JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

The Effect of Ad Hoc Announcement Sentiment on Daily Return

Text Mining

Winter Semester 20/21

– Term Paper –

Submitted to

Prof. Dr. Nicolas Pröllochs
Chair for Data Science and Digitization

at

Faculty of Economics and Business Studies
University of Giessen

Submission date: 26.02.2021

Submitted by: Julius Wissel & Daniel Kinkel
Student ID: 603951, 6005391
E-Mail: julius.wissel@wirtschaft.uni-giessen.de, daniel.kinkel@wirtschaft.uni-giessen.de

Abstract

Ad hoc announcements contain information with highly important implications for the financial market. Since this information impacts the price of an underlying equity, there is a strong incentive to apply text mining to extract the relevant information. In this paper we investigate how the sentiment within those ad hoc announcements influences the daily return. Sentiment is analyzed using a dictionary-based approach and expressed by the net optimism score. For this purpose, we estimate different regression models in the main analysis and the robustness checks. Based on the analysis we can partially confirm our hypothesis that sentiment has a positive effect on the daily return. While we find a significant effect in our main model, the effect size is very small. Also, the robustness checks indicate that this effect is mainly driven by the frequency of negative terms in ad hoc announcements and only present in some company sectors. We address existing limitations and conclude with suggestions for future research.

Contents

1. Introduction

In today's world, increasingly large amounts of data exist in the form of text. In fact, so much text containing valuable information is produced every day that it would be impossible to analyze it all manually. Also in the finance context information in the form of text is published at a very high frequency (Lewis and Young 2019). One specific type of text data published regularly by companies is ad hoc announcements. As already noted by Frazier et al. (1984), announcements like these contain information relevant to the financial market and should therefore impact the price of an underlying equity. Since this information is of substantial value, there is a strong incentive to apply text mining to extract the relevant information (Hagenau et al. 2013). Text mining combines techniques like data mining, machine learning and computational linguistics (Gupta et al. 2020). One common method in the analysis of financial news is to focus on the language used in the ad hoc announcement and thereby capture and measure the prevalent sentiment (Schumaker et al. 2012). However, this procedure poses a few challenges. Das (2014) states, for example, that for some facts it may not be possible to express them in a quantitative form, as they are intrinsically qualitative. Therefore, the question arises whether a sentiment analysis succeeds in extracting the financially relevant information. For this purpose, our paper aims to investigate the relationship between the sentiment captured from ad hoc announcements and the corresponding daily return. This is based on the idea that in sufficiently efficient markets the daily return should already be a fitting indicator of how market participants perceive the content of the announcements. Consequently, we hypothesize that sentiment has a positive impact on daily return. In order to investigate this, we utilize a dataset with a large number of ad hoc announcements, which is described in Section 2 of this paper. Section 3 deals with the sentiment analysis as well as the necessary preprocessing steps. For the sentiment analysis, we apply a dictionary-based approach that allows the identification of positive and negative terms in the announcement. The terms are classified by the dictionary of Loughran and McDonald (2011) and measured using the net optimism score as proposed by Demers and Vega (2010). In Section 4, we investigate how this sentiment score relates to the daily return. For this purpose, we estimate several regression models after performing descriptive analyses. We also evaluate the robustness of the initial findings in Section 5. Here, we consider outliers, sparse words, splitting up the sample and a different sentiment measure. Section 6 contains a summary of our results as well as an outlook on possible future research. We finish with some concluding thoughts in Section 7. All steps mentioned above are carried out with the statistics software R and the *SentimentAnalysis* package by Feuerriegel and Proellochs (2019).

2. Dataset

Our dataset consists of ad hoc announcements of public companies. These mandatory announcements are regulated by the German Federal Financial Supervisory Authority (BaFin) since they are relevant to the financial market. This has the advantage that the content of the announcements is quality-checked. From ad hoc announcements published between January 2004 and June 2011, those written in English and containing at least 50 words were selected. The resulting dataset of 13,135 announcements also includes additional information about the daily stock market return, the price-to-book value, the company sector and the date of publication. The distribution of the daily stock market returns is visualized in Figure 1. The plot indicates the presence of high positive kurtosis, which would imply that the returns are not normally distributed. This inference is supported by the calculated values for skewness and kurtosis, which are displayed in Table 1. As it is characteristic for distributions with high kurtosis, the return data exhibits large outliers. Even though these outliers can lead to inaccuracies in the regression, excluding these observations also means a loss of information. Thus, we decide to keep them included for now, but investigate their impact in the robustness check. Similar characteristics of the distribution are present for the price-to-book variable. In consequence, we treat this variable accordingly. Additionally, it can be observed that 358 observations exhibit negative price-to-book ratios. Since the interpretation of this metric does not hold for negative ratios, we exclude them from our dataset. Furthermore, we exclude all observations with missing values. While only 380 data points are missing for the price-to-book ratio, the variable company sector is not specified in 4,980 cases. Still, we argue that the resulting dataset after all exclusions is sufficiently large with 7,876 observations. As an additional control variable, we construct dummy variables for the respective years.

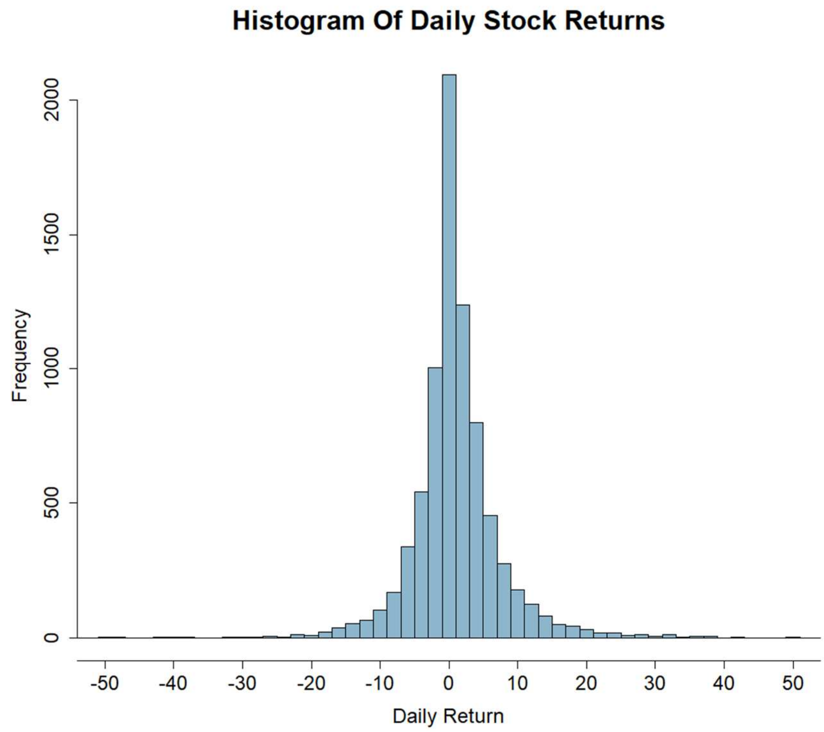

Figure 1: Frequencies of daily stock market returns grouped in intervals.

Variable	Type	NAs	Min.	Median	Mean	Max.	Skewness	Kurtosis
Return	continuous	0	-81.07	0.06	0.86	183.33	5.10	97.13
Price-To-Book	continuous	390	-385.66	1.59	2.24	281.91	-3.61	550.56
Company Sector	nominal	4,980						

Table 1: Descriptive statistics of variables with variable type, missing observations, minimum, median, mean, maximum, skewness and kurtosis.

3. Preprocessing & Sentiment Analysis

To be able to perform sentiment analysis for the ad hoc announcements, they first must be converted into a format that can be processed. Since the raw announcement data contains elements of markup language, those have to be removed. Furthermore, unnecessary spaces within the text are eliminated. Because we aim to analyze the sentiment on the word-level, the punctuation does not contain any relevant information and is thus also removed. For consistency and harmonization purposes, upper-case letters are converted to lower-case letters. Additionally, we eliminate both numbers and stopwords, to focus the analysis on the relevant parts of the text. Stopwords refer to the most common words in a language, which consequently are not a factor texts can be distinguished by. Those are usually identified using a predefined list and subsequently removed. To achieve even further harmonization of the data, stemming is performed. This process relies on algorithms to reduce words to a truncated form, i.e. their stem, as words with the same stem are assumed to have a similar meaning. All of the above mentioned preprocessing steps are first carried out manually with the *tm* R package of Feinerer and Hornik (2020). In the further course, the steps are conducted as part of the *analyzeSentiment* function of the *Sentiment Analysis* package. In a next step, the ad hoc announcement is transformed into a document-term matrix. This matrix exhibits the frequency of terms present in each announcement and is the necessary format for the sentiment analysis. In total our document-term matrix exhibits 25,893 unique words. Another possible preprocessing step involves the removal of sparse terms. This would mean that terms that do not appear in most of the ad hoc announcements are removed from the document-term matrix. Since this step is not necessary for our analysis and it might also result in a loss of information, we refrain from this for the time being and instead perform a robustness check later on. A visualization of the most frequent terms in the ad hoc announcements is provided in Figure 2. Based on the size of the words, it becomes apparent that the five most frequent terms are *million*, *eur*, *year*, *will* and *share*. As this is performed after preprocessing, these words are already reduced to their stems. Besides from typical terms from the financial context, the word cloud also shows a few meaningless character strings that were not removed by the preprocessing. However, these should not pose a problem for our further analysis.

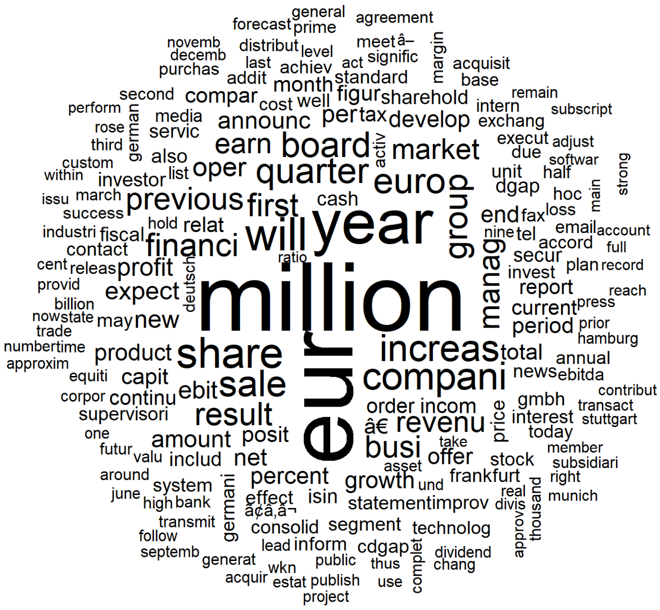

Figure 2: Word cloud of the most popular words in the ad hoc announcements after preprocessing.

For the sentiment analysis, we apply a dictionary-based approach. This kind of approach is popular in financial text mining since the results are straightforward to interpret and reproduce (Feuerriegel and Neumann 2013). It aims to identify the frequency of words within a text fitting into a defined category. The categorization of the words is performed based on a dictionary. Such dictionaries are composed of lists of words that are assigned to a particular sentiment like positive or negative. Numerous of those dictionaries have been constructed; four of which are included in the *Sentiment Analysis* package. We utilize the positive and negative word lists of the Loughran-McDonald dictionary because it was developed specifically for the finance context. These lists contain 145 and 885 different words, respectively. In our dataset 133 of the positive words and 510 of the negative words from the LM dictionary can be found. Based on these lists, positive and negative words are identified and counted for each ad hoc announcement. In order to obtain a comparable metric for the sentiment within the announcements, we use the *Sentiment Analysis* package to calculate the net optimism scores *(NOS)* as proposed by Demers and Vega (2010):

$$NOS = \frac{N_{pos} - N_{pos}}{N_{total}} \tag{1}$$

Here, N_{pos} is the number of positive words within the announcement, N_{neg} is the number of negative words and N_{total} represents the total word count of the announcement. The distribution of the resulting NOS values is visualized in Figure 3. It shows that for our dataset a similar number of positive and negative sentiments scores are calculated. Furthermore, the values vary around the mean value of -0.002 with a standard deviation of 0.031.

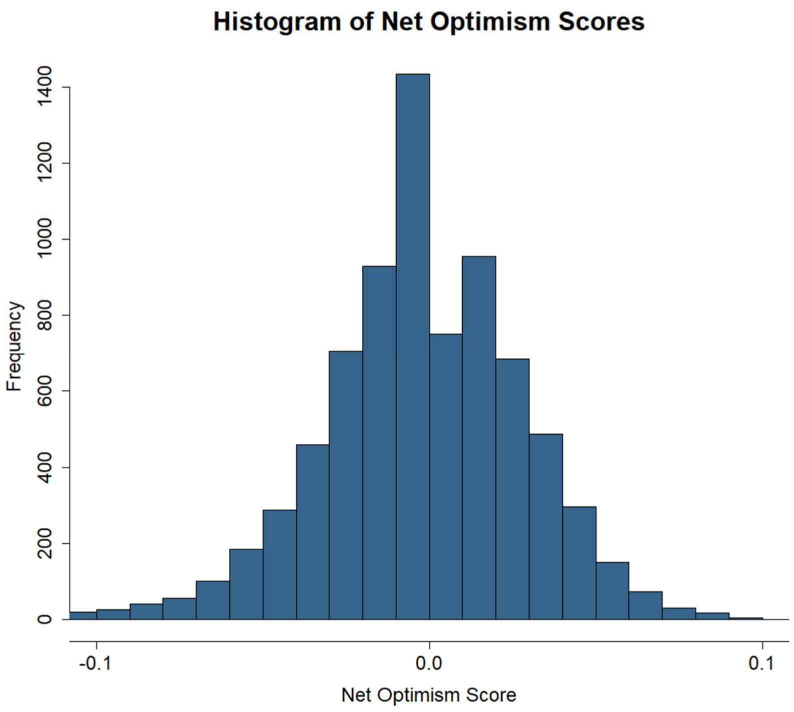

Figure 3: Frequencies of net optimism scores grouped in intervals.

4. Analysis of the Effect of Sentiment on Daily Return

In order to gain a first impression of the relationship between sentiment and daily return, we conduct a sample split for positive, neutral and negative sentiment scores. As displayed in Table 2, the mean and median show the highest values for announcements with positive sentiment and the lowest for ones with negative sentiment. This could indicate a positive impact of sentiment on the daily return. However, these values must be interpreted with caution because the corresponding standard deviations are very high. A more meaningful indicator is the Pearson correlation coefficient, which takes on the value of 0.07 and is significant on the 0.1% level. Although the coefficient is not particularly high, it also indicates a positive relationship between the variables.

	Mean	Median	SD
positive Sentiment	1.39	0.74	6.55
neutral Sentiment	1.16	0.36	7.09
negative Sentiment	0.51	0	8.95

Table 2: Descriptive statistics for the split samples.

In a next step, we use regression analysis to further explore how the sentiment of the ad hoc news influences the daily return and test the corresponding hypothesis. Therefore, we first estimate the following regression model, where *i* stands for the different ad hoc announcements:

$$Return_i = \alpha + \beta_1 * NOS_i + \varepsilon_i \tag{2}$$

This, however, would overestimate the impact that sentiment has on the return due to the omitted variable bias. In order to isolate the effect attributable to the sentiment factor, we include several control variables in the regression. These are the price-to-book variable, dummy variables for the different company sectors denoted as *k* and dummy variables for each year denoted as *v*.

$$\begin{aligned} Return_i = \alpha + \beta_1 * NOS_i + \beta_2 * PriceToBook_i + \sum_{k=2}^{34} \beta_{1+k} * CompanySector_i^k \\ + \sum_{v=2}^{8} \beta_{34+v} * Year_i^v + \varepsilon_i \end{aligned} \tag{3}$$

It is reasonable to assume that the impact of the price-to-book value might be dependent on the sector of the company. This idea is incorporated into the regression model by the addition of interaction terms for the two variables:

$$\begin{aligned} Return_i = \alpha + \beta_1 * NOS_i + \beta_2 * PriceToBook_i + \sum_{k=2}^{34} \beta_{1+k} * CompanySector_i^k \\ + \sum_{v=2}^{8} \beta_{34+v} * Year_i^v + \sum_{k=2}^{34} \beta_{41+h} * PriceToBook_i * CompanySector_i^k + \varepsilon_i \end{aligned} \tag{4}$$

Estimating the three regression models, we obtain the results presented in Table 3. It can be observed from the regression outputs that the coefficient for net optimism is positive and significant on the 0.1 % level in all three models. The coefficient of sentiment remains fairly unchanged with the value in Model 3 being 19.122. The interpretation of the coefficient of sentiment in Model 3 would be that if the NOS increases by one standard deviation of 0.031, c. p. the daily return on average increases by 0.590 percent points. The Cohen's f^2 of sentiment in model 3 of 0.006 is considerably lower than the threshold of 0.02 that is defined for a small effect. The majority of the control variables are not significant on the 5 % level. The R squared and also the adjusted R squared still improve with increasing complexity of the models. However, the explanatory power is very low, with the values for the third model being 0.029 and 0.019 respectively.

	Model 1	Model 2	Model 3
Intercept	0.996^{***}	1.772	7.992
	(0.088)	(3.488)	(19.465)
Sentiment	18.149^{***}	18.863^{***}	19.122^{***}
	(2.842)	(2.903)	(2.908)
Control Variables	No	Yes	Yes
Interaction Terms	No	No	Yes
R^2	0.005	0.014	0.029
Adj. R^2	0.005	0.009	0.019
Cohens f^2	0.005	0.005	0.006
Num. obs.	7876	7876	7876

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3: Output for linear regressions of sentiment on daily return based on equations (2), (3) and (4).

5. Robustness Check

Since daily return and price-to-book data exhibit large outliers we remove those observations, as outliers can have a disproportionate impact on the estimates. Additionally, we perform a log-transformation for the price-to-book variable as this results in a more linear relationship with the dependent variable. After performing these steps, we repeat our regression analysis as the first robustness check and report the results in Table 4. Here, the coefficients for sentiment are slightly smaller but still positive and significant on the 0.1 % level. While the effect sizes of sentiment marginally increase compared to Models 1 to 3, they remain very low.

To further check the robustness of our findings we remove sparse terms present in the document-term matrix. For this purpose, we choose thresholds of 95, 90 and 85 percent. This means that words that appear in less than these proportions of the ad hoc announcements are removed. Starting from equation (3), we estimate regression models for each case and compare the results with Model 3. As displayed in Table 5, the coefficient for the sentiment measure remains significant on the 0.1 % level for all sparsity thresholds. The size of the coefficient naturally decreases as more sparse words are removed. However, the effect sizes also decrease slightly with lower sparsity thresholds.

	Model 4	**Model 5**	**Model 6**
Intercept	0.840***	1.108	9.375
	(0.063)	(2.500)	(16.439)
Sentiment	16.230***	15.950***	15.136***
	(2.070)	(2.119)	(2.144)
Control Variables	No	Yes	Yes
Interaction Terms	No	No	Yes
R^2	0.008	0.015	0.022
Adj. R^2	0.008	0.010	0.012
Cohens f^2 (Sentiment)	0.008	0.007	0.007
Num. obs.	7714	7714	7714

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 4: Regression output of robustness check 1 after outlier correction and log transformation of the price-to-book variable. Regression models correspond to equations (2), (3) and (4).

	Model 7	**Model 8**	**Model 9**	**Model 10**
Sparsity	100 %	95 %	90 %	85 %
Intercept	7.992	8.119	8.104	8.107
	(19.465)	(19.477)	(19.483)	(19.486)
Sentiment	19.122***	10.664***	7.700***	6.009***
	(2.908)	(1.844)	(1.445)	(1.176)
Control Variables & Interactions Terms	Yes	Yes	Yes	Yes
R^2	0.029	0.028	0.027	0.027
Adj. R^2	0.019	0.018	0.018	0.017
Cohens f^2 (Sentiment)	0.006	0.004	0.004	0.003
Num. obs.	7876	7876	7876	7876

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 5: Regression output of robustness check 2 after applying different levels of sparse term removal. Regression models correspond to equation (4).

Company Sector	N	Sentiment	Std. Error	t value	Pr(> | t|)	Cohens f^2
Aerospace and Defense	5	12.71	66.563	0.191	0.880	-
Alternative Energy	411	28.807	16.546	1.741	0.082	-
Automobiles and Parts	264	1.145	14.670	0.078	0.938	-
Banks	152	-11.177	16.868	-0.663	0.509	-
Beverages	9	4.344	64.624	0.067	0.951	-
Chemicals	169	34.774**	12.664	2.746	0.007	0.047
Construction and Materials	118	27.985	18.698	1.497	0.137	-
Electricity	13	-50.645	98.224	-0.516	0.622	-
Electronic and Electrical Equipm	437	-6.183	10.203	-0.606	0.545	-
Financial Services (Sector)	539	-9.004	9.612	-0.937	0.349	-
Food and Drug Retailers	15	26.716	37.888	0.705	0.504	-
Food Producers	45	28.841	22.252	1.296	0.203	-
Forestry and Paper	14	86.676	75.455	1.149	0.294	-
General Industrials	52	-1.895	16.083	-0.118	0.907	-
General Retailers	197	62.739***	14.316	4.383	0	0.103
Health Care Equipment and Servic	387	14.369	11.946	1.203	0.230	-
Household Goods and Home Constru	108	6.281	30.828	0.204	0.839	-
Industrial Engineering	620	25.297***	6.543	3.866	0	0.025
Industrial Metals and Mining	21	18.668	40.939	0.456	0.657	-
Industrial Transportation	58	-226.442	122.760	-1.845	0.071	-
Leisure Goods	140	2.964	24.762	0.120	0.905	-
Life Insurance	39	68.622	51.559	1.331	0.193	-
Media	633	31.11*	13.909	2.237	0.026	0.008
Mining	10	109.515	710.095	0.154	0.885	-
Mobile Telecommunications	131	34.229	24.723	1.384	0.169	-
Nonlife Insurance	65	12.644	16.942	0.746	0.459	-
Personal Goods	62	4.665	24.510	0.190	0.850	-
Pharmaceuticals and Biotechnolog	273	5.778	19.299	0.299	0.765	-
Real Estate Investment and Servi	347	38.84***	10.837	3.584	0	0.038
Real Estate Investment Trusts	151	15.737	19.832	0.794	0.429	-
Software and Computer Services	1,382	14.611*	5.837	2.503	0.012	0.005
Support Services	436	62.333***	14.229	4.381	0	0.045
Technology Hardware and Equipmen	434	28.64	16.475	1.738	0.083	-
Travel and Leisure	139	28.052*	12.675	2.213	0.029	0.038

***p < 0.001; **p < 0.01; *p < 0.05

Table 6: Coefficients of linear regressions of sentiment on daily return for sample split by company sector with control variables price-to-book and year.

	Model 11	**Model 12**	**Model 13**
Intercept	1.987***	2.950	8.391
	(0.214)	(3.489)	(19.436)
Positive Ratio	0.482	0.252	1.561
	(4.488)	(4.597)	(4.599)
Negative Ratio	-31.951***	-33.359***	-32.867***
	(3.928)	(4.016)	(4.028)
Control Variables	No	Yes	Yes
Interaction Terms	No	No	Yes
R^2	0.008	0.017	0.032
Adj. R^2	0.008	0.012	0.022
Cohens f^2 (Negative Ratio)	0.008	0.009	0.009
Num. obs.	7876	7876	7876

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Regression output of robustness check 4 after splitting up net optimism score into positive and negative ratio.

For our third robustness check, we split our dataset by company sector and perform a regression analysis for each subsample with price-to-book value and year as the remaining control variables. The size of the underlying subsamples ranges from 5 to 1382 observations. The results for the sentiment coefficient are displayed in Table 6. Of those 34 coefficients, only eight are significant on the 5 % level. The corresponding effect sizes vary considerably, with the lowest of them being 0.005 for "Software and Computer Services" and the highest being 0.103 for "General Retailers". In contrast, we do not find a significant effect of sentiment on the daily return in 26 company sectors.

As a fourth robustness check, we split up the net optimism factor into the individual ratios of positive and negative words. Subsequently, the main analysis is repeated with these ratios as explanatory variables. From the results, presented in Table 7, it can be observed that the coefficients for the ratio of negative to total words are negative and significant on the 0.1 % level. The estimation of the effect size for the negative ratio in Model 13 is 0.009. The coefficient for the positive ratio is not significant in any of the three models.

6. Discussion

Based on the results we can only partially confirm our hypothesis that the sentiment of ad hoc announcements positively impacts the corresponding daily return. Support for the hypothesis is provided by the main analysis in which we find a significant positive coefficient for sentiment. This result appears to be robust after removing outliers as well as after removing sparse terms. From the results of robustness check 4 we can observe a negative and highly significant effect of the negative ratio, but no significant effect of the positive ratio. This finding provides some support for the existence of a sentiment effect, but it also indicates that this effect is mainly driven by the frequency of negative words used in ad hoc announcements. While this does not contradict the hypothesis, it still suggests a different underlying framework of sentiment. Additionally, robustness check 3 reveals that for 26 of the 34 company sectors no significant coefficient is found. While for some sectors this might be due to the small corresponding sample size, this argument does not hold for all cases. In the main analysis, we found a very low effect size of 0.006 for 7838 observations. It is possible that the significance of the effect is in part due to the large sample size. For our hypothesis, this could mean that we cannot assume a generalizable effect of sentiment on return, but that this effect is only present in certain sectors. Another explanation could be that in some sectors the applied methodology is not able to correctly capture the sentiment of the ad hoc announcements.

The fact that no significant effect for the positive ratio can be found might stem from the identification of the positive words using the LM dictionary. As noted in Chapter 3, the positive word list is substantially smaller than the negative word list and fewer of the positive words are present in the document-term matrix. Additionally, the positive word list might feature terms that are not perceived as especially positive in ad hoc announcements. Here, future research could aim to find a dictionary that is able to identify positive words within ad hoc announcements that have a substantial effect on the daily return. Generating a new dictionary using a machine learning approach might also yield promising results. If no improvements can be obtained, a possible explanation could be that positive words are overused within ad hoc announcements and even negative news is positively framed. Hence, it would be the case that the net optimism score that weights positive and negative terms equally, is not an ideal measurement of sentiment.

It is also important to note that all of our findings depend on the included control variables, which might not be sufficient. The fact that the models can only explain a small portion of the overall variance of the daily return, suggests that other relevant variables might be missing from the model specifications. If sentiment is correlated with one or more of those variables, the resulting confounding could lead to biased coefficient estimates for sentiment. Thus, it is possible that in a more sophisticated model the effect would not be found to be significant. This could be tested by adding further control variables. Suitable controls could include well-established factors like size and momentum as well as the beta coefficient measuring the market risk.

Another essential factor that might be missing from the model specification is expectations. This is because price reactions occur only if events or announcements contradict investor expectations and thus have not been priced in. In future work, this could be incorporated in a model by relating the sentiment score with the investor expectations. Especially challenging in this context is the operationalization and measurement of this latent variable. Possible starting points for that kind of proxy could be the put-call-ratio, analyst ratings or social media sentiment prior to the announcement.

7. Conclusion

Ad hoc announcements contain information in the form of text that is highly relevant for the valuation of public companies. One common method to extract this valuable information is to focus on the prevalent sentiment within the announcement. Adequately quantifying this concept, however, is challenging. In this paper, we analyze the sentiment of ad hoc announcements by applying a dictionary-based approach and calculating the net optimism score. In order to examine, if this methodology is able to extract the relevant information, we investigate the effect of the sentiment score on the daily return of the corresponding stock. For this purpose, we conduct several regression analyses where we gradually add control variables. Our initial finding is that the sentiment has a highly significant, positive effect on the daily return. This remains true even when all available control variables are included in the model and after considering the impact of outliers and sparse words. However, all results based on the full sample indicate that the overall effect is very small. Based on the results of robustness checks 3 and 4, we argue that the initial finding of a general effect of sentiment measured by the net optimism score might be misleading. This is because the additional findings indicate that the conducted sentiment analysis is only able to extract the relevant information for some sectors and mainly by identifying negative content. One reason for this could be that the positive words are not properly identified by the word list of the LM Dictionary. However, it is also possible that relevant factors like expectations are missing from the model specification or that sentiment has indeed no effect on the return in some company sectors. This opens up exciting research opportunities to gain a better understanding of the underlying relationships. To this end, we provide possible starting points on how the research methodology could be further developed to achieve this.

References

Das, S. R. 2014. "Text and Context: Language Analytics in Finance," *Foundations and Trends® in Finance* (8:3), pp. 145-261 (doi: 10.1561/0500000045).

Demers, E. A., and Vega, C. 2010. "Soft Information in Earnings Announcements: News or Noise?" *SSRN Electronic Journal*, pp. 1-56 (doi: 10.2139/ssrn.1153450).

Feinerer, I., and Hornik, K. 2020. *Text Mining Package*: *R package version 0.7-8.* https://CRAN.R-project.org/package=tm.

Feuerriegel, S., and Neumann, D. 2013. "News or Noise? How News Drives Commodity Prices," *Thirty Fourth International Conference on Information Systems, Milan*, pp. 1-20.

Feuerriegel, S., and Proellochs, N. 2019. *SentimentAnalysis: Dictionary-based sentiment analysis*: *R package version 1.3-3*. https://CRAN.R-project.org/package=SentimentAnalysis.

Frazier, K. B., Ingram, R. W., and Tennyson, B. M. 1984. "A Methodology for the Analysis of Narrative Accounting Disclosures," *Journal of Accounting Research* (22:1), pp. 318-331 (doi: 10.2307/2490713).

Gupta, A., Dengre, V., Kheruwala, H. A., and Shah, M. 2020. "Comprehensive review of text-mining applications in finance," *Financial Innovation* (6:1), pp. 1-25 (doi: 10.1186/s40854-020-00205-1).

Hagenau, M., Liebmann, M., and Neumann, D. 2013. "Automated news reading: Stock price prediction based on financial news using context-capturing features," *Decision Support Systems* (55:3), pp. 685-697 (doi: 10.1016/j.dss.2013.02.006).

Lewis, C., and Young, S. 2019. "Fad or future? Automated analysis of financial text and its implications for corporate reporting," *Accounting and Business Research* (49:5), pp. 587-615 (doi: 10.1080/00014788.2019.1611730).

Loughran, T., and McDonald, B. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance* (66:1), pp. 35-65 (doi: 10.1111/j.1540-6261.2010.01625.x).

Schumaker, R. P., Zhang, Y., Huang, C.-N., and Chen, H. 2012. "Evaluating sentiment in financial news articles," *Decision Support Systems* (53:3), pp. 458-464 (doi: 10.1016/j.dss.2012.03.001).

Affidavit

We hereby declare that we are the sole authors of this term paper and that we did not use any other aids or resources than the ones stated. Those parts of the paper that were taken from other works, either as quote or paraphrase, are marked by respective statements of sources. Furthermore, we declare that this term paper has not been handed in for a different academic assessment by us or another person. We agree with our term paper being screened using anti-plagiarism software.

22.02.2021