STOCK ANALYSTS VS. THE CROWD: A STUDY ON MUTUAL PREDICTION

Eickhoff, Matthias, University of Göttingen, Germany, matthias.eickhoff@wiwi.uni-goettingen.de
Muntermann, Jan, University of Göttingen, Germany, muntermann@wiwi.uni-goettingen.de

Abstract

The predictive power of stock analyst reports has been used to relate report contents to stock returns or describe herding behavior of analysts themselves. In this paper, the sentiment of analyst reports is related to that of a large social media data set via Granger Causality testing on the basis of wisdom of crowds theory based considerations, in order to investigate whether the two types of content are inherently related or not. Results show strong significance for a large number of the tested time series, indicating that the two types of content are indeed suitable for mutual prediction. In addition, we elaborate on the conditions under which cognitive diversity of the crowd matters. Furthermore, a second analysis stage provides evidence for which type of company and news environment a particular direction of granger cause arises between the two types of content.

Keywords: Wisdom of Crowds, Stock Analysts, Sentiment-Analysis, Predictive Analytics

1 Introduction

The impact of financial analyst reports has been subjected to increasing scientific scrutiny. In particular, the herding behavior of financial analysts, i.e. how some analysts seem to impact the opinion of others (Twedt & Rees 2012, p. 2), has been analyzed extensively (Clement & Tse 2005; Hong & Kubik, et al. 2000; Trueman 1994). Also, the reaction of the capital market has been analyzed. The traditional way of analysis either concerns itself with the analysis of the buy/hold/sell recommendation itself, available via the Institutional Broker Estimate System (I/B/E/S). More recently, researchers have begun to apply text mining methods to analyst reports, in order to automatically extract more information than previously available. As noted by Twedt & Rees (2012) the analysis of the entire report is desirable because it may yield information beyond the constrained categorization of the stock into one of recommendation categories (Twedt & Rees 2012, p.2). In contrast to these prior studies, this article is not concerned with inter-dependencies of the reports themselves or their (direct) relation to the stock market but rather with how the opinions of professional stock analysts relate to those of social media users and if the two related content types can be used to predict each other's sentiment. While other studies have shown that social media content can be used to forecast stock returns in a similar manner to analyst recommendations, the fact that this is possible does not necessarily imply that the two mediums prediction power is somehow inherently related. Such a relation is of interest because, as illustrated in section 2, professional analysts are assumed to have privileged access to relevant information but are also faced with a number of constraints, influencing their recommendation. Based on the assumption that the average social media user is not faced with the same constraints, such a relation could allow social media content to be incorporated in models traditionally using stock analysts recommendations for predictive purposes, serving as a control variable for the analysts biases. Therefore, it is the goal of this research to investigate whether such a relationship can be found between the content types, in either direction. To this end, social media and analyst report data is collected covering the year 2013 and the 30 component companies of the Dow Jones Industrial Average (DJIA) index. A two-staged analysis was conducted to determine if and for which time period (quarterly and annual) and company combinations Granger Causality (GC) can be established between social media and analyst sentiment. In a second stage of analysis we aim at determining the reasons for different cases of GC direction. The paper is structured as follows: After this introduction the theoretical background of the analysis is introduced by contrasting theoretical insight related to the information value of analyst reports and the concept of Wisdom of Crowds (WoC) in section 2, focusing on inefficiencies known to be incorporated in analyst opinions but likely to be mitigated by social media content. Based on these foundations, hypotheses are developed and tested. In order to test these hypotheses, Granger Causality (GC) is introduced as a methodological foundation for the analytic part of the paper in section 3. Consequently, the data used in the analysis, as well as its acquisition and pre-processing, is described in section 4 before subjecting it to the previously described GC testing in section 5. Finally, section 5 aims to provide evidence for the question which types of company and public interest foster which direction of GC using binary response models, providing statistically significant answers for cases of analyst reports being granger caused by social media content. The conclusion summarizes the results of the analysis and elaborates on their theoretical value for WoC research.

2 Theoretical background

Two research streams are especially relevant for the work conducted here. The first stream is concerned with the information value of analyst reports, while the second stream explores the principle of social media user sentiment through the lens of crowd wisdom. This research aims to combine insights from these two fields.

2.1 Information value of analyst reports

Analyst reports have been the focus of continued research for many years and a large portion of this research is conducted striving to answer one of the following two questions: 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. The underlying assumption is that analysts either have privileged access to relevant information about companies, either through thorough research generating the information or their close relation to companies. They are therefore viewed as information providers who can improve information efficiency (Frankel et al. 2006, p. 32) and can shorten the time elapsed between publication of information and its incorporation into stock prices (Elgers et al. 2001; Hong & Lim, et al. 2000). The answer to this fundamental question has been mixed. While earlier studies suggest that investments based upon analyst recommendations can indeed be profitable (Barber et al. 2001; Womack 1996), more recent research sheds doubt on the impartiality of stock analysts (Barniv et al. 2009; Bradshaw 2009). 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. This refers to the tendency to provide recommendations close to the consensus estimate (Twedt & Rees 2012, p. 2), thus introducing a bias towards the status quo. Reasons for this tendency to stick to the herd include career concerns (Clement & Tse 2005), which are especially relevant for younger analysts who are prone to stick to the consensus estimate because they fear termination if they make bold predictions and fail (Hong & Kubik, et al. 2000). Another potential source for biased analyst recommendations is examined by Groysberg et al. (2011), who investigate how the analysts compensation structure can influence their recommendation and provide evidence that compensation schemes are designed to increase brokerage and investment-banking revenues. Consequently, as stock analyst recommendations are potentially biased, other data sources offering insight into the current opinion about companies are desirable. Due to the rise of social networks in the last decade, social media content is widely available for a large number of companies. Furthermore, social media users are not faced with the same types of repercussions or incentives as stock analysts are. Therefore, social media users might be able to provide less biased opinions about a company's current state or its future developments. Research by Bollen et al. (2011) offers indication that social media content (Twitter), can be used to predict stock returns, indicating that the sentiment of social media users and the sentiment of stock analysts, may perhaps be used in a similar manner. The latter question is of particular interest to the research presented here because the deficiencies of stock analyst recommendations provide a need for alternate data sources providing information about companies, which are of interest for investors and researchers alike. One possible source of data, social media content, will be investigated here with regard to its connection to analyst reports.
The sum of these known deficiencies of the information value of analyst reports create a need for alternate sources of information that can help to mitigate the known biases. Social media users should, as a group, not be faced by the same problems as professional analysts are. They are unlikely to be punished if their opinions regarding a company turn out to be wrong and there are no compensation concerns involved. The next section explores the mechanism that might allow the research community and practitioners alike to augment models currently relying on analyst opinions alone.

2.2 Wisdom of Crowds

As the average social media user is not likely to have the finance background of stock analysts, another factor than domain expertise seems to provide the insight evidenced by Bollen et al. (2011), enabling social media users to arrive at mood states predicting the stock market. This novel source of expertise has been described as the wisdom of crowds (WoC). The term was coined by Surowiecki (2005) and WoC theory proposes that large independent and heterogeneous groups can outperform small expert communities in their assessments. Following the definition of Poetz & Schreier (2012, p. 4), who define the crowd as *"potentially large and diverse"*, no further assumptions about the makeup of this group will be made here. In contrast to this diverse group, our expert group can be described more clearly. Following the definition of Nofer & Hinz (2014, p. 306), an expert – in the context of this research – is *"a professional analyst from a bank or research company who has a lot of experience in his area of expertise: publishing share recommendations and predicting the stock market development"*. As Surowiecki (2005) elaborates, there are three key conditions a group should satisfy for crowd wisdom to arise: Diversity, independence and decentralization (Surowiecki 2005, p. 9).[1] Given a large sample size, social media users should satisfy these criteria. Furthermore, aggregation of the crowds diverse opinions is required to arrive a consensus decision (Surowiecki 2005, p. 9). However, in contrast to groups such as the 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, not to arrive at a consensus. The text mining methods applied in this study shall serve as a humble substitute. Whether the groups lack of intent to arrive at a common goal diminishes its ability to create a WoC effect will be investigated by means of the following analysis. On the other hand, as Poetz & Schreier (2012) point out, expert knowledge can lead to superior skills and problem solving within a given domain, an assertion supported by long standing research (Anderson 1981; Larkin et al. 1980), but the extend of this superiority is especially limited with regard to the predictive accuracy of expert opinions (Johnston & McNeal 1967). While such earlier sources mainly stem from the psychological research domain, the rise of the networked society has enabled crowd-driven projects on a scale previously unknown. A popular example of such a project is Wikipedia, which has been shown to be as accurate as the Encyclopedia Britannica (Giles 2005). Poetz & Schreier (2012) continue to point out that, within the financial domain, analysts and fund managers are well known to rarely beat the market (Carhart 1997; Jensen 1968; Malkiel 1995), while previous studies show that stock prediction communities can achieve higher performance than the market in general (Hill & Ready-Campbell 2011), their own study concludes that they can outperform stock analysts. Based upon these previous findings two conclusions are drawn: To begin with, analyst recommendations are known to be biased because of the payment incentive structures inherent to the system professional analysts work in. In contrast, social media users are not faced by the same type of incentives. In unison these two factors create a need and possible source for supplementary sources of information about companies. Based upon the discussed research in the areas of analyst reports and recommendations, as well as WoC, there may be no uniform answer to the question whether professional analysts or social media users are quicker when incorporating new information into their opinion because both groups are faced with different obstacles. Against this background, the following hypothesis is proposed:

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

Against the theoretical background that cases exist where (a) their financial expertise is especially relevant and/or (b) they have privileged information, we hypothesize:

- **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.

While against the theoretical background of WoC research that cases exist where (a) the information is not privileged and therefore accessible to social media users and (b) the user group satisfies the three key conditions for crowd wisdom proposed by (Surowiecki 2005) to arise, we hypothesize the opposite:

- **Hypothesis 1c:** Social media users' mean sentiment changes prior to professional analysts if the discussed conditions for crowd wisdom are satisfied.

[1] In Surowiecki (2005) see p. 23, for independence p. 41 and for decentralization p. 66 for more detail on these conditions.

For which of these conflicting theories evidence can be provided will be investigated in the analytic part of this paper, for which the next section provides a methodological introduction.

3 Method

The question whether analysts or social media users react quicker to new information will be analyzed via GC testing. First of all, this method is easily misunderstood because of its name. A test for GC is essentially a WALD-test, comparing a model explaining a time series present value using its lagged values to another model, which also adds lagged values of a second time series, or formally according to (Kirchgässner et al. 2012, p. 97) x is Granger causal to y if:

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

i.e. the forecast error σ^2 is reduced by including the past values of x. Thus, despite of the methods name *no causal relationship is implied* if GC is discovered and asserting such a relationship purely on the basis of GC would be a *Post Hoc Fallacy* (Damar 2005, p. 180). Consequently, three outcomes of a test for GC between a pair of time series can occur: Either no GC relationship is discovered, meaning that neither series can statistically improve the prediction of the other, or a GC relationship is discovered in one direction, or a GC relationship is discovered in both directions. For each company pair of social media and analyst report series, five different models are estimated per possible direction of the GC. One model for the entire year 2013 and four quarterly models, each following the model specification with the null hypothesis that there is no GC, i.e. that M2 does not reduce the forecast error.

$$M1 : Y_t = \alpha + \sum_{i=1}^{n} \beta_i Y_{t-i} + \varepsilon_t, \tag{2}$$

$$M2 : Y_t = \alpha + \sum_{i=1}^{n} \beta_i Y_{t-i} + \sum_{i=1}^{n} \gamma_i X_{t-i} + \varepsilon_t, \tag{3}$$

where n refers to the number of lags included in each model. The selection of the lag-length parameter in the GC model is a balance between including few lags, which potentially misses a present relation between the time series, and including many lags, which can lead to spurious results. The selection of this critical parameter is outline in section 5, while the next section will establish the hypotheses the following analysis aims to address. Now that method used here has been established, the question what insights can be gained by the following analysis can be answered. Revisiting the hypotheses proposed in section 2.2, the aim of the analysis is to establish whether analyst report and social media sentiment may be used to predict one another and, if so, under which circumstances. 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, as well as the importance of expert knowledge. However, 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 is taken as evidence that in such cases events occurred that were not foreseen by the domain experts and were consequently incorporated into the public opinions of social media users in a more timely manner. Furthermore, such a relation could indicate that, besides the known herding behavior of analysts, they also follow the public opinion about a given company.

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
CocaCola	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
Homedepot	253	144,170	3
IBM	337	146,425	135
...	...	...	...

Company Name	Analyst Reports	Social Media	Public News
Intel	525	147,317	157
JPMorgan	331	140,364	71
Johnson&Johnson	361	141,091	1900
McDonalds	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 Cos.	212	4,413	167
Unitedhealth	281	41,876	14
Unitedtechnologies	268	44,247	1
Verizon	405	155,642	204
Visa	291	137,386	401
Walmart	333	106,735	154
Σ	9,439	3,814,839	21,278

Table 1. Observation counts for analyst report (TRAA), social media (SDL) and news (Guardian Open API) data.

- **Case 3 – Analyst reports are found to GC social media content:** This provides evidence that, due to their superior domain knowledge or privileged access to information, analysts are able to assess situations before the crowd can arrive at a similar conclusion. Since it is unlikely that a significant portion social media users have direct access to analyst reports, any impact of such reports on the opinion of the crowd has to be by proxy, either via traditional media channels reporting on the professionals opinions or via individual *star users* within a social community.
- **Case 4 – No GC is discovered between the two types of media content:** The information contained therein appears to be independent from one another. While such a result indicates that the two types of data are not interchangeable this does not by itself provide a contraindication to their predictive power with regard to other data, such as stock returns.

4 Data acquisition & processing

Three data sets will be used in the following analysis, each containing information about the companies included in the *DIJA* in 2013, providing a sample of companies from a variety of industries. Table 1 gives an overview of the data. Each data set is collected for the year 2013 starting on January 1st and ending on December 31st. The first data set consists of analyst reports requested from Thomson Reuters Advanced Analytics (TRAA) platform. After extracting the textual content from the reports, this results in a total of 9,439 observations. The second data set contains a broad selection of social media content, including web logs, forums and product reviews and was obtained from SDLs SM2 database, which is a database primarily intended for marketing oriented use and therefore contains content of these types. For each company, a maximum of 40,000 observations per quarter was requested from the database, resulting in a total of 3,814,839 observations after all processing steps, about 127,000 per company. Finally, a public news data set is 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”. 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 presented in section 5.[2] Table 1 contains an overview of the number of observations available for each company. Figure 1 gives an overview of the data retrieval process for the social media and analyst report data, which are used for GC testing.
As the figure illustrates, the two content sources are treated analogously when preparing them for the following analysis. Sentiment scores are calculated for each document. This is done using the General

[2] Furthermore, some common company specifications, such as industry dummies, will be used in this second analysis step. This data is described in section 5.

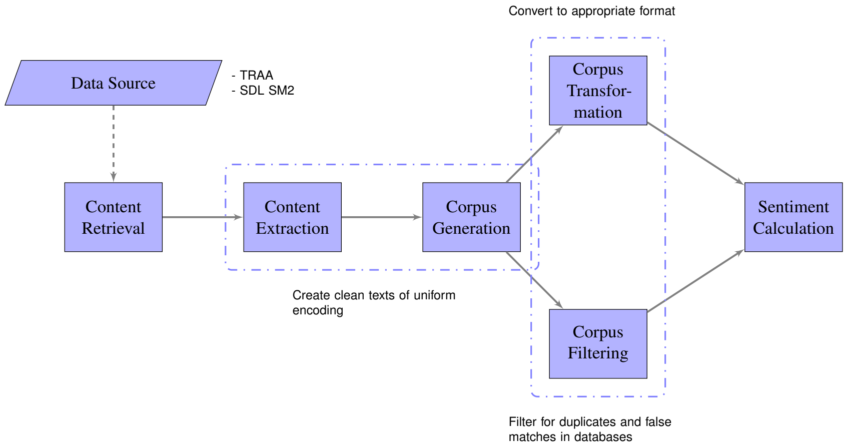

Figure 1. Content extraction workflow from data retrieval to sentiment calculation.

Inquirer (GI) software using the *Positiv* and *Negativ* categories from the Harvard IV-4 dictionary (Stone et al. 1966).[3] The basic assumption made when using such dictionaries is that the words contained therein have a *prior polarity* (Wilson et al. 2005, p. 347), e.g. the word "good", when considered without context, will be perceived as positive by most people. It is this *prior polarity* that is used to assign words to a sentiment category. Of course, 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". Where such violations of the assumption occur, they introduce a bias to the analysis. The sentiment score for each document i of company j is consequently calculated using a positivity measure:

$$Positivity_{i,j} = \frac{pos_{i,j}}{pos_{i,j} + neg_{i,j}} \tag{4}$$

Observations are dropped unless the analyzed text contains more than 50 words and a positivity score could be calculated.[4] This results in a 4% reduction of social media data. After the sentiment scores for each document have been calculated, the resulting social media sentiment and analyst report sentiment time series need to be scaled to a common frequency. In principle, a higher common frequency seems desirable in order to provide a larger number of observations to the test for GC. Given the large number of observations in the social media data, the analyst report data dictates the achievable frequency. As the number of available reports ranges from 166 (General Electric) to 541 (Cisco) and, as can be expected, there are quarterly heaps of report releases, a higher frequency than daily aggregates of report sentiment would not be supported by the available data. Consequently, daily means of the positivity measure are calculated for both analyst reports and social media content for each company in the sample. Furthermore, missing values in the report series, are added by linear interpolation. Figure 2 illustrates notable aspects of the data. An important point is illustrated by the left plot in the figure. Stock analysts seem to be much more cheerful than the average social media user. Indeed, prior research indicates that this might be a result of the fact that an analysts compensation is, at least in part, determined by the investment banking business generated after a report, instead of the accuracy of the recommendations or forecasts included in the report (Groysberg et al. 2011). On the other hand, Twedt & Rees (2012) argue that this effect might be lessened by examining the full content of a report, as done here, instead of focusing on the buy/hold/sell recommendation or a forecast measure (Twedt & Rees 2012, p. 6). However, there

[3] The source refers to the original GI publication. For the current dictionaries see `http://www.wjh.harvard.edu/~inquirer/`.

[4] An example for a dropped document would be one without any dictionary matches because this results in a division by zero in the positivity measure.

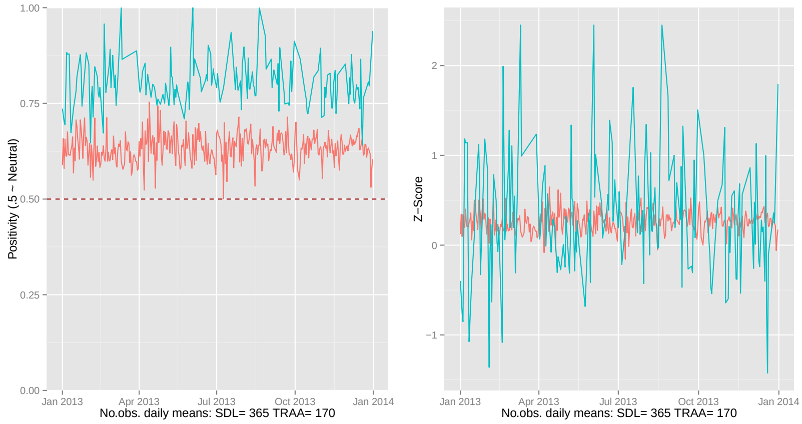

Figure 2. *Daily sentiment means (left) and z-scores (right) of daily sentiment means for Cisco. The daily mean counts on the charts refer to the number of available daily data points prior to linear interpolation of missing values.*

seems to be an indication that analysts are hesitant to use negative language. One possible way to address the apparently domain-specific language of stock analyst would be to compile a sentiment dictionary specifically for the domain. In the meantime, following the method applied by Bollen et al. (2011), z-scores of both time series are used to normalize the two series around a common level (Bollen et al. 2011, p. 3), resulting in the centered series illustrated in figure 2 (lower right).[5]

$$Z = \frac{x - \mu}{\sigma} \tag{5}$$

Since 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.

5 Results

In this section the statistical results of the analysis are presented in two steps. First, the results of GC testing are developed and interpreted. Consequently, these results are used in a second analysis aiming to provide evidence about the circumstances leading to a particular direction of GC. Here, a lag-length selection method following Bollen et al. (2011), who evaluate model performance based upon the p-values of the additional time series in M2, is chosen because of the large number of models in the analysis.[6] For each company (30) in the sample and both possible directions of the GC (2) models are estimated for quarterly subsets, as well as the entire annual time span of 2013, models from lag-length $n = 1$ to $n = 15$ days are estimated. This results in $30 * 2 * 15 * 5 = 4500$ models. There are two possible scenarios regarding the question in what kind of GC relationship the two mediums at hand might be connected: The two mediums might (a) have an inherently different reaction-speed to new information resulting in a uniform direction GC between the content types or (b) the direction of GC may depend on the particular

[5] Bollen et al. (2011) use a sliding window Z-Score, here the time-series is normalized over the entire year.
[6] Another approach would be to compare the Aikake Information Criterion (AIC) for all lag-lengths of a given series pair and select the model with the smallest AIC. As the individual models are not of interest here, the chosen approach suffices.

type of information and company at hand. While the latter scenario conforms with the hypotheses outlined in section 2.2, it seems worth exploring the alternative. To this end, the results that will be reported here first are those of the number of models with a significant relationship between the series pairs, i.e. those models with a p-value smaller than 10%, for each of the 15 lag-lengths. If one of the two mediums is inherently faster in its reaction than the other, this should be reflected in higher counts of GC in one direction. Table 2 gives an overview these results, the first row refers to models in which social media sentiment was used to augment analyst report sentiment and the second row offers the results of the opposite direction.

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
Σ	44	40	33	36	37	37	43	45	39	**57**	50	56	52	48	45

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

As the table illustrates, the number of lags resulting in the largest amount of significant models is conveniently identical for both prediction directions. While this is a convenient result, using another sample, e.g. requiring another minimal text length or including minimal dictionary hit counts, different optimal parameters for the two directions could result. Another interesting observation is the fact that there seems 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, therefore providing evidence that GC between the two mediums is indeed driven by more circumstances than an inherent imbalance in reaction-speed between the two mediums. Although a lag-length of 10 periods creates the largest number of significant relationships, there is no indication that this number of lags is preferable for all series pairs since other lag-lengths also produce a similar magnitude of significant models. Therefore, another set of models is estimated, in which the lag-length parameter is not chosen for all models at once, but individually for each of the 300 models, again looking to minimize the p-value of the model. The results of this second selection of models are reported in table 3. As illustrated, the model specific p-value based lag-length selection procedure improves the number of significant models from 57 (19%) to 128 (42.7% Case 1 and 57% Case 4). Of these 127 significant models, 71 are models predicting analyst report sentiment by social media sentiment, while 57 models predicting social media sentiment via analyst report sentiment are statistically significant, which corresponds to 45.3% evidence for Case 2 and 39.3% evidence for Case 3. These results provide a strong indication that both directions of prediction are feasible. Note that since the pre-processing of the input documents was performed identically for both types of content and all companies, further improvements of the counts of significant models could possibly be made by introducing a case specific pre-processing logic. As the question of interest in this work is not case specific, the applied uniform logic seems sufficient. Of course, these results raise the question which situation results in which case of GC relationship, i.e. what kind of company is more prone to either direction of GC or what situation fosters it. This question will now be addressed. This is achieved using binary response models, estimating the chance that a given set of circumstances lead to GC in either direction with a separate model for each direction. The reason why two models per situation are estimated is that, as discussed earlier, the two directions of GC are not mutually exclusive and therefore a single binary encoding of the GC direction would miss the cases where both directions are simultaneously significant. In particular, Generalized Linear Models (GLM) using Log-Link functions are estimated (Logit). The independent variables included in these models include common financial measures, as well as industry dummies. These two kinds of variables are inherent to the company in question. Therefore, if one of them can be shown to have significant impact on the direction of the GC, a specific type of company would be more suitable for such predictions. The second type of dependent variable is not

Sample	Q1		Q2		Q3		Q4		Annual		
Predicted	Analyst	Social	Analyst	Social	Analyst	Social	Analyst	Social	Analyst	Social	
Explanatory	Social	Analyst	Social	Analyst	Social	Analyst	Social	Analyst	Social	Analyst	Σ
3M	0.692	**0.042****	0.372	0.282	0.253	0.118	**0.066***	**0.066***	0.157	0.505	3
AmericanExpress	0.16	0.334	0.405	**0.032****	0.3	**0.099***	0.491	0.22	0.428	0.138	2
ATT	**0.089***	0.545	0.196	0.28	0.222	0.129	0.115	0.417	0.226	0.131	1
Boeing	**0.039****	**0.06***	0.37	0.319	0.293	**0.031****	0.288	**0.089***	0.307	**0.004*****	5
Caterpillar	**0.021****	0.617	0.464	0.242	**0.094***	0.281	**0.007*****	**0.017****	**0.004*****	**0.063***	6
Chevron	**0.024****	0.174	0.25	0.315	0.105	**0.037****	**0.035****	**0.047****	**0.099***	0.208	5
Cisco	**0.019****	0.386	0.101	**0.016****	**0.021****	0.125	0.266	**0.095***	**0.045****	**0.024****	6
CocaCola	0.183	**0.022****	**0.027****	**0.002*****	**0.041****	0.504	**0.032****	0.245	**0.026****	0.219	6
Disney	**0.036****	0.119	0.317	**0.047****	0.48	0.166	**0.064***	0.338	0.28	**0.023****	4
DuPont	**0.011****	0.282	**0.048****	0.697	0.43	**0*****	**0.041****	**0.004*****	0.2	**0.007*****	6
Exxon	**0.02****	**0.093***	**0.047****	0.16	0.108	0.304	0.132	0.128	**0.007*****	0.19	4
GE	**0*****	**0*****	**0.095***	0.644	0.511	0.163	**0.031****	**0.029****	**0.019****	0.229	6
Goldman	0.262	0.39	**0.09***	0.237	**0.038****	**0.099***	0.278	**0.047****	0.108	0.122	4
Homedepot	0.154	0.104	0.232	0.6	**0.088***	**0.032****	0.125	**0.028****	**0.051***	**0.045****	5
IBM	**0.058***	0.19	**0.001*****	0.138	0.605	**0.044****	**0.032****	**0.03****	0.37	0.599	5
Intel	**0.015****	0.161	**0.036****	**0.022****	0.255	0.189	0.361	0.107	0.812	0.307	3
JohnsonJohnson	**0.088***	**0.078***	0.271	0.194	**0.022****	0.127	**0.01*****	0.147	0.613	0.143	4
JPMorgan	**0.001*****	**0.029****	0.222	0.741	**0*****	0.317	0.128	**0.009*****	0.283	**0.007*****	5
MCDonalds	0.122	**0.013****	**0.054***	0.438	**0.095***	**0.08***	**0.057***	0.341	0.101	0.267	5
Merck	0.277	0.671	0.148	0.115	**0.022****	**0.05****	0.555	0.353	0.188	0.276	2
Microsoft	0.134	0.266	0.307	0.397	0.289	0.325	0.104	**0.001*****	**0.026****	0.295	2
Nike	**0.071***	0.224	**0.043****	0.511	**0.006*****	0.413	**0*****	0.597	**0.034****	0.712	5
Pfizer	**0.002*****	**0.014****	0.114	0.22	**0.037****	0.194	**0.015****	0.485	**0.004*****	**0.014****	6
ProcterGamble	0.172	0.339	0.3	0.196	0.202	**0*****	**0.039****	0.131	0.313	0.106	2
Travelers	0.161	0.296	0.483	**0.024****	**0.011****	0.373	0.591	**0.001*****	0.347	**0.023****	4
Unitedhealth	0.598	0.101	0.301	0.46	**0.005*****	**0.084***	0.761	0.185	**0.03****	0.359	3
Unitedtech	**0.074***	0.238	0.266	0.115	0.23	**0.011****	0.105	**0.026****	0.273	**0.08***	4
Verizon	**0.005*****	0.166	0.306	0.619	**0.022****	0.253	0.134	**0.09***	**0.073***	0.776	4
Visa	0.244	**0.057***	0.166	**0.027****	**0.043****	0.149	0.659	0.651	**0.047****	0.163	4
Walmart	**0.08***	0.238	**0.024****	**0.071***	**0.021****	**0.007*****	0.437	0.18	**0.006*****	**0*****	7
Σ	18	10	10	8	16	13	13	15	14	11	Σ 128

Table 3. *P-Values (3-digit rounding) of GC tests for quarterly sub samples and annual data. Optimal lag-length for each individual time series pair based upon lag-lengths between 1 and 15 (p-value < 0.01: ***, > 0.05: **, < 0.1: *). The* Predicted *column indicates the test direction, e.g. Predicted = Analyst means that reports are predicted via social media content. The sums refer to the number of significant models in columns and rows respectively.*

directly determined by the companies operations but rather the reporting of these activities by the press. To this end, using the Guardian API data mentioned in table 1, the most common (median) news category for each time frame and company is extracted and introduced to the models in dummy coding. All other variables are constructed from Thomson Reuters Datastream. Thus, the following models are estimated. For the social media model the dependent variable is 1 if the analyst report data was found to GC the social media data in a given period and firm;

$$Social_{i,j} = \alpha + \beta_0 * R\&D + \beta_1 * Parent + \delta * IND_{i,j} + \gamma * Newscat_{i,j} \quad \text{with} \quad Social_{i,j} \sim Log-Link, \tag{6}$$

while for the analyst report model the dependent variable is 1 if the social media data was found to GC the analyst report data in a given period and firm;

$$Analyst_{i,j} = \alpha + \beta_0 * R\&D + \beta_1 * Parent + \delta * IND_{i,j} + \gamma * Newscat_{i,j} \quad \text{with} \quad Analyst_{i,j} \sim Log-Link, \tag{7}$$

for a given company (i) and time frame (j), while *Parent*, *IND* and *Newscat* refer to the number of subsidiaries of the company, industry dummies and news categories respectively. The estimated models are reported in table 4. As illustrated, there are a number of significant dummy variables allowing to identify a situation in which analyst reports sentiment is granger caused by social media sentiment. In particular, companies in the energy, financial, industrial or telecommunication sector are less prone to this type of relationship. Looking at the news dummies, we observe that environmental news seem to spark

	Analyst ← Social	Social ← Analyst
(Intercept)	**1.85** (0.80)*	0.17 (0.74)
IND_Consumer Staples	−0.07 (1.72)	−0.93 (1.58)
IND_Diversified	−0.86 (0.93)	−0.06 (0.91)
IND_Energy	−**4.62** (2.24)*	0.16 (2.20)
IND_Financials	−**1.92** (0.88)*	0.11 (0.78)
IND_Health Care	0.27 (1.00)	−0.18 (1.05)
IND_Industrials	−**2.50** (1.33)$^{\cdot}$	1.39 (1.31)
IND_Technology	0.04 (1.69)	0.13 (1.71)
IND_Telecommunication	−**1.90** (0.93)*	−1.61 (1.25)
R&D	−0.00 (0.00)	−0.00 (0.00)
Parent	−0.00 (0.00)	−0.00 (0.00)
News_commentisfree	−**1.73** (0.85)*	−0.70 (0.76)
News_environment	**4.01** (2.08)$^{\cdot}$	−0.87 (2.06)
News_film	−1.88 (1.16)	−0.48 (1.12)
News_football	−2.69 (2.03)	−0.52 (1.92)
News_media-network	−**1.92** (1.16)$^{\cdot}$	0.31 (1.13)
News_music	−0.95 (1.29)	0.22 (1.09)
News_sport	0.05 (0.78)	−**2.18** (1.17)$^{\cdot}$
News_sustainable-business	−0.05 (1.26)	0.43 (1.26)
News_technology	−0.03 (1.71)	−0.74 (1.76)
News_world	−0.36 (0.91)	−0.30 (0.86)
Log Likelihood	-88.22	-85.26
Deviance	176.43	170.52
Num. obs.	145	145

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 4. Model summaries for both directions of GC. Analyst ← Social indicates that analyst reports were GC by Social Media data and vice versa. Five observations are lost due to unavailable R&D data. Coefficients can be interpreted as $exp(coeff)$ in % c.p. on average, keeping in mind the intercept as a baseline given it is significant.

interest in the social media content before professional analysts can react. Surprisingly, the likelihood for the relationship decreases if media-network news are the most common news category. It seems that such news do not create much social media impact in the sample. Finally, news are less likely to let the social media sentiment GC the analyst report sentiment if the comment section of the article is open. This might be explained by the fact that comment sections are more likely to be closed (or closely moderated) for controversial issues and therefore are open for news that do not spark controversy. In the opposite direction, social media content granger caused by analyst reports, only one significant relationship can be established. Sport news decrease the likelihood for such a relationship. As social media channels are often advertised during sporting events, this does not surprise. It should be noted that while other model specifications can slightly change the picture by changing which categories are significant in either direction, the overall result does not change: There are still more significant coefficients for the social media model than for the analyst report model. Finally, note the insignificant *R&D* and *Parent* variables. These refer to the research and development budget of the company in 2013, as well as its ultimate number of subsidiaries. These are included as examples of variables only available on an annual level. Because there is no variance in such variables over the different quarterly periods, no significant effect can be established. Also, because all companies included in the index are by definition "large caps", another factor inherent to the data limits variance in these variables.
Regarding the proposed hypotheses the results can be interpreted as follows:

- **H1a (Situational direction of GC):** Results indicate that this is indeed true, no uniform direction of GC can be established between the two types of content.
- **H1b (Analysts expertise and privileged access matter):** This hypothesis is supported by the negative coefficients in the industry dummies (Analyst model) regarding energy, financial and industrial companies, indicating that these types of company require the expertise that analysts are able to offer.
- **H1c (When relevant information is public and diverse, independent opinions can be aggregated, the crowd has an advantage):** This is supported by the positive coefficient of the environmental

news category when analyst reports are predicted and the negative coefficient of the sport news category in the opposite direction. Due to the statistically insignificant intercept in the *Social* model, the interpretation of these results is difficult.

Thus, the analysis provides evidence for H1 and H2, while evidence in support of H3 is sparse. This could be improved by lifting the aforementioned limitations of the data using a longer sample period and a more diverse set of companies.

5.1 Limitations

These results should be considered with both methodological and theoretical limitations in mind. First of all, there is no reliable way to determine which portion of social media users are professional analysts. A more controlled experiment using a single social network would be an interesting venue for future research. Also, it could be argued that analysts could exhibit some WoC effects themselves because of their known tendency to exhibit herd behavior. Furthermore, reactions falling out of the $n = 15$ lag-length included in the analysis may be missed. As large lag-lengths can lead to spurious correlation, this methodological trade-off has to be accepted. Of course, using a larger and more diverse set of companies and a multi-year sample, such balance sheet based variables could provide interesting results, thus offering interesting questions for future research.

6 Conclusion

The goal of this research was to provide evidence regarding the question of a possible inherent relation between the prediction power of stock analysts' sentiment and that of "the crowd", i.e. a large set of social media users. To this end, GC testing between the two types of content showed significant statistical relations in a large number of cases in the sample. Since this indicates that the two types of content can indeed be used to predict one another in many cases, this does provide evidence that the two types of content can indeed be used in a similar manner, or at least to complement one another (keeping in mind the arising multicollinearity issue). Practical applications for such relationships include fields, such as algorithmic trading, news reporting and customer relations. Furthermore, the second stage of statistical analysis (logit-models) provides evidence for the question in which circumstances social media contents sentiment can be used to predict analyst reports and vice versa. Results are promising for the prediction of analyst reports, while little significant relationships can be established in the opposite direction. This might well be mitigated by a larger sample, which either introduces more variance regarding the type of company or spans a significantly longer period, thus offering an interesting possibility for future research. The theoretical contribution of these results is given by the evidence for WoC theory provided by the cases where social media users are able to arrive at a (collective) opinion before professional stock analysts are able to include changes in the environment into their own reports to the public. Results indicate that the conditions (diversity, independence and decentralization) for crowd wisdom are met and that the crowd can outperform experts when information is readily available to the crowd. On the other hand, professional analysts seem to be able to react quicker to technical issues, such as changes in a company's financial situation. Therefore, the findings of this study support the basic assumptions of WoC theory, while professional analysts are able to persevere within their niche of expertise. Opportunities for future research are given by reactions outside the lag-scope of the presented analysis and more granular analyses of specific social networks or user types.

References

Anderson, John Robert (1981). *Cognitive skills and their acquisition*. Psychology Press.

Barber, Brad et al. (2001). "Can investors profit from the prophets? Security analyst recommendations and stock returns." *The Journal of Finance* 56 (2), 531–563.

Barniv, Ran et al. (2009). "Do analysts practice what they preach and should investors listen? Effects of recent regulations." *The Accounting Review* 84 (4), 1015–1039.

Bollen, Johan et al. (2011). "Twitter mood predicts the stock market." *Journal of Computational Science* 2 (1), 1–8.

Bradshaw, Mark T (2009). "Analyst information processing, financial regulation, and academic research." *The Accounting Review* 84 (4), 1073–1083.

Carhart, Mark M (1997). "On persistence in mutual fund performance." *The Journal of finance* 52 (1), 57–82.

Clement, Michael B & Senyo Y Tse (2005). "Financial analyst characteristics and herding behavior in forecasting." *The Journal of finance* 60 (1), 307–341.

Damar, TE (2005). *Attacking Faulty Reasoning: A Practical Guide to Fallacy-Free Arguments*. Belmont, CA: Thomson Wadsworth.

Elgers, Pieter T et al. (2001). "Delayed security price adjustments to financial analysts' forecasts of annual earnings." *The Accounting Review* 76 (4), 613–632.

Frankel, Richard et al. (2006). "Determinants of the informativeness of analyst research." *Journal of Accounting and Economics* 41 (1), 29–54.

Giles, Jim (2005). "Internet encyclopaedias go head to head." *Nature* 438 (7070), 900–901.

Groysberg, Boris et al. (2011). "What Drives Sell-Side Analyst Compensation at High-Status Investment Banks?" *Journal of Accounting Research* 49 (4), 969–1000.

Hill, Shawndra & Noah Ready-Campbell (2011). "Expert stock picker: the wisdom of (experts in) crowds." *International Journal of Electronic Commerce* 15 (3), 73–102.

Hong, Harrison & Jeffrey D Kubik, et al. (2000). "Security analysts' career concerns and herding of earnings forecasts." *The Rand journal of economics*, 121–144.

Hong, Harrison & Terence Lim, et al. (2000). "Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies." *The Journal of Finance* 55 (1), 265–295.

Jensen, Michael C (1968). "The performance of mutual funds in the period 1945–1964." *The Journal of finance* 23 (2), 389–416.

Johnston, Roy & Benjamin F McNeal (1967). "Statistical versus clinical prediction: Length of neuropsychiatric hospital stay." *Journal of Abnormal Psychology* 72 (4), 335.

Kirchgässner, Gebhard et al. (2012). *Introduction to modern time series analysis*. Springer.

Larkin, Jill et al. (1980). "Expert and novice performance in solving physics problems." *Science* 208 (4450), 1335–1342.

Malkiel, Burton G (1995). "Returns from investing in equity mutual funds 1971 to 1991." *The Journal of Finance* 50 (2), 549–572.

Nofer, Michael & Oliver Hinz (2014). "Are crowds on the internet wiser than experts? The case of a stock prediction community." *Journal of Business Economics* 84 (3), 303–338.

Poetz, Marion K & Martin Schreier (2012). "The value of crowdsourcing: can users really compete with professionals in generating new product ideas?" *Journal of Product Innovation Management* 29 (2), 245–256.

Stone, Philip J et al. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* MIT press.

Surowiecki, James (2005). *The wisdom of crowds*. Random House LLC.

Trueman, Brett (1994). "Analyst forecasts and herding behavior." *Review of financial studies* 7 (1), 97–124.

Twedt, Brady & Lynn Rees (2012). "Reading between the lines: an empirical examination of qualitative attributes of financial analysts' reports." *Journal of Accounting and Public Policy* 31 (1), 1–21.

Wilson, Theresa et al. (2005). "Recognizing contextual polarity in phrase-level sentiment analysis." In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, pp. 347–354.

Womack, Kent L (1996). "Do brokerage analysts' recommendations have investment value?" *The Journal of Finance* 51 (1), 137–167.