“Deliberated Intuition for Groups”: An Explanatory Model for Crowd Predictions in the Domain of Stock-Price Forecasting

Tobias Endress
University of Gloucestershire
tobiasendress@connect.glos.ac.uk

Tony Gear
University of Gloucestershire
agear@glos.ac.uk

Abstract

Crowd predictions in the domain of stock-price forecasting is a fascinating concept. Several special-interest online communities were founded following this idea. However, there is a limited body of literature about the domain of stock-price predictions based on such a crowdsourced approach. This paper presents an empirical study in the form of a two-phase, sequential mixed-methods experiment.

Data from purposefully designed groups, consisting of lay people and professional financial analysts, were examined to inform the understanding of the prediction process. The findings led to an explanatory model, which we introduce as ‘deliberated intuition for groups’. The model of deliberated intuition for groups, which is proposed here, views prediction as a process of practice which will be different for each individual and group. The model proposes that a predictor will decide, consciously or semi-consciously, either to rely on gut-feeling or to undertake more analysis.

1. Introduction

Traditionally, banks and specialized research companies create company analysis reports, calculate price targets and provide them to potential investors. However, despite the professionalism and typically high qualifications of these experts, their forecasts are still wrong in many cases, and uncertainty remains in the investment decisions.

Recently, a new approach has appeared beside traditional research, a kind of crowdsourced or collective intelligence research. Not least through the increasing popularity of digital social networks, also known as “online communities”, the topic of collective intelligence has become widely discussed. Surowiecki highlights [1] some very interesting cases where collective intelligence approaches led to more accurate results than the forecasts of experts. The collective approach with groups consisting of members from different locations and working at the same time and with combined forces on a solution to a problem is a fascinating concept [1, 2]. Following this idea, many special interest online communities, such as marketocracy.com, sharewise.com, predictwallstreet.com or stockjaeger.de, were founded. These communities are essentially service providers with different applications of open-collaboration crowdsourcing focused on the topic of equity trading and on different kinds of company analysis, mostly leading to forecasts of stock prices. Despite this popularity in business, there is a very limited body of literature available covering this domain with empirical rigour. This study aims to contribute using domain-specific insights on the underlying key mechanisms in the context of crowd-based equity predictions. The research question asks what the underlying key mechanisms are that influence the predictive process of crowds, such as groups in online communities, and how might the process be improved.

2. Literature Review

Even though some recent authors argue that they have observed a “new collaborative economy” [3], the literature suggests that the idea of letting a group decide is not as new as the popularity of the book The Wisdom of Crowds [1] might indicate. In fact, there has been a lot of research about group-versus-individual decision-making. This “new” approach might contain or combine parts of established group-individual decision-making-procedures, such as the Delphi methodology [4, 5], the nominal group technique [6], prediction markets [7, 8], the social psychology of groups [9]–[11] and group support systems [12, 13], which are covered quite well by academic research.

Further research on group decisions has shown more and more difficulties and hurdles in decision-making [14]–[16]. Some authors argue that there is a body of empirical and theoretical evidence indicating an advantage in combining different forecasts [17]–[19]. The modality of crowdsourcing [20] applied by special interest online communities which inspired this experiment can be seen as open collaboration [21].
Aside from the aggregation algorithm, expertise and cognition were found to influence the prediction process [22]–[24]. Numerous authors work to enhance the concept of crowd wisdom [20], [25]–[27]. However, there is a limited body of literature about the domain of stock price predictions based on a crowdsourced approach. While it seems to be a popular claim that the “wisdom of the crowds” is applied in the context of open collaboration equity predictions [28]–[30], it might be questionable that the approach of crowdsourced equity predictions is an application of the very specific phenomenon called “crowd wisdom” [31] or collective intelligence [32]. One key paper in the domain of stock-price predictions was presented by Kaplan [33]. He described the design and the first tests of a prototype system that was supposed to create stock trading recommendations based on input from a crowd. He claimed that “there is a growing body of evidence that the key to forecasting the stock market lies neither in value analysis nor in technical analysis. Rather, investor psychology seems to be the critical factor” [33, p. 1]. In 1936, Keynes had already stated that many professional investors did not determine the proper value of an investment, but anticipated how the crowd of investors might act during optimistic periods in so-called bull markets, with their expectations and hopes being ‘castles in the air’. To be successful, an investor only has to buy before the crowd builds the castle too high [34]. There are several models in place to explain investor psychology, such as the “prospect theory” [35] or the “behavioural model of rational choice” [36], [37]. Simon’s model takes into consideration that access to information and the computational capacities of man are limited. It might be true that forecasting “is a skill that can be cultivated” [38, p. 4]. However, it is certainly not an easy exercise to promote such cultivation. The literature suggests that one important factor is to understand the underlying role of intuition in the process [39].

3. The Assessment of Equity Predictions in Online Communities

Many special interest communities focus on decision-making, using a remote group process to create equity price predictions, but the literature review suggested that so far no academic evaluation of when or if this practice is effective has been conducted. This study assesses the practice in terms of the conditions that may enable it to outperform equity research experts. The aim of this two-phase, sequential mixed-methods study is to develop an explanatory schema and create a theory to begin understanding why and when it happens.

3.1. Methodology and Methods

This research project is based on empirical data and appropriate methods [40]. The data for this research were gathered primarily from two sources: a controlled experiment and interviews.

In existing stock trading communities on the Internet, the data availability and their quality might not be reliable. For this reason, a field-based experiment was conducted in a defined and controlled environment. Financial analysts and a group of lay people using a defined process (e-Delphi) generated the data needed. The field experiment was conducted following an e-Delphi approach as a simple variation a crowdsourcing modality in open collaboration. Every e-Delphi cycle in this experiment consisted of a first round for data collection. These data were compiled and distributed among the panel. In a second round, participants were allowed to give different answers in respect to the feedback they got from the group’s decision in the first round.

The design and approach of the main experiment was similar in principle to the design of a preceding pilot run. The pilot run allowed the improvement of the process and structure of the main experiment. The main experiment was conducted using a larger sample, a longer period, more shares and a wider range of different group designs, including setups where the members of the crowd never interacted with each other and more collaborative approaches with feedback loops. The main run was performed with 59 participants distributed in five groups:

- The analyst group (AG), with a group size of 5 participants.
- The e-Delphi group (EDG), with a group size of 21 participants (lay person).
- The interactive group (IG), with a group size of 7 participants (lay person).
- The non-feedback group (NFG), with a group size of 21 participants (lay person).
- The professional investors group (PG), with a group size of 5 participants.

Financial professionals who were not investment experts, including regular bank clerks, insurance brokers etc., were excluded from the sampling, because they are an indefinite middle-ground group. The
primary data collection followed an adapted e-Delphi method. Each e-Delphi cycle in this experiment consisted of a first stage for data collection regarding predictions. These data were compiled and distributed back to the group. In a second round, participants could provide different responses; 10 cycles were conducted during the main experiment. The shares were selected from five different companies in different sectors and during 3 periods (1-week, 1-month, and 3-month predictions) to allow further differentiation. In all, the main experiment was set up to gather up to 17,700 individual judgements about equity predictions (i.e. 5,900 individual judgements about equity predictions for each period).

Data from financial analysts were used to compare the predictions from the groups with the actual results of share prices. All participants were individually assessed, including the determination of variables, such as age, gender, education level, profession and decision-making type (PID-score) [41].

Additionally, all participants were interviewed in parallel to the e-Delphi rounds. The e-Delphi online questionnaire and the accompanying in-depth interviews included questions which aimed to identify patterns of information influence on the decision-making process of the individuals and groups. Eventually, this contributed towards crowd science backed by a domain-specific empirical research in the area of equity predictions based on an open collaboration crowdsourcing approach.

3.2. Data and Analysis

A quantitative data analysis was conducted using a sequential approach. The first step was to use a univariate analysis. In the second step, a multi-criteria analysis and data reduction techniques were applied (with SPSS and Excel). Both approaches aimed to support an understanding of the factors that influence the decision-making process and forecast quality. The absolute performance of the aggregated predictions from lay groups (52% correct predictions, see Table 1) was slightly above the value of 50% correct predictions, which would be the expected value with a purely random distribution (correct/wrong in the ratio 1:1). However, this outperformance is not significant (Chi-square 0.688; p-value = 0.406).

Table 1. Aggregated Predictive Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Sum (All Lay Groups)</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>407</td>
<td>54</td>
</tr>
<tr>
<td>Wrong</td>
<td>403</td>
<td>36</td>
</tr>
<tr>
<td>Excluded</td>
<td>60</td>
<td>10</td>
</tr>
<tr>
<td>Correct (%)</td>
<td>52.0%</td>
<td>63.3%</td>
</tr>
<tr>
<td>Wrong (%)</td>
<td>48.0%</td>
<td>36.7%</td>
</tr>
</tbody>
</table>

Even though we made the assumption that the prediction accuracy might follow a purely random distribution, this assumption may not be correct. The data from the experiment also allowed relative comparisons between the different groups and group members. The relative performance on an aggregated level from the lay groups compared with the expert recommendation were significantly different. The predictions by financial analysts were significantly better than those by lay groups (Chi-square 10.55; p-value = 0.001).

Table 2. Aggregated 3-Month Main Run Predictions

<table>
<thead>
<tr>
<th></th>
<th>AG</th>
<th>EDG</th>
<th>IG</th>
<th>NFG</th>
<th>PG</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>41</td>
<td>70</td>
<td>42</td>
<td>59</td>
<td>62</td>
<td>60</td>
</tr>
<tr>
<td>Wrong</td>
<td>46</td>
<td>27</td>
<td>51</td>
<td>37</td>
<td>22</td>
<td>30</td>
</tr>
<tr>
<td>Excluded</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Correct (%)</td>
<td>47.1%</td>
<td>72.2%</td>
<td>45.2%</td>
<td>61.5%</td>
<td>73.8%</td>
<td>66.7%</td>
</tr>
<tr>
<td>Wrong (%)</td>
<td>52.9%</td>
<td>27.8%</td>
<td>54.8%</td>
<td>38.5%</td>
<td>26.2%</td>
<td>33.3%</td>
</tr>
</tbody>
</table>

The 3-month prediction results (at least in part) contradicted the suggestion that crowds of lay people predict shared price movements better than the experts (see Table 2). From 100 predictions (m = 100), the EDG had 70 (72.2%) correct predictions, and the financial AG had only 41 (47.1%) correct predictions; however, the IG had just 42 (45.2%) correct predictions, the NFG had 59 (61.1%) correct predictions and the single expert had 60 (66.7%) correct predictions. The best performance in the main run was from the PG, with a frequency of 62 (73.8%) correct predictions.

An analysis of the predictions to test the equality of group means (for groups with correct and incorrect predictions) with SPSS discriminant analysis tools indicated that some variables revealed significant differences between correct and wrong predictions, but
there are also a number of variables with no significant differences. However, the explanatory power of the analysed variable assignment is limited (as it explains only a small percentage of the variability, from 15.1% for 1-week predictions and 20.5% for 1-month predictions to 29.8% for 3-month predictions).

Information processing and prediction approach

Further analysis of the participants’ data provided during the survey to the question about the basis of their decision and influencing factors helped identify mechanisms and patterns. The data on the use of information sources and basis during information gathering, processing and prediction approach were different between participants.

The analysis of the data from the experiment showed that there were indeed a number of participants who appeared to be especially good at predicting stock prices, as compared with the average of other members of a given group, of which they are members. All participants were interviewed to gain a better understanding of the differences between the best and the less good participants. At first glance, it appears that the top predictors don’t have a lot in common. However, an in depth-analysis reveals some striking patterns. A few variables have proven to be of significant importance. E.g. a test of the aggregated predictions grouped by gender suggests a highly significant correlation with a chi-square of 14.681 and p-value < 0.001. Apparently, female participants were significantly more accurate with their predictions in the experiment. Also, the predictions of people with a preference for both intuition and deliberation (PID-S-plus) participants are apparently of significantly higher accuracy (see Table 3).

<table>
<thead>
<tr>
<th>PID</th>
<th>Participants</th>
<th>Correct</th>
<th>Wrong</th>
<th>Sum</th>
<th>Percentage of correct answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID-D</td>
<td>22</td>
<td>2610</td>
<td>2514</td>
<td>5124</td>
<td>50.9%</td>
</tr>
<tr>
<td>PID-I</td>
<td>11</td>
<td>1320</td>
<td>1245</td>
<td>2565</td>
<td>51.5%</td>
</tr>
<tr>
<td>PID-S minus</td>
<td>17</td>
<td>2098</td>
<td>1967</td>
<td>4065</td>
<td>51.6%</td>
</tr>
<tr>
<td>PID-S plus</td>
<td>9</td>
<td>1082</td>
<td>883</td>
<td>1965</td>
<td>55.1%</td>
</tr>
</tbody>
</table>

Table 3. Comparison of predictions grouped by PID scale score

The interview sessions were recorded, and the audio files were completely transcribed. All interviews were imported to MAXQDA, where all coding of the interviews was conducted. The codes were clustered in several categories to allow for a more systematic analysis. The categories used in the systematic qualitative interview analysis were inspired by Kuckartz’s [42] proposed spectrum of categories, but

Are people who are interested in the stock market able to provide better predictions?

The data from the experiment doesn’t support the assumption that people who are interested in the stock market are able to provide better predictions. Participants who stated in the interviews that they had no or very little interest in stock markets had an average predictive accuracy of 51.5% (N = 24), while participants who stated that they were interested in stock markets had an average predictive accuracy of 51.1% (N = 29). Six participants did not answer the question; their average accuracy was 52.3%.

All the experts stated that they are interested in stock markets. An analysis excluding the experts shows that lay participants who stated in the interviews that they have no or very little interest in stock markets had an average predictive accuracy of 51.5% (N = 24), while participants who stated that they are interested in stock markets had an average predictive accuracy of 50.0% (N = 19).

This result suggests the conclusion that whether people are interested or not at all interested/show very little interest in the stock market has no impact on their ability to predict stock price movements. This finding is quite interesting in light of the idea that a driver for crowd intelligence is diversity in the group design. While self-selection might be an important factor in people’s participation in Internet groups, there might be also a tendency for like-minded people to meet in these groups and create a reconfirmation of existing prejudice instead of an exchange of different opinions. In certain situations, it might be beneficial to the overall group decision-making to have people with different mindsets.

Insight from the interview analysis

The interview sessions were recorded, and the audio files were completely transcribed. All interviews were imported to MAXQDA, where all coding of the interviews was conducted. The codes were clustered in several categories to allow for a more systematic analysis. The categories used in the systematic qualitative interview analysis were inspired by Kuckartz’s [42] proposed spectrum of categories, but
the categories have been selected and adjusted to fit the research question.

In summary, the interview analysis revealed that it was equally difficult for good and poor predictors to answer. There was also no perceivable difference in the use of the Internet. All participants used the Internet at least on a daily basis. However, there are differences between the top and the worst predictors in other domains. It seems that for good predictors, the ideal is to make “rational” decisions. One of the top predictors even mentioned explicitly that he tries to be rational because “emotions are not good”. All but one top predictor mentioned that they are interested in the stock market, while at least two of the poor predictors mentioned that they are not interested or have very little interest in this topic. A particularly striking difference is in the bases on which they make decisions: While poor predictors mentioned that they rely on news (and their gut feelings) to a large extent, there was a much more differentiated picture for the top predictors. The top predictors also emphasized slightly more that they were more sensitive to news related to companies in the experiment.

4. Conclusion

The experiment provided a rich data set with qualitative and quantitative components [43]. The analysis of these data sets provided insights into the decision-making process and predictive qualities of stock-price predictions from online groups (to simulate crowd intelligence) and professional equity analysts. The research study indicated that equity predictions by Internet groups are not, per se, superior to predictions by professional equity analysts. However, in addition to the personal predisposition of the predictor, more factors contribute to good predictive quality. Superior predictions were observed from participants who used an approach which could be described as informed intuitive prediction. In-depth knowledge and training with deliberative forecasting methods, like formal prediction models and market experience, enhances predictive quality.

Even though no single factor makes a good predictor, it can be concluded that several variables impact the accuracy of stock-price predictions. The findings indicate also that intuition plays a significant role in the prediction process. It was found that

---

Figure 1: Deliberated Intuition Model for Groups
particularly good predictors base their intuition on several factors—even including fundamental and macroeconomic considerations. Different variables in terms of the individual characteristics of the participants indicated a significant impact on the quality of equity predictions. These are in particular educational level, gender, decision-making type, but some relevant variables are also related to a particular prediction, such as confidence, upside or downside potential and decision-making approach. The findings led to an explanatory model, which we introduce as ‘deliberated intuition’ [43]. The model proposes that predictors will decide, consciously or semi-consciously, when they feel ready to rely on gut-feeling or to undertake more analysis. It suggests three clusters of antecedents of the quality of prediction and comprises the personality traits of the individual, their individual experience and training and the situation in terms of risk and social context. The concept of ‘bounded rationality’ [36] is compatible with ‘deliberated intuition’, implying that the decision to decide ‘enough is enough’ will depend on personal traits and perceived risks coupled with social context, training and experience. The model of deliberated intuition for groups, which is proposed here, is an enhancement of this model and views prediction as a process of practice which will be different for each individual and group. It includes group factors, such as influence through the group setting, approach and the interaction of group members. Crucial factors include the discussed factors and group interaction [44], trust [45], [46], peer pressure [47] and aggregation methods [20]. The model proposes that a predictor will decide, consciously or semi-consciously, when they feel ready to rely on gut-feelings or to undertake more analysis. Generally, it appears to contribute to a good prediction to think about the problem in different ways and with various techniques.

The application of this ‘deliberated intuition model for groups’ has implications for the predictive quality of online or face-to-face groups making predictions. A careful group design with the consideration of the mix of personality traits and the training and experience of the participants might improve the predictive quality of group output. Further research is needed to identify the optimal design, size and composition of such groups. However, diversity in terms of intuition and deliberative perspectives within the group and a process that facilitates both perspectives being considered in the group decision-making process might be the key to reaching superior predictive quality (see Figure 1). Still, these findings are from an experiment with a limited number of participants, and the study should be repeated with a larger sample size and in different settings. The next step should be purposeful sampling to allow for an in-depth assessment of the impact of the individual factors of Figure 1 and their interdependencies in a group setting.

The next steps include improving the reliability and usability of the deliberated intuition model for groups, conducting an experiment on a larger scale and experimenting with variations of group design and composition to identify those that are most effective. The deliberated intuition model for groups might also be useful for selecting members for crowds [48], [49]. Further research should also include variations in cultural context, investment instruments and market conditions, as well as the assessment of additional variables.

5. References


