Do Company Related Newspaper Articles Have an Impact on the Share Prices? The Case of the German Market

Jyoti Gupta[1] & Benjamin Graubner[1]

[1] ESCP Europe, Paris, France

Correspondence: Jyoti gupta, ESCP Europe, Paris, France. E-mail: gupta@escpeurope.eu

Abstract

The paper looks at the impact of information on stock prices within the context of the German Market. Using data set from the Thomson Reuters, a new platform using a self-written Java Program, between the time period of 27 August and 29 September 2013, we analysed the impact of information on stock prices in the German Market. We developed an Information Based Return Model (IBRM) to analyse how information drive stock prices. We counted certain words within newspaper articles to understand their meaning. We analyse the impact of those word-clusters on different trading intervals. Our Information Based return Model shows that stock prices anticipate news from the non-trading time within the first minute of trading. We also analysed the time drifts between news release and personal reception. Our results show that the German Market anticipates new information as effectively as the US Market.

Keywords: D53, D82, G14, G15, G19

1. Introduction

Since Fama's famous statement, "a market in which prices always 'fully reflects' available information is called 'efficient'" Fama (1970). The discussion of exactly what we mean by information level has been summarized by Sewell (2011). An increase in the level of information available to investors implies a change in the valuation process of an asset. Especially extreme events like natural catastrophes or systematic shocks imply a high volatility of a stock after the event occurs; Basse (2013), Aharony and Swary (1980), Roll (1984). This shows the adjustment of the stock price based on the new information. For the past decades, researches try to explain this stock price behaviour ex post. In this paper we conduct an event based study using data from the German stock Market.

Roll (1988) first carried out an in depth-analysis, focused on the stock price movement based on news documents. His market model regression to explain daily stock price movement based on newspaper releases can be considered as the basis for this research. Several researches followed, which led to an evolution in the field of news based event studies. Antweiler and Frank (2005) argued and proved that the correlation between stock price and newspaper articles can also be more specifically evaluated when the contents of the article itself is analysed. Tetlock (2007) measured the impact of corporate specific news-events on stocks analysed. The latest and most sophisticated research had been done by Feldman et al. (2013). By analysing and clustering news documents they were able to explain almost 50% of the daily stock price movement of US stocks throughout the past 10 years. Although those results seem very promising, one major criticism throughout the decades is highly legitimate: To evaluate the nature of the underlying information from a news document, one must be able to write an algorithm that fully covers wording and text-style analyse. It is questionable iby further understanding the news document the explanatory power will increase. Loughran and McDonald (2011) discussed this issue in their research paper and constructed a dictionary that contains different words and their implication on investor behaviour, in order to be more precise than the previous textual analyses have been. Still the problems of text-style analyse and information deviation remains. Additionally researchers have only focused on the impact of stock prices on the daily or monthly level.

Based on the above mentioned research papers and other researches (Cutler, Poterba, & Summers, 1989), Davis, Piger, and Sedor (2006) one can generalise that an event of any kind injects new information in the system. The information about this event can be either public or private. Private information is by definition not accessible to the public but can be transformed. Newspapers are primary agents within this transformation process. Berry and

Howe (1994) Frazzini (2006), Demers and Vega (2008) and Groß-Klußmann and Hautsch (2011) closely observed this transformation and found an information-drift throughout the process. Information needs time to travel through different channels to be transformed and received.

We develop an Information Based Return Model based on the arrival of new information in the system. Our model explains the price building process based on the information available to the public. On these bases we conduct an event study on newspaper articles with the implications on the firm specific stock prices to further understand the above mentioned drift and the implication of information of article on a stock. Our research differs from these major papers by focusing on the intraday stock price movement. Furthermore we include the recommendations of Loughran and McDonalds (2011) concerning the textual analyse of the news documents. We focus on the major companies within the German stock market for three reasons: firstly, we observe that the German market has not been a focus for major researches on information based stock price movement so far. Secondly, we can compare our results with previous researches done on the US market and determine if those markets behave in the same way. Thirdly, pervious research (Mandelker, 1974; Roll, 1988; Antweiler & Frank, 2004) showed a correlation between size of a company and the amount of articles published about it.

Another dimension for our research beside the stock returns is the news sources we include in the research. It cannot be a German newspaper agency because it might not be as influential as an international one and the distribution of the information would not be global. Furthermore, the information needs to be clearly identifiable for the reader and the computer program we code in order to download and process all information. This analysis leads to the observation of different levels of information. To distinguish between relevant events for the change of level of information within the German stock market, we identified three major types of news: macro environmental, industry and company-specific. Throughout the research we discuss the impact of those layers on the stock price and how we determine their effect on the price buying process.

Finally, our research needs to identify the tone of a newspaper document to conclude if there is a correlation between the natures of the new event and the way the stock prices moves. Aligned with the research of Feldman et al. (2013) and Roll (1988), our regression model is founded on the basis of an asset pricing framework. Therefore, our research evaluates the influence of press releases on the intraday and daily trading level of the German stock market and how different types of news documents influence the stock price.

2. The Information Based Return Model – Testing Hypothesis

In the following section we define our hypotheses to build the foundation for our testing model. The past decades of research have shown the link between information and asset pricing (cf Basse, 2013; Feldman, Govindaraj, Livnat, & Segal, 2010; Capelle-Blancard & Laguna, 2010; Antweile & Frank, 2005; Fama, 1969). This general doctrine is the basis for our "Information Based Return Model" (in short IBRM). Therefore we need to define our first hypothesis as follows:

H_I: A change in the level of information available to the public implies a change in the asset price

To test this assumption we need to further investigate the information variable in this hypothesis. According to Fama (1970) "available information" is the level of information at point *t* while *t+1* includes all new information in the system between those periods until now: *T*. This change of information between those two periods is therefore somehow identifiable and can only lead to an increase of information level. In this context information is the sum of all data, thoughts, theories and whatever content influences the price building process of an individual. Aligning this discussion with respect to the previous research to our model we define the total information level as:

$$I_T = \sum_t^T \varphi_t + \sum_t^T \chi_t \tag{1}$$

I_T is the total level of information within the system to the time t;

φ_t is the historical public information to the time t;

χ_t is the private information to the time t;

T   is the time period.

As private information is by definition not available to the public we can denote $\chi_t = 0$. This leaves us with only public data. The transformation from private to public is done by newspapers or other types of exposure (for further discussion on this topic we refer to Maier 2002) for example financial statements etc. Within this information set we define, in line with Roll (1988), the three layers of information: macro, industry and company-specific-information. Adding those components to our model we define the information set more precisely as:

$$\text{I}_T = \sum_{t=0}^{T} macro_t + industry_t + company_t \qquad \forall: \chi_t = 0 \tag{2}$$

I_T is the total level of information within the system to the time t;

T   is the time period.

with

$$\sum_{t}^{T} \varphi_t = \sum_{t=0}^{T} macro_t + industry_t + company_t \tag{3}$$

φ_t is the historical public information to the time t;

T   is the time period.

Note the time interval from period 0 until *T*. Due to this change we can say that in our theory all historical data since time 0 is included. As a conclusion those three layers are having an impact on the pricing of an asset *i*:

$$p_{i,T} = f(\text{I}_T) + \varepsilon \tag{4}$$

$p_{i,T}$ is the impact on the pricing of an asset *i* to the time t;

I_T is the total level of information within the system to the time t;

ε is the error term representing all variables;

T   is the time period.

This error term ε represents all other variables. Those could be the personal behavioural aspects during a purchasing process, the liquidity of an individual etc. One can see that this error term is not quite correct because it itself inherits private information. This information had been excluded in our argumentation $\forall: \chi_t = 0$ on the information level equation but to be precise we need to include this term in order to fully cover all sources of the price building process.

Furthermore research has shown (for example Demers, Mitchell, & Mulherin, 1994; Li & Bloomflied, 2008; Frazzini, 2006; Tetlock et al., 2008b) that information needs to travel through the system in order to be anticipated.

Information Drift

Vega (2008) indicated that the timeframes of news anticipation need to be further investigated. Following our first hypothesis we can specify it with the following hypothesis:

H_{II}: A change in the level of information available to the public needs time to travel through the system

Therefore we have to include a variable that indicates this information anticipation. In our model we will define different stages of the drift variable in line with the research done by Demers and Vega (2008) in order to evaluate how fast information travels through the system. We therefore determine between 1-, 5-, 15-, 30-, 60-minute and 1-day drift intervals. Those cases can be included into the testing as follows:

$$\int_{n=1}^{N} p_{i,T} = f(\text{I}_T) + \varepsilon \tag{5}$$

$p_{i,T}$ is the price function of an asset *i* until time *T*;

I_T is the total level of information within the system to the time t;

ε is the error term representing all variables;

N   is the different intervals depending on the period of observation;

T   is the time period.

2.1 Newspaper Content, Tone and Influence

The recent study of Feldman et al. (2013) showed a strong increase of R2 when one is able to understand the origin of the underlying information. To put this into our research framework we want to analyse, whether the tone of the underlying newspaper document influences the stock price:

H_{III}: Newspaper articles have a certain tone and influence the stock price accordingly

Using this approach, we can infer if this news arrival has a negative or positive influence on the stock price. But not only a negative or a positive outcome is critical to determine but also if an article is written with strong or weak words or has an uncertain meaning. As for our model all those categories can influence the price building process. Hence, we need to deepen our understanding of the underlying macro-, industry- and company-specific-information.

We will examine the wording impact of those levels by leveraging equation 2 as given above.

$$I_T = \sum_{t=0}^{T} macro_t + industry_t + company_t \quad \forall: \chi_t = 0 \tag{2}$$

I_T is the total level of information within the system to the time t;

T   is the time period.

where

$$macro_t, industry_t, company_t = P_{positive} - N_{negative} + U_{uncertain} + S_{strong} - W_{weak} \tag{6}$$

$P_{positive}$ is the relative number of positive words to the length of an article;

$N_{negative}$ is the relative number of negative words to the length of an article;

$U_{uncertain}$ is the relative number of uncertain words to the length of an article;

S_{strong} is the relative number of strong words to the length of an article;

W_{weak} is the relative number of weak words to the length of an article.

and

$$P = (\sum p)/Q \tag{7}$$

P   is the relative number of positive words to the length of an article;

p   is the number of positive words within an article in period t;

Q   is the total sum of words within all articles in period t.

$$N = (\sum n)/Q \tag{8}$$

N   is the relative number of negative words to the length of an article;

n   is the number of negative words within an article in period t;

Q   is the total sum of words within all articles in period t.

$$U = (\sum u)/Q \tag{9}$$

U   is the relative number of uncertain words to the length of an article;

u   is the number of uncertain words within an article in period t;

Q   is the total sum of words within all articles in period t.

$$S = (\sum s)/Q \tag{10}$$

S   is the relative number of strong words to the length of an article;

s   is the number of strong words within an article in period t;

Q   is the total sum of words within all articles in period t.

$$W = (\sum w)/Q \tag{11}$$

W   is the relative number of weak words to the length of an article;

w   is the number of weak words within an article in period t;

Q   is the total sum of words within all articles in period t.

where

$$Q \supseteq (P, N, U, S, W) \tag{12}$$

Q   is the total sum of words within all articles in period t;

P   is the relative number of positive words to the length of an article;

N is the relative number of negative words to the length of an article;

U   is the relative number of uncertain words to the length of an article;

S   is the relative number of strong words to the length of an article;

W   is the relative number of weak words to the length of an article.

This hypothesis implies a major condition. We defined information of the specific level (macro-, industry- and company) being the underlying basis for our I_T. Therefore our information equals the content of newspaper documents. This is logical since the nature of a newspaper article is to distribute specific information through the system and convert private to public data. Therefore we can imply the above described setup.

3. The Information Based Return Model – Model Construction

In this section we construct our final IBRM which is the basis for the regression analyse. First we define the price function of an asset *i* being the sum of all underlying information until time *T*:

$$p_{i,T} = f(I_T) + \varepsilon \tag{5}$$

$p_{i,T}$ is the price function of an asset *i* until time *T*;

I_T is the total level of information within the system to the time t;

ε is the error term representing all variables;

T   is the time period.

As we continue, we derive that this information is based on three dimensions:

$$I_T = \sum_{t=0}^{T} macro_t + industry_t + company_t \quad \forall: \chi_t = 0 \tag{2}$$

I_T is the total level of information within the system to the time t;

T   is the time period.

with the underlying source of newspaper articles. The articles transmits a certain message that is positive, negative, uncertain, strong or weak towards the reader as per the information dimensions ton

$$m_t, i_t, c_t = P_{positive} - N_{negative} + U_{uncertain} + S_{strong} - W_{weak} \tag{13}$$

m is macro;

i is industry;

c   is company;

t is the time period;

$P_{positive}$ is the relative number of positive words to the length of an article;

$N_{negative}$ is the relative number of negative words to the length of an article;

$U_{uncertain}$ is the relative number of uncertain words to the length of an article;

S_{strong} is the relative number of strong words to the length of an article;

W_{weak} is the relative number of weak words to the length of an article.

By inserting equation 13 into 2 we obtain the equation on the emphasis of the information dimensions tone

$$I_T = \sum_{t=0}^{T} (P_{m,t} - N_{m,t} + U_{m,t} + S_{m,t} - W_{m,t}) + \sum_{t=0}^{T} (P_{i,t} - N_{i,t} + U_{i,t} + S_{i,t} - W_{i,t}) + \sum_{t=0}^{T} (P_{c,t} - N_{c,t} + U_{c,t} + S_{c,t} - W_{c,t}) \quad \forall: \chi_t = 0 \tag{14}$$

I_T is the total level of information within the system to the time t;

T   is the time period;

P   is the relative number of positive words to the length of an article;

N is the relative number of negative words to the length of an article;

U   is the relative number of uncertain words to the length of an article;

S   is the relative number of strong words to the length of an article;

W   is the relative number of weak words to the length of an article;

m is macro;

i is industry;

c   is company.

All three layers have the same unique construction of words. A specific word is therefore made of underlying information towards a certain layer of information. To distinguish the target layer of the information, we allocate certain dummy variable to cover this issue. Our program to download and analyse the newspaper articles can distinguish if the name of the company is in the title. If this is the case the article is related to the company level. If not it is related to the industry level. Therefore we define one dummy variable i_t to be either zero if it the article is company related or one if its related to the industry. The macro level will be reflected by the index of the respective company. Since we are focusing on the DAX we will use its return to describe the macro level of information. By inserting the information levels into the return function we conclude with the equation of

substituted information based pricing model written out

$$p_{i,T} = \sum_{t=0}^{T} m_t + \sum_{t=0}^{T} i_t + \sum_{t=0}^{T} P_t - \sum_{t=0}^{T} N_t + \sum_{t=0}^{T} U_t + \sum_{t=0}^{T} S_t - \sum_{t=0}^{T} W_t + \varepsilon \qquad (15)$$

$p_{i,T}$ is the price function of an asset i until time T;

T   is the time period;

m is macro;

i is industry;

P   is the relative number of positive words to the length of an article;

N is the relative number of negative words to the length of an article;

U   is the relative number of uncertain words to the length of an article;

S   is the relative number of strong words to the length of an article;

W   is the relative number of weak words to the length of an article;

ε is the error term representing all variables.

And since we want to identify the return of an asset i at time T we use the delta notation between t and t-1 leveraging the substituted Information based return model

$$r_{i,t} = m_{i,\Delta t} + i_{i,\Delta t} + P_{i,\Delta t} - N_{i,\Delta t} + U_{i,\Delta t} + S_{i,\Delta t} - W_{i,\Delta t} + \varepsilon \qquad (16)$$

$r_{i,t}$ is the return of an asset i at time T;

T   is the time period;

m is macro;

i is industry;

P   is the relative number of positive words to the length of an article;

N is the relative number of negative words to the length of an article;

U   is the relative number of uncertain words to the length of an article;

S   is the relative number of strong words to the length of an article;

W   is the relative number of weak words to the length of an article;

ε is the error term representing all variables.

As a next step we derive an APT model in the following setting referred to as the information based return model - basic

$$E(r_{i,t}) = \beta_1 m_{i,\Delta t} + \beta_2 i_{i,\Delta t} + \beta_3 P_{i,\Delta t} - \beta_4 N_{i,\Delta t} + \beta_5 U_{i,\Delta t} + \beta_6 S_{i,\Delta t} - \beta_7 W_{i,\Delta t} \qquad (17)$$

$r_{i,t}$ is the return of an asset i at time T;

T   is the time period;

m is macro;

i is industry;

P   is the relative number of positive words to the length of an article;

N is the relative number of negative words to the length of an article;

U   is the relative number of uncertain words to the length of an article;

S   is the relative number of strong words to the length of an article;

W   is the relative number of weak words to the length of an article.

For testing reasons we are also including the length l of the article and a dummy variable e to determine if there is a news event or not. This leaves us with our Information Based Return Model (IBRM):

$$E(r_{i,t}) = \beta_1 m_{\Delta t} + \beta_2 i_{\Delta t} + \beta_3 P_{\Delta t} - \beta_4 N_{\Delta t} + \beta_5 U_{\Delta t} + \beta_6 S_{\Delta t} - \beta_7 W_{\Delta t} + \beta_8 e_{\Delta t} + \beta_9 l_{\Delta t} \qquad (18)$$

$r_{i,t}$ is the return of an asset i at time T;

T   is the time period;

m is macro;

i is industry;

P   is the relative number of positive words to the length of an article;

N is the relative number of negative words to the length of an article;

U   is the relative number of uncertain words to the length of an article;

S   is the relative number of strong words to the length of an article;

W   is the relative number of weak words to the length of an article;

e is a dummy variable;

l is the length of the article.

The testing variables are: macro-, industry/company-, positive-, negative-, uncertain-, strong- and weak news, as well as event: yes or no, and the length of the news document. We identify the following results to be the indicators to verify our hypotheses and therefore our results:

▪ H_I: The impact of information on the pricing of an asset, can be confirmed if the R2 is significant..

▪ H_{II}: The information drift can be confirmed if the R2 changes throughout the intervals and increases over time.

▪ H_{III}: The newspaper tone influence can be confirmed if all textual analysis variables of the IBRM explain the stock movement as expected.

4. Data and Regression

In this section we evaluate our dataset und describe the regression methodology used. Since information is not only limited to the German news reader but should include internationally highly relevant and frequently visited news sources. Therefore we consulted the Thomson Reuters news platform. Based on a self-written Java-program we were able to download all newspaper articles for the time period between 27th of August and 29th of November 2013. Furthermore our program assessed all newspaper articles based on the criteria we needed to conduct our IBRM (the name of the company, the publishing date and time of the article, the title, the content of the article itself, the URL, to check if the name of the company is mentioned in the title and number of words).

The German Stock data was downloaded from Thomson Reuters for all 30 companies within the German DAX. The index represents the 30 biggest listed companies in Germany. We choose the 10 most mentioned companies based on the newspaper articles we could download for two reasons: First the high visibility of those sample companies can reflect better the international relation towards the population. And secondly since we only extract one news source we need to cover as much news events as possible related to one stock to cover all the news events that are theoretically happening every second.

The sampling period is between the 28th of August 2013 and the 28th of November 2013. The trading day in Germany starts at 9:00 am and ends at 5:30 pm. We cover all price and volume ticks on a minute basis. We do not include pre and after trading prices in our data set, because those prices are adjusted within the first minute of the trading day. Hence we noticed the highest volatility to be around 9:00 am.

The Table 1 gives the newspaper data and Sample Companies. We can observe several noteworthy facts from the data.

First, concerning the newspaper documents in relation to the stock data we identified (Figure 1, 2 and 3) that the data of our news documents is consistent. The deviation over time and the amount of published data is logical. Therefore we see no problems for this data set to be the basis for our regression model.

Table 1. Newspaper data and sample companies

	Total	Name In Title	Sum Word Count	Average Length
Deutsche Telekom	176	27	548,642	3,117.28
Commerzbank	173	20	479,409	2,771.15
Lufthansa	163	52	466,201	2,860.13
Deutsche Post	158	29	499,027	3,158.40
SAP	156	38	455,548	2,920.18
Volkswagen	110	7	262,479	2,386.17
ThyssenKrupp	87	28	210,630	2,421.03
Daimler AG	86	19	214,336	2,492.28

|			
Deutsche Börse	80	21	238,329	2,979.11
ADIDAS	79	29	199,044	2,519.54
Fresenius	73	17	163,946	2,245.84
Münchner Rück	59	11	133,518	2,263.02
KBSAG	56	0	54,559	974.27
Allianz	51	2	138,657	2,718.76
Deutsche Bank	50	21	103,581	2,071.62
Bayer	47	15	120,514	2,564.13
Hoch Tief	39	12	90,236	2,313.74
Continental	32	16	144,637	4,519.91
Merck	31	16	90,577	2,921.84
Siemens	28	10	58,801	2,100.04
EONSE	18	0	54,421	3,023.39
Infineon	18	9	29,802	1,655.67
Fresenius Med Care	15	3	32,051	2,136.73
Beiersdof	14	5	46,768	3,340.57
Lanxess	14	6	33,198	2,371.29
REW-DEA	13	8	29,374	2,259.54
Henkel AG	9	5	10,934	1,214.89
Linde AG	7	0	16,229	2,318.43
BMW	5	3	10,534	2,106.80
BASF SE	4	0	44,533	11,133.25
Heidelberg Cement	2	2	535	267.50
Total	**1,853.00**	**431.00**	**4,981,050.00**	**2,714.40**
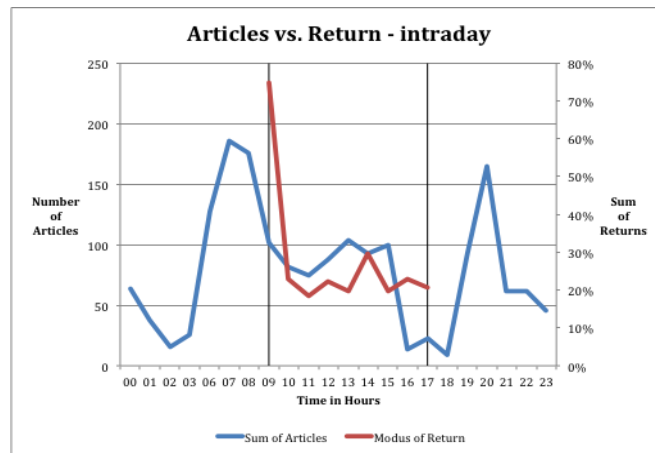
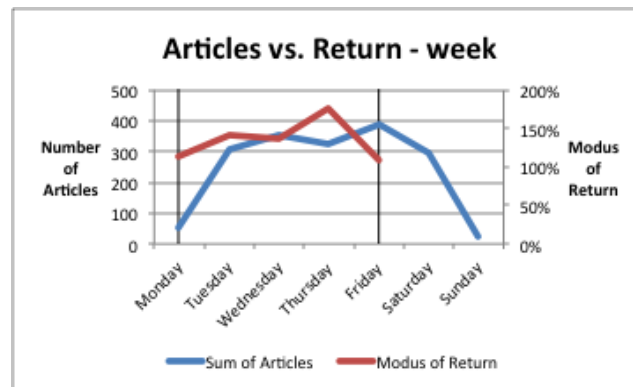Figure 1. Distribution of articles during the day vs. absolute sum of return

Figure 2. Distribution of articles during the day vs. modus of return – week
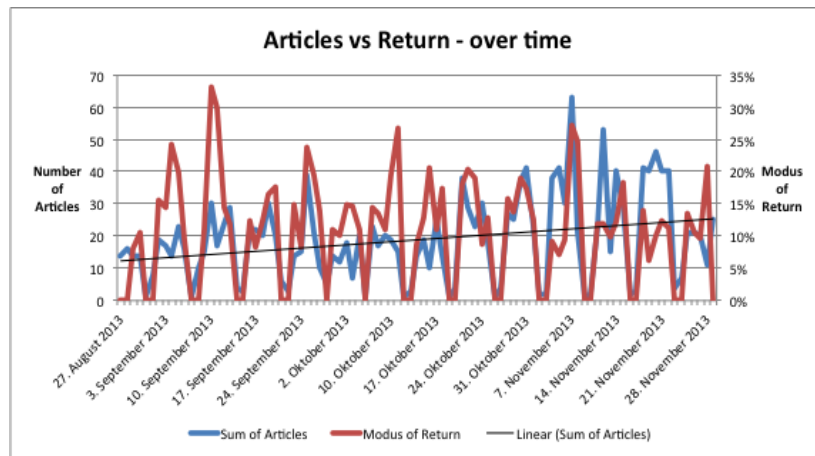
Figure 3. Distribution of articles during the day vs. absolute sum of return - 3-month

Second, Figure 4 and Tables 2, 3 and 4 show the overall stock return throughout the observation period. We observe an overall positive trend in the return. The stock return for the 1-minute interval seems to be normally distributed. The data set for the stock return itself is consistent for the regression but the returns might be related. This is logical since we assume that a change in the macro environment triggers a change in all stocks.
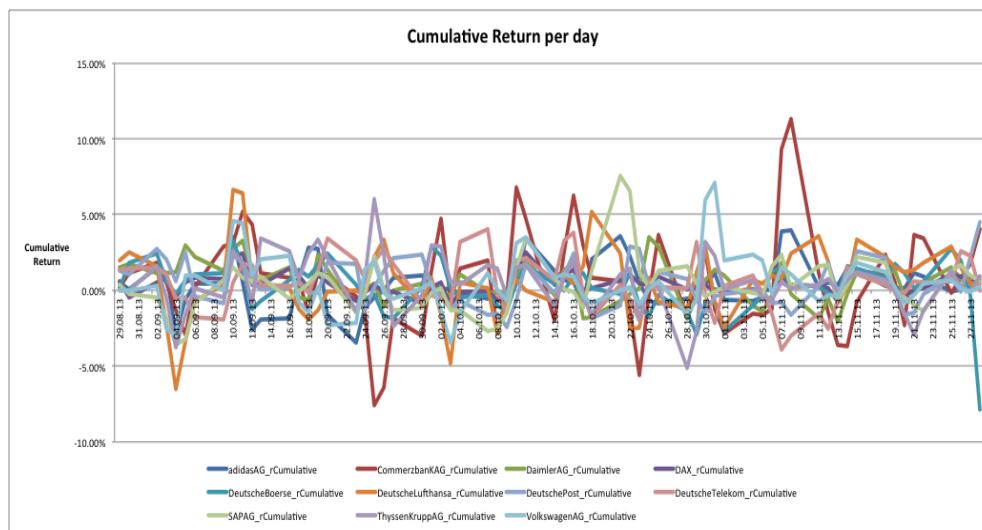

Figure 4. Cumulative returns per day throughout the observation period

Table 1. Return – min, max, mean & standard deviation at 1 minute-interval

Variable	Min (1-Min)	Max (1-Min)	Mean (1-Min)	Std Dev (1-Min)
Adidas~r	-5.7%	1.8%	0.00%	0.06%
Commerbank~r	-2.5%	5.6%	0.00%	0.09%
Daimler~r	-1.6%	2.0%	0.00%	0.06%
DAX~r	-1.2%	1.3%	0.00%	0.03%
Dt. Börse~r	-7.5%	2.0%	0.00%	0.07%
Dt. Lufthansa~r	-4.4%	2.6%	0.00%	0.08%
Dt. Post~r	-1.8%	2.3%	0.00%	0.06%
Dt. Telecom~r	-3.0%	1.8%	0.00%	0.06%
SAP~r	-1.5%	4.4%	0.00%	0.06%
ThyssenKrup~r	-2.6%	3.0%	0.00%	0.07%
Volkswagen~r	-1.0%	3.5%	0.00%	0.06%

Table 2. Return – min, max, mean & standard deviation at 1 day-interval

Variable	Min (1-Day)	Max (1-Day)	Mean (1-day)	Std Dev (1-Day)
Adidas~r	-2.86%	3.90%	0.18%	1.08%
Commerbank~r	-5.73%	9.52%	0.33%	2.33%
Daimler~r	-1.91%	3.43%	0.25%	1.02%
DAX~r	-0.89%	2.11%	0.21%	0.60%
Dt. Börse~r	-7.60%	2.48%	0.06%	1.31%
Dt. Lufthansa~r	-3.51%	5.87%	0.24%	1.56%
Dt. Post~r	-1.97%	3.22%	0.23%	1.11%
Dt. Telecom~r	-2.76%	3.37%	0.22%	1.14%
SAP~r	-2.17%	5.29%	0.17%	1.16%
ThyssenKrup~r	-3.21%	3.81%	0.23%	1.33%
Volkswagen~r	-2.85%	5.38%	0.24%	1.26%

Table 3. Signal words count in articles of sampling items

Companies	Strong	Weak	Negative	Positive	Uncertainty
Adidas	0.05%	0.05%	0.48%	0.34%	0.12%
Commerbank	0.04%	0.06%	0.57%	0.20%	0.14%
Daimler	0.04%	0.04%	0.45%	0.15%	0.08%
Dt. Börse	0.05%	0.03%	0.76%	0.15%	0.07%
Dt. Lufthansa	0.05%	0.06%	0.53%	0.18%	0.13%
Dt. Post	0.04%	0.05%	0.45%	0.17%	0.10%
Dt. Telecom	0.04%	0.06%	0.52%	0.20%	0.11%
SAP	0.04%	0.05%	0.52%	0.20%	0.13%
ThyssenKrup	0.05%	0.06%	0.54%	0.15%	0.12%
Volkswagen	0.05%	0.04%	0.53%	0.21%	0.11%
Average	**0.04%**	**0.05%**	**0.54%**	**0.20%**	**0.11%**

Third (Figure 5 and 6), textual analysis is consistent throughout the time span with a minor absence of negative, uncertain or positive words. Also this data set is consistent for the regression analysis and the model.
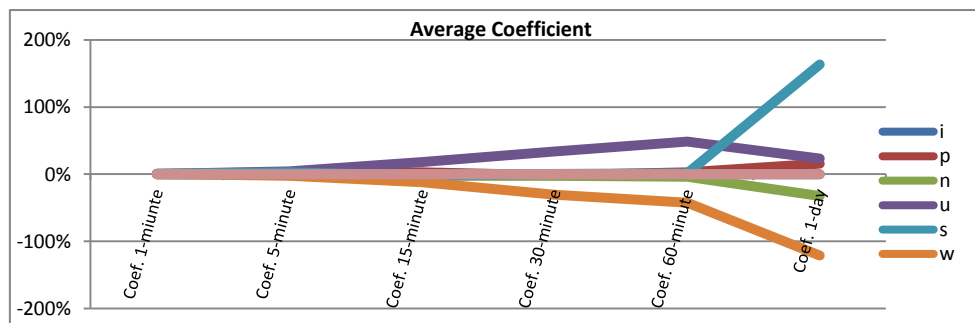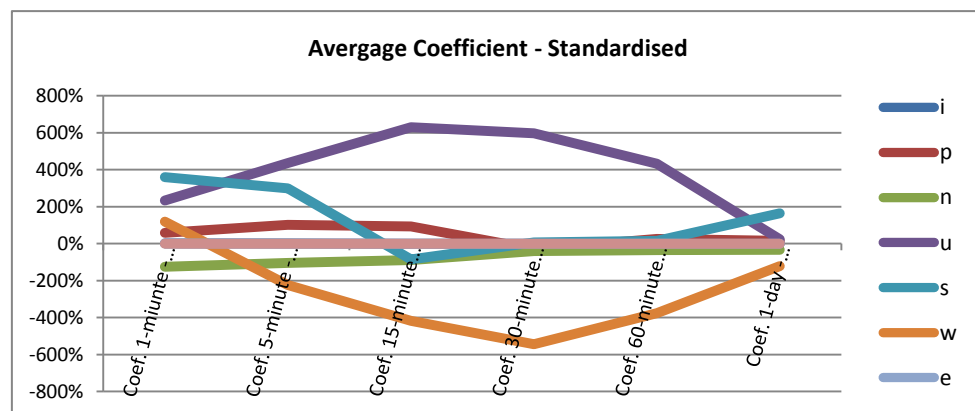

Figure 5. Average coefficient

Figure 6. Average coefficient standardised

As indicated above we can observe the possibility of relating returns: Following our argumentation about the private data that is included into our error terms and the fact that the macro information triggers reactions between the stocks, we need to adjust our regression methodology. Since we identified additional information within the error terms when using the normal regression methodology we decided to use the seemingly unrelated regression (SUR) to fully convert all information within and between the company returns. By doing so we can incorporate the information in the single error term into the SUR estimation. For further discussion on this rather mathematical and statistical topic we refer to Cadave and Henningsen (2012). The correlation matrix based on the Breusch-Pagan-test shows exactly this circumstance:

▪ On a 1-minte interval the correlation is below 11% between the different equations. We notice that the Breusch-Pagan test of independence is highly significantly rejected.

▪ On a 1-day interval the correlation is up to 50% between the different equations. We notice that the Breusch-Pagan test of independence is highly significantly rejected.

Lastly, we need to discuss the data-pooling for the intervals to test the H2 hypothesis. To create those intervals we need to cluster our data set. We analysed the data and noticed that we cannot just sum up all the relative data because this would distort the sample's correctness towards the population. Therefore we came up with the following summarization process and identified the mathematically and logically correct summarising process: For all dummy variables – zero/one- we sum all observation points in the interval to hold zero/one condition. For all other variables we average all observation points above zero in the interval. With this data set we have consistency throughout the sample and it represents the population as accurately as possible.

5. Results and Interpretation

We take a close look on the regression outcome of the variables, the intervals and the R^2 and test the validity of our hypotheses. In this discussion we talk about different dimensions of the model. The first dimension is the sample size of 10 companies. The second dimension is the time-interval. Again the results can be either on the individual or on an average across all intervals. The third dimension is the overall explanatory power - R^2 - of the first and second interval. Again this can also be interpreted on a single or average level. .The results can be interpreted on the basis of a single company or of an average across all companies. To simplify the description we will only focus on the averaged values and indicate noteworthy observations for single stocks.

5.1 Company & Time-Interval Results

For a better interpretation of the first dimensions we average the regressing variables for the specific companies. To determine the second dimension we look at the time-interval development over time (Figure 7).
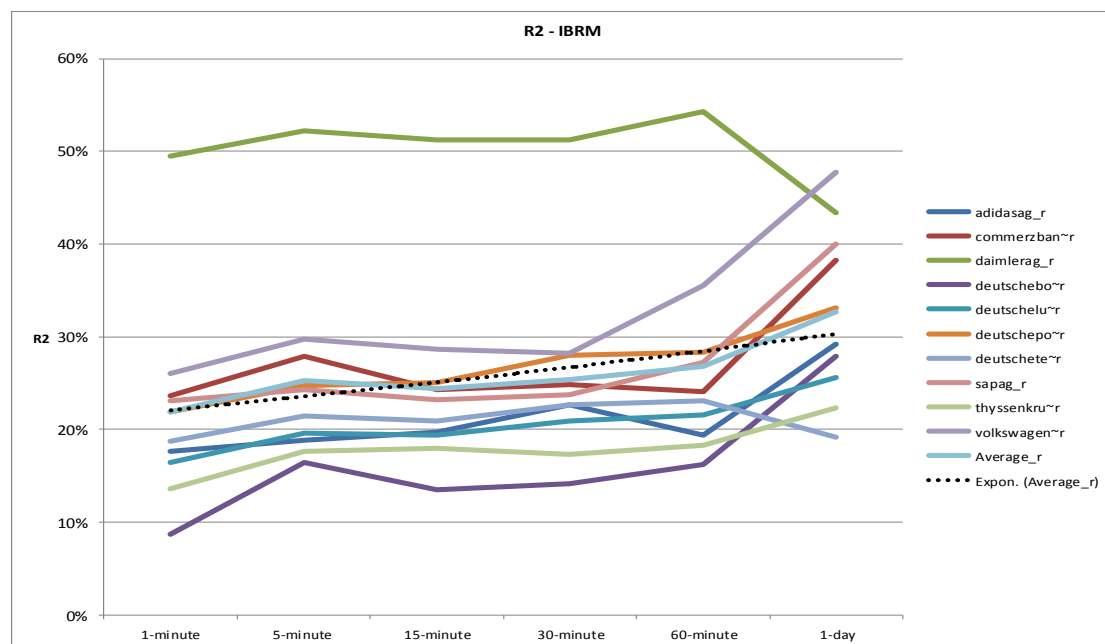

Figure 7. R2 IBRM

One can observe a certain interaction between the uncertain and weak variables. We examine those two variables and noticed a multicollinearity. To find the source of this multicollinearity we examine the data set again. After analysing the dictionary we identified words within both dictionaries. This clearly explains our data set and leads to an interesting observation: The coefficient react almost exactly in the reverse. This shows that both dictionaries identify the correct words to extract the overall information from the newspaper articles and explain the movement correctly. Furthermore, weak words tend to have a negative impact on the stock price themselves uncertain words tend to have a positive impact. Therefore the word clusters are reliable. As a conclusion the uncertain words diminish the positive reaction within the data set. Hence we can argue that the uncertain words would even have a higher impact within the regression. As a result we can cluster those two variables while analysing the regression results. Again the multicollinearity has no impact on our results as a whole but we are able to cluster the uncertain and weak variables. We furthermore performed a test on omitted variables across all dimensions. For this test we used the OLS-Regression in order to test the estimators. As a result we calculated a Prob > F of 73.54%. This rejects the assumption that our model misses relevant variables to our existing sample.

The increase of the coefficient for all variables throughout the time intervals proves the above-mentioned linearity. As for our results we see that the positive, negative and strong variables are increasing only in the daily interval whereas the other variables weak and uncertain are increasing in the intraday intervals. Positive words have a positive influence towards the average stock return, just like negative words have a negative impact on the average stock return. But this can only be observed clearly in the daily interval. If those words are mixed with strong words the influence on the average daily stock return can be significantly higher. Appendix 8 shows the same effect in the first three intervals. Comparing our results with Tetlock (2007) and Li (2008) we draw the same conclusion as them: negative news are not as influential as we might have predicted. The same is applicable for positive news.

The weak and uncertain variables are significantly higher throughout the intervals. In respect to the population, we can derive that weak words imply an immediate negative influence on the average stock return. Uncertain words have an immediate positive influence on the average stock return. As we discussed above the overall market trend during our observation phase was positive. This ties back to our observation: If there is a newspaper document with uncertainty the market remains in a bullish stage. In contrast, weak news implies a bearish movement because individuals do not trust the news as such and react counter-intuitively.

Furthermore the event, length and industry variables do not have an impact on the average stock return at all. For our interpretation we see no influence if there is a news event or not. This could have two reasons. First, either the change in the information level itself is not significant towards the expectation of the market or, secondly, the news event needs to be clearly analysed and our variable simply cannot measure this event. We neither prove nor deny the first argument because, on one hand the specific wording has an influence on the stock return as proven above, but on the other hand there is no clear reaction from the market to this single variable 'change in information level yes or no'. The second argument is probably the most plausible of the two. We could prove that positive words have a positive influence on the average stock price. Furthermore the positive words are part of this news event. Therefore only if the news event is analysed, will it lead to a reaction by the stock return. When taking a look at reality, this is very logical. Only because an individual sees a newspaper article without knowing the content they will not change their investment structure without further information.

We can derive two major explanations from this observation:

First, there are two different clusters of words with different types of impact on the stock return: emotional words and supportive words. Such emotional words would be strong, negative and positive. As we showed above these word groups have a high impact on the daily interval. The second cluster - supportive words - is made of weak and uncertain word groups and has an impact on the intraday basis. Those words also have a higher overall impact on the average stock return.

Second, there are two different trading horizons: intraday vs. outraday. This distinction can be derived from the word clusters. Supportive words tend to trigger intraday trading whereas emotional words have an impact on the daily trading. This could lead to the conclusion that there are two different trading universes or horizons. The first intraday universe is more focused on the non-fundamental stock price evaluation. This can be determined by looking at the supportive wording. These types of words tend more towards general trends than towards the fundamental wording analysis. Thus the second universe would be the 'outraday' (Note 1) trading where fundamental estimations are calculated for the anticipated stock price. The emotional word cluster supports this theory because positive and negative words tend to show additional information about the underlying assumption when it comes to estimating the fundamental development of a company's future growth and profit.

5.2 R2

The results indicated in Appendix 9 show an increase in the explanatory power from somewhere around 20% in the 1-minute interval to around 30% in the 1-day interval. Clearly the overall trend is upwards but the explanatory power within the intervals is minor in the intraday dimension but increases significantly in the outraday dimension.

Comparing those results with Feldman et al. (2013) and Roll (1988) we see the same trend in our regression: the explanatory power increases towards a larger time interval and with the ability to identify the nature of the news event. The dotted line implies this trend and we can see a significant increase in almost all companies from the 60-minute to the 1-day interval. We also see that the increase to an average R2 across all 10 companies is 33,1%. Comparing this number with Roll, Feldman identifies a R2 for his single variable model (CAPM) of 16.28% and for his multi-variable model (APT) of 20.53%. Feldman could increase this result by identifying the nature of the news event to 48%. In line with Roll's (1988) observation that the R2 drops from a monthly to a daily observation for around 10%, we can clearly see that this decrease also continues on the intraday level. Furthermore other researches like Antweiler and Frank (2004) could also only calculate an R2 of 5% for their news document stock return regression model.

Generally speaking, the variables do explain the movement of the stock return well yet the higher the chosen time interval is, the higher the explanatory power of the information level becomes to the public. Consequently we see once again the above-identified intra vs. outraday trading horizon. We can derive from our data that beside our chosen variables there are other variables that explain the movement of stock prices. While the information driven movements tend to be more effectual in the outraday sphere, we cannot draw this conclusion for the intraday stock price movement: the R2 remains on a bases at around 20% and increases sharply in the after the 60-minute interval to somewhere around 31%. Hence, there need to be other variables that might explain intraday stock price movement besides the information level available to the public.

Finally, compared to our results, the results of previous research are slightly higher in compared to our results. We anticipated our R2 being lower than Feldman's due to the not so sophisticated textual analysis we used. They were able to cluster 14 event categories for all news documents and identify the tone of the documents. Their regression model contained 8 of the 14 categories with the characteristic positive vs. negative news and new news event vs. old news event vs. no news event. Finally they regressed the results against the daily stock prices. Clearly their textual analysis software is more sophisticated and more precise than our word count model. Never the less our model explains the movement as precisely as their model. Therefore we can draw the conclusion that news events have an influence to a certain extent but are not the major driver of the stock price.

To further investigate the hypothesis above we determined the influence of the macro economical level. Based on our initial model we adjusted the IBRM and excluded the variable of the underlying index DAX. The adjusted IBRM therefore incorporates only the industry- and company specific information. Figure 8 shows the decrease of the explanatory power towards the information of around 20%. This observation leads to the following conclusions.
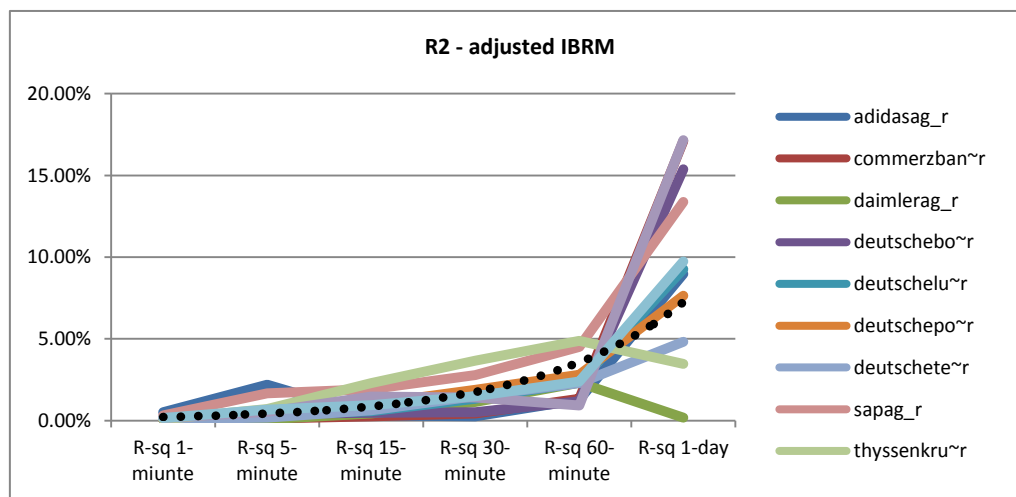

Figure 8. R2 adjusted IBRM

The explanatory power in the 1-minute interval is somewhere around 1% and increases to around 10% in the 1-day interval. Clearly the overall trend is upwards but the explanatory power within the intervals is minor. One explanation for this observation might be the existence of the information drift. This argument is quite logical because an individual has to discover the news, read the article, understand its meaning and then react accordingly. The question remains how long this interval is until news is anticipated within the price building process.

Following up on this observation we carried out further analysis on our data. After plotting the absolute sum of all returns against the number of published articles, we could observe intraday news drift of one hour (Appendix 2). An increase of newspaper articles lead to an increase of the absolute sum of the returns after one hour. Appendix 1 mathematically proves this one hour time drift. Combined, this proves the hypothesis about personal information anticipation. Following the above mentioned argumentation we cannot agree with Groß-Klußmann and Hautsch's (2011) argumentation of news anticipation right after its occurrence due to computer trading. Our results clearly show that we can observe an information drift. The question remains of the length of the drift.

Additionally the increase of the R2 by 10% between the intraday and outraday interval clearly indicates another drift. Frazzini (2006), Garfinkel and Sokobin (2006), Francis et al. (2007) and Tetlock et al. (2008a) could observe this anticipation time frame that is needed for the stock price to in cooperate the new information. We also identify this increased explanatory power throughout the time intervals. This leads to the observation of an systematic time drift of at least 1 day.

6. Implication on Hypotheses

In this last section we conclude if the results imply a negative or positive interpretation towards our hypothesis and therefore if our IBRM is valid.

6.1 Conclude: Information Level and Asset Pricing

We assumed that "*a change in the level of information available to the public implies a change in the asset price*". Our objective towards the regression is to analyse how increase in the explanatory power of our model. Therefore we need to consider the R2.

Our explanatory power may be low but in compared to other research it lies in the acceptable range. The analyse of the IBRM shows two trading horizons and manage to explain the impact of information. Therefore we can conclude that our H1 hypothesis holds to the extent that the German stock market anticipates a change in the level of information within a day or more. Intraday movement cannot be explained by an increase of information based on newspaper articles but rather based on the macro environmental information level. In this discussion one could argue that the information might have been available to the public even before publishing and this would lead to a distortion of our data and regression because the information level might have increased before the newspaper article had been published. As a consequence we would not see a change in the stock return because this change might have happened before the newspaper article was published and after the information was made available to the public. We argued in above that in our terms a newspaper article is the transformation from private information to public information. Therefore every newspaper article increases the public data. Hence even if the information would be available before the event of publication, our model understands the publication of this information as the transformation from private to public data. Clearly if another news agency does this transformation before Reuters does, we have a temporal lack concerning the anticipation of this information. But this ties back to our observation that news documents are anticipated on a daily or more frequent level.

As a conclusion, our hypothesis stating that "*a change in the level of information available to the public implies a change in the asset price*" can only be observed on a daily level and not during the intraday trading.

6.2 Conclude: Information Drift

Our second hypothesis is related to the information drift. We argue that "*a change in the level of information available to the public needs time to travel through the system*".

To test this assumption we need to observe a change in the stock price after a certain amount of time. We could observe this personal drift during our data analyse by plotting the time of released articles against the absolute sum of all returns.

We further more identified a systematically time drift by conducting all intervals - 1-minute, 5-minutes, 15-minutes, 30- minutes, 60-minutes and 1-day - to analyse if the previous hypotheses are valid and if the R2 increases. Looking at our results above, we can infer that the 1- to 60-minute intervals can only prove the

personal drift, though not as precisely as we can prove it for the 1-day interval. We can prove the implication of news documents on the asset price in the 1-day interval. Additionally, only some information do have an influence on the stock price in the intraday horizon.

We can derive that by clustering the information the explanation for why the stock price moves is more accurate. Because we can observe this information drift from the chart and a 1-day interval accuracy within the regression, we can draw an average of this number and conclude that "*a change in the level of information available to the public needs a systematic time of at least 1 day to travel through the system and a personal time of 1 hour to travel through the system*". Our research has proven this hypothesis.

6.3 Conclude: Newspaper Content and Tone

According to our fourth hypothesis, "*Newspaper articles have a certain tone and influence the stock price accordingly*". This is true if our explanatory variables have an impact on the stock return according to their expectation.

As we have shown above, positive and negative words imply a stock movement according to their assumption in the daily level. Again we can observe the split between intraday and outraday stock behaviour. This split implies a reversal of the behaviour of the variables. Furthermore we identified two types of words: emotional (strong, negative and positive) and supportive (weak and uncertain) words. Only the supportive words have an impact on the intraday and outraday level.

Finally we can conclude, *newspaper articles have a certain tone and influence on the stock price accordingly*" but only on the outraday level. Intraday only supportive words have an influence on the stock price.

6.4 General Conclusions

Newspaper articles influence German stock price movement in a minor way during intraday trading but affect the price building process on a daily basis. This research therefore answers the question whether we can explain the stock price movement ex post by analysing information of a newspaper document. We reviewed the literature for information based event studies and developed four testing hypotheses:

- A change in the level of information available to the public implies a change in the asset price.
- A change in the level of information available to the public needs time to travel through the system.
- Newspaper articles have a certain tone and influence the stock price accordingly.

On this basis we created the information based return model on the theoretical APT framework. It is constructed by different intervals: 1-minute, 5-minutes, 15-minutes, 30-minutes, 60-minutes and 1-day with different variables (positive-, negative-, weak-, strong-, uncertain-words, length of the article, name in header of the article and event in the interval) in order to explain the return of a stock. After describing the stock prices, newspaper articles and the method of textual analysis, we identified the appropriate regression model to test our hypotheses.

Our work has some limitations. The 3 months observation phase from the 28th of August 2013 until the 28th of November 2013 is too short in order to have bullish and bearish overall market conditions. Thus, for the time being, our observation is only applicable to a bullish market. Furthermore, the form of textual analysis is not as sophisticated in our case as it was in previous research but it analyses the articles well enough to increase the explanatory power throughout the intervals. After analysing our results based on the above-mentioned limitations, we proved our hypothesis. Further, we identified the following facts for asset pricing based on newspaper information.

Our Information Based Return Model shows that stock prices anticipate news from the non-trading time within the first minute of the trading. We can see that the first hour of trading has the highest absolute sum of returns. The stock price therefore adapts all information that arrived before the official trading start. Earlier research also identified a trading high on a Friday. We could not prove this pattern but identified the highest sum of the absolute return occurring on Thursday for the German stock market.

This leads to the next observation: in order to react to newspaper information we identified and proved two time drifts at a time between news release and personal reception. On average, news travels for at least one day across the system before reaching a person. This person then needs another hour to read, understand and react to the new information.

This drift grants us another remarkable observation: stocks react differently within different time intervals. We discovered an every minute (intraday) and daily stock (outraday) price pattern. In terms of intraday patterns, the stock logically does not move according to newspaper information because of the delayed information drift. The

explanatory power of our model is somewhere around 20% throughout the intraday intervals. Only weak and uncertain words have an influence on the price, whereas the outraday stock price movement has an explanatory power of our IBRM somewhere around 31% and is influenced by all word groups. By eliminating the marco-environmental information layer the explanatory power of newspaper triggered stock price movement decreases by roughly 19%. Therefore intraday wise stock is driven by macro-economic news or the overall market situation. Outraday the stock is driven by fundamental information.

As a future outlook we encourage the scientific community to continue to investigate the impact of specific words within articles on stock prices. Furthermore we see the split between intraday and outraday stock behaviour as a promising field of research.

References

Aharony, J., & Swary, I. (1980). Quarterly Dividend and Earnings Announcements and Stockholders' Returns: An Empirical Analysis. *The Journal of Finance, 35*(1), 1-12. https://doi.org/10.1111/j.1540-6261.1980.tb03466.x

Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance, 59*(3), 1259-1294. https://doi.org/10.1111/j.1540-6261.2004.00662.x

Antweiler, W., & Frank, M. Z. (2005). Do US Stock Markets Typically Overreact to Corporate News Stories?

Basse, A. (2013). Contagion effects in the electric utility industry following the Fukushima nuclear accident. *Applied Economics, 45*(24), 3421-3430. https://doi.org/10.1080/00036846.2012.714072

Berry, T. D., & Howe, K. M. (1994). Public Information Arrival. *The Journal of Finance, 49*(4), 1331-1346. https://doi.org/10.1111/j.1540-6261.1994.tb02456.x

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., & Roll, S. (2013). Wich news move stock prices? A textual analysis.

Cutler, D., Poterba, J., & Summers, L. (1989). What Moves Stock Prices? Moves in stock prices reflect something other than news about fundamental values. *Journal of Portfolio Management, 15*, 4-12. https://doi.org/10.3905/jpm.1989.409212

Davis, A. K., Piger, J. M., & Sedor, L. M. (2012). Beyond the Numbers : Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research, 29*(3), 845-868. https://doi.org/10.1111/j.1911-3846.2011.01130.x

Demers, E. A., & Vega, C. (2008). Soft Information in Earnings Announcements: News or Noise? *INSEAD Working Paper No. 2010/33/AC*.

Fama E., FisherL., Jensen M., & Roll R. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review, 10*(1), 1-21. https://doi.org/10.2307/2525569

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417. https://doi.org/10.2307/2325486

Feldman, S., Govindaraj, J., Livnat, B., & Segal R. (2010) Management's Tone Change, Post Earnings Announcement Drift and Accruals. *Review of Accounting Studies, 15*(4), 915-953. https://doi.org/10.1007/s11142-009-9111-x

Francis, J., Lanfond, R., Olsson, P., & Schipper, K. (2007). Information Uncertainty and Post-Earnings-Announcement-Drift. *Journal of Business Finance & Accounting, 34*(3-4), 403-433. https://doi.org/10.1111/j.1468-5957.2007.02030.x

Frazzini, A. (2006). The Disposition Effect and Underreaction to News. *The Journal of Finance, 61*(4), 2017-2046. https://doi.org/10.1111/j.1540-6261.2006.00896.x

Garfinkel J., & Sokobin J. (2006). Volume, Opinion Divergence, and Returns: A Study of Post–Earnings Announcement Drift. *Journal of Accounting Research, 44(1),* 85-112. https://doi.org/10.1111/j.1475-679X.2006.00193.x

Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance, 18*(2), 321-340. https://doi.org/10.1016/j.jempfin.2010.11.009

Laguna, M. A., & Capelle-Blancard, G. (2010). *How Does the Stock Market Respond to Chemical Disasters?*

Retrieved from http://ideas.repec.org/p/hal/journl/halshs-00637961.html

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*(2-3), 221-247. https://doi.org/10.1016/j.jacceco.2008.02.003

Loughran, T. I. M., & Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis , Dictionaries, and 10-Ks. *The Journal of Finance, LXVI*(1), 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Mandelker, G. (1974). Risk And Return: The Case Of Merging Firms. *Journal of Financial Economics, 1*, 303-335. https://doi.org/10.1016/0304-405X(74)90012-9

Mitchell, M. L., & Mulherin, J. H. (1994). The Impact of Public Information on the Stock Market. *The Journal of Finance, 49*(3), 923-950. https://doi.org/10.1111/j.1540-6261.1994.tb00083.x

Roll, R. (1984). Orange Juice and Weather. *The American Economic Review, 74*(5), 861-880.

Roll, R. (1988). R2. *American Finance Association, 48*(3), 541-566. https://doi.org/10.1111/j.1540-6261.1988.tb04591.x

Tetlock, P. C. (2007). Giving Content to Investor Sentiment : The Role of Media in the Stock Market. *The Journal of Finance, 62*(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-tsechansky, M., & Macskassy, S. (2008a). More than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance, 63*(3), 1437-1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

Tetlock, P. E., & Mitchell, G. (2008b). Calibrating Prejudice in Milliseconds. *Social Psychology Quarterly, 71*(1), 12-16. https://doi.org/10.1177/019027250807100104

Note

Note 1. 'Outraday' is understood here as follows: an investor can either have several transactions throughout the day or only a few trades during the day or the week. This behaviour of non-frequent trading during the day is here determined with the word outraday.

Copyrights