Social Media Stocks Reviews Big Data Management Research

Nanli Zhu[1, a], Boyu Jia[1*, b], Rui Li[1, c]

[1] *Yunnan University of Finance and Economics, LianHuang WuHua Kunming China*
[a] *nanli_zhu@sina.com,* [b] *1135027156@qq.com*,* [c] *1129747389@qq.com*
*Corresponding author: Boyu Jia**

Abstract

[Purpose/Significance] In order to enhance the big data management of stock reviews by regulators, make management more scientific and efficient, and improve the screening efficiency of big data stock reviews by investors. This paper explores the influencing factors of the perceived usefulness of big data in social media stock reviews. [Methods/Process] Based on the Information adoption theory, this paper constructed a factor theory model of the helpfulness of stock review information through the signals related to reviewers and reviews and the signal environment. Using Tobit regression to empirically test the relationship between various signals and review helpfulness. [Result/Conclusion] The findings suggest that the perceived helpfulness of stock reviews is positively influenced by review images, review information entropy, review professionalism, review bilaterality, financial blogger certification, number of followers of the reviewer, and published at non-trading times.

Keywords: *stock reviews; big data management; signal theory; social media.*

1. INTRODUCTION

With the advent of the Web 2.0 era, social media has impacted the development of traditional media. Take the financial market as an example: financial analysis used to be more common in financial newspapers and TV channels. Now, however, more and more investors are seeking helpful information by using social media. [8] Bartov et al. (2017) [2] suggest that at least 34% to 70% of investors consider social media content to make an investment decision. Therefore, it is very important to study the helpfulness of stock commentary information in social media. The research on the influencing factors of the helpfulness of online reviews has become a hot topic in the field of online reviews [12]. At the same time, existing research focuses on product themes, while research on tourism [7], hotel [6], health [20], and other pieces is also increasing, while research on stock themes has not yet attracted attention. In addition, in social media, the speed of information generation is faster, coupled with the speculative nature of stocks, which makes the quantity of stock evaluation information overwhelming and the quality uneven, regulators will face complex regulatory problems, and investors will also face the problem of information overload. Therefore, it is necessary to find out the influencing factors of the helpfulness of stock evaluation information, which will not only help the regulatory agencies to supervise more scientifically and efficiently, but also help investors screen the stock evaluation information more efficiently.

Based on the information adoption theory, this paper constructs a model of the influencing factors of the perceived helpfulness of stock reviews by social media investors, in which the information entropy is used to measure the amount of information of stock reviews, which further optimizes the traditional measurement method.

2. THEORY AND HYPOTHESE

2.1 Information Adoption Theory

The Information Adoption Model (IAM) was proposed by Sussman et al., which holds that the helpfulness of information is mainly affected by the quality of information and the credibility of the information source, and information adoption is further considered after the information is determined to be helpful. As shown in Figure 1.

S. A. R. Khan et al. (Eds.): ICMSSE 2022, AHCS 12, pp. 320–326, 2023.
https://doi.org/10.2991/978-94-6463-056-5_46

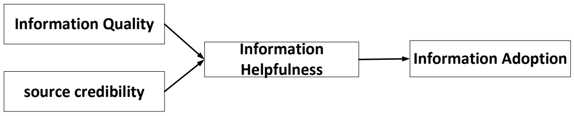

Figure 1 The Information Adoption Model.

In the context of social media, investors read online stock reviews to get more stock information and reduce irrational decision-making, which becomes the logical starting point for studying the influencing factors of the helpfulness of social media stock reviews. Based on IAM theory and previous research, we select the financial blogger certification and the number of reviewers' fans to measure the information source's reliability. We select review images, information entropy, professionalism, and bilaterality to measure review information quality.

2.2 Research Hypothesis

The pictures provided by the reviewer will enrich the product information and increase the review's authenticity [19]. According to the dual coding theory in cognitive psychology, human visual imagery has more advantages in the function of information processing, so pictures are easier to be recognized and remembered. In stocks, investors often refer to pictures to help their decisions, such as K-line charts and time-division charts, so reviews combined with pictures will have higher credibility. With the discussion above, we hypothesize:

H1: The use of pictures in reviews has a positive effect on the helpfulness of reviews.

The information entropy of stock reviews refers to the amount of information contained in stock reviews. In information theory, information entropy is a concept to measure the amount of information. With the help of the concept of thermodynamics, Shannon called the average amount of information after eliminating redundancy in information "information entropy". Jorge et al. (2020) [5], and Wu et al. (2021) [18] applied information entropy to the field of online reviews. Singh et al. (2017) [16] pointed out directly that information entropy is an essential parameter in determining the helpfulness of online reviews. Ai Shizhong et al. (2019) [1] believe that the greater the information entropy of online comments, the greater the amount of information they contain, and the more unique the reviews are. With the discussion above, we hypothesize:

H2: The information entropy of reviews has a positive effect on the helpfulness of reviews.

Most people will measure the reliability of information by the degree of professionalism it reveals [3]. Choi et al. (2020) [4] believe that reviews with high professionalism will be more recognized because readers will preferentially filter out reviews lacking the professional knowledge to reduce the search cost, which will lead to the lack of helpfulness of these reviews. In stock investment, professional analysis ability is conducive to reducing investment risk so that more professional reviews will increase the persuasion of online comments. With the discussion above, we hypothesize:

H3: The professionalism of the review has a positive impact on the helpfulness of the review.

Bilaterality means that a review contains both positive and negative statements [10]. Most studies have shown that bilateral reviews are more helpful than unilateral reviews [14]. Winter et al. (2012) [17] believe that readers who want to collect more information will be more likely to adopt bilateral reviews. When a reader is browsing a stock review, he or she usually wants to get information that can help him or her both make a profit and avoid a loss because it can get rid of more uncertainty. With the discussion above, we hypothesize:

H4: Bilateral reviews have a positive effect on the helpfulness of the review.

Siering et al. (2018) [15] Believe that the professional knowledge certification of reviewers is a crucial influence signal for readers to judge whether reviews are helpful or not. In the micro-blog platform, the authentication of financial bloggers has a particular professional threshold. Users who want to obtain the certification must complete original blog writing related to finance and economics, read more than 50,000 within 30 days, and have more than 1,000 fans. The fact that the reviewer has a professional knowledge certification reflects the ability of the reviewer to provide high-quality reviews [19]. As a result, reviews posted by users of financial bloggers are more reliable. With the discussion above, we hypothesize:

H5: Reviews posted by financial bloggers are more helpful than those posted by regular users.

Reputation and recognition are beneficial to increase the credibility of users [9]. Credibility directly affects the perceived helpfulness of a review. The number of fans of users represents the recognition of other users to the user, and the more fans the user has, the more credible it is, which affects the helpfulness of comments. Min Qingfei et al. (2017) [13] and others have used the quality of inward social networks constructed by the number of fans to predict the helpfulness of online comments. After comparison, it is found that the number of fans has more advantages in predicting helpfulness. With the discussion above, we hypothesize:

H6: The number of followers of the reviewer has a positive effect on the helpfulness of the review.

Control variables included days of exposure to reviews, number of replies, and number of retweets to reviews. In practice, the posting time of a review impacts the number of likes for the helpfulness of the review, and reviews posted earlier are more likely to get more likes. The number of replies and retweets of reviews represents the attention of review readers to reviews. The attention to reviews will affect the number of likes of reviews and then impact the helpfulness of reviews.

To sum up, this paper proposes a research model on the influencing factors of the helpfulness of microblog stock comments, as Figure 2 shows:

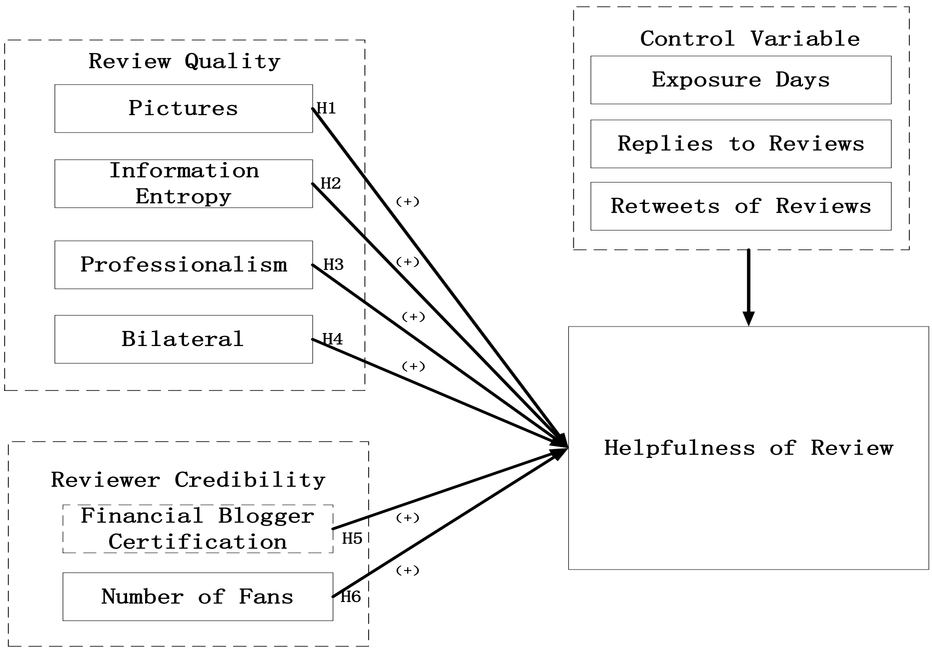

Figure 2 Research on the Factors Influencing the Helpfulness of Stock Reviews on Microblog.

3. STUDY DESIGN

3.1 Data Collection

This study collected comment data on 25,220 Weibo stock topics from January 3, 2022, to January 14, 2022. The reasons for choosing Weibo topics are as follows: first, Weibo is one of China's most mainstream social media. Its stock review information is updated quickly, with rich content and representative data. Second, the information attributes of stock reviews under the topic of Weibo stocks and the overall stock reviews of Weibo are homogeneous and easy to crawl. After removing the comments with ad-likes and the comments whose actual content has nothing to do with the stock theme, 20,664 valid samples were finally obtained.

Numerous studies have shown that likes can express attitudes of agreement and support for content [11] [21]. Therefore, the number of likes is set as a proxy variable for the perceived helpfulness of the comment. The names and explanations of the variables can be found in

Table 1. The information entropy of the comment represents the amount of information contained in the comment. Referring to the practice of Singh and Ai Shizhong, the calculation formula is as follows:

$$E = -\sum_{i,j} \mathrm{P}(\mathrm{b}_i, j)\log_2 P_{bi}(j) \qquad (1)$$

b_i is a some word, j is the word that follows b_i, $P(b_i, j)$ is the probability that words b_i and j appear at the same time, $P_{bi}(j)$ is the conditional probability that word j is followed by wordb_i, obtained from $P(b_i, j)$ /$P(b_i)$(Ai Shizhong et al. , 2019).

Table 1 Variable description table.

VARIABLE CATEGORY	VARIABLE NAME	VARIABLE MEASURE
DEPENDENT VARIABLE	Helpfulness of reviews （Helpfulness）	The number of likes a review has received
CONTROL VARIABLE	Exposure days of reviews (Day)	The number of days until the collection deadline for reviews to be published
	The number of replies to the review (Respond)	Number of replies to reviews
	The number of retweets of reviews (Retweet)	he number of retweets a review gets

VARIABLE CATEGORY		VARIABLE NAME	VARIABLE MEASURE
INDEPENDENT VARIABLE	Review Quality	Review on pictures (Picture)	Dummy variable, with pictures represented by "1", without pictures by "0"
		Review information entropy (Entropy)	Information entropy of reviews, calculated using Shannon's conditional entropy formula
		Review Professional (Professional)	The percentage of professional words contained in reviews to the total number of words
		Review on bilaterality (Two-side)	Dummy variable, which contains two kinds of emotional information, is represented by "1", otherwise it is represented by "0"
	Reviewer Credibility	Financial Blogger Certification (Blogger)	Dummy variable, which is represented by "1" if there is a financial blogger's certification, otherwise it is represented by "0"
		The number of fans of the reviewer (Fans)	The number of followers of the reviewer

3.2 Analysis Method

The Tobit model is selected to conduct empirical analysis with Stata16.0 software in this research. The reason is that the dependent variable is bounded, 57% of the data is 0, showing the overall left skew, and the Tobit model can solve the problem of selection bias. Combined with the previous assumptions, the overall regression model in this paper is:

$$Helpfulness = \alpha + \beta_1 day + \beta_2 Repond + \beta_3 Retweet + \beta_4 Picture + \beta_5 Entropy + \beta_6 Professional + \beta_7 Two\text{-}side + \beta_8 Blogger + \beta_9 Fans + \varepsilon \quad (2)$$

4. RESEARCH RESULTS

4.1 Study Variables

Descriptive statistical analysis of independent, dependent, and control variables was made. The descriptive statistical analysis results of variables are as follows

Table 2 shows.

Table 2 Variable descriptive statistics.

Name	Sample size	Minimum value	Maximum value	Average value	Standard deviation
Helpfulness	20664	0	2449	20	84
Retweet	20664	0	1447	2	13
Respond	20664	0	1517	4	28
Day	20664	0	11	6	3
Picture	20664	0	1	0	0
Entropy	20664	0	711	51	41
Professional	20664	0	8	0	0
Two-side	20664	0	1	1	0
Blogger	20664	0	1	0	0
Fans	20664	0	20148000	189249	480083

Table 3 Tobit regression results.

	Model 1	Model 2	Model 3	Tolerance	Variance inflation factor (VIF)
	Beta coefficient	Beta coefficient	Beta coefficient	0.755	1.324
Retweet	2.09***	1.47***	0.98***	0.776	1.289
Respond	1.32***	1.08***	1.00***	0.991	1.009
Day	0.03***	0.04***	0.04***	0.870	1.150
Picture		0.26***	0.20***	0.794	1.260
Entropy		0.44***	0.40***	0.886	1.129
Professional		10.44***	9.38***	0.861	1.161
Two-side		0.09***	0.07***	0.533	1.877
Fans			0.03***	0.467	2.142
Blogger			0.14***	0.755	1.324
Log likelihood	-18739.112	-14208.745	-13603.151		
Pseudo R-squared	0.1041	0.3207	0.3497		

*** p<0.01, **p<0.05, *p<0.1

4.2 Model Checking

This paper adopts the method of dividing the independent variables into the regression model.

Table 3 describes each model's independent variables and summarizes the model analysis results. We compare the fit of each model by the fit parameters of Likelihood Ratio and Efron's pseudo R-squared. Finally, with the addition of new explanatory variables, the fit of the model is better.

The detailed results of the regression are shown in Table 3. Model 1 adds three control variables: the number of retweets of comments (Retweet), the number of replies to comments (Respond), and the number of days of exposure to comments (Day). Model 2 added review-related signals on the basis of Model 1. The results showed that the reviews contained pictures ($\beta 4 = 0.26$; H1 supported), the information entropy of the reviews ($\beta 5 = 0.44$; H2 supported), and the professionalism of the reviews ($\beta 6 = 10.44$; H3 supported), and the reviews are bilateral ($\beta 7 = 0.09$; H4 supported), all have a positive effect on the helpfulness of reviews.

Based on Model 2, Model 3 added the reviewer-related signal, and the results show that the reviewer is a financial blogger ($\beta 8 = 0.14$; H5 supported) and the number of fans of the reviewer ($\beta 9 = 0.03$; H6 supported) has a positive effect on the helpfulness of the review. The variance inflation factor and tolerance index are introduced to measure multicollinearity. The results are shown in

Table 3. The VIF value of the variance inflation factor of all variables is less than 10, and the tolerance of all variables is greater than the critical value of 0.1, indicating that the model There is no multicollinearity problem.

5. CONCLUSION

The theoretical significance of this study is that, First, this study explores the influencing factors of the helpfulness of social media stock reviews for the first time, expanding the research in the field of stocks. Second, this study uses information entropy to measure the amount of information in reviews, which is better than the traditional method of measuring the amount of information by the number of words in the reviews.

The practical significance of this study is that, for regulatory authorities, the research results can be used as a reference for supervision, which is conducive to optimizing regulatory management; for review readers, the research results can help them manage and screen stock comment information more effectively.

This study also has shortcomings, which can be further discussed in the follow-up research. This study ignores the difference in the perceived value of information helpfulness by different readers; secondly, this study only considers the reviewer's liking behavior and does not consider whether the review reply and collection data are worth analyzing.

REFERENCES

[1] Ai Shizhong, Zeng Xin 2019. The influence of total sentiment and information entropy on review usefulness based on Ebay review data J. Soft Science, 33:129-132+144.

[2] BARTOV E F L, MOHANRAM P S. 2017. Can twitter help predict firm-level earnings and stock returns? J. The Accounting Review, 93: 25-27.

[3] CHEN Y-L, CHANG C-L, SUNG A-Q 2021. Predicting eWOM's influence on purchase intention based on helpfulness, credibility, Information Quality and Professionalism. J. Sustainability , 13: 7486.

[4] CHOI H S, LEON S 2020. An empirical investigation of online review helpfulness: A big data perspective. J. Decision Support Systems , 139: 113403.

[5] FRESNEDA J E, GEFEN D 2020. Gazing at the stars is not enough, look at the specific word entropy, too! J. Information & Management , 57: 103388.

[6] HU F 2020. What makes a hotel review helpful? An information requirement perspective.J. Journal of hospitality marketing & management , 29: 571-591.

[7] HU X, YANG Y 2021. What makes online reviews helpful in tourism and hospitality? a bare-bones meta-analysis. J. Journal of hospitality marketing & management , 30: 139-158.

[8] Gao Yang, Zhao Kun, Wang Yaojun 2021. Social Media Attention, Mobility and Information Asymmetry J. Statistics and Decision Making , 37: 170-173.

[9] Han Dongmei, Wang Tao 2021. Research on the influencing factors of discounted information post popularity based on signal theory. J. Information Science , 39: 68-74+84.

[10] KAMINS M A, BRAND M J, HOEKE S A, et al. 1989. Two-Sided versus one-sided celebrity endorsements: the impact on advertising Effectiveness and Credibility. J. Journal of Advertising , 18: 4-10.

[11] LIU X, LU J, WANG H 2017. When health information meets social media: exploring virality on Sina Weibo. J. Health communication , 32: 1252-1260.

[12] Li Ang, Zhao Zhijie 2019. A review of online reviews usefulness research J. Business Economics Research: 77-80.

[13] Min Qingfei, Qin Liang, Zhang Keliang 2017. A study on factors affecting the usefulness of online reviews J. Management Review , 29: 95-107.

[14] RIETSCHE R, FREI D, STOECKLI E, et al. 2019. Not all reviews are equal C. A LITERATURE REVIEW ON ONLINE REVIEW HELPFULNESS; City.

[15] SIERING M, MUNTERMANN J, RAJAGOPALAN B 2018. Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. J. Decision Support Systems , 108: 1-12.

[16] SINGH J P, IRANI S, RANA N P, et al. 2017. Predicting the "helpfulness" of online consumer reviews. J. Journal of Business Research , 70: 346-355.

[17] WINTER S, KRäMER N C 2012. Selecting science information in Web 2.0: How source cues, message sidedness, and need for cognition influence users' exposure to blog posts. J. Journal of Computer-Mediated Communication , 18: 80-96.

[18] WU C, MAI F, LI X 2021. The effect of content depth and deviation on online review helpfulness: Evidence from double-hurdle model. J. Information & Management , 58: 103408.

[19] YANG S, ZHOU Y, YAO J, et al. 2019. Understanding online review helpfulness in omnichannel retailing. Industrial J. Management & Data Systems , 119: 1565-1580.

[20] Zhang Weiwei, Jiang Xue 2020. A review of research on influencing factors of user participation behavior in online health communities J. , 64: 136-145.

[21] Zhang Yu, Zhao Haobo 2017. A Multidimensional Study of Like Behavior in Social Media J. : 8-11.