capable of predicting the crisis either. Neither did Federal Reserve System (FED) and the Treasury detect the financial bubble; after retiring from the FED chairmanship position, Alan Greenspan admitted that he failed to foresee the speculative bubble in the mortgage lending market.[2]

On the other hand, Brockman et al. (2015) find that some market participants might have been aware of the forthcoming crisis. Their results show that institutional investors significantly reduced their ownership compared with non-financial stocks, and analysts issued lower recommendations for financial stocks. Hanley and Hoberg (2019) analyze risk factors in bank 10-Ks based on computational linguistics and find that these factors predicted financial instability in 2008. Lin et al. (2020) show that change in short interest predicts

☆ We would like to thank Gerard Hoberg for the helpful comments. Michael Chau and Tse-Chun Lin gratefully acknowledge research support from the Faculty of Business and Economics at the University of Hong Kong and the Research Grants Council of the Hong Kong SAR government (project numbers: 17502715B, 17501518B). Chih-Yung Lin appreciates financial support from Taiwan Ministry of Science and Technology (MOST108-2636-H-009-001). Any remaining errors are our own.

* Corresponding author.
E-mail addresses: mchau@business.hku.hk (M. Chau), d95723009@ntu.edu.tw (C.-Y. Lin), tsechunlin@hku.hk (T.-C. Lin).

[1] S. Mitra Kalita (2009). *Americans See 18% of Wealth Vanish*. The Wall Street Journal.

[2] Edmund L. Andrews, 2008. Greenspan Concedes Error on Regulation. The New York Times. (http://www.nytimes.com/2008/10/24/business/economy/24panel.html?_r=0).

https://doi.org/10.1016/j.jfs.2020.100741
1572-3089/

banks' stock returns during the financial crisis. The evidence is thus mixed regarding whether the financial market had any knowledge of the upcoming financial crisis or how it would affect the banking sector. To shed light on the debate, our paper explores whether the wisdom of crowds could predict the performance of financial firms during the crisis period. Specifically, we conduct a textual analysis of articles published on Seeking Alpha, one of the most popular social media platforms for investors in the United States, and link the tone of these articles to the subsequent performance of financial firms.[3]

Our conjecture that investor opinions expressed via social media might contain information regarding the crisis performance of financial firms is inspired by several recent studies. Based on the word lists of Loughran and McDonald (2011), Chen et al. (2014) use investor opinions from Seeking Alpha and find that the tone of articles could predict stock returns and earnings surprises in the next quarter.[4] Their results also support the proposition that peer opinions play an increasingly important role in financial markets.[5] Similarly, Bartov et al. (2018) use Twitter data from 2009 to 2012 and show that the aggregate opinion in individual tweets about a company's prospects can predict its earnings and stock returns. Da and Huang (2019) examine the impact of herding on the accuracy of consensus earnings forecasts from a crowd-based platform (Estimize.com) and find that the wisdom of crowds can be better harnessed by encouraging independent views from participants.[6]

Given the evidence cited above that investor opinions expressed via social media can predict future firm performance, we hypothesize that the wisdom of crowds in social media might also predict the bleak outlook for some banks before the occurrence of the financial crisis. We expect that the fraction of negative words contained in Seeking Alpha articles predicts the negative crisis returns of major banks. We focus exclusively on financial firms during the 2007–2009 crisis period for the following two reasons: (1) we aim to shed light on the debate regarding whether the occurrence of a financial crisis or the influence of such a crisis on the banking sector can be foreseen and (2) the crisis began in the banking industry which suffered such excessive losses that the SEC had to issue a temporary short-sales ban to this industry. Accordingly, we propose our first hypothesis:

H1. *The fraction of negative words contained in Seeking Alpha articles in the pre-crisis period can predict bank stock returns in the crisis period.*

Chen et al. (2014) show that not only the fraction of negative words contained in Seeking Alpha articles but also that of negative words in Seeking Alpha comments can predict future stock returns. Namely, the collective wisdom of readers may also be insightful. Hence, we propose our second hypothesis:

H2. *The fraction of negative words in Seeking Alpha comments has predictive power for the crisis returns of financial firms.*

In addition, we also examine whether authors' track records are related to the predictive power of Seeking Alpha articles. We thus propose our third hypothesis:

H3. *Seeking Alpha articles written by authors with better track records have stronger predictive power for the crisis returns of financial firms.*

To test our three hypotheses, we collect investor opinions from Seeking Alpha (SA), comprising 677 banks from January 2005 to June 2007 and use textual analysis to quantify the opinions disseminated through those articles. Specifically, we use the frequency of negative words in an article to capture the tone of the report (e.g., Das and Chen, 2007; Tetlock, 2007; Tetlock et al., 2008; Li, 2008; Loughran and McDonald, 2011, 2014, 2016). We use the negative word list proposed by Loughran and McDonald (2011) to characterize the views expressed in SA articles and SA comments.

Our results show that the fraction of negative words contained in SA articles in the pre-crisis period (January 2005 to June 2007) is negatively correlated with banks' stock returns during the financial crisis (July 2007 to December 2009).[7] The results still hold when we control for bank characteristics such as previous one-year buy-and-hold returns, book-to-market ratio, beta, tangible ratio, idiosyncratic volatility, percent change in earnings per share, percent change in non-interest income to net income, percent change in book value of equity to total assets, return of equity, asset growth, leverage, and default probability, providing supporting evidence for our first hypothesis.[8]

In terms of economic magnitude, a one-standard-deviation increase in the number of negative words contained in SA articles in the pre-crisis period leads to a decrease in bank stock returns during the financial crisis by 4.11 percent (0.0074×5.5475), a magnitude that represents approximately 50 percent of the risk-culture effect in Fahlenbrach et al. (2012).[9] This finding indicates that the predictability of SA articles on bank crisis returns is both statistically significant and economically meaningful.

Regarding our second hypothesis, we find that the fraction of negative words in SA comments during the pre-crisis period is also negatively correlated with bank stock returns during the financial crisis. The result is consistent with findings in Chen et al. (2014), who show that the fraction of negative words in SA comments can also predict future stock returns in the subsequent quarter. Thus, our results not only provide a validity check to their claim on the existence of the predictability of a negative-sentiment effect in SA but also indicate that some investors were able to predict the banks' crisis performance some considerable time prior to the occurrence of the crisis.

[3] As of August 2013, Seeking Alpha had between 500,000 to 1 million unique visitors per day and, as such, was one of the biggest investment-related social media websites in the United States. See the detailed introduction and discussions in Section 2.2.

[4] Loughran and McDonald (2011) created six different word lists (negative, positive, uncertain, litigious, strong modal, and weak modal) by examining word usage in a large sample of 10-Ks.

[5] Empirical evidence suggests that the influence of peer-based advice on consumer decision making is increasing (Chevalier and Mayzlin, 2006; Chen and Xie, 2008; Zhu and Zhang, 2010). In addition, Deloitte (2007) finds that 82% of US Internet consumers report that they are directly influenced by peer reviews in their purchasing decisions.

[6] Similarly, in the media literature, Dougal et al. (2012) find that journalists associated with a more pessimistic column tone in the Wall Street Journal are linked to more negative market returns the following day. Garcia (2013) measures the tone of two financial columns in the New York Times during the period 1905–2005 and finds that newspaper sentiment is related to future stock returns. Also see the studies of Chan (2003), Fang and Peress (2009), Engelberg and Parsons (2011), Solomon et al. (2014), Liu et al. (2014), Peress (2014), Miller and Skinner (2015), and Lee et al. (2015) for detailed discussions of the media effect on stock returns.

[7] In our main results, we set article tone for banks without articles to neutral (zero). In Section 4.3.3, we exclude banks not mentioned on the SA website and find similar results. In addition, results hold if we set article tone for banks without articles to the mean value or median value of our sample. The results of them are reported in Appendix B.

[8] Adding these variables helps address an alternative explanation that banks entering the crisis in bad shape as captured by SA articles performed worse during the crisis period.

[9] Fahlenbrach et al. (2012) propose that a financial institution has a persistent culture. Bank performance in past crises could proxy for its sensitivity to a crisis and could predict it performance in a subsequent crisis. Thus, they compare banks' stock market performance in two recent crises (i.e., the financial crises of 1998 and 2007) to test their hypothesis. They find that a one-standard-deviation decrease in bank stock returns during the Long-Term Capital Management (LTCM) crisis is associated with 8.2 percent lower stock returns during the 2007–2009 financial crisis.

For the third hypothesis, we find that the interaction terms of the fraction of negative words contained in SA articles and authors' track records are negatively correlated to banks' stock returns during the financial crisis. Essentially, the better the author's track record, the stronger the predictive power of their SA articles. Overall, the results of the three tests provide convincing evidence that the collective wisdom of the authors and readers in SA provide valuable information regarding the bleak outlook of some banks prior to the 2007–2009 financial crisis.

We conduct several additional tests to corroborate our main hypothesis. First, we find that the fraction of negative words contained in SA articles in the pre-crisis period also predicts bank default risk during the crisis period. That is, banks that were described with a higher fraction of negative words in SA articles in the pre-crisis period 1) tended to have greater increases in expected default probability during the crisis and 2) had more increases in loan defaults in the crisis, indicating lower loan quality. These two alternative measures of bank performance also indicate the economic channels through which the wisdom of crowds can predict bank stock performance in a crisis. Third, our results are robust after controlling for the content of Dow Jones News Service (DJNS) articles. Moreover, the results also show that the content of DJNS articles does not significantly predict bank crisis stock performance. Hence, interestingly, the collective wisdom of contributors to SA seems to be greater than that of journalists in traditional media.

Finally, to address potential selection issues for the banks being mentioned on SA, we construct three additional tests. First, we use a propensity-score-matching methodology to mitigate the omitted variable problem.[10] We use banks that are mentioned in SA articles in the pre-crisis period as treated banks and others as control banks. The matching procedure starts with a probit regression using firm characteristics as independent variables.[11] Based on the matched samples, we then re-estimate the regressions and find consistent results. Second, we delete banks not mentioned on the SA website and re-estimate the regressions. Third, to avoid the possibility that our results are driven by the distressed banks during the crisis, we delete them and re-estimate the regressions. All the results remain supportive of our main hypothesis after correcting for the potential selection bias.

Our paper contributes to the literature in three ways. First, we provide new insights into the debate on whether the 2007–2009 financial crisis in general or, more specifically, the influence of the crisis on bank performance, was foreseen by some market participants. Existing studies suggest that, while insiders of financial firms and regulators of financial markets were not aware of the imminent crisis (Fahlenbrach and Stulz, 2011; Cheng et al., 2014), certain institutional investors and analysts exhibited some awareness (Brockman et al., 2015). Our results indicate that the negative tone of SA articles and comments prior to the financial crisis contains predictive information. This is a novel finding with practical implications, since bank managers and financial market regulators may be able to learn from the wisdom of crowds regarding the potential impact of the next financial crisis on the banking sector.

Second, our paper adds to the investment literature on textual analysis, which has shown that the expression of negative sentiments in different documents (e.g., 10-Ks, newspaper articles) can predict stock returns (Loughran and McDonald, 2011; Dougal et al., 2012; Garcia, 2013; Ahern and Sosyura, 2014; Chen et al., 2014; Lo and Chau, 2019). We complement this line of research by showing that the wisdom of crowds in social media contains useful information regarding bank performance during the financial crisis. Importantly, our design differs from and complements these studies that mainly focus on short-term return predictability, since we study the influence of investor opinions for the 30 months prior to the crisis period and their predictability on bank returns over the next 30 crisis months.[12]

In this sense, our paper differs from Chen et al. (2014) in three major respects. First, we study the influence of investor opinions for the 30 months prior to the crisis period and their predictability on bank returns over the next 30 crisis months, which is different from the short-term return predictability in their design. Second, our results indicate that the negative tone of SA articles and comments prior to the financial crisis contains predictive information, which provides a policy suggestion that financial market regulators may be able to learn from the wisdom of crowds regarding the potential occurrence of the next financial crisis. Third, we further extend the analysis to bank lending quality and default risk, which is related to the individual bank stability and bank system stability.

Third, our research also adds to the financial crisis literature.[13] Recently, a plethora of studies have focused on *why* some banks performed relatively poorly in the financial crisis. These banks include those with CEOs who had better incentives, riskier business models, more shareholder-friendly boards and fragile financing, lower quality of regulatory capital ratios, worse risk management, more assets related to securitization tranches, and overconfident CEOs.[14] We add to these studies by focusing on *who* may have possessed *what* information of the impending crisis.

2. Data and descriptive statistics

In this section, we describe our data sources and present descriptive statistics of the variables of interest.

2.1. Sample

Our sample comprises all US financial institutions with SIC codes between 6000 and 6399. These financial institutions consist of four groups: depository institutions (SIC 6000–6099), non-depository credit institutions (SIC 6100–6199), investment intermediaries (SIC 6200–6299), and insurance (SIC 6300–6399).[15] Following Fahlenbrach et al. (2012), the term "financial crisis" refers to the period from July 2007 to December 2009. We use January 2005 to June 2007 as the pre-crisis period for our analysis of investor activities in SA.[16]

The dependent variables that capture bank crisis performance are *RE09* (the annualized buy-and-hold returns from July 1, 2007 through December 31, 2009), ΔNPL (the level change in the ratio of nonperforming loans to total gross loans between crisis years

[10] See the detailed discussion of the propensity-score-matching methodology in Section 4.3.2.

[11] For the robustness of our results, we use several different matching methods: Nearest neighbors (n = 1), Nearest neighbors (n = 2), and Mahalanobis.

[12] There are several reasons why we focus on the long-term returns rather than short-term returns. First, the prior studies on the 2007–2009 financial crisis almost exclusively focus on the long-term returns (e.g., Fahlenbrach and Stulz, 2011; Fahlenbrach et al., 2012; Beltratti and Stulz, 2012; Berger and Bouwman, 2013; Ellul and Yerramilli, 2013; Erel et al., 2014; Ho et al., 2016). Since our main goal is to explore the role of wisdom of the crowd in the financial crisis, it would be easier to interpret our results if we stick to the conventional return duration of the literature. Second, related to the first point, there is no clear-cut and well agreed definition of short-term returns for the crisis performance.

[13] See Fahlenbrach et al. (2012), Beltratti and Stulz (2012), Boyd et al. (2019), Beutel et al. (2019), Park and Shin (2020), Drago and Gallo (2020), respectively.

[14] See Fahlenbrach and Stulz (2011), Fahlenbrach et al. (2012), Beltratti and Stulz, (2012), Berger and Bouwman (2013), Ellul and Yerramilli (2013), Erel et al. (2014), Ho et al. (2016), and Lin et al. (2020).

[15] For brevity, we use the term "banks" to represent "financial institutions" in this paper.

[16] We find similar results when we use different definitions of the pre-crisis period, such as January 2006 to June 2007.

2007–2009 and year 2006), and Δ*EDF* (the level change in expected default frequency (EDF) between crisis years 2007–2009 and year 2006).

The controls variables are *BHAR06* (the buy-and-hold returns from July 1, 2006, through June 30, 2007), *LnAssets* (log of total assets), *BM* (book value of common equity divided by market value of common equity), *Beta* (bank equity beta from a market model of daily returns in excess of three-month T-bills using previous two-year data, where the market is represented by the value-weighted CRSP index), *TCE ratio* (tangible common equity ratio: tangible common equity divided by tangible assets and multiplied by 100), *Idiosyncratic volatility (IDIORISK*, standard deviations of the residuals obtained from a market model of daily returns in excess of three-month T-bills using previous two-year data), Δ*EPS* (percent change in earnings per share), Δ*EA* (percent change in book value of equity to total assets), Δ*TNII_NI* (percent change in non-interest income to net income), *ASSETGROW* (percent change in total asset), *ROE* (ratio of net income to total equities), *Leverage* (ratio of assets to book value of equity), and *EDF* (expected default frequency measure of the firm, it is the percentile ranking of a firm's default risk based on its distance to default (constructed from Bharath and Shumway, 2008)).

2.2. Seeking Alpha

Seeking Alpha is one of the biggest investment-related social media websites in the United States. For instance, Seeking Alpha had between 500,000–1 million unique visitors per day in August 2013. Specifically, the website's goal is to provide "*opinion and analysis rather than news, and it is primarily written by investors who describe their personal approach to stock picking and portfolio management, rather than by journalists*" (Seeking Alpha, 2012).[17]

Following the methodology of the study of Chen et al. (2014), we download all opinion articles published on the SA website from January 2005 to June 2007. However, we focus on the bank sample. We wrote a computer program to automate the process for downloading articles from SA and extracting relevant information from the downloaded HTML files.[18] SA assigns a unique identity number to each article, and SA editors tag each article with one or more stock tickers prior to publication. Hence, one article could include more than one company. In multiple-ticker articles, the opinions for each of the tagged stocks become more difficult to calculate. To avoid this issue, we focus only on single-ticker articles. Our sample total 1021 single-ticker articles, for which we collect information including article ID, title, date of publication, author name, stock ticker, and main text.

SA allows investors not only to write and read articles but also to post comments in response. Hence, we also download all comments written in response to our sample articles. In our sample, 79.3 percent of the comments were posted on the day of article publication, an additional 13.4 percent were posted on the ensuing calendar day, with others posted sporadically over ensuing weeks. We collect 261 comments written in response to our sample articles. We assume that comments written in response to an article mostly pertained to that article, and that the relevant stock was discussed in that article. We then collect the same information for each comment as that for the articles.

Similar to the methodology in prior literature, we examine the frequency of negative words used in an article to capture the tone of authors' opinions. We then use the negative word list proposed by Loughran and McDonald (2011) to characterize the views expressed in SA articles and SA comments.[19]

The main independent variables are as follows: *NegSA* is the average percentage of negative words across all Seeking Alpha articles published about bank *i* during the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; *I(SA)* is a dummy variable that takes a value of one if bank *i* has at leastarticle published on SA in the pre-crisis period, and zero otherwise; *NegSA_Comment* is the average fraction of negative words across all SA comments on bank *i* in the pre-crisis period if there were any such comments, and zero otherwise; *Consistency* is author consistency based on each author's published articles in the past three years, based on whether the stock's three-month performance reflects the article's position. We set zero if the author has published no articles in the past three years, since the consistency of such articles cannot be calculated.

2.3. Descriptive statistics

Table 1 presents the summary statistics, which include mean, standard deviation, minimum, maximum, and quartiles. We winsorize all variables at the 1st and 99th percentiles to mitigate the impact on our results from the extreme values. The table shows that, not surprisingly, bank stock returns were quite negative during the financial crisis. The average annualized buy-and-hold returns for the period from July 2007 to December 2009 (*RE09*) is −24.41 percent. The average Δ*EDF* and Δ*NPL* are 0.5290 and 9.69 percent, respectively.

Specifically, the overall sentiment expressed in SA during the pre-crisis period is negative. The average *NegSA* and *NegSA_Comment* are 0.0037 and 0.0002, respectively. The average *I(SA)* and *Consistency* are 0.1320 and 0.0646, respectively. The averages of firm characteristics are similar to those cited in the prior literature. The average *BHAR06*, *LnAssets*, *BM*, *Beta*, *TCE ratio*, *IDIORISK*, Δ*EPS*, Δ*EA*, Δ*TNII_NI*, *ASSETGROW*, *ROE*, *Leverage*, and *EDF* are −0.0086, 13.0520, 0.9398, 0.9372, 13.8201, 0.0217, 0.5008, 0.0032, 0.1682, 0.1624, 0.1114, 9.4611, and 0.0397, respectively.

In addition, in Fig. 1, we present a scatter plot between negative sentiment (*NegSA*) and crisis return (*RE09*) in Seeking Alpha data. For brevity, we skip the samples with zero value of *NegSA* in this figure. We find that the negative relationship between two variables, supporting that the fraction of negative words contained in Seeking Alpha articles in the pre-crisis period can predict bank stock returns in the crisis period.

3. Empirical results

In this section, we first test the relationship between the fraction of negative words contained in SA articles in the pre-crisis period and bank stock returns in the financial crisis. We then examine whether the fraction of negative words in SA comments in the pre-crisis period can also predict bank stock returns in the financial crisis. Finally, we explore whether authors' track records are related to the predictive power of their SA articles.

3.1. Seeking Alpha articles and stock returns in the 2007–2009 financial crisis

We use the following ordinary least-squares (OLS) regression to investigate whether the fraction of negative words contained in

[17] See the detail introduction and discussions on their website (http://seekingalpha.com).

[18] Our program can directly access a PostgreSQL database and store the extracted information therein.

[19] http://www.nd.edu/~mcdonald/Word_Lists.html.

Table 1
Summary statistics of variables.

	Mean	S.D.	Min	25th	Median	75th	Max
RE09	−0.2441	0.2869	−0.8769	−0.4253	−0.2088	−0.0484	0.4546
ΔEDF	0.5290	0.3591	−0.3532	0.2286	0.5962	0.8441	0.9990
ΔNPL	0.0969	0.1834	−0.0041	0.0146	0.0343	0.0812	1.0424
NegSA	0.0037	0.0074	0.0000	0.0000	0.0000	0.0000	0.0304
I(SA)	0.1320	0.3387	0.0000	0.0000	0.0000	0.0000	1.0000
NegSA_Comment	0.0002	0.0022	0.0000	0.0000	0.0000	0.0000	0.0055
NegDJNS	0.0003	0.0017	0.0000	0.0000	0.0000	0.0000	0.0119
I(DJNS)	0.0367	0.1881	0.0000	0.0000	0.0000	0.0000	1.0000
Consistency	0.0646	0.1870	0.0000	0.0000	0.0000	0.0000	0.7500
BHAR06	−0.0086	23.8538	−59.8438	−12.8493	−2.9342	9.7994	82.1180
LnAssets	13.0520	1.8361	9.7625	11.6820	12.8795	14.2211	17.5703
BM	0.9398	2.2233	0.1070	0.4535	0.6063	0.7764	18.9068
Beta	0.9372	0.1790	0.1052	0.8672	0.9618	1.0403	1.3927
TCE ratio	13.8201	27.8255	−1.1407	3.6183	5.2836	10.9241	198.5350
IDIORISK	0.0217	0.0129	0.0078	0.0139	0.0181	0.0252	0.0838
ΔEPS	0.5008	3.1822	−3.6800	−0.1600	0.0800	0.4299	14.5600
ΔEA	0.0032	0.0633	−0.1623	−0.0042	0.0017	0.0097	0.1715
ΔTNII_NI	0.1682	9.7402	−14.0762	−0.0505	0.0000	0.0662	6.6233
ASSETGROW	0.1624	0.4723	−1.1600	0.0271	0.0939	0.1923	0.4000
ROE	0.1114	0.0937	−0.2535	0.0722	0.1122	0.1499	0.3735
Leverage	9.4611	5.1476	1.1599	5.4294	9.7873	12.5296	28.8156
EDF	0.0397	0.1236	0.0000	0.0000	0.0000	0.0023	0.6557

This table presents summary statistics for the variables used in this study. The definitions of variables are presented in the Appendix.

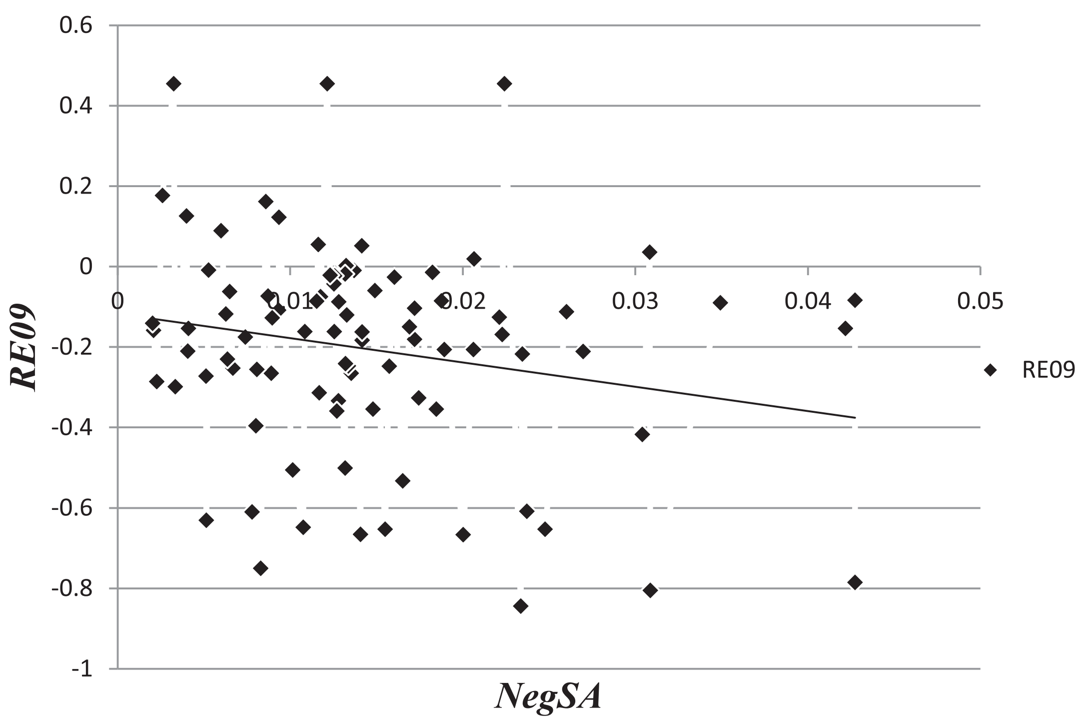

Fig. 1. Negative sentiment and crisis return.
This figure presents scatter plot between negative sentiment and crisis return in Seeking Alpha data. *RE09* represents stock returns for bank *i* in the financial crisis and *NegSA* is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007). For brevity, we skip the samples with zero value of *NegSA* in this figure.

SA articles in the pre-crisis period can predict bank stock returns during a crisis[20]:

$$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i \quad (1)$$

where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all SA articles published about bank *i* in the pre-crisis period if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about bank *i* was published in the SA in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are presented in the Appendix A. The *t*-statistics are in parentheses and are based on standard errors adjusted for

[20] We follow the study of Fahlenbrach et al. (2012) to set the regression model and the studies of Cooper et al. (2003) and Fahlenbrach et al. (2012) to choose the control variables.

Table 2
Seeking Alpha and bank crisis returns.

	(1) *RE09*	(2) *RE09*	(3) *RE09*	(4) *RE09*
NegSA	−6.6847***	−7.2188***	−7.1534***	−8.4879***
	(−3.23)	(−3.59)	(−3.81)	(−4.31)
I(SA)	0.1317**	0.1451**	0.1431**	0.1610***
	(2.20)	(2.14)	(2.19)	(2.57)
BHAR06	0.0029***	0.0026***	0.0026***	0.0022***
	(4.64)	(4.11)	(4.10)	(3.22)
LnAssets	−0.0110	−0.0058	−0.0053	−0.0011
	(−1.11)	(−0.75)	(−0.71)	(−0.18)
BM	−0.0032	−0.0059	−0.0059	−0.0003
	(−0.66)	(−1.04)	(−1.03)	(−0.04)
Beta	0.1996***	0.2073***	0.2085***	0.1952**
	(2.77)	(3.17)	(3.26)	(2.22)
TCE		0.0007	0.0007	0.0002
		(1.27)	(1.23)	(0.39)
IDIORISK			0.2601	−0.1268
			(0.24)	(−0.12)
ΔEPS				0.0032
				(0.86)
ΔEA				−0.2919**
				(−2.05)
ΔTNII_NI				−0.0016
				(−0.84)
ASSETGROW				−0.0409
				(−0.84)
ROE				−0.3163
				(−1.57)
Leverage				−0.0115***
				(−3.17)
EDF				−0.0539
				(−0.39)
Constant	−0.2891**	−0.3817***	−0.3946***	−0.3117**
	(−2.02)	(−3.27)	(−3.97)	(−2.12)
Adj-R^2	0.0636	0.0616	0.0600	0.0952
Obs.	677	568	568	565

This table presents OLS regression results for the Seeking Alpha articles and bank crisis returns. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009).[21]

Our first hypothesis argues that some authors in SA anticipated the imminent financial crisis and used a higher fraction of negative words in SA articles to describe some banks in the pre-crisis period. The banks with a higher percentage of negative descriptors in SA articles should perform worse during the crisis period. We thus expect the signs of coefficients β_1 in Eq. (1) to be negative.

Table 2 presents the regression results. We control for a number of important bank characteristics: *BHAR06*, *LnAssets*, *BM*, *Beta*, *TCE ratio*, *IDIORISK*, *ΔEPS*, *ΔEA*, *ΔTNII_NI*, *ASSETGROW*, *ROE*, *Leverage*, and *EDF*. The first specification controls for *BHAR06*, *LnAssets*, *BM*, and *Beta*. The second specification adds *TCE*. The third specification adds *IDIORISK*. The last specification adds *ΔEPS*, *ΔEA*, *ΔTNII_NI*, *ASSETGROW*, *ROE*, *Leverage*, and *EDF*. Adding these variables helps alleviate the concern that our main findings are driven by the negative performance persistence of some banks.

Across all specifications, the coefficients of *NegSA* are significantly negative. For example, in Model (4), the coefficient of *NegSA* is −4.1177 and is statistically significant. A one-standard-deviation increase in the *NegSA* is associated with a 3.05 percent (0.0074×4.1177) lower stock return during the 2007–2009 financial crisis. After controlling for more bank characteristics in Models (2) to (4), the economic magnitude of the coefficients becomes even larger. For example, in Model (4), the stock returns decrease by 4.41 percent (0.0074×5.54758) during the crisis for a one-standard-deviation increase in the pre-crisis change in *NegSA*.

This economic magnitude of 4.11 percent is approximately 50 percent of the risk-culture effect in Fahlenbrach et al. (2012), who find that a one-standard-deviation lower return during the LTCM crisis is associated with an 8.2 percent lower return during the financial crisis. Thus, the predictability of the wisdom of crowds in SA on bank stock returns during the crisis periods is both statistically significant and economically meaningful, supporting our first hypothesis.

3.2. Seeking Alpha and stock returns: comments from readers

To test our second hypothesis, we use the following OLS regression to examine whether the fraction of negative words contained in SA comments in the pre-crisis period can also predict bank stock returns during a crisis:

$$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \beta_3 NegSA\_Comment_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i \quad (2)$$

where $RE09_{i,crisis}$ represent stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all SA articles published about bank *i* in the pre-crisis period if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about bank *i* is published on SA in the pre-crisis period (January 2005 to June 2007), and zero otherwise; $NegSA\_Comment_{i,pre}$ is the average percentage of negative words across all SA comments published about bank *i* in the pre-crisis period if there were any such comments, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are presented in the Appendix A.

Our second hypothesis argues that some SA readers also anticipated the imminent financial crisis and used a higher fraction of negative words in their comments in response to articles on some banks in the pre-crisis period. We expect banks with more negative sentiments expressed in SA comments to perform worse during the crisis period. We thus anticipate the sign of coefficients β_3 in Eq. (2) to be negative.

Table 3 presents the regression results. First, across all specifications, the coefficients of *NegSA* are significantly negative, reconfirming our first hypothesis. Second, the coefficients of *NegSA_Comment* are also significantly negative, providing supporting evidence for our second hypothesis, i.e., that the fraction of negative words in SA comments also has predictive power for crisis returns of financial firms.

3.3. Seeking Alpha and stock returns: authors' track records

Our third hypothesis argues that articles written by authors with more consistent track records have stronger predictive power. Following Chen et al. (2014), we measure authors' track records by the consistency of each author. That is, $Consistency_{i,pre}$ is calculated by an author's published articles in the past three years, depending on whether the three-month performance of the discussed stock

[21] We find similar results if we perform regressions without clustering or clustering standard errors at 2-digit industry level.

Table 3
Seeking Alpha and bank crisis returns: Negative Comments.

	(1) *RE09*	(2) *RE09*	(3) *RE09*	(4) *RE09*
NegSA	−4.1678***	−4.6709**	−4.5284**	−5.5497***
	(−2.53)	(−2.43)	(−2.47)	(−4.90)
I(SA)	0.1073	0.1186	0.1151	0.1282***
	(1.71)	(1.63)	(1.63)	(2.80)
NegSA_Comment	−6.0430**	−4.5530*	−5.0058*	−4.3966*
	(−2.43)	(−1.89)	(−1.74)	(−1.71)
BHAR06	0.0029***	0.0025***	0.0025***	0.0021***
	(4.46)	(3.86)	(3.85)	(3.03)
LnAssets	−0.0116	−0.0065	−0.0056	−0.0016
	(−1.17)	(−0.86)	(−0.77)	(−0.59)
BM	−0.0034	−0.0062	−0.0062	−0.0008
	(−0.70)	(−1.08)	(−1.06)	(−0.10)
Beta	0.2077***	0.2153***	0.2184***	0.2073***
	(2.82)	(3.29)	(3.40)	(3.33)
TCE		0.0007	0.0007	0.0002
		(1.28)	(1.22)	(1.28)
IDIORISK			0.4991	0.1294
			(0.42)	(0.10)
ΔEPS				0.0042
				(0.96)
ΔEA				−0.3057***
				(−3.22)
ΔTNII_NI				−0.0015
				(−0.90)
ASSETGROW				−0.0384
				(−0.64)
ROE				−0.3163
				(−1.37)
Leverage				−0.0111***
				(−4.16)
EDF				−0.0571
				(−0.55)
Constant	−0.2904**	−0.3805***	−0.4058***	−0.3254***
	(−2.03)	(−3.30)	(−4.21)	(−3.10)
Adj-R^2	0.0596	0.0556	0.0543	0.0873
Obs.	677	568	568	565

This table presents OLS regression results for the Seeking Alpha articles, comments, and bank crisis returns. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \beta_3 NegSA\_Comment_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $NegSA\_Comment_{i,pre}$ is the average percentage of negative words across all Seeking Alpha comments published about bank *i* in the pre-crisis period if there were any such comments, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

reflects the article's position. We set zero if the author published no article in the past three years. Therefore, a higher value of author consistency means that the author has a better track record.

To test our third hypothesis, we use the following OLS regression to investigate whether authors' track records are related to the predictive power of their SA articles:

$$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \beta_6 NegSA \times Consistency_{i,pre} + \beta_7 Consistency_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i \quad (3)$$

where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $Consistency_{i,pre}$ is defined above; all other variables are similar to those in Eq. (1) and the variable definitions are presented in the Appendix A. We expect the sign of coefficient β_6 in Eq. (3) to be negative.

Table 4
Seeking Alpha and bank crisis returns: Author's track record.

	(1) *RE09*	(2) *RE09*	(3) *RE09*	(4) *RE09*
NegSA	5.4862*	−0.1134	0.0918	2.2355
	(1.74)	(−0.05)	(0.04)	(0.67)
I(SA)	−0.0055	0.0697	0.0643	0.0119
	(−0.09)	(0.76)	(0.72)	(0.20)
NegSA × Consistency	−19.1374**	−8.6230**	−8.8238*	−14.8348**
	(−2.49)	(−2.01)	(−1.89)	(−2.00)
Consistency	0.2021*	0.0787	0.0838	0.2111**
	(1.95)	(0.70)	(0.75)	(2.36)
BHAR06	0.0029***	0.0025***	0.0025***	0.0021***
	(4.47)	(3.78)	(3.77)	(2.81)
LnAssets	−0.0113	−0.0062	−0.0056	−0.0017
	(−1.12)	(−0.84)	(−0.77)	(−0.28)
BM	−0.0032	−0.0060	−0.0060	−0.0006
	(−0.66)	(−1.07)	(−1.05)	(−0.10)
Beta	0.2021***	0.2093***	0.2111***	0.2032**
	(2.80)	(3.10)	(3.19)	(2.32)
TCE		0.0007	0.0007	0.0002
		(1.28)	(1.25)	(0.39)
IDIORISK			0.3348	−0.0173
			(0.30)	(−0.02)
ΔEPS				0.0044
				(1.17)
ΔEA				−0.2778*
				(−1.91)
ΔTNII_NI				−0.0016
				(−0.83)
ASSETGROW				−0.0375
				(−0.77)
ROE				−0.3282
				(−1.51)
Leverage				−0.0113***
				(−3.17)
EDF				−0.0593
				(−0.43)
Constant	−0.2890**	−0.3785***	−0.3951***	−0.3132**
	(−2.03)	(−3.30)	(−4.01)	(−2.16)
Adj-R^2	0.0585	0.0530	0.0515	0.0854
Obs.	677	568	568	565

This table presents OLS regression results for the Seeking Alpha articles and bank crisis returns by considering the author's track record. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \beta_6 NegSA \times Consistency_{i,pre} + \beta_7 Consistency_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $Consistency_{i,pre}$ is author consistency based on author's published articles in the past three years, determined based on whether a stock's three-month performance reflects the article's position. We set zero if no article is published by the author in the past three years; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4 presents the regression results. Across all specifications, the coefficients of *NegSA × Consistency* are all significantly negative, supporting our third hypothesis. Interestingly, the coefficients of *NegSA* become insignificant after we control for the author's track record. This result indicates that our previous finding in Section 3.1 is mainly driven by authors with good track records.

In sum, our results in Tables 2–4 not only support our three hypotheses but also complement the findings in Chen et al. (2014), who show that the fraction of negative words in SA comments can predict next-quarter stock returns. Our results suggest that authors and readers on SA were aware of the bleak outlook of some banks months before the global financial crisis occurred.

4. Additional supporting evidence

In this section, we provide further supportive evidence that aligns our findings with our main hypotheses. We first investigate the correlation between the fraction of negative words contained in SA articles and two alternative bank performance measures in the financial crisis (bank default risk and loan quality). Second, we further control for the tone of Dow Jones News Service (DJNS) articles in our regressions. Third, to mitigate the endogeneity issue, we adopt the propensity-score-matching methodology, delete the banks not mentioned in the SA website, and delete distressed banks to account for the omitted variable problem and sample selection bias.

4.1. Seeking alpha and bank performance: default risk

Fahlenbrach et al. (2012) find that banks with a riskier business model suffered more in stock returns and had a higher default probability during the 2007–2009 crisis period. Similarly, Ho et al. (2016) find that overconfident banks with a higher level of risk-taking behavior before the financial crisis had a higher expected default frequency (*EDF*) during the crisis. Therefore, in this subsection, we examine whether the fraction of negative words contained in SA articles can also predict default risk of banks during the financial crisis as the following regressions:

$$\Delta EDF_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i \qquad (4)$$

where $\Delta EDF_{i,crisis}$ represents the level change in expected default frequency (EDF) between year 2009 and year 2006. The EDF is the percentile ranking of a firm's default risk based on its distance to default (Bharath and Shumway, 2008). All other variables are similar to those in Eq. (1) and the variable definitions presented are in the Appendix A. Based on our first hypothesis, we expect the sign of coefficient β_1 in Eq. (4) to be positive.[22]

Table 5 presents the results for the SA articles and bank default risk in the financial crisis. The results show a significantly positive correlation between the fraction of negative words contained in SA articles and the level change ratio of expected default frequency during the financial crisis. For example, in Model (2), a one-standard-deviation increase of the number of negative words contained in SA articles in the pre-crisis period is associated with a 4.02 percent (0.0074 × 5.4373) increase in the level change ratio of expected default frequency. The result is consistent with the notion that some authors of articles on SA could foresee the higher default risk potential of some banks before the financial crisis.

Table 5
Seeking Alpha and default risk.

	(1) ΔEDF	(2) ΔEDF	(3) ΔEDF	(4) ΔEDF
NegSA	5.2071**	5.4373***	4.9737**	4.5250**
	(2.14)	(2.72)	(2.41)	(2.18)
I(SA)	0.0190	−0.0155	−0.0039	0.0038
	(0.26)	(−0.25)	(−0.06)	(0.06)
BHAR06	−0.0031**	−0.0025*	−0.0026*	−0.0024**
	(−2.48)	(−1.94)	(−1.93)	(−2.34)
LnAssets	−0.0155	−0.0196	−0.0260	−0.0213
	(−0.75)	(−0.70)	(−0.91)	(−0.70)
BM	−0.1704***	−0.1965***	−0.2311***	0.0693
	(−2.68)	(−3.05)	(−3.25)	(0.56)
Beta	−0.1642	−0.1634	−0.2033	−0.0805
	(−0.94)	(−0.78)	(−0.99)	(−0.38)
TCE		−0.0003	−0.0001	0.0006
		(−0.21)	(−0.09)	(0.67)
IDIORISK			−4.4910**	−3.5030*
			(−2.09)	(−1.82)
ΔEPS				−0.0036
				(−1.64)
ΔEA				0.5332***
				(2.91)
ΔTNII_NI				0.0072
				(1.47)
ASSETGROW				0.0563
				(1.55)
ROE				0.2808
				(1.07)
Leverage				0.0207***
				(4.11)
EDF				−0.9336***
				(−10.21)
Constant	0.9630***	1.0519***	1.2815***	0.7507***
	(7.15)	(5.65)	(5.28)	(2.79)
Adj-R^2	0.0411	0.0307	0.0433	0.1767
Obs.	497	440	440	439

This table presents OLS regression results for the Seeking Alpha articles and bank default risk in the financial crisis. The crisis period is from July 1, 2007, through December 31, 2009.
$\Delta EDF_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $\Delta EDF_{i,crisis}$ represents the change in the expected default frequency of bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

4.2. Seeking alpha and bank performance: loan quality

In addition to default risk, we further examine whether the fraction of negative words contained in SA articles also predicts the loan quality of banks during the 2007–2009 financial crisis as additionally supportive evidence to support our first hypothesis. Specifically, we perform the following regression:

$$\Delta NPL_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i \qquad (5)$$

where $\Delta NPL_{i,crisis}$ represents the level change in ratio of non-performing loans (NPL) to total gross loans between crisis years (2007–2009) and year 2006. Nonperforming loans are defined as loans with interest payments and principal more than 90 days overdue. All other variables are similar to those in Eq. (1). We expect the sign of coefficient β_1 in Eq. (5) to be positive.[23]

Table 6 presents the results for the SA articles and bank loan quality during the financial crisis. A significantly positive correlation exists between the fraction of negative words contained in SA articles and the level change ratio of non-performing loans during the financial crisis.

In sum, the results in Tables 5 and 6 support the argument that some authors in SA were aware that certain banks would perform poorly in the forthcoming crisis. Moreover, these results also demonstrate the economic channels through which authors of SA articles could detect the worst-performing banks in the crisis period.

[22] We find similar results if we use the level of EDF as the dependent variable in the regressions.

[23] We find similar results if we use the level of NPL as the dependent variable in the regressions.

Table 6
Seeking Alpha and bank loan quality.

	(1) ΔNPL	(2) ΔNPL	(3) ΔNPL	(4) ΔNPL
NegSA	2.8036***	2.4192***	2.5948***	3.0921***
	(20.61)	(14.53)	(17.15)	(18.73)
I(SA)	0.0027	0.0219	0.0192	−0.0102
	(0.07)	(0.99)	(0.91)	(−0.38)
BHAR06	−0.0015***	−0.0011**	−0.0010**	−0.0009*
	(−2.95)	(−2.47)	(−2.27)	(−1.85)
LnAssets	−0.0110**	−0.0159***	−0.0148***	−0.0217***
	(−2.00)	(−11.32)	(−11.33)	(−17.98)
BM	−0.0028***	−0.0032**	−0.0031**	−0.0065***
	(−2.78)	(−2.22)	(−2.03)	(−4.68)
Beta	0.0197	0.0630***	0.0800***	0.0112
	(1.04)	(5.81)	(5.74)	(0.28)
TCE		0.0021	0.0020	0.0017
		(1.03)	(0.98)	(1.14)
IDIORISK			0.9326***	0.0793
			(3.44)	(0.19)
ΔEPS				−0.0066***
				(−4.24)
ΔEA				0.0724
				(0.29)
ΔTNII_NI				−0.0004
				(−0.39)
ASSETGROW				0.1762***
				(11.29)
ROE				0.8136***
				(3.84)
Leverage				−0.0026
				(−1.44)
EDF				0.1230
				(0.72)
Constant	0.2050**	0.2139***	0.1666***	0.4344***
	(4.33)	(12.09)	(5.14)	(11.03)
Adj-R²	0.0142	0.0154	0.0151	0.0474
Obs.	403	367	367	366

This table presents OLS regression results for the Seeking Alpha articles and bank loan quality in the financial crisis. The crisis period is from July 1, 2007, through December 31, 2009.
$\Delta NPL_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $\Delta NPL_{i,crisis}$ represents the change in the nonperforming loan ratio of bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

4.3. Robustness checks

4.3.1. Controlling for the articles of Dow Jones News Service

One may be concerned that the predictive power of SA articles is merely a reflection of what has been reported in traditional media outlets such as newspapers. Similar to Chen et al. (2014), we mitigate this concern by controlling for the information revealed through Dow Jones News Service (DJNS) articles. We collected DJNS articles for each stock covered by the single-ticker SA articles from the Factiva database. Because DJNS articles are not tagged by company name or stock ticker, we had to use a search query to find the matched news articles for each stock from January 2005 to June 2007.[24]

In a manner similar to the method we used for the SA articles, we collected the information of each DJNS article regarding article title, date of publication, and main text. We then constructed two DJNS related variables: *NegDJNS* is the average fraction of negative words across all articles published in the DJNS about bank *i* in the pre-crisis period if there were any such articles, and zero otherwise. *I*(*DJNS*) is a dummy variable that takes a value of one if at least one article about bank *i* was published in the Dow Jones News Service in the pre-crisis period, and zero otherwise. Specifically, we perform the following regression:

$$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \beta_3 NegSA\_Comment_{i,pre} + \beta_4 NegDJNS_{i,pre} + \beta_5 I(DJNS)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i \quad (6)$$

where *NegDJNS* and *I*(*DJNS*) are defined above and other variables are similar to those in Eq. (1). We expect that the sign of coefficient β_1 in Eq. (6) remains negative after we control for the news released in traditional media outlets, such as DJNS articles.

Table 7 presents the results. First, in all models, the correlations between the fraction of negative words contained in SA articles and the bank's stock returns during the financial crisis are significantly negative, confirming our first hypothesis. Interestingly, the correlations between the fraction of negative words contained in DJNS articles and stock returns during the financial crisis are insignificant. This result suggests that the collective wisdom of the authors on SA seems more insightful concerning the bleak future of certain banks, compared with those in DJNS articles.

4.3.2. Propensity-score-matching methodology to address potential omitted variables and selection bias

To address the potential concerns regarding potential omitted variables and selection bias, we use the propensity-score-matching approach and construct an optimal control firm sample in the same spirit of Houston et al. (2014) and Hasan et al. (2014).[25]

We use banks mentioned in SA articles in the pre-crisis period as the treated group and other banks as the control. The matching procedure starts with a probit regression model using all firm characteristics (*BHAR06*, *LnAssets*, *BM*, *Beta*, *TCE ratio*, *IDIORISK*, Δ*EPS*, Δ*EA*, Δ*TNII_NI*, *ASSETGROW*, *ROE*, *Leverage*, and *EDF*) as control variables. For the robustness of this exercise, we adopt three different matching methods: Nearest neighbors (n = 1), Nearest neighbors (n = 2), and Mahalanobis.

Table 8 shows that the results remain qualitatively similar. In all models, the correlation between the fraction of negative words contained in SA articles and stock returns during the financial crisis is significantly negative. This alleviates the concern that our main finding might be driven by potential omitted variables.

4.3.3. Subsample analysis for addressing sample selection bias

Our results might stem from the selection bias of authors in SA. To mitigate the concern, we further re-do our analysis based on the subsample that excludes banks not mentioned on the SA website. By doing so, we reduce the concern that only the bad news is selected from SA articles.

Table 9 shows a significantly negative correlation between the fraction of negative words contained in SA articles and bank crisis returns. Hence, even when we exclude banks not mentioned on the SA website, the fraction of negative words contained in SA articles in the pre-crisis period can still predict the poor stock performance of banks.

Another concern is that our main results cloud been driven by the distressed banks during the crisis. In our sample, there are 15

[24] Following Chen et al. (2014), we use each company's name in the CRSP database and require that these names be mentioned at least once in the first 50 words of the DJNS articles.

[25] This approach is developed by Rosenbaum and Rubin (1983) and Heckman et al. (1997, 1998).

Table 7
Seeking Alpha and bank crisis returns: Controlling DJNS articles.

	(1) *RE09*	(2) *RE09*	(3) *RE09*	(4) *RE09*
NegSA	−4.6768***	−5.0947***	−4.9014***	−6.0678***
	(−2.90)	(−2.85)	(−2.66)	(−3.13)
I(SA)	0.1200*	0.1293	0.1264	0.1421**
	(1.81)	(1.62)	(1.63)	(1.98)
NegSA_Comment	−5.7700**	−4.5847*	−5.0419*	−4.3647
	(−2.35)	(−1.89)	(−1.70)	(−1.63)
NegDJNS	9.6366	11.3128	11.0527	14.2762
	(1.15)	(1.30)	(1.20)	(1.70)
I(DJNS)	−0.0928	−0.1137*	−0.1167*	−0.1478**
	(−1.31)	(−1.83)	(−1.86)	(−2.04)
BHAR06	0.0030***	0.0026***	0.0026***	0.0022***
	(4.42)	(3.98)	(3.98)	(2.99)
LnAssets	−0.0117	−0.0069	−0.0058	−0.0017
	(−1.25)	(−1.01)	(−0.86)	(−0.29)
BM	−0.0036	−0.0064	−0.0064	−0.0010
	(−0.71)	(−1.04)	(−1.02)	(−0.15)
Beta	0.2097***	0.2206***	0.2237***	0.2137**
	(2.80)	(3.35)	(3.51)	(2.47)
TCE		0.0007	0.0007	0.0002
		(1.22)	(1.15)	(0.34)
IDIORISK			0.5047	0.1049
			(0.39)	(0.09)
ΔEPS				0.0041
				(1.10)
ΔEA				−0.2778*
				(−1.76)
ΔTNII_NI				−0.0014
				(−0.75)
ASSETGROW				−0.0378
				(−0.82)
ROE				−0.3356
				(−1.54)
Leverage				−0.0112***
				(−3.16)
EDF				−0.0597
				(−0.42)
Constant	−0.2897**	−0.3803***	−0.4069***	−0.3253***
	(−2.10)	(−3.71)	(−4.91)	(−2.69)
Adj-R^2	0.0579	0.0536	0.0523	0.0862
Obs.	677	568	568	565

This table presents OLS regression results for Seeking Alpha articles and bank crisis returns by controlling the effect of Dow Jones News Service (DJNS) articles. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \beta_3 NegSA\_Comment_{i,pre} + \beta_4 NegDJNS_{i,pre} + \beta_5 I(DJNS)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in the Seeking Alpha in the pre-crisis period, and zero otherwise;; $NegSA\_Comment_{i,pre}$ is the average percentage of negative words across all Seeking Alpha comments published about the bank *i* in the pre-crisis period if there were any such articles, and zero otherwise; $NegDJNS_{i,pre}$ is the average percentage of negative words across all DJNS articles published about bank *i* in the pre-crisis period if there were any such articles, and zero otherwise; $I(DJNS)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank was published in the DJNS in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

distressed banks during the crisis. Thus, we delete these samples and re-estimate the regressions. Table 10 again shows a significantly negative correlation between the fraction of negative words contained in SA articles and bank crisis returns. Hence, even when we exclude distressed banks during the crisis, our main results still hold.

Table 8
Seeking Alpha and bank crisis returns: Propensity-score-matching approach.

Matching method	Nearest neighbors (n = 1) (1) *RE09*	Nearest neighbors (n = 2) (2) *RE09*	Mahalanobis (3) *RE09*
NegSA	−4.8380***	−4.1814**	−4.8380***
	(−2.82)	(−2.30)	(−2.82)
I(SA)	0.1460**	0.1030*	0.1460**
	(2.02)	(1.65)	(2.02)
BHAR06	0.0032**	0.0023	0.0032**
	(2.23)	(1.25)	(2.23)
LnAssets	−0.0284***	−0.0228*	−0.0284***
	(−3.13)	(−1.65)	(−3.13)
BM	0.0030	0.0102***	0.0030
	(0.37)	(2.71)	(0.37)
Beta	0.2166	−0.2292	0.2166
	(0.90)	(−0.94)	(0.90)
TCE	−0.0003	−0.0007	−0.0003
	(−0.23)	(−0.78)	(−0.23)
IDIORISK	0.3673	2.1216	0.3673
	(0.23)	(0.74)	(0.23)
ΔEPS	−0.0115	−0.0030	−0.0115
	(−1.35)	(−0.30)	(−1.35)
ΔEA	0.1138	−0.1407	0.1138
	(0.31)	(−0.64)	(0.31)
ΔTNII_NI	0.0183	−0.0132**	0.0183
	(0.49)	(−2.28)	(0.49)
ASSETGROW	0.0024	−0.0092	0.0024
	(0.09)	(−0.29)	(0.09)
ROE	−0.0255	−0.3468	−0.0255
	(−0.07)	(−0.83)	(−0.07)
Leverage	−0.0092**	−0.0106**	−0.0092**
	(−2.10)	(−2.24)	(−2.10)
EDF	0.0783	−0.3344	0.0783
	(0.26)	(−1.15)	(0.26)
Constant	0.0019	0.4318	0.0019
	(0.01)	(1.50)	(0.01)
Adj-R^2	0.0665	0.0492	0.0665
Obs.	138	140	138

This table presents OLS regression results for Seeking Alpha articles and bank crisis returns by using a propensity-score-matching approach. We use the banks mentioned in Seeking Alpha articles in the pre-crisis period as representing the treatment banks and the others as control banks. Matching starts with a probit regression using various all firm characteristics (*BHAR06, BM, Beta, TCE,* and *IDIORISK*) as control variables. For the robustness of our results, we use several different matching methods: Nearest neighbors (n = 1), Nearest neighbors (n = 2), and Mahalanobis. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

5. Conclusion

Existing studies have shown that investor opinions expressed in social media can predict stock returns and earnings surprises in the next quarter. Our study sheds additional light on this literature by focusing on whether the wisdom of crowds expressed through articles and comments published on Seeking Alpha (SA) before the financial crisis predicted bank stock returns during the crisis. Moreover, we explore whether the negative sentiments expressed in SA comments and authors' track records also play a role in return predictability.

Our results provide convincing evidence that the fraction of negative words contained in SA articles in the pre-crisis period (January 2005 to June 2007) is negatively correlated to bank stock returns

Table 9
Seeking Alpha and bank crisis returns: Reduced subsample.

	(1)	(2)	(3)	(4)
	RE09	*RE09*	*RE09*	*RE09*
NegSA	−3.5039*	−4.0369***	−3.6299**	−3.5378**
	(−2.06)	(−2.55)	(−2.33)	(−2.13)
BHAR06	0.0018	0.0013	0.0014	0.0025
	(1.16)	(0.72)	(0.70)	(1.45)
LnAssets	−0.0263**	−0.0230	−0.0205	−0.0156
	(−2.40)	(−1.62)	(−1.38)	(−1.66)
BM	0.0027	0.0019	0.0022	0.0054
	(0.63)	(0.45)	(0.51)	(0.65)
Beta	0.0115	0.0757	−0.0244	−0.3568
	(0.07)	(0.26)	(−0.07)	(−1.05)
TCE		−0.0004	−0.0004	−0.0001
		(−0.37)	(−0.36)	(−0.09)
IDIORISK			1.6193	2.1774
			(0.76)	(1.17)
ΔEPS				−0.0105**
				(−2.00)
ΔEA				0.0259
				(0.10)
ΔTNII_NI				−0.0801
				(−1.27)
ASSETGROW				0.0234**
				(2.10)
ROE				−0.4778
				(−1.13)
Leverage				−0.0018
				(−0.35)
EDF				−0.0917
				(−0.25)
Constant	0.2257	0.1320	0.1575	0.4971
	(0.99)	(0.28)	(0.33)	(1.05)
Adj-R^2	0.0699	0.0554	0.0610	0.1486
Obs.	90	69	69	69

This table presents OLS regression results for the Seeking Alpha articles and bank crisis returns by using a reduced sample. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

in the financial crisis (July 2007 to December 2009). We also find that the fraction of negative words in SA comments in the pre-crisis period contains valuable information regarding bank stock returns in the financial crisis. The results also indicate that author track record is positively related to the predictive power of SA articles.

We find consistent results when we use expected default frequency and non-performing loans as alternative measures of bank performance during the crisis. This finding also demonstrates economic channels through which the authors of articles published in SA would be able to predict bank stock returns. Moreover, our results are robust to controlling for the information released in Dow Jones News Service (DJNS) articles. The results also hold when we adopt the propensity-score-matching methodology, exclude banks not mentioned in SA, and delete distressed banks during the crisis.

Since bank insiders and regulators did not foresee the 2007–2009 global financial crisis, our paper provides an intriguing and novel perspective regarding the wisdom of crowds: Market participants and regulators may be able to learn from the collective wisdom of contributors to social media regarding the performance of the financial sector in the next financial crisis.

Table 10
Seeking Alpha and bank crisis returns: Survival subsample.

	(1) *RE09*	(2) *RE09*	(3) *RE09*	(4) *RE09*
NegSA	−0.7195***	−0.6946***	−0.6923***	−0.7612***
	(−14.13)	(−15.19)	(−15.36)	(−14.37)
I(SA)	0.0273	0.0277	0.0259	0.0293
	(0.67)	(0.58)	(0.54)	(0.69)
BHAR06	0.0028***	0.0025***	0.0025***	0.0023***
	(4.50)	(4.04)	(4.03)	(3.17)
LnAssets	−0.0075	−0.0008	−0.0000	0.0034
	(−0.78)	(−0.10)	(−0.00)	(0.49)
BM	−0.0038	−0.0067	−0.0067	−0.0017
	(−0.75)	(−1.14)	(−1.12)	(−0.26)
Beta	0.2134***	0.2180***	0.2198***	0.2110**
	(2.73)	(3.04)	(3.09)	(2.37)
TCE		0.0006	0.0006	0.0003
		(1.17)	(1.12)	(0.49)
IDIORISK			0.3849	0.0677
			(0.36)	(0.06)
ΔEPS				0.0022
				(0.59)
ΔEA				−0.2931**
				(−2.03)
ΔTNII_NI				−0.0003
				(−0.18)
ASSETGROW				−0.0299
				(−0.65)
ROE				−0.2966
				(−1.33)
Leverage				−0.0088**
				(−2.35)
EDF				0.0194
				(0.13)
Constant	−0.3327**	−0.4381***	−0.4580***	−0.3962***
	(−2.80)	(−4.98)	(−5.20)	(−3.49)
Adj-R^2	0.0670	0.0690	0.0676	0.0854
Obs.	661	553	553	550

This table presents OLS regression results for the Seeking Alpha articles and bank crisis returns by using a survival sample. The crisis period is from July 1, 2007, through December 31, 2009.
$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$
where $RE09_{i,crisis}$ represents stock returns for bank *i* in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank *i* in the year 2006. The variable definitions are given in the Appendix. The *t*-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Appendix A. Variable definitions

Variable	Definition	Data Source
Panel A: Crisis performance variables		
RE09	The annualized buy-and-hold returns from July 1, 2007 through December 31, 2009.	CRSP
ΔEDF	Level change in expected default frequency (EDF) between crisis years (2007–2009) and year 2006. The EDF is the percentile ranking of a firm's default risk based on its distance to default (constructed from Bharath and Shumway, 2008).	Compustat and CRSP
ΔNPL	Level change in ratio of nonperforming loans to total gross loans between crisis years (2007–2009) and year 2006. Nonperforming loans are defined as loans with interest payments and principal more than 90 days overdue.	Compustat
Panel B: Seeking Alpha (SA) related variables		
NegSA	The average percentage of negative words across all Seeking Alpha articles published about bank *i* in the pre-crisis period (2005/1–2007/6) if there were any such articles, and zero otherwise.	
I(SA)	A dummy variable that takes a value of one if at least one article about a bank published in Seeking Alpha in the pre-crisis period (January 2005 to June 2007), and zero otherwise.	
NegSA_Comment	The average fraction of negative words across Seeking Alpha comments regarding bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such comments, and zero otherwise.	
Consistency	Author consistency based on author's published articles in the past three years, determined based on whether a stock's three-month performance reflects the article's position. We set zero if no article is published by the author in the past three years.	
Panel C: Bank characteristics		
BHAR06	The buy-and-hold returns from July 1, 2006, through June 30, 2007.	CRSP
LnAssets	Log of total assets (U.S. billion) on December 31, 2006.	Compustat
BM	Book value of common equity divided by market value of common equity on December 31, 2006.	Compustat and CRSP
Beta	Bank's equity beta from a market model of daily returns in excess of three-month T-bills from January 2004 to December 2006, where the market is represented by the value-weighted CRSP index.	CRSP
TCE ratio	Tangible common equity ratio: tangibly common equity divided by tangible assets and multiplied by 100.	Compustat
Idiosyncratic volatility (IDIORISK)	Standard deviation of the residuals obtained from a market model of daily returns in excess of three-month T-bills from January 2004 to December 2006, where the market is represented by the value-weighted CRSP index.	CRSP
ΔEPS	Percent change in earnings per share.	Compustat
ΔEA	Percent change in book value of equity to total assets.	Compustat
ΔTNII_NI	Percent change in non-interest income to net income.	Compustat
ASSETGROW	Percent change in total asset.	Compustat
ROE	Ratio of net income to total equities.	Compustat
Leverage	Ratio of assets to book value of equity.	Compustat
EDF	Expected default frequency measure of the firm. It is the percentile ranking of a firm's default risk based on its distance to default (constructed from Bharath and Shumway, 2008).	Compustat and CRSP
Panel D: Dow Jones News Service related variables		
NegDJNS	The average fraction of negative words across all articles published in Dow Jones News Service about bank *i* in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and zero otherwise.	
I(DJNS)	A dummy variable that takes a value of one if at least one article about a bank was published in the Dow Jones News Service in the pre-crisis period (January 2005 to June 2007), and zero otherwise.	

Appendix B. Robustness checks for Seeking Alpha in bank crisis returns

This table presents OLS regression results for the for Seeking Alpha in bank crisis returns. The crisis period is from July 1, 2007, through December 31, 2009. In our prior main results, we set article tone for banks without articles to zero. In this table, we set article tone for banks without articles to the mean value or median value of our sample.

$$RE09_{i,crisis} = \alpha + \beta_1 NegSA_{i,pre} + \beta_2 I(SA)_{i,pre} + \gamma Z_{i,t-1} + \varepsilon_i$$

where $RE09_{i,crisis}$ represents stock returns for bank i in the financial crisis; $NegSA_{i,pre}$ is the average percentage of negative words across all Seeking Alpha articles published about bank i in the pre-crisis period (January 2005 to June 2007) if there were any such articles, and otherwise are setting at mean value or median value of our sample; $I(SA)_{i,pre}$ is a dummy variable that takes a value of one if at least one article about a bank were published in Seeking Alpha in the pre-crisis period, and zero otherwise; $Z_{i,t-1}$ is a vector of control variables for bank i in the year 2006. The variable definitions are given in the Appendix. The t-statistics are in parentheses and are based on standard errors adjusted for heteroskedasticity (White, 1980) and industry clustering at three-digit SIC code (Petersen, 2009). The superscripts *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: Set article tone for banks without articles to the mean value of our sample (1) *RE09*	Panel B: Set article tone for banks without articles to the median value of our sample (2) *RE09*
NegSA	−5.8565**	−5.8753**
	(−2.25)	(−2.29)
I(SA)	0.0484	0.0541
	(1.06)	(1.14)
BHAR06	0.0021***	0.0021***
	(2.77)	(2.77)
LnAssets	−0.0021	−0.0021
	(−0.34)	(−0.34)
BM	−0.0006	−0.0006
	(−0.10)	(−0.10)
Beta	0.2007**	0.2005**
	(2.30)	(2.30)
TCE	0.0002	0.0002
	(0.41)	(0.41)
IDIORISK	0.0099	0.0061
	(0.01)	(0.01)
ΔEPS	0.0042	0.0042
	(1.11)	(1.11)
ΔEA	−0.2779*	−0.2786*
	(−1.93)	(−1.93)
ΔTNII_NI	−0.0016	−0.0016
	(−0.83)	(−0.84)
ASSETGROW	−0.0385	−0.0386
	(−0.80)	(−0.80)
ROE	−0.3173	−0.3169
	(−1.48)	(−1.49)
Leverage	−0.0114***	−0.0114***
	(−3.13)	(−3.13)
EDF	−0.0566	−0.0565
	(−0.41)	(−0.41)
Constant	−0.2242	−0.2301
	(−1.34)	(−1.39)
Adj-R^2	0.0877	0.0878
Obs.	565	565

References

Ahern, K., Sosyura, D., 2014. Who writes the news? Corporate press releases during merger negotiations. J. Finance 69, 241–291.
Bartov, E., Faurel, L., Mohanram, P.S., 2018. Can Twitter help predict firm-level earnings and stock returns? Account. Rev. 93, 25–57.
Beltratti, A., Stulz, R.M., 2012. The credit crisis around the globe: why did some banks perform better? J. Financ. Econ. 105, 1–17.
Berger, A.N., Bouwman, C.H., 2013. How does capital affect bank performance during financial crises? J. Financ. Econ. 109, 146–176.
Beutel, J., List, S., von Schweinitz, G., 2019. Does machine learning help us predict banking crises? J. Financ. Stab., 100693.
Bharath, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. Rev. Financ. Stud. 21, 1339–1369.
Boyd, J.H., De Nicolò, G., Rodionova, T., 2019. Banking crises and crisis dating: disentangling shocks and policy responses. J. Financ. Stab. 41, 45–54.
Brockman, P., Nikolic, B., Yan, X., 2015. Anticipating the 2007-2008 financial crisis: who knew what and when did they know it? J. Financ. Quant. Anal. 50, 647–669.
Chan, W.S., 2003. Stock price reaction to news and no–news: drift and reversal after headlines. J. Financ. Econ. 70, 223–260.
Chen, Y., Xie, J., 2008. Online consumer review: word-of-mouth as a new element of marketing communication mix. Manage. Sci. 54, 477–491.
Chen, H., De, P., Hu, Y., Hwang, B.H., 2014. Wisdom of crowds: the value of stock opinions transmitted through social media. Rev. Financ. Stud. 27, 1367–1403.
Cheng, I.H., Raina, S., Xiong, W., 2014. Wall Street and the housing bubble. Am. Econ. Rev. 104, 2797–2829.
Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Mark. Res. 43, 345–354.
Colander, D., Goldberg, M., Haas, A., Juselius, K., Kirman, A., Lux, T., Sloth, B., 2009. The financial crisis and the systemic failure of the economics profession. Crit. Rev. 21, 249–267.
Cooper, M.J., Jackson, W.E., Gary, A.P., 2003. Evidence of predictability in the cross-section of bank stock returns. J. Bank. Financ. 27, 817–850.
Da, Z., Huang, X., 2019. Harnessing the wisdom of crowds. Manage. Sci., forthcoming.
Das, S.R., Chen, M.Y., 2007. Yahoo! For Amazon: sentiment extraction from small talk on the web. Manage. Sci. 53, 1375–1388.
Deloitte, 2007. Most Consumers Read and Rely on Online Reviews; Companies Must Adjust.
Dougal, C., Engelberg, J., Garcia, D., Parsons, C., 2012. Journalists and the stock market. Rev. Financ. Stud. 25, 639–679.
Drago, D., Gallo, R., 2020. The impact of financial crises on the syndicated loan spreads applied to public and private firms. J. Financ. Stab., 100718.
Ellul, A., Yerramilli, V., 2013. Stronger risk controls, lower risk: evidence from US bank holding companies. J. Finance 68, 1757–1803.
Engelberg, J., Parsons, C., 2011. The causal impact of media in financial markets. J. Finance 66, 67–97.
Erel, I., Nadauld, T., Stulz, R.M., 2014. Why did holdings of highly rated scuritization tranches differ so much across Banks? Rev. Financ. Stud. 27, 404–453.
Fahlenbrach, R., Stulz, R.M., 2011. Bank CEO incentives and the credit crisis. J. Financ. Econ. 99, 11–26.
Fahlenbrach, R., Prilmeier, R., Stulz, R.M., 2012. This time is the same: using bank performance in 1998 to explain bank performance during the recent financial crisis. J. Finance 67, 2139–2185.
Fang, L., Peress, J., 2009. Media coverage and the cross-section of stock returns. J. Finance 64, 2023–2052.
Garcia, D., 2013. Sentiment during recessions. J. Finance 68, 1267–1300.
Hanley, K.W., Hoberg, G., 2019. Dynamic interpretation of emerging risks in the financial sector. Rev. Financ. Stud. 32, 4543–4603.
Hasan, I., Hoi, S., Wu, Q., Zhang, H., 2014. Beauty is in the eye of the beholder: the effect of corporate tax avoidance on the cost of bank loans. J. Financ. Econ. 113, 109–130.
Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. Rev. Econ. Stud. 64, 605–654.
Heckman, J.J., Ichimura, H., Todd, P.E., 1998. Matching as an econometric evaluation estimator. Rev. Econ. Stud. 65, 261–294.
Ho, P.H., Huang, C.W., Lin, C.Y., Yen, J.F., 2016. CEO overconfidence and financial crisis: evidence from bank lending and leverage. J. Financ. Econ. 120, 194–209.
Houston, J., Jiang, L., Lin, C., Ma, Y., 2014. Political connections and the cost of bank loans. J. Account. Res. 52, 193–243.
Lee, L.F., Hutton, A.P., Shu, S., 2015. The role of social media in the capital market: evidence from consumer product recalls. J. Account. Res. 53, 367–404.
Li, F., 2008. Annual report readability, current earnings, and earnings persistence. J. Account. Econ. 45, 221–247.
Lin, C.Y., Bui, D.G., Lin, T.C., 2020. Do short sellers exploit risky business models of banks? Evidence from two banking crises. J. Financ. Stab., 100719.
Liu, Y.J., Zhang, Z., Zhao, L., 2014. Speculation spillovers. Manage. Sci. 61, 649–664.
Lo, K.K., Chau, M., 2019. A penny is worth a thousand? Information from stock opinions in social media for penny stocks. ACM Trans. Manag. Inf. Syst. 9 (4), 14.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Finance 66, 35–65.
Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. J. Finance 69, 1643–1671.
Loughran, T., McDonald, B., 2016. Textual analysis in accounting and finance: a survey. J. Account. Res. 54, 1187–1230.
Miller, G.S., Skinner, D.J., 2015. The evolving disclosure landscape: how changes in technology, the media, and capital markets are affecting disclosure. J. Account. Res. 53, 221–239.
Park, C.Y., Shin, K., 2020. Contagion through national and regional exposures to foreign banks during the global financial crisis. J. Financ. Stab., 100721.
Peress, J., 2014. The media and the diffusion of information in financial markets: evidence from newspaper strikes. J. Finance 69, 2007–2043.
Petersen, M., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Rev. Financ. Stud. 22, 435–480.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Seeking Alpha, 2012. About Seeking Alpha. http://seekingalpha.com/page/about_us.
Solomon, D., Soltes, E., Sosyura, D., 2014. Winners in the spotlight: media coverage of fund holdings as a driver of flows. J. Financ. Econ. 113, 53–72.
Tetlock, P.C., 2007. Giving content to investor sentiment: the role of media in the stock market. J. Finance 62, 1139–1168.
Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: quantifying language to measure firms' fundamentals. J. Finance 63, 1437–1467.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.
Zhu, F., Zhang, X.M., 2010. Impact of online consumer reviews on sales: the moderating role of product and consumer characteristics. J. Mark. 74, 133–148.