Stock analysts vs. the crowd: Mutual prediction and the drivers of crowd wisdom

Matthias Eickhoff*, Jan Muntermann

University of Göttingen, Chair of Electronic Finance and Digital Markets, Germany

ABSTRACT

We examine the drivers of crowd wisdom in the financial domain by relating analyst report and social media sentiment via Granger causality (GC) testing based on the wisdom of crowds (WoC) theory. The significance of a large number of the tested time series indicates that analyst reports and social media content are suitable for mutual prediction. We elaborate on the conditions under which crowd cognitive diversity matters, and we derive related measures. The results suggest that the WoC theory can partially explain the GC between the two media types and that both professional analysts and the crowd can outperform one another under favorable circumstances.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The impact of financial analyst reports has been subject to increasing scientific scrutiny. In particular, the herding behavior of financial analysts, that is, how some analysts seem to affect the opinion of others [44], has been analyzed extensively [11,25,43]. The reaction of the capital market has also been analyzed. Traditional analyses have dealt with buy/hold/sell recommendations, from the Institutional Brokers' Estimate System (I/B/E/S). More recently, researchers have been applying text-mining methods to analyst reports to automatically extract more information than previously available. The analysis of the entire report is desirable, as it may yield further information beyond the constrained categorization of the stock [44]. In contrast to prior studies, this study focuses on how the opinions of professional stock analysts, that is, individuals who are paid to provide regularly updated opinions about certain companies, relate to those of social media users. The study also analyzes whether the two related content types can be used to predict each other's sentiment. Although studies have shown that social media content can be used to forecast stock returns similar to analyst recommendations, this does not necessarily imply that the prediction powers of the two mediums are inherently related. Such a relation is of interest because professional analysts have access to privileged information, as illustrated in Section 2.

However, analysts also face a number of constraints that can influence their recommendations. These constraints include incentives to generate in-house brokerage volume [27] and the tendency to stick to established recommendations [11,43]. The average social media user does not face these constraints.

Therefore, such a relation could allow the incorporation of social media content in models traditionally using traditional stock analyst recommendations for predictive purposes. This incorporation could serve as a control variable for the analysts' biases. Therefore, the aim of this research is to investigate the presence of such a relationship between content types in either direction. Social media and analyst report data regarding the 30 component companies of the Dow Jones Industrial Average (DJIA) index are collected from the year 2013. First, we perform sentiment analysis on analyst reports and social media data to establish time periods (quarterly and annual) and company combination Granger causality (GC) between social media and analyst sentiment. We then aim to determine reasons for different cases of GC direction. Afterthis section, the theoretical background of the analysis is introduced. Theoretical insight on the information value of analyst reports and the concept of wisdom of crowds (WoC) is compared in Section 2, focusing on inefficiencies in analyst opinions likely to be mitigated by social media content. Based on these foundations, hypotheses are developed and tested. In Section 3, methodological foundations are introduced and comprise sample generation, GC, sentiment analysis, and variable operationalization related to WoC constructs. We then present our analysis and empirical results in Section 4. Using binary response models, we provide evidence regarding which types of companies and public interest foster GC

* Corresponding author.
  E-mail address: Matthias.Eickhoff@wiwi.uni-goettingen.de (M. Eickhoff).

directions. This section also examines drivers of WoC in our social media sample by introducing variables for measuring WoC-related conditions proposed by Surowiecki [42]. The conclusion summarizes the results of the analysis and elaborates on their theoretical value for WoC and analyst research.

2. Theoretical background

This research aims at combining insights from research regarding the information value of analyst reports with the principle of social media user sentiment through the lens of crowd wisdom.

2.1. Information value of analyst reports

Analyst reports have been researched, largely to answer the following question: How do stock analysts influence the stock market, and how do stock analysts arrive at their conclusions?

The former question addresses the information value of analyst reports and their recommendations (i.e., whether they can be used as a basis for supporting investment decisions). It is assumed that analysts have privileged access to relevant information about companies by thorough research or close relationships with these businesses. Therefore, they are viewed as information providers capable of improving information efficiency [18] and shortening the time between publication and incorporation of information into stock prices [15,26]. Answers to this question have been mixed. While earlier studies suggested that investments based on analyst recommendations can be profitable [3,46], recent research doubts the impartiality of stock analysts [4,8]. These doubts motivate the second question (i.e., upon which information stock analysts base their recommendations).

Several studies have provided evidence for herding behavior among stock analysts. Herding behavior refers to the tendency to provide recommendations close to those of the consensus [44]. This behavior introduces a bias toward the status quo. Career concerns are largely responsible for this tendency [11]. Career concerns are especially relevant for younger analysts who fear termination if they make bold predictions and fail [25]. Groysberg et al. [22] found that compensation schemes are designed to increase brokerage and investment-banking revenues. Therefore, other data sources offering insight into companies are desirable. Social media content is widely available for a large number of companies, and social media users are not faced with the same repercussions or incentives as stock analysts. Thus, social media users might be able to provide less biased opinions about a company's current state or future developments. Earlier research indicates that social media content (Twitter) can be used to predict stock returns [6]. The sentiment of social media users and stock analysts may be used in a similar manner [38]. The latter issue is of particular interest to this study, because deficiencies of stock analyst recommendations necessitate alternate data sources about companies. Social media content is a possible data source that is investigated in this paper.

Alternate sources of information can help mitigate known biases. Social media users should not be faced with the same problems as professional analysts and are unlikely to be punished if their opinions are wrong. The next section explores how models that currently rely solely on analyst opinions can be augmented.

Poetz and Schreier note that expert knowledge can lead to superior skills and problem solving within a given domain [40]. This assertion is supported by previous research [1,33]. However, this superiority is imited in its predictive accuracy of expert opinions [28]. While earlier sources mainly stem from psychology, crowd-driven projects have increased due to advances in networking. A popular example of such a project is Wikipedia, which has been shown to be as accurate as the Encyclopedia Britannica [20].

The efficient market hypothesis suggests that individual actors cannot outperform the market as the current price of a stock should incorporate all available information [14,36]. However, conventional market theory may not hold true in the context of crowd wisdom. Recent studies indicate that stock prediction communities, which are small special-purpose social networks, can achieve higher performance than the market in general [23].

2.2. Wisdom of crowds

The average social media user is not likely to have the finance background of stock analysts. Prior studies show that social media indicates mood and may assist in making stock market predictions [6,38]. Surowiecki describes this novel source of expertise as the WoC. The WoC theory proposes that large independent and heterogeneous groups can outperform smaller groups in their assessments even if the smaller group consists of subject matter experts [42].

Poetz and Schreier [40] define a crowd as "potentially large and diverse". Beyond these characteristics, no further assumptions will be made on the makeup of crowds.

In contrast to a crowd, our expert group can be described more clearly. According to Nofer and Hinz [37], an expert is "a professional analyst from a bank or research company who has experience in his area of expertise: publishing share recommendations and predicting the stock market development." In order for crowd wisdom to emerge, a group should satisfy three key conditions: diversity, independence, and decentralization [42]. Aggregation of the crowd's diverse opinions is required to reach a consensus decision [42]. In contrast to authors of Wikipedia articles, social media users have no explicit intention to aggregate their collective sentiments. Their intent is to share their opinions with the community. The text-mining methods applied in this study shall serve as a humble substitute. This study investigates the extent to which groups lacking the intent to reach a common goal are not able to create a WoC effect easily.

Previous studies generally assume that these three conditions are satisfied. In this study, we derive measures for each condition and assess their influence on the crowds' predictive power.

In order to confirm that these conditions are satisfied, we examine substantial social media data in each of these categories. Building on previous research, we develop measures to operationalize these conditions. We treat them as factors that may explain situations from which crowd wisdom actually arises.

2.2.1. Diversity

The presence of group diversity is the central concept of the WoC theory. A group consisting of some informed and uninformed individuals will tend to outperform those consisting only of experts, even if experts are more informed than any member of the other group [42]. The argument for a positive effect of group diversity hinges on two separate effects.

First, diversity ensures that the group's opinion is based on sufficiently varying individual perceptions. Sorowiecki argues that "[diversity] expands a groups' set of possible solutions and allows the group to conceptualize problems in novel ways" [42]. This argument does not refute expert opinion. Instead, Surowiecki suggests that exposing experts to less informed individuals challenges their opinions and forces them to make stronger arguments.

Second, larger and more diverse groups make voicing dissenting opinions easier. This reduced risk is due to the increased possibility of finding allies supporting a novel point of view. Even if a new opinion does not find support, the increased number of

perspectives makes failing to convince the group more acceptable. Group diversity creates an atmosphere that is open to debate.

2.2.2. Independence

Surowiecki [42] suggests independence is a counterweight for herding behavior (i.e., the tendency to conform to the opinion of others). This factor reduces the risk of information cascades [42] that threaten to override individual opinions over the group's opinions. This threat results in the loss of added information value that dissenting opinions can offer. Independence reduces the correlation between individual opinions, reducing the risk of spreading erroneous judgment from one individual to the next. In addition, independent individuals contribute their own perspectives instead of conforming to predominant points of view [42].

2.2.3. Decentralization

A centralized group should face fewer problems coordinating and aggregating decisions. However, decentralization is critical for group diversity and independence. Decentralization also makes the group more permeable. This allows members familiar with particular problems to participate in discussions when their expertise is needed. This ability is "at the heart of decentralization" [42]. However, the price of decentralization is increased communication and coordination costs and the possibility of redundancy. The increase in flexibility and the support decentralization offers may outweigh its cost.

The conditions under which crowd wisdom can arise are neither easily separated from one another nor directly observable. Therefore, a later section of this study translates these theoretical constructs into operational definitions, that is, into associated variables and their attributes that are observable and measurable.

Hypotheses. Previous findings lead us to the conclusion that analyst recommendations may be biased because of the payment incentive structures in which professional analysts work. Social media users do not face these types of incentives. Social media and analyst recommendations may be viewed as supplementary sources of information about companies.

Alternatively, social media users may not have access to all relevant information available to stock analysts and have no coordinated way to aggregate their opinions. Thus, there may be no uniform answer to the question of whether professional analysts or social media users are quicker when incorporating new information into their opinion. This would also be the case if both groups are equally good at the task. Finally, this would be observed if there are situations or contexts in which either groups' advantages outweigh the other groups. Against this background, hypothesis 1a is proposed:

Hypothesis 1a. There is no uniform direction of lead–lag between the two content types, and neither professional stock analysts nor social media users always shift their mean sentiment quicker than the other group.

However, their financial expertise and domain-specific training may enable professional analysts to develop superior information processing capabilities. Moreover, when analysts are able to communicate directly with corporations, they might be able to extract information that is of particular value to their analysis. Even if this information is made public immediately, others might not be able to gain insights from it. Against this background, hypothesis 1b is proposed:

Hypothesis 1b. Professional analysts are able to incorporate information into their assessment prior to social media users, therefore shifting their sentiment quicker when their expertise is relevant or they operate on superior information.

However, against the theoretical background of WoC research, we hypothesize the opposite, that is, that cases exist where (a) the information is not privileged and therefore accessible to social media users and (b) the user groups satisfy the three key conditions for crowd wisdom proposed by Surowiecki (2005) to arise. The combination of decentralization, independence, and group diversity may enable social media users to arrive at conclusions that professional analysts either are unable to reach on their own or may take longer to process public information newly available. In example, an *age-diverse* group of users may be able to predict the impact of a product announcement, such as a new generation of smartphone, in a more timely fashion if the product will be used by consumers of all ages. This is because a single analyst lacks insight into what each age group expects from this type of product. A *decentralized* group may be able to judge a product announcement's implication for different markets in a similar manner, if a sufficient number of users from each market is represented. A single analyst may lack the cultural insight necessary to understand how different communities will perceive the product. Finally, *independence* within the social media user group may improve the aggregate of social media users' opinions by enabling them to utilize their own reasoning instead of following an established opinion. If one or more of these effects are present in related social media data, we should be able to see that social media users adjust their opinions more timely compared to individual analysts. Against this background, we propose hypothesis 1c:

Hypothesis 1c. Social media users' mean sentiment changes faster than that of professional analysts if the discussed conditions for crowd wisdom are satisfied.

These conflicting theoretical considerations will be investigated in the analytical part of this paper. In addition, we examine the drivers of crowd wisdom in our social media sample. The aim of this analysis is to examine whether the WoC theory can help explain the situation in which a particular group of people possesses superior information processing capabilities. To this end, we derive measures for each of the three conditions and propose hypothesis 2:

Hypothesis 2. There is a positive relationship between the extent to which a group satisfies the conditions of crowd wisdom and the likelihood of a successful prediction of analyst opinion by social media users.

3. Methodology

3.1. Sample

Three datasets are used, each containing information about the companies included in the DIJA in 2013. Table 1 shows an overview of the data.

Each dataset is collected for the year 2013, starting 1 January and ending 31 December. The first dataset consists of analyst reports from the Thomson Reuters Advanced Analytics (TRAA) platform. After having extracted textual content, 9439 observations remained. The second dataset contains a broad selection of social media content, including web logs, forums and product reviews. These data were obtained from the SDLs SM2 database, which is primarily intended for marketing. A maximum of 40,000 observations per quarter was requested from the database for each company, resulting in 3,814,839 observations (approximately 127,000 per company). A public news dataset was obtained from the Guardian open API, which contains 21,278 news articles from the same period. These articles are annotated with their respective news category, such as "financial," "technology," or 'environment.'

Table 1

Observation counts for analyst report (TRAA), social media (SDL), and news (Guardian Open Platform) data.

Company Name	Analyst Reports	Social Media	Public News
3M	177	127,547	435
AT&T	345	83,053	101
American Express	356	152,781	6889
Boeing	510	141,357	136
Caterpillar	362	133,019	78
Chevron	249	144,790	57
Cisco	541	152,742	81
Coca Cola	190	127,533	257
Disney	250	153,809	461
DuPont	272	148,654	44
Exxon	204	133,028	65
General Electric	166	140,357	7531
Goldman	232	142,807	293
Home Depot	253	144,170	3
IBM	337	146,425	135
Intel	525	147,317	157
JP Morgan	331	140,364	71
Johnson&Johnson	361	141,091	1900
McDonald's	375	142,714	109
Merck	397	140,500	27
Microsoft	464	156,697	759
Nike	236	129,887	241
Pfizer	252	142,273	49
Procter & Gamble	264	111,625	458
Travelers	212	4,413	167
UnitedHealth	281	41,876	14
United Technology	268	44,247	1
Verizon	405	155,642	204
Visa	291	137,386	401
Walmart	333	106,735	154
Total	9,439	3,814,839	21,278

Their daily median (i.e., the most common news category for a given firm on a specific day) will serve as a basis for the second stage of the analysis.

3.2. Granger causality

The question whether analysts or social media users react faster to new (exogenous) information will be analyzed via GC testing. This method is easily misunderstood because of its name. A GC test compares a model explaining a time series value using lagged values with one that adds lagged values of a second time series [30]. Thus, x is Granger causal to y if

$$\sigma^2_{M2}(y_{t+1}|I_t) < \sigma^2_{M1}(y_{t+1}|I_t - \bar{x}_t),$$

that is, the forecast error σ^2 is reduced by including the past values of x. No causal relationship is implied if GC is discovered. Asserting such a relationship purely on the basis of GC would be a post hoc fallacy [13]. Pearl does not classify GC as a causal, but as a statistical methodology [39].

Three outcomes can occur for a GC test between a pair of time series. First, no GC relationship is observed. This outcome would entail that neither series could statistically improve the prediction of the other. Second, a GC relationship is observed in one direction. Third, a GC relationship is noted in both directions. For each company's pair of social media and analyst reports, five different models are estimated per possible direction of the GC. One model for the entire year 2013 and four quarterly models are presented. Each follows the model specification with the null hypothesis that there is no GC (i.e., that $M2$ does not reduce the forecast error):

$$M1 : Y_t = \alpha + \sum_{i=1}^{n} \beta_i Y_{t-i} + \varepsilon_t$$

$$M2 : Y_t = \alpha + \sum_{i=1}^{n} \beta_i Y_{t-i} + \sum_{i=1}^{n} \gamma_i X_{t-i} + \varepsilon_t$$

n refers to the number of lags included in each model and t to a specific period. If only a small number of lags are included in the selection of the lag-length parameter, a present relation between the time series is potentially missed. Including many lags can lead to spurious results. The selection of this critical parameter is outlined in the section presenting our analysis and empirical results. With the establishment of the method, the question as to which insights can be gained by the following analysis can be answered. The aim is to analyze whether and under which circumstances analyst reports and social media sentiments may be used to predict one another. Regarding the first question, the following cases might occur as a result of GC testing between the two types of content:

- Case 1–GC exists, in both directions between the two types of content. This provides mixed evidence supporting WoC considerations and the importance of expert knowledge. Due to the nature of GC testing, such cases are expected and the two theoretical foundations are not mutually exclusive.
- Case 2–Social media content is found to GC analyst reports. This proves that events were not foreseen by domain experts and were incorporated into the public opinions of social media users in a more timely manner. Such a relation could indicate that, besides the known herding behavior of analysts, they also follow public opinion about a given company.
- Case 3–Analyst reports are found to GC social media content. This confirms that, due to superior knowledge, analysts are able to assess situations before the crowd can arrive at a similar conclusion. As it is unlikely that social media users have direct access to analyst reports, the impact of such reports on the opinion of the crowd has to be by proxy. This proxy occurs via traditional media channels reporting on the professionals' opinions or individual star users within a social community.
- Case 4–No GC is found between the two types of media content. The contained information appears to be independent from one another. This result indicates that the two types of data are not interchangeable. Still, this does not contraindicate their predictive power with regard to other data, such as stock returns.

3.3. Sentiment analysis

Sentiment scores are calculated for each document as an input for our GC testing. This is performed with the General Inquirer (GI) software and "Positiv" and "Negativ" categories from the Harvard IV-4 dictionary [41]. The use of such dictionaries assumes that the contained words have a prior polarity [45] (e.g., the word "good," when considered without context, will be perceived as positive by most people). This prior polarity is used to assign words to a sentiment category. However, a word's prior polarity will not always coincide with its contextual polarity (e.g., 'fast' might be contained as a positive word in a dictionary for the automobile domain and a text might contain the phrase "it broke fast"). In the case of such violations, bias is introduced. The sentiment score for each document i of company j is consequently calculated using a positivity measure:

$$Positivity_{i,j} = \frac{pos_{i,j}}{pos_{i,j} + neg_{i,j}}$$

Unless the analyzed text contained >50 words, and a positivity score could be calculated, observations were dropped. This resulted in a 4% reduction of social media data. After sentiment scores for each document have been calculated, the resulting social media and analyst report sentiment time series need to be scaled to a common frequency. A higher common frequency seems desirable for providing a large number of observations to test for GC. The large number of social media data observations allowed the analyst report data to dictate the achievable frequency. The number of available reports ranges from 166 (General Electric) to 541 (Cisco). As expected, many such quarterly report releases are published. A higher frequency of daily aggregates of report sentiment is not supported by the available data. Consequently, daily means of the positivity measure are calculated for both analyst reports and social media content. Missing values in the report series are added by linear interpolation. Fig. 1 illustrates notable aspects of the data.

The left plot in the figure shows that stock analysts seem to be much more cheerful than the average social media user. Prior research indicates that, instead of recommendations or accuracy, analyst compensation is partially determined by the investment banking's business generated after a report [22]. Twedt and Rees argue that this effect might be diminished when considering the full content of a report instead of focusing on the buy/hold/sell recommendation or forecast measure [44]. Analysts seem to be hesitant to use negative language. A possible way of addressing the apparent domain-specific language of stock analysts would be to compile a sentiment dictionary specifically for the domain. Z-Scores of both time series are used to normalize the two series around a common level [6]. This is achieved by subtracting the mean (μ) of the observations from each data point (x) before dividing them by their standard deviation (σ), resulting in the centered series illustrated in Fig. 1 (right):

$$Z = \frac{x - \mu}{\sigma}$$

There are fewer analyst reports than social media observations. This results in more volatile time series for the reports. However, this should be of no immediate consequence for GC testing.

3.4. WoC measures

In order to examine social media data, we introduce a number of measures intended to operationalize each WoC condition described by Surowiecki [42].

3.4.1. Diversity

Diversity is a driver of team performance in many disciplines. Diversity cannot always be measured easily in social media observations. Therefore, we rely on two established dimensions of diversity, including age and gender. Lee and Farh [34] found a positive interaction between gender diversity and group self-efficacy outcomes. In addition, gender diversity in the boardroom has been shown to have a positive influence on firm performance [9].

We follow Blau's definition of group heterogeneity [5] describing nonhierarchical (nominal, inherently unranked) distinctions within groups. This definition of heterogeneity stands in contrast to a hierarchy emphasizing inequality, such as income differences or the "glass ceiling." We derive Blau indices for heterogeneity, which are calculated as one minus the sum of squared group fractions per measured category type [5]:

$$\text{Blau Index} = 1 - \sum_{i=1}^{N} Prop_i^2,$$

$Prop_i$ refers to the fraction of a specific group in the total population, i describes individual categories regarding a property (e.g., male or female for a gender index), and N refers to the total number of categories for the property. The value range of these indices varies depending on the number of categories included in a property, but it is always bound between zero and one [32]. A perfectly homogeneous group would result in an index equal to zero (the sum is 1). A "perfectly diverse" (i.e., distributed evenly across the categories) group receives a score of 0.5 for a property with two categories. A larger value corresponds to a more diverse group. Such measures are widely used to capture group diversity in multiple disciplines. Similar measures are available across several disciplines, such as the Gini [21], Gibbs–Martin [19], and Herfindahl–Hirschman indices [24]. Gender composition measures the male to female ratio of the data.
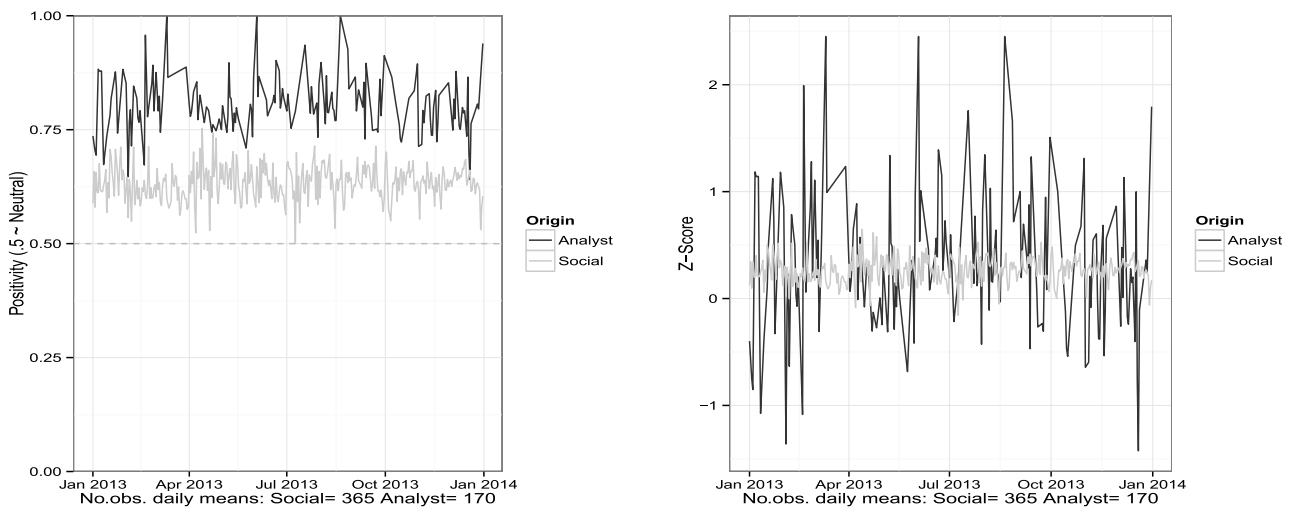

Fig. 1. Daily sentiment means (left) and z-scores (right) of daily sentiment means for Cisco. The gray line denotes social media and the black line analyst sentiment. The daily mean counts on the charts refer to the number of available daily data points prior to linear interpolation of missing values.

3.4.2. Decentralization

Decentralization has been studied in many different systems. Its role as a determinant of health-care system performance has been examined [7]. Furthermore, decentralization has been studied in the context of government performance [17]. Kim and Burton found decentralized groups to perform better regarding cost and time, whereas centralized groups achieved better quality [29].

No common studies have tested decentralization as a condition for WoC. In this analysis, we construct a geospatial measure to determine whether the geographical concentration of the crowd influences their predictive power regarding analyst sentiment.

The data contains 1163 unique location annotations. We calculate the concentration measure among those locations. The Platform Concentration measure refers to the type of social media platform (e.g., blog, microblog, forum, or social network) using a Blau index.

An additional location measure contains the average distance between users posting about a specific company within a specific period. Depending on availability, location annotations in the data mostly refer to the country, state, or city level. We geocode those locations to longitude–latitude coordinates and calculate the average distance between posts for a given period and company. Distance is calculated as the path on the surface of a unit sphere (radius = 1) between two sets of coordinates and scaled up to kilometers with the factor 1:6371 (earth radius in kilometers).

3.4.3. Independence

Independence is the most abstract of the three conditions proposed by Surowiecki [42]. Independence depends on a user-to-user relationship within the crowd that is not = directly observable. Common social networks and social media sites are not structured in hierarchical tiers, and the data used here do not include the structure of the social network. Independence is expected to be the most fleeting condition to test. We construct a measure of author tone to capture how assertive each social media text is written.

The Semantic Authority measure is calculated similar to a text sentiment, using the "Modalweak" and "Modalstrong" categories from the Louhgran and McDonald 10-K dictionary [35]. These modal categories contain words commonly used to increase or decrease the assertiveness of a statement. For example, "always" or "definitely" are included in the "Modalstrong" category, while "maybe" or "could" are examples of the "Modalweak" category. We calculate the measure for the ith document and the jth company as follows:

$$Authority_{i,j} = \frac{ModStrong_{i,j} - ModWeak_{i,j}}{ModStrong_{i,j} + ModWeak_{i,j}}$$

The variable is a tonal measure for an author's certainty. An authority figure uses more strong modal words than someone who is influenced by the opinion of others. We use this variable to measure social media users' certainty and not that of analysts.

4. Analysis and empirical results

4.1. Granger causality

Statistical results of the analysis are presented in two steps. GC testing results are first developed and interpreted. Second, these results are used in an analysis of circumstances leading to a particular GC direction. Following Bollen et al., who evaluated model performance based on p-values of the additional $M2$ time series [6], a lag-length selection method was used. This method was chosen because of the large number of models in the analysis. Models for each company (30) in the sample and direction of the GC (2) are estimated for quarterly subsets and the entire annual time span of 2013. Models from lag-length $n = 1$ to $n = 15$ days are estimated. This results in $30 \times 2 \times 15 \times 5 = 4500$ models. The two mediums might have an inherently different reaction speed to new information, resulting in a uniform-direction GC between the content types. The direction of GC may also depend on the type of information and company at hand. While the latter scenario corresponds to the hypotheses outlined in section 2.3, it is worth exploring the alternative. For each of the 15 lag lengths, the number of models with a significant relationship between the series pairs (i.e., those models with a p-value smaller than 10%) are reported here. If one of the two mediums is inherently faster than the other, higher numbers of GC going in one direction will be observed. Table 2 demonstrates these results. The first row shows models in which social media sentiment was used to augment analyst report sentiment. The second shows results of the opposite direction.

The number of lags resulting in the largest amount of significant models is identical in both prediction directions. Another sample (e.g., requiring another minimal text length or including minimal dictionary hit counts) may result in different optimal parameters for the two directions. There appears to be an actual maximum (i.e., both smaller and longer lag selections reduce the number of significant models). However, no strong imbalance between the two directions of GC is observed. This proves that the GC between the two mediums is driven by more circumstances than an inherent imbalance in reaction speed between the two mediums. A lag length of 10 periods creates the largest number of significant relationships. However, there is no indication that this number of lags is preferable for all series pairs, as other lag lengths produce a similar amount of significant models. Therefore, another set of models is estimated in which the lag-length parameter is not chosen simultaneously for all models. Instead, the estimation is made individually for each of the 300 models. The aim is to minimize the p-value of the model. The results of this second selection of models are reported in Table 3.

The p-value-based lag-length selection procedure improves the number of significant models from 57 (19%) to 127 (42%). Of the 127 significant models, 68 (45.3%) predict analyst report sentiment by social media sentiment and 59 (39.3%) predict social media sentiment via analyst report sentiment. Both directions of prediction appear to be feasible. The preprocessing of input documents was performed similarly for both types of content and all companies. Improvements can be made by introducing a case-specific preprocessing logic. As the question of interest in this work

Table 2
Number of significant models in both prediction directions for social media and analyst report sentiment, n = 1–15. Maxima are highlighted.

n-lag	1	2	3	4	5	6	7	8	9	**10**	11	12	13	14	15
Analyst ← Social	21	20	13	18	19	17	23	20	20	**30**	27	29	29	26	24
Social ← Analyst	23	20	20	18	18	20	20	25	19	**27**	23	27	23	22	21
Sum	44	40	33	36	37	37	43	45	39	**57**	50	56	52	48	45

Table 3

p-Values (three-digit rounding) of GC tests for quarterly subsamples and annual data. Optimal lag length for each individual time series pair based on lag lengths between 1 and 15 (p-values ≤ 10% are highlighted). The first row indicates the test direction (e.g., the Analyst heading indicates that reports are predicted via social media content). The sums refer to the number of significant models in columns and rows.

Company	Analyst ← Social					Social ← Analyst					Sum
	Q1	Q2	Q3	Q4	Annual	Q1	Q2	Q3	Q4	Annual	
3M	0.759	0.385	**0.073**	**0.064**	0.113	**0.05**	0.215	**0.09**	**0.044**	0.416	5
American Express	0.297	0.305	0.216	0.355	0.264	0.341	**0.025**	**0.06**	0.277	0.168	2
AT&T	0.155	0.176	0.3	0.225	0.229	0.512	0.31	**0.082**	0.279	0.124	1
Boeing	**0.066**	0.176	0.423	0.269	0.312	**0.008**	0.195	**0.082**	0.24	**0.012**	4
Caterpillar	**0.027**	0.336	0.217	**0.039**	**0.001**	0.338	0.532	0.25	**0.036**	**0.044**	5
Chevron	**0.096**	0.239	0.116	**0.046**	0.109	0.103	0.331	**0.006**	0.133	0.381	3
Cisco	**0.062**	0.351	**0.004**	0.17	**0.079**	0.576	**0.024**	0.115	0.188	**0.033**	5
Coca Cola	0.233	**0.065**	**0.045**	**0.062**	**0.079**	**0.031**	**0.001**	0.745	0.194	0.125	6
Disney	**0.005**	0.107	0.216	0.167	0.23	0.195	**0.019**	0.145	0.164	**0.017**	3
DuPont	**0.056**	**0.06**	0.187	**0.02**	0.357	0.223	0.73	**0.005**	**0.001**	**0.071**	6
Exxon	**0.013**	**0.059**	0.161	0.257	**0.021**	**0.062**	**0.051**	0.162	**0.036**	0.147	6
General Electric	**<0.01**	0.125	0.514	**0.008**	**0.009**	**<0.01**	0.912	**0.03**	0.289	0.14	5
Goldman	0.124	**0.023**	**0.018**	0.622	0.112	0.735	0.133	0.306	**0.052**	0.493	3
Home Depot	**0.068**	0.151	0.21	0.285	0.101	**0.06**	0.699	**0.042**	**0.05**	**0.028**	5
IBM	**0.007**	**0.007**	0.454	**0.013**	0.111	0.24	0.241	**0.007**	**0.035**	0.288	5
Intel	**0.016**	**0.075**	0.234	0.56	0.683	0.117	**0.01**	0.11	**0.059**	0.227	4
Johnson&Johnson	**0.052**	0.369	**0.014**	**0.066**	0.427	**0.053**	0.167	0.12	0.161	0.216	4
JP Morgan	**0.02**	0.242	**0.029**	**0.046**	0.403	**0.013**	0.724	0.153	**0.011**	**0.032**	6
McDonald's	**0.009**	0.118	**0.095**	**0.079**	**0.065**	**0.048**	0.477	**0.012**	0.484	0.452	6
Merck	0.459	0.184	**0.041**	0.658	0.771	0.576	0.329	**0.002**	0.698	0.475	2
Microsoft	0.347	0.212	0.218	0.106	**0.012**	0.568	0.239	0.446	**<0.01**	0.462	2
Nike	**0.035**	**0.019**	**0.014**	**0.001**	**0.056**	0.213	0.522	0.174	0.245	0.67	5
Pfizer	**0.025**	0.13	**0.037**	**0.032**	**0.007**	**0.025**	0.162	0.16	0.707	**0.039**	6
Procter & Gamble	0.174	0.254	0.147	**0.018**	**0.071**	0.295	**0.078**	**0.001**	0.101	**0.056**	5
Travelers	**0.045**	0.267	0.158	0.872	0.406	**0.034**	**0.061**	0.334	**0.067**	0.193	4
UnitedHealth	0.189	0.247	**0.001**	0.475	**0.042**	0.15	0.304	0.126	0.187	0.504	2
United Technology	0.168	0.139	**0.028**	0.324	0.205	0.447	**0.051**	**0.097**	**0.005**	**0.019**	5
Verizon	**0.008**	0.47	**0.007**	0.179	**0.059**	0.259	0.733	0.196	0.192	0.688	3
Visa	0.232	0.133	**0.068**	0.766	**0.094**	0.122	**0.043**	0.167	0.73	0.238	3
Walmart	0.127	**0.092**	**0.014**	0.585	**0.052**	0.282	**0.023**	**0.027**	0.238	**<0.01**	6
Sum	18	8	15	13	14	11	11	14	12	11	127

is not case specific, an applied uniform logic seems sufficient. Given these results the question arises as to which situation leads to which kind of GC relationship (i.e., what kind of company is more prone to either direction of GC or what situation fosters this tendency).

4.2. Drivers of crowd wisdom

We create binary response models to explore the drivers of crowd wisdom. The chance that a given set of circumstances leads to GC in either direction was estimated with a separate model for each direction.

We estimate one model for each direction of GC. There are two reasons why two models per situation are estimated. First, as Table 3 indicates, the two directions of GC are not mutually exclusive. Second, the Social model (*M2*) can serve as a benchmark for the WoC model. Therefore, a single binary encoding of the GC direction would omit cases where both directions are significant. Generalized linear models (GLMs) using log-link functions are estimated (Logit). For the "Analyst" model (*M1*), the dependent variable is 1 if the social media data were found to GC analyst reports in a given period and firm (i.e., if the *p*-value in Table 3 is <0.1). In the opposite case, the dependent variable in the "Social" model (*M2*) is 1. If the measured variables derived from the theoretical WoC constructs proposed by Surowiecki [42] help us understand the situation in which crowd wisdom emerges, *M1* should be the better overall model. The theoretical argument for our measures only holds true in this direction. Consequently, *M2* can serve as a benchmark for the results of *M1*. The independent

variables included in these models are split into four groups. We provide a brief overview of the intention and composition of each group in the next section, before discussing the results for each of these categories (Table 4).

4.3. Discussion and interpretation

4.3.1. WoC measures

In this category, we group the main WoC measures of this study. The first five variables refer to the previously derived Blau index measures for age, geolocations, platform diversity, and author name diversity. Authority refers to the tonal authority measure. This variable is followed by the average age of social media users in a given period–company combination and age variance.

Platform diversity increases the likelihood of GC between the two types of content in either direction. The significant coefficient in the "Analyst ← Social" model supports the WoC theory. By contrast, the negative coefficient of age diversity in the same model contradicts the WoC theory, suggesting that any form of cognitive diversity improves the group consensus. An increase in the average age of users decreases their predictive capabilities. A slight depreciation of cognitive ability with age might be expected. However, this depreciation seems counterintuitive, as older users are expected to become more experienced in their assessments. The negative coefficient for authority suggests that users who are sure of their opinion decrease the quality of the group's mean sentiment. This is in line with the expectation of the WoC theory, which suggests that an independent crowd should perform better. The corresponding coefficient in the "Social ← Analyst" case

Table 4
Analyst ← Social describes the case in which analyst reports are Granger-caused by social media content and vice versa. Quarterly data were used (i.e., 4 quarter times 30 companies equals No. Obs.). The dependent variable is given by a binary coding of the results in Table 3 ($p \leq 0.1$ or not). All noncategorical variables are standardized.

		M1: Analyst ← Social			M2: Social ← Analyst		
		Estimate	Std. Error	Pr(>|z|)	Estimate	Std. Error	Pr(>|z|)
WoC Measures	(Intercept)	10.061	3.972	**0.011**	1.454	2.607	0.577
	Age Blau	−1.097	0.768	0.153	0.729	0.612	0.234
	Geo Blau	1.86	0.979	**0.058**	−0.28	0.566	0.621
	Platform Blau	3.876	1.409	**0.006**	1.564	0.926	**0.091**
	Name Blau	0.616	1.404	0.661	0.107	0.344	0.755
	Authority	−5.815	1.702	**0.001**	−2.269	0.835	**0.007**
	Avg. Age	−3.498	1.42	**0.014**	−0.359	0.633	0.57
	Age Var.	0.809	0.706	0.252	−0.9	0.53	**0.09**
	Avg. Distance	1.335	0.795	**0.093**	0.879	0.577	0.128
	% Male	−0.816	0.968	0.399	−0.05	0.601	0.934
Industry Dummies	Consumer Staples	−5.699	3.326	**0.087**	3.701	2.492	0.138
	Diversified	−2.983	2.352	0.205	3.607	1.887	**0.056**
	Energy	−31.704	8.834	**<0.001**	−0.144	4.693	0.975
	Financial	−15.442	4.862	**0.001**	0.932	2.194	0.671
	Health Care	−19.204	5.854	**0.001**	−4.154	3.073	0.177
	Industrials	−13.796	4.386	**0.002**	0.028	2.625	0.991
	Technology	−8.183	3.627	**0.024**	−3.014	3.061	0.325
	Telco	10.23	6.422	0.111	4.557	4.141	0.271
Firm Specific	Revenue	−2.685	1.044	**0.01**	0.504	0.616	0.413
	Operating Income	−1.613	1.729	0.351	−1.615	1.198	0.178
	R&D Budget	4.23	1.916	**0.027**	2.572	1.218	**0.035**
	S&P Long-term Rating	1.192	0.949	0.209	−0.231	0.607	0.704
	No. Subsidiaries	−3.796	1.152	**0.001**	−0.898	0.668	0.179
News Dummies	Business	−2.6	3.192	0.415	−4.193	2.249	**0.062**
	Allows Comments	−1.673	3.73	0.654	−0.551	2.631	0.834
	Environment	19.442	6.079	**0.001**	−0.841	3.545	0.812
	Film	−5.445	4.05	0.179	−1.644	2.987	0.582
	Football	−17.691	6.509	**0.007**	−6.156	3.973	0.121
	Media-Network	−13.513	4.965	**0.006**	−2.171	3.196	0.497
	Music	−14.777	1356.967	0.991	−0.413	3.553	0.908
	Sport	−9.414	5.018	**0.061**	−5.405	3.24	**0.095**
	Sustainability	−7.139	4.876	0.143	−6.152	3.449	**0.074**
	Technology	−9.564	5.3	**0.071**	−2.774	3.745	0.459
	World	−7.767	3.787	**0.04**	−3.435	2.589	0.185
Controls	URL Blau	−3.747	1.432	**0.009**	−1.163	0.849	0.171
	COL Readability	1.778	0.899	**0.048**	0.528	0.504	0.295
	Authority Var.	−7.494	2.259	**0.001**	−1.765	1.026	**0.085**
Model Summary	No. Obs.			120			120
	AIC			168.45			196.82
	Family			Binomial			Binomial
	Link			Logit			Logit

suggests that certainty in social media posts decreases the likelihood of analysts predicting their sentiment.

Overall, the results of the WoC measures for the theoretical constructs described by Surowiecki [42] are promising. The model indicates that there is indeed a measurable connection between the makeup of the crowd and its ability to explain analyst opinion. Although it is difficult to compare results between different content domains, this is in line with previous WoC research, such as the efforts to explain the content quality of Wikipedia [2] or its comparable quality to classical encyclopedias [20]. Interestingly, this suggests that WoC can arise without a system specifically designed to allow the crowd to aggregate their opinion. Previous research highlights the importance of group coordination for content quality [31]. In our case, this aggregation only takes place after the fact using sentiment analysis. Examining how the support of the crowd's coordination may improve their information processing capabilities is an interesting question for future research.

4.3.2. Industry dummies

The main industry classification of a company may be of use when searching for companies for which GC is present between the two content types. We observe intuitive results in several industries. Both energy and industrial companies are unfavorable for the "Analyst ← Social" case, whereas consumer staples, diversified and financial companies, favor the "Social ← Analyst" case. The financial industry dummy deserves special attention: Although this industry does favor the "Social ← Analyst" case, no significant reduction of prediction quality in the "Analyst ← Social" case can be observed. Overall, the industry dummies confirm previous studies indicating that stock analysts' recommendations do indeed carry inherent value [46]. At the same time, no significant support for any industry in the sample adds explanatory power to the crowd's opinion.

4.3.3. Company specific

We include revenue and operating income, research and development (R&D) budgets, issuer ratings and the (ultimate) number of subsidiaries of a given company. We observe a decrease in the capabilities of the crowd to predict analyst sentiment for companies with a higher number of subsidiaries. This might be because large multinational companies are too complex to be summarized by a single measure of crowd opinion. The crowd may

have a positive opinion of one division of a firm, while expressing a negative one about another. The "Social ← Analyst" case demonstrates a positive effect of revenue, suggesting that expert opinion is more informative for larger companies. Another possible method of interpreting this result is to suggest that experts be more careful in their evaluation of larger companies. This would be an interesting question for future research, because all companies in the sample are comparatively large (in order to make sure there are enough social media posts and analyst reports).

A larger R&D budget also exhibits a similar effect. Company size overall has a negative effect on the comparative quality of the "Analyst ← Social" case possibly due to the aggregation of all crowd opinions to a single opinion. This weakness can be eliminated using topic-mining methods and calculating topic-specific sentiments in order to capture more nuances of the crowd's opinion on different divisions of larger companies.

4.3.4. News dummies

Using the Guardian API data mentioned in Table 1, the most common (median) news category for each period and company is extracted and introduced to the dummy coding models. We include the median news category to see whether specific types of novel information favor one of the two groups' information processing capabilities. The results indicate that business news decreases the likelihood of analysts' ability to predict social media users. This supports the known tendency of analysts sticking to prior assessments [43], whereas social media users are free to make bolder changes in their opinions. Media network news decreases the predictive power of social media users. Many social media posts coinciding with this median news category may be concerned with product debates, rather than company evaluations. Only the environmental news dummy provides a significant increase in the likelihood of WoC. This finding seems intuitive if negative events, such as chemical spills, occur. Overall, the news categories provide intuitive results, suggesting that the chosen sentiment measure for the two content types indeed provides an adequate aggregation of the groups' opinion.

4.3.5. Controls

We aggregate three variables in this group. The first variable "URL Blau" serves as a verification of the diversity in URLs, rather than platform diversity. It can be used to assess the decentralization of the observations. We do not group this variable with the WoC measures because, in contrast to platform diversity, which is based on a curated field generated from our social media data, this measure is less certain to capture the diversity concept as proposed by the WoC theory. Interestingly, this measure exhibits a negative coefficient. This may be explained by an effect resulting from crowds being too diverse or decentralized, thus preventing crowd wisdom to emerge. This interesting finding can be explored in future research.

The second variable "COL Readability" is computed as the average Coleman–Liau Index for the social media posts in each period [12]. Such readability measures intend to rate the complexity of speech in a document and may serve as a proxy for author education. The positive coefficient indicates that more advanced syntax correlates with better crowd judgment.

The third variable "Authority Var." computes the variance of our authority measure instead of its period mean. We choose to report it in the controls group because a variance measure is less distinguishable from a diversity measure than the mean specification. It is less compatible with the theoretical foundation of the analysis. However, as the two measures are not highly correlated and the variance measure adds power to the model, practitioners may be interested in including both specifications in their analyses. From the WoC theoretical standpoint of the study, this measure

may be interpreted as a diversity, rather than an independence measure, the theoretical integration of which can be studied in future.

Altogether, the two models suggest that the WoC theory indeed provides useful constructs that can be operationalized to explain when the crowd can arrive at opinions prior to the availability of expert assessment. The following provides a brief overview of the results of both stages of our analysis regarding the initial hypotheses:

H1a (Situational direction of GC). The results indicate that this is indeed true; no uniform GC direction can be established between the two types of content.

H1b (Analyst expertise and privileged access matter). This hypothesis is supported by the negative coefficients in the industry dummies (analyst model) regarding energy and industrial companies. In addition, the positive coefficients in M2 for financials and consumer staples indicate that these types of companies require analyst expertise.

H1c (When relevant information is public and diverse, independent opinions can be aggregated, and the crowd has an advantage). This is supported by the positive coefficient of the environmental news category when analyst reports are predicted, as well as the negative coefficient of the sport news category in the opposite direction.

Thus, the analysis provides evidence for H1a and H1b, whereas evidence in support of H1c is sparse.

H2 (Satisfaction of WoC conditions should improve social media users' opinion quality). Our results provide mixed evidence regarding this hypothesis. The hypothesis is supported by the positive effect of platform diversity. However, both age diversity measures show negative effects on the dependent variable. Gender diversity shows no significant effect.

5. Implications

As discussed, our results add to the growing body of work suggesting that WoC as a phenomenon can be used to explain the sometimes surprising quality of the content created by large groups [20,31] as well as the value of stock opinions of social media users [10]. Our contribution to this body is twofold. First, the comparison to stock analysts allows us to benchmark the crowd wisdom against an expert group. The results of our GC analysis suggest that in some situations the crowd can add information in a more timely manner than experts. Second, our analysis supports the WoC theory as proposed by Surowiecki [42]. Through the operationalization of the independence, decentralization, and diversity constructs, which constitute the central pillars of the theory, the contribution of each condition to the WoC is examined. Furthermore, we examine the conditions beyond those proposed by WoC theory under which the crowd is wise. The results indicate that crowd wisdom exists, although it is highly dependent on the conditions described by the WoC theory and the degree to which particular subjects appeal to the crowd's interest. This suggests that more constructs may be needed to fully explain when a crowd is wise than are included in the current WoC theory. Future research should focus on discovering additional determinants of crowd wisdom. Although sentiment as an aggregate measure for crowd opinion works in principle, the results suggest additionally that complex or conflicting topics, which may be especially prevalent for larger companies, may require more complex aggregation methods for the crowd's diverse opinions.

We divide the practical implications of this research into implications for the financial sector and those for social media users and platforms. Within the financial sector, our results inform the customers of analysts about the conditions under which analyst research is especially valuable, but also when it may be wise to resort to social media monitoring tools to gauge the crowd's opinion. The results also inform the customers how to aggregate the opinions of social media users. Similarly, stock analysts are informed of the circumstances under which it may be wise to listen to social media users' opinions as an additional source of information, but also when they are unlikely to provide valuable information.

In addition, the operationalization of WoC-related constructs can help companies refine their social media monitoring tools to better reflect the diverse opinion of the crowd. After companies, this should also be of interest to social media content aggregators who need to know the kind of data on social media users that interest their customers.

Finally, special-purpose social networks such as stock recommendation communities[40] and social lending communities [16] are fundamentally based on the assumption of crowd wisdom. The members of such communities expect these platforms to provide them with insights gained from this crowd wisdom. Our WoC theory-based results provide the administrators of such communities with evidence on the makeup of wise crowds. While such communities are unlikely to actively control the characteristics of the user group making up their community, our results may prove useful regarding the user selection for the samples of users that are chosen to compute their crowd-based recommendations.

5.1. Limitations

This study has methodological and theoretical limitations, which warrant discussion. There is no reliable method of determining which portion of social media users are professional or "hobby" analysts. A more controlled experiment using a single social network would be an interesting avenue for future research. In addition, because of their herding tendency, analysts could exhibit some WoC effects. Reactions falling out of the $n = 15$ lag length included in the analysis may be missed. Large lag lengths can lead to spurious correlations. This methodological trade-off has to be accepted. Using a larger and more diverse set of companies and multiyear samples could provide interesting results. The available data do not provide insight into the social hierarchy of social media users, which may provide a more suitable measure of independence as a precursor to WoC. Finally, the aggregation of social media users' opinions to the singular sentiment measure sacrifices the diversity of opinions of the group. Other measures may well be more suited to capture this diversity. Beyond these methodological considerations, it is also important to keep in mind that the domain of company-specific opinions may not be comparable to other areas of WoC research. The reproduction of a similar analysis for a noneconomic domain could provide an interesting comparison.

6. Conclusion

The aim of this research was to investigate a possible relation between the prediction power of stock analysts' sentiment and that of "the crowd" (i.e., a large set of social media users). Earlier studies have identified inefficiencies in professional analysts' decision processes. The WoC theory suggests that the crowd might be able to mitigate these problems.

GC testing between the two types of content showed statistically significant relations for a large number of cases in this sample. This finding indicates that the two types of content can be used to predict the other in many cases. However, evidence for the similar use of the two types of content is lacking. Similarly, no evidence is provided as to the contents' complementarity (keeping in mind the emerging issue of multicollinearity). The practical applications of such relationships include algorithmic trading, news reporting, and customer relations.

Logit models provide information on the circumstances under which social media content sentiment can be used to predict analyst reports, and vice versa.

There is mixed evidence supporting WoC theory. Platform diversity in the social media sample increases the crowd's success, whereas age diversity decreases it. This finding might be mitigated by a larger sample. A larger sample introduces more variance in company type, spanning a significantly longer period. Evidence for the WoC theory is provided by cases in which social media users arrived at a (collective) opinion, before professional stock analysts were able to include environmental changes into their reports.

These results suggest that crowd wisdom can outperform experts if information is instantly available. However, the drivers of crowd wisdom might not be sufficiently explained by the current WoC theory.

Professional analysts seem to react quicker to technical issues, such as changes in financial situations. Therefore, this study supports the WoC theory (i.e., the general possibility of crowd wisdom), although the drivers of this wisdom have not been fully understood yet.

References

[1] J.R. Anderson, Cognitive skills and their acquisition, Psychology Press, 1981.
[2] O. Arazy, W. Morgan, R. Patterson, Wisdom of the crowds: decentralized knowledge construction in wikipedia, 16th Annu Work. Inf. Technol. Syst. Pap. (2006).
[3] B. Barber, R. Lehavy, M. McNichols, B. Trueman, Can investors profit from the prophets? Security analyst recommendations and stock returns, J. Finance 56 (2001) 531–563.
[4] R. Barniv, O.-K. Hope, M.J. Myring, W.B. Thomas, Do analysts practice what they preach and should investors listen? Effects of recent regulations, Account. Rev. 84 (2009) 1015–1039.
[5] P.M. Blau, Inequality and heterogeneity: A primitive theory of social structure, Free Press New York, 1977.
[6] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, J. Comput. Sci. 2 (2011) 1–8.
[7] T. Bossert, Analyzing the decentralization of health systems in developing countries: decision space, innovation and performance, Soc. Sci. Med. 47 (1998) 1513–1527.
[8] M.T. Bradshaw, Analyst information processing financial regulation, and academic research, Account. Rev. 84 (2009) 1073–1083.
[9] K. Campbell, A. Mínguez-Vera, Gender diversity in the boardroom and firm financial performance, J. Bus. Ethics. 83 (2007) 435–451.
[10] H. Chen, P. De, Y. (Jeffrey) Hu, B.-H. Hwang, Wisdom of crowds: the value of stock opinions transmitted through social media, Rev. Financial Stud. 27 (2014) 1367–1403.
[11] M.B. Clement, S.Y. Tse, Financial analyst characteristics and herding behavior in forecasting, J. Finance 60 (2005) 307–341.
[12] M. Coleman, T.L. Liau, A computer readability formula designed for machine scoring, J. Appl. Psychol. 60 (1975) 283.
[13] T.E. Damar, Attacking Faulty Reasoning: A Practical Guide to Fallacy-Free Arguments, Thomson Wadsworth, Belmont, CA, 2005.
[14] E. Dimson, M. Mussavian, A brief history of market efficiency, Eur. Financial Manage. 4 (1998) 91–103.
[15] P.T. Elgers, M.H. Lo, R.J. Pfeiffer, Delayed security price adjustments to financial analysts' forecasts of annual earnings, Account. Rev. 76 (2001) 613–632.
[16] C.R. Everett, Group membership, relationship banking and loan default risk: the case of online social lending, SSRN Electron. J. 7 (2010).
[17] J.-P. Faguet, Decentralization and local government performance: improving public service provision in Bolivia, Rev. Econ. Del Rosario. 3 (2000) 127–176.
[18] R. Frankel, S.P. Kothari, J. Weber, Determinants of the informativeness of analyst research, J. Account. Econ. 41 (2006) 29–54.
[19] J.P. Gibbs, W.T. Martin, Urbanization, technology, and the division of labor: international patterns on JSTOR, Am. Sociol. Rev. 27 (1962) 667–677.
[20] J. Giles, Internet encyclopaedias go head to head, Nature 438 (2005) 900–901.
[21] C. Gini, 1912. Variabilità, mutabilità, Repr. Mem. Di Metodol. Stat. (Ed. Pizetti E, Salvemini, T). Rome Libr. Eredi Virgilio Veschi. 1.
[22] B. Groysberg, P.M. Healy, D.A. Maber, What drives sell-Side analyst compensation at high-Status investment banks? J. Account. Res. 49 (2011) 969–1000.

[23] S. Hill, N. Ready-Campbell, Expert stock picker: the wisdom of (experts in) crowds, Int. J. Electron. Commerce 15 (2011) 73–102.

[24] A.O. Hirschman, The paternity of an index, Am. Econ. Rev. 76 (1964) 1–76 (2).

[25] H. Hong, J. Kubik, A. Solomon, Security analysts' career concerns and herding of earnings forecasts, RAND J. Econ. 31 (2000) 121.

[26] H. Hong, T. Lim, J.C. Stein, Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies, J. Finance 55 (2000) 265–295.

[27] P.J. Irvine, Do analysts generate trade for their firms? Evidence from the Toronto stock exchange, J. Account. Econ. 30 (2000) 209–226.

[28] R. Johnston, B.F. McNeal, Statistical versus clinical prediction: length of neuropsychiatric hospital stay, J. Abnorm. Psychol. 72 (1967) 335.

[29] J. Kim, R.M. Burton, The effect of task uncertainty and decentralization on project team performance, Comput. Math. Organ. Theory 8 (2002) 365–384.

[30] G. Kirchgässner, J. Wolters, U. Hassler, Introduction to Modern Time Series Analysis, Springer, 2012.

[31] A. Kittur, R.E. Kraut, Harnessing the wisdom of crowds in wikipedia, Proc. ACM 2008 Conf. Comput. Support. Coop. Work–CSCW'08, ACM Press, New York, New York, USA, 2008, pp. 37.

[32] Alison M. Konrad, Pushkala Prasad, J.K. Pringle, Handbook of Workplace Diversity, Sage, London, 2006.

[33] J. Larkin, J. McDermott, D.P. Simon, H.A. Simon, Expert and novice performance in solving physics problems, Science 80- (208) (1980) 1335–1342.

[34] C. Lee, J.-L. Farh, Joint effects of group efficacy and gender diversity on group cohesion and performance, Appl. Psychol. 53 (2004) 136–154.

[35] T. Loughran, B. McDonald, When is a liability not a liability? Textual analysis dictionaries, and 10-Ks, J. Finance 66 (2011) 35–65.

[36] B.G. Malkiel, E.F. Fama, Efficient capital markets: a review of theory and empirical work, J. Finance 25 (1970) 383–417.

[37] M. Nofer, O. Hinz, Are crowds on the internet wiser than experts? The case of a stock prediction community, J. Bus. Econ. 84 (2014) 303–338.

[38] M. Nofer, O. Hinz, Using twitter to predict the stock market, Bus. Inf. Syst. Eng. 57 (2015) 229–242.

[39] J. Pearl, Causality, Cambridge University Press, 2009.

[40] M.K. Poetz, M. Schreier, The value of crowdsourcing: can users really compete with professionals in generating new product ideas? J. Prod. Innov. Manage. 29 (2012) 245–256.

[41] P.J. Stone, D.C. Dunphy, M.S. Smith, D.M. Ogilive, The General Inquirer, The M.I.T Press, Cambridge, Massachusetts, 1966.

[42] J. Surowiecki, The Wisdom of Crowds, Random House LLC, 2005.

[43] B. Trueman, Analyst forecasts and herding behavior, Rev. Financial Stud. 7 (1994) 97–124.

[44] B. Twedt, L. Rees, Reading between the lines: an empirical examination of qualitative attributes of financial analysts' reports, J. Account. Public Policy. 31 (2012) 1–21.

[45] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, Proc Conf. Hum. Lang. Technol. Empir. Methods Nat. Lang. Process. (2005) 347–354.

[46] K.L. Womack, Do brokerage analysts' recommendations have investment value? J. Finance 51 (1996) 137–167.

Matthias Eickhoff is a Ph.D. student at the Chair of Electronic Finance and Digital Markets in the Faculty of Economic Sciences, University of Göttingen. His research interests include the analysis of unstructured data, text mining and business intelligence in the financial domain. He holds a M. Sc. in Finance, Accounting and Taxes. His work has appeared in the proceedings of the conferences on Design Science Research in Information Systems and Technology and the Pacific Asian Conference on Information Systems.

Jan Muntermann holds the Chair of Electronic Finance and Digital Markets in the Faculty of Economic Sciences, University of Göttingen. His research interests include the design and analysis of information systems, business intelligence and analytics, IT governance and research methodology. He has published in journals such as Information Systems Research, Decision Support Systems, and European Journal of Information Systems.