



OPEN ACCESS

EDITED BY

Ioannis Pavlidis,
University of Houston, United States

REVIEWED BY

Yousra Trichilli,
University of Sfax, Tunisia

*CORRESPONDENCE

Xueqing Zhou
liuq2012@lzu.edu.cn

SPECIALTY SECTION

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

RECEIVED 31 May 2022

ACCEPTED 27 June 2022

PUBLISHED 28 July 2022

CITATION

Liu Q, Zhou X and Zhao L (2022) View
on the bullishness index and
agreement index.
Front. Psychol. 13:957323.
doi: 10.3389/fpsyg.2022.957323

COPYRIGHT

© 2022 Liu, Zhou and Zhao. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

View on the bullishness index and agreement index

Qing Liu^{1,2}, Xueqing Zhou^{3*} and Lili Zhao⁴¹College of Economics and Management, Huainan Normal University, Huainan, China, ²Graduate School of Management of Technology, Pukyong National University, Busan, South Korea, ³Communist Youth League, Huainan Normal University, Huainan, China, ⁴School of Mathematics, Shandong University, Jinan, China

KEYWORDS

bullishness index, agreement index, comment, mathematical description, data simulation

Introduction

Antweiler and Frank's (2004) paper "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards" published in *The Journal of Finance*, is an important piece of literature on the relationship between investor sentiment and stock markets based on natural language processing. More than 2,400 research publications have acknowledged its techniques, background, or conclusions (Google Scholar, 2022.5.17), making it an outstanding academic resource for future researchers. This article is not intended to challenge the paper's interpretation or conclusions, but rather to provide further references to the "bullishness index" and "agreement index."

Antweiler and Frank denote the number of bullish, bearish, and neutral messages in a period t using the symbols M_t^{BUY} , M_t^{SELL} , and M_t^{HOLD} , respectively. After aggregating the messages, they created three functions to calculate the bullish signal, using the formula

$$M_t = M_t^{BUY} + M_t^{SELL} \quad (1)$$

to represent the sum of messages related to the sentiment metric. The one used in the paper's report is the second function formula,

$$B_t^* = \ln \left[\frac{1 + M_t^{BUY}}{1 + M_t^{SELL}} \right]. \quad (2)$$

Antweiler and Frank also created the "agreement index" to measure investor disagreement, which is calculated as follows:

$$A_t = 1 - \sqrt{1 - \left(\frac{M_t^{BUY} - M_t^{SELL}}{M_t^{BUY} + M_t^{SELL}} \right)^2} \quad (3)$$

In a recent study, we used the above formula to create a bullishness index and an agreement index to measure the volatility cycle of investor sentiment on social media.

While visualizing the data, we discovered that the distribution of the “bullishness index” and “agreement index” is so neat that we find it difficult to believe that this is due to the data itself. As a result, we suspect that there is a strong relationship between Equations (2) and (3). We confirmed our suspicions with mathematical proof and data simulation. We discover that this association is absolute at larger, M_t , regardless of the data examined in the study. When using the above formula to study investor sentiment, if a variable is correlated with the “bullishness index,” it is inevitably correlated with the “agreement index.” This is due to the formula, not the object under study. We’re not sure if this is in conflict with the original intent of creating and analyzing the “bullishness index” and “agreement index,” respectively.

Traditional approaches to sentiment proxies are divided into three categories: market indices, survey indices, and special events. Market indicator proxies measure investor sentiment indirectly by using market indicators such as trading volume, closed-end fund discounts, first-day returns on initial public offerings (IPOs), and so on. Survey index-based proxies quantify investor sentiment by collecting optimistic or pessimistic expectations regarding the stock market from investors *via* surveys, such as the Consumer Confidence Index (Brown and Cliff, 2005), the Business Confidence Index (Liston-Perez et al., 2018), and the UBS/GALLUP Investor Optimism Index (Lemmon and Portniaguina, 2006). In a special event-based approach, special social events are frequently used as emotional proxies, and COVID-19 may be the most illustrative current example. Naseem et al. (2021) look at how COVID-19 affects the minds of investors and how that affects the stock market. Market-based measures have the advantage of being readily available at a relatively high frequency, but they also have the disadvantage of being the equilibrium result of many economic forces other than investor sentiment. The data acquisition of the proxy approach based on market indices is very slow, and it is usually done monthly or quarterly. The agent approach based on special things, on the other hand, is often used to analyze the impact of special events and is not universal.

“Google search queries” based investor sentiment proxies are easy to obtain and highly credible (Da et al., 2015), and are gaining more attention from scholars (Trichilli et al., 2018, 2020a,b,c). In addition to the emotional proxy methods mentioned above, natural language processing (NLP) technologies provide new ways to measure how investors feel. NLP is a research basis for social media-based investor sentiment research because it can tap into investor sentiment embedded in text and social networks. It also has the advantages of easy data availability, real-time access, and high credibility. It also gives us the chance to predict stock market returns based on high-frequency sentiment (Renault, 2017; McGurk et al., 2020). The research conducted by Antweiler and Frank (2004) was one of the very first to investigate the relationship between investor sentiment and the stock market based on social media.

Researchers rely on their “bullishness index” and “agreement index” to construct investor sentiment (Rao and Srivastava, 2014; Checkley et al., 2017; Liu et al., 2017; Chernozhukov et al., 2018; Fallahgoul, 2021). These studies typically measure daily investor sentiment, resulting in a large M_t in formula (1) and a high correlation between the “bullishness index” and the “agreement index”. Xiong et al. analyzed the correlation between investor sentiment and the stock market. According to the paper’s disclosures, the average daily total number of messages in that study was 10,443, and the average M_t was $\sim 3,550$ (Xiong et al., 2017).

In this age of exploding social media, natural language processing-based approaches to sentiment agents are growing in importance. Therefore, the research of Antweiler and Frank is increasingly cited by scholars. Due to the strong correlation between Equations (2) and (3), however, some of the researchers’ conclusions regarding investor sentiment may be susceptible to systematic errors. They may not be able to confirm whether these conclusions are caused by the characteristics of the study subjects or by the formulas themselves. Disclosure of the relationship between the characteristics of the “bullishness index” and the “agreement index” will assist researchers in better constructing the investor sentiment index, thereby preventing analytical misunderstandings. This is essential for analyzing investor sentiment, disagreement, and their relationship.

In the subsequent section of the paper, we use mathematical functions and data simulation to describe the relationship between Equations (2) and (3).

Mathematical description

To improve the readability of the paper, x is used below to denote the “bullishness index” B_t^* and y is used to denote the “agreement index” A_t .

From Equation (2)

$$x = \ln \left[\frac{1 + M_t^{BUY}}{1 + M_t^{SELL}} \right]$$

we can get

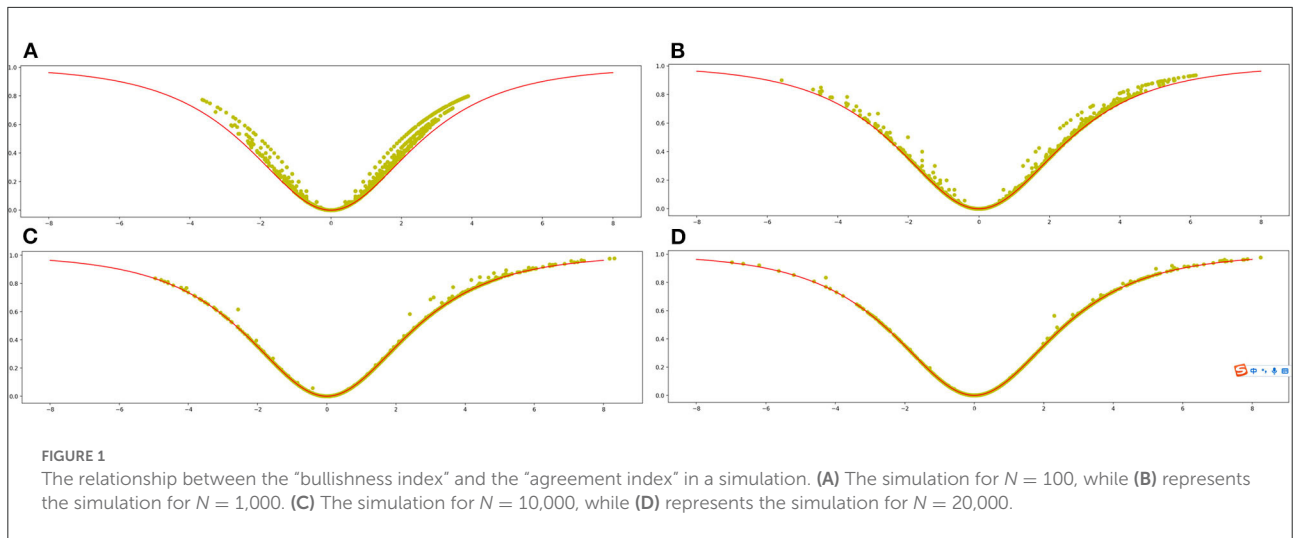
$$e^x (1 + M_t^{SELL}) = 1 + M_t^{BUY}.$$

Substituting Equation (1) into the above formula, we can get

$$e^x + e^x M_t - e^x M_t^{BUY} = 1 + M_t^{BUY},$$

then

$$\begin{aligned} M_t^{BUY} &= \frac{e^x + e^x M_t - 1}{e^x + 1} = \frac{e^x + (e^x + 1)M_t - M_t - 1}{e^x + 1} \\ &= M_t - \frac{-e^x + M_t + 1}{e^x + 1}, \end{aligned} \quad (4)$$



and

$$M_t^{SELL} = M_t - M_t^{BUY} = \frac{-e^x + M_t + 1}{e^x + 1}. \tag{5}$$

Combining Equations (4) and (5), we can get

$$\begin{aligned} M_t^{BUY} - M_t^{SELL} &= M_t - \frac{-2e^{B_t} + 2M_t + 2}{e^{B_t} + 1} \\ &= \frac{e^{B_t}M_t + 2e^{B_t} - 2 - M_t}{e^{B_t} + 1} = \frac{(e^{B_t} - 1)(M_t + 2)}{e^{B_t} + 1}. \end{aligned} \tag{6}$$

In Equation (3), introducing Equations (6) and (1) can obtain

$$\begin{aligned} y &= 1 - \sqrt{1 - \left(\frac{M_t^{BUY} - M_t^{SELL}}{M_t^{BUY} + M_t^{SELL}}\right)^2} \\ &= 1 - \sqrt{1 - \left(\frac{(e^x - 1)(M_t + 2)}{(e^x + 1)M_t}\right)^2} \\ &= 1 - \sqrt{1 - \left(\frac{e^x - 1}{e^x + 1} \times \frac{M_t + 2}{M_t}\right)^2}, \end{aligned}$$

and approximate formula

$$y \approx 1 - \frac{2e^{\frac{x}{2}}}{e^x + 1}. \tag{7}$$

Data simulation

By controlling the total number of messages, N_t in period t , we indirectly control the number of valid data, M_t . With data simulation, you can see how the “bullishness index” and the “agreement index” are related to each other.

Let the minimum value of investor message volume be 3 and the maximum value be N in this experiment. The message volume N_t of period t is a random number in the range [3,

$N]$, M_t^{BUY} is a random number in the range $[0, N_t]$, and M_t^{SELL} is a random number in the range $[0, N_t - M_t^{BUY}]$. In each experiment, 1095 simulated samples are made, and the samples are numbered from 0 to 1094. If (x_t, y_t) stands for the “bullishness index” and “agreement index” of period t , and x_t and y_t are found by using Equations (2) and (3), then the set of all sample points is defined as

$$(X, Y) = \bigcup_{t=0}^{1094} (x_t, y_t).$$

Figure 1A shows the scatter plot of the sample set (X, Y) when the control variable N is equal to 100. The black curve in Figure 1A is the plot of the approximate function

$$y = 1 - \frac{2e^{\frac{x}{2}}}{e^x + 1}.$$

The ratio of M_t+2 to M_t is relatively large when the number of messages in time period t is small, as shown in Figure 1A, and most random sample points are above the approximation curve. The shape of the sample distribution, on the other hand, is consistent with the approximation curve, and the “bullishness index” and the “agreement index” show statistical correlation.

Except for $N = 1,000$, the settings in Figure 1B are identical to those in Figure 1A. The random sample points fit the approximation curve better as the volume of investor messages increases in period t , and the correlation between the “bullishness index” and the “agreement index” increases.

The simulated graphs for $N = 10,000$ and $N = 20,000$ are shown in Figures 1C,D, where the relationship between the “bullishness index” and the “agreement index” is well fitted to the curve of Equation (7). With a very small error, the value of y_t can be deduced from x_t using Equation (7) at this point.

In conclusion, the “bullishness index” created by Antweiler and Frank demonstrates a strong correlation with the

“agreement index” even when the total sample size is small, indicating a statistical association.

Conclusion

In the age of artificial intelligence and big data, the sentiment agent approach based on natural language processing is gaining importance (Farzindar and Inkpen, 2015). At the same time, sentiment measurement based on natural language processing has become an essential tool for governments, research institutions, and financial institutions to formulate industry policies and manage financial risks (Ku et al., 2008; Fisher et al., 2016). The “bullishness index” and “agreement index” developed by Antweiler and Frank are becoming increasingly cited by academics. In this paper, we use a mathematical formula to approximate the relationship between the “bullishness index” and the “agreement index” and examine the effect of the range of the total sample on the error of the formula through data simulation. Due to the correlation, the correlation analysis of investor sentiment and investor opinion disagreement using Equations (2) and (3) may produce conclusions that are unrelated to the underlying data. Although we do not consider the paper’s conclusion, “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards” to be problematic, we advise avoiding the use of both Equations (2) and (3). Disclosure of the approximate relationship between the “bullishness index” and the “agreement index” is crucial for the study of investor sentiment, disagreement, and behavioral finance.

References

- Antweiler, W., and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *J. Fin.* 59, 1259–1294. doi: 10.1111/j.1540-6261.2004.00662.x
- Brown, G. W., and Cliff, M. T. (2005). Investor sentiment and asset valuation. *J. Bus.* 78, 405–440. doi: 10.1086/427633
- Checkley, M., Higón, D. A., and Alles, H. (2017). The hasty wisdom of the mob: how market sentiment predicts stock market behavior. *Expert Syst. Appl.* doi: 10.1016/j.eswa.2017.01.029
- Chernozhukov, V., Hardle, W., Huang, C., and Wang, W. (2018). LASSO-driven inference in time and space. *Ann. Statistics.* 49, 1702–35. doi: 10.2139/ssrn.3188362
- Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Rev. Fin. Stud.* 28, 1–32. doi: 10.1093/rfs/hhu072
- Fallahgoul, H. A. (2021). Inside the mind of investors during the COVID-19 pandemic: evidence from the StockTwits data. *J. Fin. Data Sci.* 3, 134–148. doi: 10.2139/ssrn.3583462
- Farzindar, A., and Inkpen, D. (2015). Natural language processing for social media. *Synth. Lect. Human Lang. Technol.* 8, 1–166. doi: 10.2200/S00659ED1V01Y201508HLT030
- Fisher, I. E., Garnsey, M. R., and Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: a synthesis of the literature with a roadmap for future research. *Intell. Syst. Account. Fin. Manag.* 23, 157–214. doi: 10.1002/isaf.1386
- Ku, C. H., Iriberry, A., and Leroy, G. (2008). “Natural language processing and e-government: crime information extraction from heterogeneous data sources,”

This paper demonstrates and confirms the relationship between the “bullishness index” and the “agreement index” and the possibility of analytical error. However, we do not offer a superior alternative. In future studies, we will propose potential substitutes.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

in *Proceedings of the 2008 International Conference on Digital Government Research*, 162–170.

Lemmon, M., and Portniaguina, E. (2006). Consumer confidence and asset prices: some empirical evidence. *Rev. Fin. Stud.* 19, 1499–1529. doi: 10.1093/rfs/hhj038

Liston-Perez, D., Torres-Palacio, P., and Bayram, S. G. (2018). Does investor sentiment predict Mexican equity returns? *Int. J. Manag. Fin.* 14, 484–502. doi: 10.1108/IJMF-05-2017-0088

Liu, Y., Qin, Z., Li, P., and Wan, T. (2017). “Stock volatility prediction using recurrent neural networks with sentiment analysis,” in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 192–201. doi: 10.1007/978-3-319-60042-0_22

McGurk, Z., Nowak, A., and Hall, J. C. (2020). Stock returns and investor sentiment: textual analysis and social media. *J. Econ. Fin.* 44, 458–485. doi: 10.1007/s12197-019-09494-4

Naseem, S., Mohsin, M., Hui, W., Liyan, G., and Penglai, K. (2021). The investor psychology and stock market behavior during the initial era of COVID-19: a study of China, Japan, and the United States. *Front. Psychol.* 12, 16. doi: 10.3389/fpsyg.2021.626934

Rao, T., and Srivastava, S. (2014). “Twitter sentiment analysis: how to hedge your bets in the stock markets can,” in *State of the Art Applications of Social Network Analysis*, eds T. Özyer and F. Polat (Berlin: Springer International Publishing), 227–47. doi: 10.1007/978-3-319-05912-9_11

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *J. Bank. Fin.* 84, 25–40. doi: 10.1016/j.jbankfin.2017.07.002

Trichilli, Y., Abbas, M. B., and Masmoudi, A. (2020a). Islamic and conventional portfolios optimization under investor sentiment states: Bayesian vs Markowitz portfolio analysis. *Res. Int. Bus. Fin.* 51, 101071. doi: 10.1016/j.ribaf.2019.101071

Trichilli, Y., Abdelhédi, M., and Boujelbène Abbas, M. (2018). Googling Investor's sentiment: powerful measure in conventional and Islamic MENA financial markets. *Int. Econ. J.* 32, 454–469. doi: 10.1080/10168737.2018.1522055

Trichilli, Y., Abdelhédi, M., and Boujelbène Abbas, M. (2020b). The thermal optimal path model: does Google search queries help to predict dynamic relationship between investor's sentiment and indexes returns? *J. Asset Manag.* 21, 261–279. doi: 10.1057/s41260-020-00159-0

Trichilli, Y., Boujelbène Abbas, M., and Masmoudi, A. (2020c). Predicting the effect of Googling investor sentiment on Islamic stock market returns: a five-state hidden Markov model. *Int. J. Islamic Middle Eastern Fin. Manag.* 13, 165–193. doi: 10.1108/IMEFM-07-2018-0218

Xiong, X., Chunchun, L. U. O., and Ye, Z. (2017). Stock BBS and trades: the information content of stock BBS. *J. Syst. Sci. Math. Sci.* 37, 2359.