Crowd Wisdom: The Impact of Opinion Diversity and Participant Independence on Crowd Performance

Hong Hong
Department of Management Science
Xiamen University
honghongtju@126.com

Qianzhou Du
Business Information Technology
Virginia Tech
qiand12@vt.edu

Alan G. Wang
Business Information Technology
Virginia Tech
alanwang@vt.edu

Weiguo Fan
Accounting and Information Systems
Virginia Tech
wfan@vt.edu

Di Xu
Department of Management Science
Xiamen University
dxu@xmu.edu.cn

Abstract

Recent advances in information technologies such as Web 2.0 have dramatically expanded the use of crowd wisdom in dealing with a wide range of problems. Prior research confirms the importance of crowd wisdom in the context of stock markets, but fails to investigate the impact of crowd characteristics on crowd performance. We study the influence of opinion diversity and participant independence on the crowd performance in the context of online investment communities. Instead of using a survey research methodology, which is usually time-consuming and limited to a small sample, we develop text mining based measures for the variables in our research model. The empirical results show that opinion distance is positively related to crowd performance. Opinion content similarity and participant dependence are negatively associated with crowd performance. Crowd size significantly moderates the relationships between crowd characteristics and crowd performance. Related theoretical and practical implications are also discussed.

Keywords
Wisdom of crowds, Stocktwits, Social media, Crowd performance, Moderating effect

Introduction

Online investment communities (OICs, e.g., Seeking alpha, Yahoo!Finance, Stocktwits) allow community members to interact with other community members online, making OICs a popular venue for individual investors to exchange and share their opinions (Chen et al. 2014; Gu et al. 2014). Taking Stocktwits as an example, approximately 11 million messages about 9,303 stocks, marketing indices, and exchange traded funds were posted by 63,445 unique users in 2014 alone.

Peer opinions from online investment communities capture “crowd wisdom” and have begun to play an important role in financial markets, which traditionally are the domain of professional financial intermediaries or experts (Chen et al. 2014). The Wisdom of Crowd (WoC) theory is put forward by
Surowiecki, who argues that a crowd of individuals with diverse knowledge is likely to make better predictions than professional experts when working independently (Surowiecki 2005). Hong and Page (2004) conduct computational experiments to confirm that a team of randomly selected diverse agents outperformed a team comprised of the most intelligent agents. The wisdom of crowd effect was also supported by many examples from stock markets (Antweiler et al. 2004; Chen et al. 2014; Nofer et al. 2014; Ready-Campbell et al. 2011; Sprenger et al. 2014). Compared to traditional media, social media provide investors with rapidly updated information ahead of other channels and have a faster and stronger predictive relationship with the firm equity value (Luo et al. 2013). Peer-based financial advice posted on the Internet significantly impacts stock returns (Antweiler et al. 2004; Chen et al. 2014), stock market volatility (Antweiler et al. 2004), and earnings surprises (Chen et al. 2014). The disagreement degree among the posted messages is associated with increased trading volume (Antweiler et al. 2004). According to the Seer-Sucker theory, which states “no matter how much evidence that seers do not exist, suckers will pay for the existence of seers,” the crowd can outperform experts (Armstrong 1980). Evidence also shows that the crowd even outperforms professional analysts in financial predictions. Investors can achieve a higher return based on the recommendations of the crowd rather than those of the analysts (Nofer et al. 2014). The Internet crowd is also able to outperform the S&P 500 companies in terms of both overall returns and risk-adjusted returns (Ready-Campbell et al. 2011).

Prior research confirms the importance of crowd wisdom in the context of stock markets. However, very few studies have investigated the impact of crowd characteristics on crowd performance. Arazy et al. (2006) propose a crowd wisdom model and investigate the influence of crowd size and author diversity on crowd performance in the context of Wikipedia. Nofer et al. (2014) study the impact of opinion independence and diversity on the crowd performance of stock predictions. They use stock performance to measure crowd performance, age and gender difference to measure diversity, and the presence of professional analysts and ranking systems to measure opinion independence. Different from Nofer et al. (2014), our study uses text mining based crowd diversity measures. In addition, we use crowd prediction accuracy to measure crowd performance, which is a more accurate measure than stock performance. In the wisdom of crowd theory, Surowiecki explains diversity and independence being two important factors of crowd performance as the best crowd decisions are the product of disagreement and contest, not consensus (Surowiecki 2005). The diversity of crowd members’ opinions means that each member in the crowd should have some private information or distinct interpretation and independence of crowd participants means that individuals’ opinions are not determined by others around them (Surowiecki 2005). Online investment communities provide us with a good context to study whether or not OIC crowds can make better investment decisions and how crowd characteristics influence the crowd performance. Hence, we study the influence of opinion diversity and participant independence on crowd performance in the context of online investment communities.

Our study enriches extant empirical research literature on crowd wisdom performance. Instead of using a survey research methodology, which is usually time-consuming and limited to a small sample, we develop text mining based measures, which can be readily applied to a large amount of data collected in an online community. We use sentiment distance and posting content similarity to measure the diversity of crowd opinions and use the difference between the number of followings and the number of followers for those individuals who involved in a stock discussion to measure participant independence. In addition, the moderating effect of crowd size between crowd wisdom determinants (i.e., crowd diversity and participant independence) and crowd performance is investigated.

The rest of this paper is structured as follows. The next section provides the theory background and hypotheses development. In the third section we present the research methodology and measures adopted. We then describe the assembled data for this study and discuss research results. In the final section, we conclude the paper discussing the findings, implications.

Theoretical Background and Hypotheses Development

Crowd Wisdom

Recent advance in information technologies has dramatically expanded the use of crowd wisdom in order to deal with a wide range of problems (e.g., Wikipedia, Amazon Mechanical Turk, Online Investment Communities). Surowiecki’s crowd wisdom theory provides a useful framework for the study of the crowd
Opinion Diversity, Participant Independence and Crowd Performance

performance. He summarizes three conditions for being a wise crowd: (1) diversity of crowd members' opinions: each person should have some private information or personal interpretations of the same facts; (2) independence of crowd participants' decision making: people's opinions are not determined by others' opinions; (3) decentralization of the crowd: crowd members have their own specialization and learn from their own local knowledge sources (Surowiecki 2005). Diversity and independence are two important influencing factors for crowd performance because the best decisions are the product of disagreement and contest, not consensus or comprise (Surowiecki 2005). In this study we exclude the consideration of decentralization in our research model because it is difficult to be measured using automated text mining approaches. Therefore, we will be focused on the impact of crowd diversity and independence on crowd performance in this study.

**Impact of Opinion Diversity on Crowd Performance**

Surowiecki (2005) argues that diversity is not in a sociological sense, but rather in a conceptual and cognitive sense. In general, diversity of a crowd refers to the members' differences in terms of demographic characteristics, cultural identities, ethnicity, training, and expertise (Hong et al. 2004). Cheruvell et al. (2014) identify five team diversity dimensions, including career stage, degree of familiarity, interaction mode, discipline type, and viewpoints. A diverse group is a group of people who possess varying degrees of knowledge and insight (Surowiecki 2005). Davis-Stober et al. (2015) confirm that diversity of individual predictions is essential for an optimal performance in team forecasting. Cheruvell et al. (2014) suggest that high-performance collaborative research teams are enabled by team diversity. The computational experiment results of Hong et al. (2004) confirm that diversity of group members enhances the group’s collective performance. Thus, we propose the following hypotheses:

H1: Increased opinion diversity among crowd members will lead to higher crowd performance.

**Effect of Participant Independence on Crowd Performance**

Independence does not mean isolation and it means that crowd members cannot be easily influenced by others (Surowiecki 2005). Independence is important to crowd wisdom as it helps avoid crowd mistakes becoming correlated and makes easier for crowd to have new information (Surowiecki 2005). Independence is often considered as a positive determinant for crowd performance (Nofer et al. 2014). According to the experimental results of Lorenz et al (2011), crowd wisdom is undermined by social influence among participants. The more influence the crowd exert on each other, the more likely crowd participants will have similar ideas, inducing the same mistakes cannot be avoided (Surowiecki 2005). Opinions in a large social networks converge to the truth when the influence of opinion leaders vanishes (Golub et al. 2010). Dependence among participants narrows the diversity of opinions and then impairs crowd wisdom (Lorenz et al. 2011). Therefore, we posit the following hypothesis:

H2: A higher level of participant dependence will decrease crowd performance.

**Moderating Effect of Crowd Size**

Crowd size is an important factor for a wise crowd with sufficient diversity and expertise (Wagner et al. 2014). As the crowd size becomes larger, the participants in the crowd may become more diverse, inducing the crowd to perform better (Hong et al. 2004). At the same time, the crowd size cannot be so large as to prevent the group of the best problem solvers from becoming similar (Hong et al. 2004). In practice, samples of 20, or even 10 lead to convergence (Wagner et al. 2014). The quality of collective estimates would depend on the size of the crowd, with diminishing marginal returns for size increases (Wagner et al. 2014). Fostering diversity and keeping independence are more important in small groups than larger groups, as it is easy for a few biased individuals to exert undue influence and skew the group's collective decision in small groups (Surowiecki 2005). The size of the crowd may impact the relation between crowd diversity and crowd performance or crowd independence and crowd performance. Thus, we put forward the following hypotheses:

H3a: Crowd size moderates the effect of crowd opinion diversity on crowd performance.

H3b: Crowd size moderates the effect of crowd participant dependence on crowd performance.
Figure 1 summarizes our research model.

![Research Model Diagram]

**Measures for Variables and Data Collection**

We chose S&P 100 firms as our research sample and collected our data from three sources. First, we collected daily stock discussions about the S&P 100 firms from Stocktwits, a popular online investment community. We collected 971,236 postings in total from Stocktwits. A nice feature provided by Stocktwits is that it allows participants to indicate a bullish or bearish prediction for a particular stock. Similar to Twitter, Stocktwits also allows its participants to follow or be followed by others. Second, we obtained daily stock price data from the Center for Research in Security Prices (CRSP) database, which is a commonly used financial database. Third, quarterly firm-level total assets data and industry data were obtained from the Compustat database. The time period for which data were collected is from January 1, 2014 to December 31, 2014. Firms’ total assets data were released on a quarterly basis. Therefore, we assume that a firm’s daily total assets value equals to the quarterly value for the days that the quarterly release covers. We removed 6 firms for which financial or online investment community data could not be obtained. Finally, our sample data contains 94 S&P firms. We summarize the variable measures and data sources in Table 1. We also provide explanation about our measures in more detail in the rest of this section. As we mentioned earlier, we will use automated text mining methods to automatically measure the variables in our research model. Compared to the survey research methodology, our approach is more efficient when applying to a large amount of data and a much larger sample.

<table>
<thead>
<tr>
<th>Variable Type</th>
<th>Variable</th>
<th>Measure</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>Crowd performance</td>
<td>A dummy variable of 1 indicating that crowd prediction has the same trend with the real stock market performance for each stock, 0 otherwise. Crowd prediction is determined by taking the major vote from daily bullish and bearish votes in the community.</td>
<td>CRSP, Stocktwits</td>
</tr>
</tbody>
</table>

---

*Twenty-second Americas Conference on Information Systems, San Diego, 2016*
Independent variables | Opinion diversity | (1) Opinion distance: variance of sentiment scores measured from daily postings about a given stock. (2) Content similarity: average text similarity among daily postings for a given stock. | Stocktwits |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant dependence</td>
<td>The difference between the number of followings and number of followers for those individuals who make postings about a given stock on a given day.</td>
<td>Stocktwits</td>
<td></td>
</tr>
</tbody>
</table>

Moderating variable | Crowd size | Number of individuals who made postings about a given stock on a given day. | Stocktwits |

Control Variables | Posting experience | The median number of historical postings made by those individuals who made postings about a given stock on a given day. | Stocktwits |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average sentiment</td>
<td>Average sentiment value of all postings about a given stock on a given day.</td>
<td>Stocktwits</td>
<td></td>
</tr>
<tr>
<td>Stock volatility</td>
<td>Absolute difference between the highest price and lowest price of a given stock during a day.</td>
<td>CRSP</td>
<td></td>
</tr>
<tr>
<td>Firm size</td>
<td>Total assets value</td>
<td>Compustat</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>Firms are divided into ten main industries according to the Standard Industrial Code (SIC).</td>
<td>Stocktwits</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Variable Measures and Data Sources

**Dependent Variable: Crowd Performance**

In the light of the importance of crowd wisdom, it is essential to study factors influencing crowd performance. Specifically, we use crowd prediction accuracy to measure crowd performance in the research context of online investment community. The dependent variable is a dummy variable indicating whether a crowd predicts the stock trend is in accordance with the actual stock performance. Stocktwits allows its participants to label their postings with a prediction indicator being either “Bullish” or “Bearish”. There are about 16% of messages that have such a label in the whole dataset. In this study, we used the labeled dataset as ground-truth in order to evaluate the crowd prediction performance. We consider crowd prediction on a given stock to be bullish if there are more bullish stock postings than bearish ones on a given day. Similarly, we believe crowd prediction to be bearish if there are more bearish stock postings than bullish ones. The prediction of the crowd is neutral if the number of bullish tweets equals to that of bearish tweets. A dummy variable is used to measure the prediction accuracy of the crowd: 1 indicating crowd prediction for a given stock is in accordance with the actual stock performance, and 0 otherwise.

**Independent Variables**

*Opinion diversity.* In order to assess the content of stock postings extracted from Stocktwits, we develop text mining based measures. Posting opinion distance (Gu et al. 2014) and content similarity are used to measure opinion diversity. We apply a Naïve Bayes classifier to detect posting sentiment by calculating sentiment scores. Stock postings with existing “Bullish” and “Bearish” labels are used as our training set, which is used to train the optimal parameters of the classifier. Sentiment scores range from 0 to 1 with 0
indicates bearish sentiment and 1 indicates bullish sentiment. If a posting has a sentiment score closer to 1 than 0, we consider it to be bullish. Otherwise, we consider it to be bearish. Opinion distance is the variance of sentiment scores across all postings about a given stock on a given day, which can be used to measure the difference among stock posting sentiments. Content similarity is the average content similarity among all postings for a given stock on a given day. The content similarity between any two postings is calculated using the vector-based cosine similarity measure, which is commonly used in text mining research (Koudas et al. 2004). It is reasonable to use opinion distance and content similarity to measure opinion diversity. Opinion distance measures the sentiment difference among stock postings while content similarity measures the content difference among stock postings. We argue that crowd with a high level of opinion diversity is mad up with people who have a high level of opinion distance and a low level of content similarity.

Participant dependence. Community participants who provide influential opinions usually have more followers (Sprenger et al. 2014). In Stocktwits, one person can follow others' postings and be followed by others. The number of following (i.e., the number of people one follows) indicates the degree of influence that the person receives from those he/she follows. The more people a person follows, the more likely the user's opinion is influenced by those who he/she follows (i.e., higher dependence for the focal user). The number of followers (i.e., the number of people who follow the focal user) indicates the degree of influence that a person influences his/her followers. The less followers that a user has, the less likely the user can influence others in the community (i.e., lower dependence for other users). Therefore, we calculated opinion dependence as the difference between the number of followings and the number of followers for those individuals who are involved in stock discussions.

Moderating and Control Variables

We measured crowd size as the number of individuals who were involved in stock discussions. A user may post more than one post per day, but we count him or her only once. We treated crowd size as a moderating variable.

To robustly test the hypothesized effects, we included a comprehensive set of industry-, firm-, and posting-level control variables. They are: Industry, the firms are divided into ten main industries based on Standard Industrial Classification (SIC) codes; Firm size, measured as the total firm assets (from Compustat); Stock volatility, measured as the absolute difference between stock daily high price and low price; Posting experience, measured as the number of postings participants have posted on Stocktwits; Sentiment average, measured as the average sentiment value of all postings for a given stock each day.

We estimate the following model to test our research hypotheses:

\[
\text{Logit (Crowd performance)} = \beta_0 + \beta_1 \text{Crowd size} + \beta_2 \text{Sentiment distance} + \beta_3 \text{Content similarity} + \beta_4 \text{Participant dependence} + \beta_5 \text{Crowd size} \times \text{Sentiment distance} + \beta_6 \text{Crowd size} \times \text{Content similarity} + \beta_7 \text{Crowd size} \times \text{Participant dependence} + \beta_8 \text{Control variables} + \xi
\]

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crowd performance</td>
<td>0.389</td>
<td>0.487</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Crowd size (ln value)</td>
<td>2.176</td>
<td>1.121</td>
<td>0.267</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Opinion distance</td>
<td>0.042</td>
<td>0.029</td>
<td>0.233</td>
<td>0.366</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Content similarity</td>
<td>0.068</td>
<td>0.080</td>
<td>-0.107</td>
<td>-0.227</td>
<td>-0.202</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>5. Participant dependence</td>
<td>6.159</td>
<td>1.872</td>
<td>-0.125</td>
<td>-0.360</td>
<td>-0.293</td>
<td>0.111</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 2. Descriptive Statistics and Correlation

Table 2 provides the descriptive statistics and correlation matrix of the variables in our research model. All correlations are below 0.4, indicating less multicollinearity concern.
Empirical Models and Results Analysis

Stata 12.0 is used to conduct the logit regression in this study. Table 3 reports the final empirical results. Model 1 is a baseline model, which tests the direct effects of opinion diversity and poster influence. Model 2 includes the moderating effects of crowd size. Interpretation of all hypotheses is based on Model 2 results.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crowd size</td>
<td>0.429(0.017)***</td>
<td>0.532(0.059)***</td>
<td></td>
</tr>
<tr>
<td>Sentiment distance</td>
<td>13.908(0.603)***</td>
<td>12.639(0.638)***</td>
<td>H1(supported)</td>
</tr>
<tr>
<td>Content similarity</td>
<td>-1.270(0.266)***</td>
<td>-2.048(0.468)***</td>
<td></td>
</tr>
<tr>
<td>Participant dependence</td>
<td>-0.494(0.051)***</td>
<td>-0.248(0.052)***</td>
<td>H2(supported)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Crowd size × Sentiment distance</td>
<td></td>
<td>-11.259(0.682)***</td>
<td>H3a(supported)</td>
</tr>
<tr>
<td>Crowd size × Content similarity</td>
<td></td>
<td>-1.281(0.414)***</td>
<td></td>
</tr>
<tr>
<td>Crowd size × Participant dependence</td>
<td></td>
<td>-0.397(0.039)***</td>
<td>H3b(supported)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock volatility</td>
<td>0.008(0.005)'</td>
<td>0.007(0.005)</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td>0.0005(0.008)</td>
<td>0.010(0.008)</td>
<td></td>
</tr>
<tr>
<td>Posting experience</td>
<td>-0.165(0.018)***</td>
<td>-0.102(0.019)***</td>
<td></td>
</tr>
<tr>
<td>Sentiment average</td>
<td>4.834(0.159)***</td>
<td>5.034(0.163)***</td>
<td></td>
</tr>
<tr>
<td>Firm size</td>
<td>0.043(0.014)***</td>
<td>0.014(0.014)</td>
<td></td>
</tr>
<tr>
<td>Pseudo R² (%)</td>
<td>11.30</td>
<td>13.56</td>
<td></td>
</tr>
<tr>
<td>LR chi²</td>
<td>3083.33'***</td>
<td>3702.54'***</td>
<td></td>
</tr>
<tr>
<td>Change in R²</td>
<td></td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td>Wald test statistics for interactions</td>
<td>580.10'***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Results of the Hierarchical Regression Modeling**

Notes: N=20409. 'Significant at 0.10; **significant at 0.05; ***significant at 0.01. Standard errors are reported in parentheses.

We find that all main effects are significant after controlling the moderating effect of crowd size. The results of Model 2 show that the direct effect of sentiment distance on crowd performance is positively significant and posting content similarity is negatively associated with crowd performance, providing support for H1. Consistent with our hypothesis, a crowd with high participant dependence is negatively associated with crowd performance, supporting H2.

We use the Wald test to test the moderating effect of crowd size and investigate whether the moderating effect affects crowd decision making (Greene 2003). The Wald test statistic in Model 2 shows that three interaction terms (i.e., $\beta_5, \beta_6, \beta_7$) between crowd size and crowd characteristics are jointly significant ($\chi^2(3) = 580.10, p<0.000$), indicating that crowd size does exert a significant moderating effect between crowd characteristics and crowd performance.

As shown in Model 2, all interactions between crowd size and independent variables are significant, confirming H3a and H3b. Crowd size mitigates the positive influence of sentiment distance on crowd performance.
Opinion Diversity, Participant Independence and Crowd Performance

performance and magnifies the negative impact of content similarity and participant dependence on crowd performance.

Conclusions and Discussions

With the popularity of the Internet, online investment communities have become one of the most popular approaches for individual investors to share and interact with other investors. Investor generated content on OICs is updated rapidly and spreads at an unprecedented speed, providing first-hand information to other investors ahead of other sources. Investors involved in the discussion of the same stock can be seen as a crowd confronted with the same problem. Thus, online investment communities provide us with a good context to study how online crowds make decisions and what are the influences of crowd characteristics on crowd performance.

Our findings show that opinion distance is positively related to crowd performance; content similarity and participant dependence are negatively associated with crowd performance; and crowd size significantly moderates the influence of opinion diversity and participant dependence on crowd performance. This study has both theoretical and practical implications. From a theoretical perspective, our study enriches extant research by focusing on the relationship between crowd characteristics and crowd performance. Specifically, we developed text mining based measures to quantify opinion diversity, which can be readily applied to a large amount of data collected from an online community. From a practical perspective, our findings help individual investors better evaluate user generated stock trend predictions, based upon which they can make better investment decisions.

This study has several inherent limitations due to the sampling methods and measurements used. First, a convenience sampling method was used. We used S&P100 as our research sample, which is small, future studies should increase the sample size in order to improve generalizability. Second, we did not control time heterogeneity in this study, and we will consider the time factor in our future study. Third, crowd decentralization is not considered in our research model. Further research should be conducted to objectively measure crowd decentralization and identify the influence of crowd decentralization on crowd performance.

Acknowledgements

The authors would like to thank the editors and reviewers for their helpful and constructive suggestions. This research was supported by the Natural Science Foundation of China with Grant 71531013 and China Scholarship Council with Grant 201506310121.

REFERENCES


