Sentiment spillover effects for US and European companies.

April 24, 2017

Francesco Audrino ¹, Anastasija Tetereva ²

Abstract

The fast-growing literature on the news and social media analysis provide empirical evidence that the financial markets are often driven by information rather than facts. However, the direct effects of sentiments on the returns are of main interest. In this paper, we propose to study the cross-industry influence of the news for a set of US and European stocks. The graphical Granger causality of the news sentiments - excess return networks is estimated by applying the adaptive Lasso procedure. We introduce two characteristics to measure the influence of the news coming from each sector and analyze their dynamics for a period of 10 years ranging from 2005 to 2014. The results obtained provide insight into the news spillover effects among the industries and the importance of sentiments related to certain sectors during periods of financial instability.

Introduction

The influence of the news and social media in politics and economics has grown consistently over the last decades. The availability of real-time online sources and recent developments in machine learning algorithms have made the news relevant for the area of quantitative finance. The news influences the opinions and expectations of investors, which find expression in sentiments. A growing number of agencies are developing news indices, which can potentially help improve trading strategies, as recent research shows that the behavior of market participants may be more highly influenced by the news than by reality.

Some authors have explored the sensitivity of stock returns to stock-related news; see, for example, Fang and Peress (2009), Peress (2014), Akyıldırım et al. (2015), Narayan and Bannigidadmath (2015), Ding et al. (2015) and Luss and d'Aspremont (2015). Akyıldırım et al. (2015) describe the role of firm-specific public announcements on liquidity, price and the volatility of individual stocks. Allen et al. (2015) have analyzed how the performance of the GARCH, GJR and EGARCH can be improved by including sentiment data in the model. Additional, general conclusions have been obtained by Cahan et al. (2009) who have empirically shown that the information coming from the news can be seen as an additional factor in the Fama French factor models. The most recent research by Borovkova et al. (2016) makes an effort to construct a systematic risk indicator based on the news related to the biggest financial companies. They show that the proposed risk measure outperforms the conditional capital shortfall measure of systemic risk (SRISK) by Brownlees and Engle (2015) and the CBOE volatility index (VIX)

¹Chair of Mathematics and Statistics, University of St Gallen, Bodanstrasse 6, 9000 St Gallen, Switzerland, francesco.audrino@unisg.ch

²Chair of Mathematics and Statistics, University of St Gallen, Bodanstrasse 6, 9000 St Gallen, Switzerland, anastasija.tetereva@unisg.ch

by Brenner and Galai (1989) in signalling the periods of financial stress. Most of the above-mentioned studies concentrate, however, on the direct effects of the asset-related news of the price and do not consider spillover effects.

In this paper, we study the news spillover effects based on the news data provided by Thomson and Reuters News Analytics. The daily sentiment indices for 10 US sectors, 10 non-US sectors and 5 countries are considered together with the prices of more than 100 stocks. The analyzed time interval ranges from the January 1, 2005 till December 31, 2014 and covers the global financial crisis, the US debt-ceiling crisis, and the European sovereign debt crisis.

We conduct a rolling window analysis of the cross-industry influence of news sentiments. The influence of news on the stock excess returns is defined by means of the graphical Granger causality, which is estimated by constructing a sparse network. In order to reduce the number of false positive edges of the network, the adaptive Lasso methodology is applied for the estimation of the networks with a related testing procedure introduced in Audrino and Camponovo (2015). Two characteristics describing the relevance and the strength of the news coming from an individual sector or a specific country are suggested, and their dynamic behavior is studied. The news sentiments coming from US and non-US industrial sectors show similar behavior and are highly correlated; for this reason, the spillover effects are analyzed separately for the US and the European companies, and the results are compared.

In this study, we provide strong empirical evidence that the class of stock-relevant sentiments is wider than just the stock-specific announcements and the news coming from the related sector. We show that the returns of the whole industry are driven by the news coming from several sectors. Moreover, we investigate how the importance of the news changes over time, getting stronger just before periods of financial turbulence, which can be seen as an early warning signal for investors.

The paper is organized as follows: The first section features an introduction to the Thomson and Reuters MarketPsych news data, and section 2 discusses how the signal can be extracted from the noisy news. The methodology related to the estimation of the graphical Granger causality by means of sparse regression models is provided in section 3. Full results are presented in section 4. Section 5 contains an empirical illustration of how the predictive power of time series models can be improved by augmenting the conditional mean equation by specific sentiment indices. Finally, we summarize the main contribution of the paper.

1 TRMI construction

The sentiment data used in the current paper are provided by Thomson and Reuters News Analytics and include the MarketPsych indices (TRMI) ranging from 2005 till 2014. In this section, we provide more details on the construction of the MarketPsych index and its characteristics.

In analogy to a variety of approaches in the sentiment literature, the TRMI is constructed using the lexical analysis or the so-called "bag of words" technique: the frequency of a series of predefined words in the text is counted and the quantitative index is extracted by means of sentiment dictionaries such as code libraries by Tim Loughran and Bill McDonald and Harvard General Inquirer. For more details on text analysis as applied to economic news we refer to, for example, Loughran et al. (2009), Loughran and McDonald (2011), Loughran and McDonald (2013) and Loughran and McDonald (2014). In contrast to the most popular approaches, the TRMI is sensitive to grammatical structures and accounts for correlations among the words.

TRMI is a multidimensional index which considers the sentiments beyond positive and negative. Each TRMI index is constructed based on a set of over 4 000 variables (for example, *Ambiguity*, *EarningsUpFuture*, *AccountingBad*, *AccountingGood*) which are generated by the

linguistic machine learning algorithms. In the next step, a numerical value which considers the tense, proximity and many other multipliers is assigned to each variable. Finally, the variables are associated with the assets. To illustrate this procedure, consider the following example from Peterson (2016):

Analysts expect Mattel to report much higher earnings next quarter.

The linguistic analysis of the sentence will be performed in the following steps:

1. The entity "Mattel" will be associated with the ticker *MAT*.
2. Word "earnings" is associated with the variable *Earnings*.
3. Word "expect" assigns future tense to the phrase.
4. Word "higher" is an Up-Word.
5. Word "higher" is multiplied by 2 due to the presence of the word "much."
6. Word "higher" is associated with the word "earnings" due to proximity.

The above-described procedure leads to the score of 2 for the variable *EarningsUpFuture* for the ticker *MAT*.

In the next step, the variables (*PsychVar*) are combined into the TRMI sentiment indexes which are computed as a ratio of the sum of all relevant variables to the absolute values of all TRMI-contributing variables called *Buzz*. In addition, each variable is classified as additive or subtractive. Thus, the index is computed in the following way:

$$TRMI = \frac{\sum_{c \in C(A), v \in V} [I(v, t) \times PsychVar_v(c)]}{Buzz(A)}, \quad (1.1)$$

where $I(v, t) = 1$, if the variable is additive and $I(v, t) = -1$, if the variable is subtractive, $Buzz(A) = \sum_{c \in C(A), v \in V} |PsychVar_v(c)|$, V is the set of all variables *PsychVar*, A denotes a specific asset. $C(A)$ is the set of all constituents of A , i.e. set of all entities which are relevant for the specific asset; for example, Mattel is a constituent of the Nasdaq 100 index. It is worth noting that a single variable can contribute to multiple TRMI; for example, the above-mentioned variable *earningsUpFuture* is a constituent of such TRMI sentiment indexes as *Optimism* and *fundamentalStrength*, and Mattel is a constituent of the Consumer Goods sector and the Nasdaq 100. Additionally, such characteristics as relevance and novelty are taken into account when the variables are summed up in the index. The weights of all constituents are constantly recalculated, giving bigger coefficients to the more influential companies.

The common valence-arousal affective circumplex by Russel (1980) shown in Fig. 1.1 is employed for the classification of emotions and the construction of the sentiment indices. The TRMI captures the arousal dimension in such indices as Fear, Stress, and Urgency and the valence dimension in Optimism, Sentiment, and Trust. The Sentiment values range from -1 to 1.

MarketPsych provides the separate indexes based on the professional news and the social media. The first class is based on such sources as The New York Times, The Wall Street Journal, Financial Times, and many others. The second class includes, for example, Twitter, Yahoo Finance, and Seeking Alpha. As a result, around 2000000 articles are processed each day and classified into TRMI for 32 currencies, 132 countries, 12 agricultural commodities, 22 materials and energy, 8000 global equities including individual stocks, 17 global equity indices, 10 sectors, and 25 industries. The TRMI provided for each asset class is shown in Fig. 1.1. In addition, each asset class includes scores on asset-specific topics, for example, inflation and budget deficit for the countries, production volume for commodities, and price forecast for

Fig. 1.1: TRMI plotted on the affective circumplex (from Peterson (2016), p.25).

currencies. The amount of processed information is constantly growing. The number of daily observations for some assets allows for the analysis of the media's impact on the tick-by-tick level over the past several years.

In the current paper, we make use of the TRMI sentiment index for 10 industries and 5 countries which are listed in Table 1.1 and TRMI sentiment for USA, China, Germany, Italy, and Greece. This analysis considers a daily frequency, as we are interested in the analysis of the global financial crisis which goes back to 2008 when tick-by-tick data were not available for all the asset classes we are interested in.

2 Extracting signal from the news sentiment

As can be seen from Fig. 2.1, the news data are very noisy and cannot be directly used for modelling. Two main approaches for extracting the information from the noise which appeared in the recent sentiment literature could be potentially used.

The first approach is the moving average convergence/divergence oscillator (MACD). It was first developed by Appel (2003) and has been applied to sentiment data by, for example, Peterson (2016), Kirange et al. (2016) and Lugmayr and Gossen (2013). The intuition behind the MACD is that two moving averages are turned into momentum by subtracting the longer moving average from the shorter. As a result, MACD captures both the trend and the momentum. Negative values mean that the shorter moving average diverges further from the longer one; thus, the downside momentum is increasing. From Fig. 2.2 it can be seen that when the 30-day average of the TRMI sentiment for China rises above the 90-day average, the green shading represents the positive expectations about the Shanghai-Hong Kong Connect. Fig. 2.3 shows similar results for the S&P 500 price index and the TRMI sentiment: positive 200-500 MACD of the sentiment is a proxy for a positive trend in the price. Other examples are available in the literature. Becker (2016a) shows that 10-30 MACD of the TRMI Anger sentiment about Starbucks has

Asset Code	Description
MPTRXUSENE/MPTRXENE	US/non-US Energy (ENE)
MPTRXUSMAT/MPTRXMAT	US/non-US Basic Materials (MAT)
MPTRXUSIND/MPTRXIND	US/non-US Industrials (IND)
MPTRXUSYCY/MPTRXYCY	US/non-US Cyclical Consumer Goods and Services (YCY)
MPTRXUSNCY/MPTRXNCY	US/non-US Non-Cyclical Consumer Goods and Services (NCY)
MPTRXUSFIN/MPTRXFIN	US/non-US Financials (FIN)
MPTRXUSHLC/MPTRXHLC	US/non-US Healthcare (HLC)
MPTRXUSTEC/MPTRXTEC	US/non-US Technology (TEC)
MPTRXUSCOM/MPTRXCOM	US/non-US Telecommunications Services (COM)
MPTRXUSUTL/MPTRXUTL	US/non-US Utilities (UTL)
CN	China
DE	Germany
GR	Greece
IT	Italy
US	USA

Table 1.1: Thomson Reuters MarketPsych sentiment indexes used for the analysis.

an influence on its price. Becker (2016b) points out the connection between Volkswagen share prices and the 30-200 MACD of the TRMI Media sentiment. However, different studies consider different time windows for the long and the short components of the MACD with a lack of a clear statistical and economic reasoning for the (data-driven) choice of the time window.

In the current study, we follow instead a more rigorous approach which employs the Kalman filter first introduced by Kalman (1960) and discussed by Borovkova and Mahakena (2015) and Borovkova et al. (2016) for applications to sentiment data. The real unobserved sentiment y_t is extracted from the noisy news μ_t by applying the Local Level model by Durbin and Koopman (2001). The starting point of the model is the system of autoregressive equations (2.1):

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ \mu_{t+1} &= \mu_t + \eta_t, \eta_t \sim N(0, \sigma_\eta^2). \end{aligned} \quad (2.1)$$

In the sentiment literature, it is assumed that economic agents are at least as intelligent as animals and aggregate the news as a function of time decay. This fact is motivated by the findings in the behavioral literature; see Nickerson (2011) and Reilly et al. (2011), for example. The news can be considered to follow some autoregressive process.

In (2.1), the first equation is the signal equation which corresponds to the unobserved real state equation y_t , whereas the second equation corresponds to the noisy observed news sentiment μ_{t+1} , $t = 1, \dots, T$. Both variables are assumed to follow an ARMA process. The state moments $\tilde{\mu}_t = E(\mu_t | \mathcal{F}_{t-1})$ and $P_t = \text{Var}(\mu_t | \mathcal{F}_{t-1})$ are computed recursively by solving the following equations:

$$\begin{aligned} v_t &= y_t - \tilde{\mu}_t, F_t = P_t + \sigma_\varepsilon^2, \\ K_t &= \frac{P_t}{F_t}, \\ \tilde{\mu}_{t+1} &= \tilde{\mu}_t + K_t v_t, P_{t+1} = P_t(1 - K_t) + \sigma_\eta^2, \\ \tilde{\mu}_1 &= \mu_1, P_1 = e^\gamma. \end{aligned} \quad (2.2)$$

Thus, the unobserved state μ_t is updated each time a new noisy observation y_t arrives. The state at time t is calculated by exponentially weighting the previous states.

In the next step, the states are estimated by applying the Kalman smoother and solving the

Fig. 2.1: News sentiment for the US Energy and Non-Cyclical Consumer Goods and Services and the Kalman smoothed news sentiment.

following backward recursion equations:

$$\begin{aligned}
 \hat{\mu}_t &= \tilde{\mu}_t + P_t r_{t-1} \\
 r_{t-1} &= \frac{v_t}{F_t} + L_t r_t; \quad N_{t-1} = F_t^{-1} + L_t^2 N_t, \\
 L_t &= 1 - K_t, \\
 V_t &= P_t - P_t^2 N_{t-1},
 \end{aligned} \tag{2.3}$$

where $t = 1, \dots, T$, $r_T = 0$ and $N(\hat{\mu}_t | \mathcal{D})$ is the conditional density of μ_t with $\hat{\mu}_t = E(\mu_t | \mathcal{F}_T)$ and $V_t = \text{Var}(\mu_t | \mathcal{F}_T)$. For the technical details of the estimation procedure we refer to Durbin and Koopman (2001). In the current study, the Kalman smoother is applied to the daily averages of the news sentiment. The original data and the Kalman smoothed versions for US and Energy and Non-Cyclical Consumer Goods and Services sectors are shown in the Fig. 2.1. It is worth noting, that for most sectors news sentiments of the US and non-US sectors behave in a very similar way. The smoothed US and non-US financial news sentiments are given in Fig. 2.4. It is evident that the correlation between the two series is very high and it would be difficult to disentangle the influence of the US and non-US news on the different industries under investigation in a linear regression framework. We decided, therefore, to analyze the US and the European markets separately.

3 Penalized estimation of the sentiment networks

In the current study, we are interested in analyzing if and in which cases the individual time series of the companies' prices can be potentially influenced by the news on sectors and countries. Cross-industry news spillover effects have not been directly addressed in the literature thus far. Most of the recent studies investigate the direct influence of the sentiment about an asset on

Fig. 2.2: Shanghai composite versus social and news media sentiment (from Peterson (2016), p.226).

Fig. 2.3: S & P 500 versus its sentiment (from Peterson (2016), p.21).

the asset itself and do not consider that, for example, news coming from the energy sector can influence financial companies. Borovkova and Mahakena (2015) and Borovkova (2015) have shown that extreme positive and extreme negative sentiment days influence the future price momentum of natural gas and that there is a complex relationship between the arrival of the news and the price jumps. Similar results have been developed for the energy markets in Borovkova and Lammiman (2010). Erawan (2015) uses the sentiment data for different sectors as an input for classification and regression trees and finds empirical evidence that the trading strategy can be improved by including the sector-specific news data into the model. However, no spillover effects are studied. Borovkova and Mahakena (2015) show that abnormal stock returns calculated from Fama-French and Carhart factor models might be explained by the specific sentiment data. The recent study by Borovkova et al. (2016) makes an effort to construct the risk measure based on the financial news sentiment data only and tests for the ability of this indicator to forecast the stress in the market.

As mentioned above, the aim of this work is to conduct a systematic empirical investigation of the spillover effect of the sentiment coming from different industries and countries on individual stock prices. For this purpose, we employ sparse graphical models in order to obtain insight

Fig. 2.4: Kalman smoothed news sentiment for US (red solid) and non-US (blue dotted) financial sector.

into the joint causal relationship between the individual time series.

First, we introduce the notation specific to the network literature. In the first step, a graph $\mathcal{G} = \langle V, E \rangle$ is considered, where V is the set of p nodes corresponding to each variable $X^1, X^2, \dots, X^p$ and $E \subset V \times V$ is the set of the edges corresponding to the pairwise association of all variables. Causal relations are usually represented by directed graphs: for example, if the variable X^i is assumed to be influenced by a set of variables X^j , $j = 1, \dots, p$, $j \neq i$, the associations $j \rightarrow i$ are studied. Thus, the dependence of X^i and a set of the nodes is stated as:

$$X^i = f_i(X^1, X^2, \dots, X^p) + \varepsilon_i, \quad i = 1, \dots, p, j \neq i. \quad (3.1)$$

In (3.1), ε_i is an error term. The function f_i is usually assumed to be linear. In this way, the estimation of the network can be reduced to the estimation of the individual regressions

$$X^i = \sum_{j \neq i} \beta^{ij} X^j + \varepsilon_i, \quad (3.2)$$

where the associations $j \rightarrow i$ are expressed in terms of the coefficients β^{ij} .

The individual regressions can be estimated by OLS (ordinary least squares) or more sophisticated methods. For example, Peng et al. (2012) suggest estimating nonzero partial correlations by joint sparse regression models.

Where the causal relationship among the individual time series is important, it is useful to consider the concept of Granger causality, first introduced by Granger (1980) and now widely discussed in the literature. Per definition, the time series process $\{X_t^j\}$, $t = 1, \dots, T$ Granger causes the time series process $\{X_t^i\}$, $t = 1, \dots, T$ if the regression of X_t^i on the past values of X_t^j and X_t^i gives a better fit than the regression of the past values of X_t^i .

One way to define Granger causality for a network where X_t^i is the response time series which is caused by a high-dimensional time series $X^{[-i]} = X_t^1, X_t^2, \dots, X_t^p$, $t = 1, \dots, T$, $j \neq i$ is mentioned in Lozano et al. (2009) and Arnold et al. (2007) in the context of bioinformatics. It has been proposed to test for Granger causality within the network by applying some kind of variable selection procedure. In particular X_t^j is Granger causing X_t^i if the lags of X_t^j are selected by some sparse estimation procedure for any time lag $l \in L$, $j = 1, \dots, p$, $t = 1, \dots, T$.

Formally speaking, the following regression model is estimated:

$$X_t^i = \sum_{j=1}^p \sum_{l \in L} \beta_l^{i,j} X_{t-l}^j + \varepsilon_t^i \quad i = 1, \dots, p. \quad (3.3)$$

In practical applications, the networks often appear to be high-dimensional and the number of nodes can exceed the sample size. This fact leads to inaccuracy and overfitting of the estimates. The remaining part of this section discusses how these drawbacks can be overcome by means of sparse estimation methods.

A natural way to improve the performance of the estimator in a high-dimensional regression model is to introduce a regularization penalty. Many new regularized methods have been developed in the literature over the last decades, including the least absolute shrinkage and selection operator (Lasso) estimator by Tibshirani (1996), SCAD by Fan and Li (2001), elastic net by Zou and Hastie (2005), fused Lasso by Tibshirani et al. (2005) and extensions of the Lasso models, such as adaptive Lasso by Zou (2006) and group Lasso by Yuan and Lin (2006). The application of sparse estimation methods for the estimation of high-dimensional networks in computer biology is widely discussed in the literature; see, for example, Gustafsson et al. (2005), Shimamura et al. (2007), Li and Li (2008), Friedman et al. (2008) and Jacob et al. (2009). The present work employs the Lasso approach originally formulated by Tibshirani (1996), i.e.

$$\hat{\beta}(\lambda) = \underset{\beta}{\operatorname{argmin}} \left(\frac{\|X^i - X_{t-l}\beta\|_2^2}{n} + \lambda \|\beta\|_1 \right), \quad (3.4)$$

where $\|X^i - X_{t-l}\beta\|_2^2 = \sum_{j=1}^p \sum_{l \in L} \sum_{t=1}^T [X_t^i - X_{t-l}^j]^2$, $\|\beta\|_1 = \sum_{j=1}^p \sum_{l \in L} |\beta_l^{i,j}|$ and $\lambda > 0$ is a penalty parameter. As a result of l_1 penalization, the Lasso solution is sparse, i.e. some coefficients are set exactly to zero. However, Lasso has been shown to lack the consistency for selecting the relevant variables and to produce small false positive non-zero coefficients.

The two-stage adaptive Lasso procedure introduced by Zou (2006) corrects the behavior of the Lasso and reduces the number of false positives by re-weighting the penalty function, i.e.:

$$\hat{\beta}_{\text{adaptive}}(\lambda) = \underset{\beta}{\operatorname{argmin}} \left(\frac{\|X^i - X_{t-l}\beta\|_2^2}{n} + \lambda \sum_{j=1}^p \sum_{l \in L} \frac{|\beta_l^{i,j}|}{|\hat{\beta}_{\text{initial}, l}^{i,j}|} \right), \quad (3.5)$$

where $\hat{\beta}_{\text{initial}, l}^j$, $j = 1, \dots, p$, $l \in L$ is the initial estimator from the first step of the procedure. Thus, if $\hat{\beta}_{\text{initial}, l}^{i,j}$ is large, the adaptive Lasso employs a small penalty for the $\beta_l^{i,j}$, $j = 1, \dots, p$ and improves the estimation of the effective variables. The simple OLS, the ridge regression or any other consistent estimator can be used to obtain the initial estimates. The theoretical properties and the details of the estimation algorithms can be found in Bühlmann and Van De Geer (2011).

In the current work, we include several lags of each regressor in the regression model in order to check for Granger causality. In particular, we are interested in whether the news for a given sector and not the specific lags are Granger causing the price of the asset. We implement the adaptive Lasso procedure with the OLS coefficients as the initial estimators. In addition, we employ the testing procedure based on the finite sample properties of the adaptive Lasso developed by Audrino and Camponovo (2015) to reduce the number of false positive selected variables.

In the current framework, the causality of the news data and its lags on the prices of the assets in different sectors needs to be estimated. For this reason, we define two sets of variables. Let $\{X_t^i\}$, $t = 1, \dots, T$, $i = 1, \dots, p_1$ be the set of the prices or excess returns and let $\{X_t^j\}$, $t = 1, \dots, T$, $j = p_1 + 1, \dots, p$ be the set of the news sentiment and $\{X_{t-l}^i\}$, $t = 1, \dots, T$, $i = 1, \dots, p$ be the corresponding lags. Thus, the network consists of p nodes. As we are interested

in Granger causality of the news rather than returns of other assets, the edges connecting the price variable X_t^i to the lags of other prices X_{t-l}^j , $j = 1, \dots, p_1$, $j \neq i$ are set to zero. This restrictive assumption is introduced because the sentiment data of the industries contain the price information of the biggest companies. If both price and sentiment lags are included in the regression, the estimated coefficients might be misleading. The combination of the Granger causality concept with the adaptive Lasso variable selection procedure results in the following algorithm:

1. Define the input as the p_1 -dimensional vectors of returns $\{X_t^i\}_{t=1, \dots, T}$, $j = 1, \dots, p_1$ and p_2 -dimensional vector of sentiment $\{X_t^j\}_{t=1, \dots, T}$, $j = p_1 + 1, \dots, p$.
2. Set the adjacency matrix corresponding to the network $G = \langle V, E \rangle$ equal to the zero matrix.
3. For $i = 1, \dots, p_1$:
 - apply the adaptive lasso variable selection with its related testing procedure for false positives to the model

$$X_t^i = \sum_{l \in L} \beta_{t,l}^{i,i} X_{t-l}^i + \sum_{j=p_1+1}^p \sum_{l \in L} \beta_{t,l}^{i,j} X_{t-l}^j + \varepsilon_t^i, \quad (3.6)$$

where L is the set of the predefined lags.

- If for X_t^i at least one lag of X_{t-l}^j , select $l \in L$ as significant, place the edge $X^j \rightarrow X^i$ into E .

4 Results

In this section, the news spillover effects of 10 industrial sectors are analyzed using the methodology of the graphical Granger causality. The data set contains the stock excess returns and TRMI sentiment indices on 10 sectors listed in Table 1.1 and 5 countries, namely, US, China, Germany, Italy, and Greece. The country information plays the role of a control variable. 78 US companies and 78 European companies are investigated, which corresponds to approximately 8 companies per sector. The full list of the companies can be found in Appendix 1. In order to extract the signal from the noisy sentiment data, the Kalman smoothing approach described in section 2 is applied. All the data are standardized after applying the Kalman smoother. As mentioned above, the sentiments for the US and non-US industries show similar trends over time. Therefore, the analysis is performed separately for the US and European market. The rates of returns in excess are calculated by applying the CAPM model by Sharpe (1964) using as a benchmark the S&P 500. The preliminary analysis showed that the results for the prices and excess returns using different benchmarks coincide, which is consistent with the discussion in Erawan (2015). Therefore, the same benchmark is used for the US and European companies. The daily data ranging from the January 1, 2005 to December 31, 2014 are used. A daily rolling window approach is employed in order to analyze how the connectedness between the news and the excess returns changes over time. The size of the rolling window is set to be equal to 200 trading days, which is the shortest possible size addressed in the literature; see Audrino and Knaus (2016) for more details. The small size of the rolling window is motivated by the possible time breakpoints.

Fig. 4.1: The graphical Granger network for the set of US assets on November 3, 2010.

Employing the notation of section 3, we say that the news from one industry is Granger causing the excess returns in the other industry if the lags of the sentiment on one sector are selected by the adaptive Lasso procedure (3.6) for the excess returns of the assets from the other industry, i.e. $X^j \rightarrow X^i$, $i = 1, \dots, p_1$, $j = i, p_1 + 1, \dots, p$. In the current studies, we distinguish between direct effects and spillover effects, i.e. if the sentiment of a particular sector is Granger causing the excess returns of the assets from the same sector, we call them direct effects, whereas in case the assets belong to different sectors we call them spillover effects.

This study focuses on the information spillover effect due to sentiment. For this reason, the network at time t , $t = 1, \dots, T$ is constructed by regressing the individual excess returns X_t^i on their own lags X_{t-l}^i , $l \in L$ and the set of the news sentiment lags on the sectors and the countries X_{t-l}^j , $j = p_1 + 1, \dots, p$, $l \in L$. As explained above, lags of excess returns of other stocks are not taken into consideration in the individual regressions. The estimation is performed using the R package `glmnet` by Friedman et al. (2009). In this study, we suggest including 1 day, 1 week and 1 month lags in the individual regressions, i.e. $l \in \{1, 5, 22\}$. Sentiment data of the same day (0 lag) are not included in the analysis because they might contain the information about the prices and thus overestimate the causality of the news. The choice of 1 month as mentioned

Fig. 4.2: The graphical Granger network for the set of European assets on November 3, 2010.

above is motivated by the recent studies in the sentiment literature which reveal that the shocks in the sentiment can drive the price of the underlying during the period of the next month; see Borovkova et al. (2016), for instance.

The examples of the networks obtained by the adaptive Lasso procedure for the US and European stocks on November 3, 2010 are presented in the Figures 4.1 and 4.2, respectively. The industries and countries are grey-colored in the graph, the abbreviations are given in Table 1.1, and the companies are white-colored. The whole network is constructed by regressing excess returns of each company on the past values of excess returns and lags of the sentiments of the sectors and countries. Therefore, the connections coming from the sentiment (grey nodes) to the returns (white nodes) represent the Granger causality of the sentiment indices to the excess returns. If the node is self-connected, for example Amazon (AMZ) in Fig. 4.1, the lags of the excess returns of the company itself were selected by the adaptive Lasso procedure. The date always corresponds to the last day of the rolling window. The US network (4.1) is more connected than the European network (4.2). However, in general European networks contain more edges; that is, more coefficients are estimated as non-zero by the Lasso procedure, on average 14.6 % nonzero coefficients for Europe versus 13.9 % for USA. The countries that have

more outgoing edges in Fig. 4.1 (Germany, Greece and Italy) are considered to have more influence on the excess returns during this time period. The most relevant industries during this time period are Financials and Industrials. Similar results are observed for the European companies in the network in Fig. 4.2.

In order to analyze the spillover effects over time, one needs to introduce some measures of connectedness among the excess returns and the news sentiments of the sectors and countries. Dynamic analysis of these measures can provide insight into the importance of particular sectors and countries during the global financial crisis starting in 2008 and the subsequent European sovereign debt crisis. In this study, we are interested in two characteristics of the sentiments: the relevance and the strength. We propose to measure the relevance of each sentiment by the number of outgoing edges. This measure reflects the number of regressions which have selected the considered sentiment index as an active variable. Therefore, the relevance is not informative for expressing the strength of the causality. For this reason, we define the second characteristic (strength) as the mean absolute value of the outgoing edges.

Formally speaking, the overall relevance of the sentiment variable X_t^j can be defined from the estimated coefficients over p_1 regressions of the form (3.6):

$$R(X_t^j \rightarrow \{X_t^1, \dots, X_t^{p_1}\}) = \frac{\sum_{i=1}^{p_1} \mathbb{1}\{|\hat{\beta}_{t,1}^{i,j}| + |\hat{\beta}_{t,5}^{i,j}| + |\hat{\beta}_{t,22}^{i,j}|\}}{p_1}, \quad (4.1)$$

$j = p_1 + 1, \dots, p$ and $t = 1, \dots, T$,

$$\mathbb{1}\{x\} = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{otherwise} \end{cases}$$

is an indicator function.

The overall strength of the l -th lag of the sentiment variable X_t^j can be defined as the average absolute value of the coefficients of the outgoing edges:

$$S(X_{t-l}^j \rightarrow \{X_t^1, \dots, X_t^{p_1}\}) = \frac{\sum_{i=1}^{p_1} |\hat{\beta}_{t,l}^{i,j}|}{R(X_t^j \rightarrow \{X_t^1, \dots, X_t^{p_1}\})}, \quad l \in \{1, 5, 22\}, \quad (4.2)$$

if $R(X_t^j \rightarrow \{X_t^1, \dots, X_t^{p_1}\})$ is different than zero and is zero if the variable is irrelevant. Thus, this characteristic represents the average absolute value of the coefficients $\hat{\beta}_{t,l}^{i,j}$ in the regressions of the excess returns of company i for the sentiment j . If only the sum of the absolute values were considered, it would be impossible to distinguish between the nodes with many links with small coefficients and the nodes with a small number of links and bigger coefficients. Speaking in terms of graphical representation, the relevance characterizes the average number of the outgoing links and the strength characterizes the average width of the link. If the sentiment (node) has small relevance and high strength, it is selected as significant in a small number of regressions but estimated with relatively high coefficients. In definition 4.2, the strength of the regulator is defined separately for each lag.

Similarly, the same measures can be defined for particular groups of companies. In this case, the average is taken over the companies contained in the specific group. For example, if the average is taken over the companies corresponding to the financial sector with the Industrials sentiment as the variable under consideration, the relevance and the strength will show the spillover effect of the Industrials-related news to the financial sector.

	FIN	TEC	IND	MAT	UTL	COM	HLC	NCY	ENE	YCY
asset	0.72	0.77	0.72	0.69	0.61	0.71	0.64	0.68	0.66	0.71
MPTRXUSFIN	0.16	0.17	0.17	0.19	0.17	0.17	0.16	0.18	0.22	0.15
MPTRXUSTEC	0.11	0.13	0.15	0.12	0.17	0.20	0.16	0.14	0.23	0.14
MPTRXUSIND	0.12	0.16	0.17	0.13	0.14	0.19	0.13	0.14	0.15	0.14
MPTRXUSMAT	0.10	0.07	0.15	0.13	0.13	0.11	0.12	0.12	0.18	0.11
MPTRXUSUTL	0.09	0.10	0.14	0.15	0.10	0.17	0.12	0.10	0.21	0.09
MPTRXUSCOM	0.13	0.12	0.11	0.10	0.11	0.18	0.11	0.11	0.18	0.10
MPTRXUSHLC	0.15	0.14	0.21	0.14	0.12	0.16	0.12	0.14	0.13	0.13
MPTRXUSNCY	0.13	0.12	0.20	0.17	0.14	0.17	0.14	0.14	0.20	0.13
MPTRXUSENE	0.14	0.13	0.16	0.14	0.11	0.16	0.17	0.15	0.18	0.19
MPTRXUSYCY	0.12	0.10	0.12	0.11	0.12	0.14	0.11	0.12	0.16	0.11
US	0.14	0.16	0.16	0.13	0.11	0.17	0.15	0.17	0.17	0.11
IT	0.12	0.14	0.12	0.12	0.12	0.13	0.12	0.15	0.13	0.11
GR	0.17	0.15	0.16	0.16	0.15	0.17	0.11	0.16	0.14	0.14
DE	0.14	0.12	0.15	0.12	0.14	0.15	0.13	0.12	0.21	0.13
CN	0.09	0.10	0.14	0.12	0.09	0.10	0.08	0.13	0.11	0.12

Table 4.1: Mean relevance of the news sentiments of the US sectors and selected countries for the US companies by sector ranging from Jan. 1, 2005 to Dec. 31, 2014.

4.1 US results

To begin with, we discuss the average news spillover effects within the US market. In Table 4.1 we present the cross-industry mean relevance of the news over the full time period under consideration. For example, the number 0.12 in the first column and the tenth row means that YCY news are Granger causing the excess returns of 12 companies from 100 in the financial sector on average. It is worth noting that the stocks are mostly self-connected, i.e. the own lags of the excess returns of the stocks are chosen to be relevant for the future excess returns by the adaptive Lasso procedure (see the numbers in the first row). From Table 4.1, one can observe that the relevance of the country-related news spreads almost uniformly among the sectors. There is weak evidence of a bigger influence of US news on NCY and Energy. A similar effect is observed for the industries; however, a slightly higher relevance of the financial news for all sectors is evident. It is worth mentioning that the results of Table 4.1 should be interpreted with caution, as the average is taken over a long period of time. The time interval of almost 10 years might contain several structural breakpoints in the relevance of the news.

The dynamics of the relevance over time becomes clearer from Fig. 4.3 where the rolling window relevance is presented for the selected countries and industries. It can be concluded that, in general, the relevance is not constant and has a cyclical behavior. For some sectors, several breakouts can be observed. First, the relevance of the own lags of the stock drops during the period of the global financial crisis. This finding coincides with the results shown by Audrino and Knaus (2016) in the context of the HAR volatility model. Second, the relevance of Technology and Industrials rises right after the crisis.

Similarly, the relevance of the country-specific news shows a cyclical behavior. The exception is the growing relevance of the US-related news around 2008 and the high relevance of the Germany-related news around 2008-2010 and 2013. These results support the common belief of the global leading role played by the US information flow during the financial crisis and by the Germany-related news (as the leading European country) during the financial and subsequent European sovereign debt crises.

In contrast to the analysis of relevance, a closer look at the strength of the sector-related news gives better insight into the cross-industry news spillover effects. The mean cross-industry strength of the 1, 5 and 22 lags of the news sentiments is presented in Table 4.2. From Fig. 4.4 it is clearly observed that the news of Financials and Energy have the strongest causal influence

	FIN	TEC	IND	MAT	UTL	COM	HLC	NCY	ENE	YCY
asset _{t-1}	0.24	0.27	0.25	0.24	0.23	0.24	0.24	0.26	0.23	0.27
asset _{t-5}	0.22	0.23	0.23	0.24	0.21	0.23	0.22	0.22	0.22	0.23
asset _{t-22}	0.19	0.18	0.20	0.18	0.18	0.18	0.19	0.18	0.20	0.23
MPTRXUSFIN _{t-1}	1.33	1.08	1.28	1.53	0.98	1.02	0.83	1.14	1.52	1.35
MPTRXUSFIN _{t-5}	0.84	0.87	1.11	1.32	0.79	1.15	1.21	0.95	1.15	0.88
MPTRXUSFIN _{t-22}	0.90	0.94	0.94	1.28	0.95	1.01	1.07	0.78	1.10	0.87
MPTRXUSTEC _{t-1}	0.29	0.36	0.39	0.39	0.44	0.39	0.37	0.39	0.41	0.45
MPTRXUSTEC _{t-5}	0.41	0.33	0.38	0.49	0.28	0.40	0.26	0.29	0.54	0.46
MPTRXUSTEC _{t-22}	0.30	0.33	0.34	0.39	0.39	0.41	0.41	0.28	0.45	0.38
MPTRXUSIND _{t-1}	0.43	0.54	0.51	0.51	0.49	0.62	0.45	0.54	0.59	0.52
MPTRXUSIND _{t-5}	0.38	0.35	0.53	0.50	0.42	0.50	0.44	0.33	0.51	0.54
MPTRXUSIND _{t-22}	0.57	0.47	0.46	0.64	0.46	0.61	0.43	0.51	0.57	0.52
MPTRXUSMAT _{t-1}	0.35	0.38	0.41	0.55	0.33	0.42	0.37	0.38	0.38	0.35
MPTRXUSMAT _{t-5}	0.37	0.32	0.36	0.45	0.34	0.35	0.32	0.36	0.38	0.30
MPTRXUSMAT _{t-22}	0.32	0.35	0.42	0.45	0.38	0.45	0.41	0.28	0.43	0.29
MPTRXUSUTL _{t-1}	0.45	0.41	0.48	0.60	0.41	0.53	0.46	0.49	0.79	0.56
MPTRXUSUTL _{t-5}	0.53	0.30	0.36	0.56	0.26	0.47	0.28	0.33	0.55	0.47
MPTRXUSUTL _{t-22}	0.48	0.31	0.41	0.47	0.42	0.46	0.39	0.36	0.50	0.36
MPTRXUSCOM _{t-1}	0.45	0.46	0.47	0.41	0.37	0.36	0.41	0.48	0.59	0.42
MPTRXUSCOM _{t-5}	0.32	0.33	0.40	0.37	0.33	0.56	0.31	0.29	0.51	0.39
MPTRXUSCOM _{t-22}	0.33	0.43	0.46	0.63	0.34	0.41	0.32	0.29	0.48	0.44
MPTRXUSHLC _{t-1}	0.41	0.40	0.52	0.47	0.38	0.44	0.38	0.39	0.52	0.32
MPTRXUSHLC _{t-5}	0.41	0.42	0.41	0.49	0.40	0.38	0.39	0.32	0.49	0.43
MPTRXUSHLC _{t-22}	0.40	0.38	0.38	0.49	0.34	0.40	0.48	0.34	0.44	0.33
MPTRXUSNCY _{t-1}	0.29	0.32	0.32	0.36	0.30	0.41	0.30	0.42	0.48	0.39
MPTRXUSNCY _{t-5}	0.21	0.26	0.43	0.37	0.31	0.43	0.34	0.27	0.42	0.28
MPTRXUSNCY _{t-22}	0.30	0.34	0.35	0.35	0.28	0.39	0.26	0.29	0.37	0.36
MPTRXUSENE _{t-1}	0.82	0.94	1.03	2.04	0.92	1.26	1.67	0.83	2.06	0.83
MPTRXUSENE _{t-5}	1.01	0.77	1.12	0.91	0.89	0.93	1.07	0.74	1.33	1.13
MPTRXUSENE _{t-22}	0.78	0.70	0.88	0.88	0.87	1.04	0.87	0.92	0.93	1.21
MPTRXUSYCY _{t-1}	0.59	0.42	0.38	0.77	0.39	0.64	0.34	0.48	0.42	0.64
MPTRXUSYCY _{t-5}	0.36	0.35	0.45	0.37	0.32	0.51	0.33	0.38	0.49	0.44
MPTRXUSYCY _{t-22}	0.39	0.30	0.44	0.54	0.54	0.52	0.52	0.40	0.53	0.39
US _{t-1}	0.71	0.52	0.66	0.80	0.50	0.69	0.48	0.50	0.68	0.53
US _{t-5}	0.57	0.32	0.60	0.50	0.34	0.63	0.44	0.44	0.51	0.54
US _{t-22}	0.41	0.46	0.59	0.68	0.45	0.66	0.67	0.47	0.56	0.41
IT _{t-1}	0.35	0.45	0.38	0.48	0.31	0.38	0.39	0.34	0.41	0.41
IT _{t-5}	0.34	0.34	0.29	0.56	0.32	0.36	0.37	0.31	0.50	0.64
IT _{t-22}	0.34	0.38	0.38	0.58	0.35	0.51	0.63	0.34	0.54	0.43
DE _{t-1}	0.45	0.48	0.49	0.59	0.54	0.62	0.41	0.40	0.48	0.37
DE _{t-5}	0.37	0.38	0.41	0.37	0.47	0.49	0.39	0.38	0.58	0.54
DE _{t-22}	0.47	0.39	0.47	0.43	0.48	0.56	0.43	0.35	0.60	0.44
CN _{t-1}	0.36	0.44	0.38	0.41	0.35	0.42	0.38	0.37	0.41	0.43
CN _{t-5}	0.29	0.26	0.48	0.36	0.37	0.31	0.33	0.23	0.35	0.32
CN _{t-22}	0.34	0.31	0.37	0.55	0.26	0.35	0.34	0.35	0.51	0.34
GR _{t-1}	0.55	0.47	0.57	0.69	0.60	0.62	0.51	0.58	0.78	0.66
GR _{t-5}	0.56	0.41	0.47	0.47	0.54	0.55	0.53	0.53	0.64	0.49
GR _{t-22}	0.51	0.54	0.56	0.51	0.37	0.58	0.48	0.48	0.51	0.41

Table 4.2: Mean strength of the lags of the news sentiments on the US sectors and selected countries for the US companies by sector ranging from Jan. 1, 2005 to Dec. 31, 2014.

Fig. 4.3: The overall relevance of the news sentiments of the selected sectors and countries on the US stocks ranging from Jan. 1, 2005 to Dec. 31, 2014.

on the excess returns in all sectors. Such a high strength of the financial news could be explained by the fact that the global financial crisis and the European sovereign debt crisis are part of our sample period. Interestingly, the lags related to the stock itself are estimated with smaller coefficients than the lags of the sentiments. This finding supports the result that the past values of the news contain additional information about the future stock returns; moreover, the importance of the news rises during financial turbulence and macroeconomic recessions.

To get a deeper understanding, we present the rolling window results averaged over all stocks in Fig. 4.4 for one-day lag variables: results for other lags are similar. The strength of the financial news starts rising in 2007, showing that the growing influence of the news can be considered an early warning of future instability. Higher average strength of YCY news is observed during the first half of the sample and decreases thereafter. This might be explained by the fact that the aggregated consumption of US households started to grow rapidly at the end of 2005, went through recession in 2008, and continued to increase in 2009. Consumption has shown stable upward trends since 2009; therefore, the news on the YCY sector is less important in the second half of the period. Moreover, it is observed that the strength of Technology and Industrials grows right after the recession period, as innovations in these sectors contribute notably to the recovery process and gain additional attention of the media. The smooth decline in the strength of the energy sector-related news on the stock returns might be caused by the introduction of US energy independence politics. The reduction in imports might be transferred to greater confidence in the market and, as a result, less weight for the energy-related news. Fig.

Fig. 4.4: The overall strength of the 1 day lags of the news sentiments on the selected US sectors and countries for the US companies ranging from Jan. 1, 2005 to Dec. 31, 2014.

4.4 depicts the strength of selected countries, and it appears that the US news shows relatively high overall strength and relativity. On the other hand, short peaks with low persistence in the strength of the news related to Italy and Germany are observed. The first peak in the strength of Italy corresponds to the currency crisis of 2006 in Europe: there were some concerns that several countries might leave the Eurozone, which could have potentially lead to the total collapse of the Euro. This event would influence the currency market worldwide, in particular in the US. The second peak could potentially be caused by the European debt crisis, which severely affected such countries as Italy and Greece. 2010 in Italy was characterized by a rapid decrease in GDP and increasing unemployment. The fluctuations in Germany's strength can be explained by the global financial crisis of 2008, the European debt crisis of 2010 and the decision of Germany in 2012 to provide more support to its European partners for more centralized control over the Eurozone. The results for other sectors and countries are presented in Appendix 2. A detailed analysis of the strength by sector is given in Appendix 3.

4.2 EU results

Comparing the news spillover effects in the European market to the US market, we observe the similarity in terms of the relevance of the news. However, the average strength of the news sentiments is less strong on average; see the results summarized in Table 4.3. The cyclical

	FIN	TEC	IND	MAT	UTL	COM	HLC	NCY	ENE	YCY
asset	0.70	0.74	0.70	0.71	0.59	0.68	0.69	0.78	0.70	0.76
MPTRXFIN	0.13	0.17	0.17	0.16	0.12	0.11	0.19	0.17	0.14	0.14
MPTRXTEC	0.13	0.16	0.08	0.15	0.09	0.14	0.18	0.09	0.18	0.10
MPTRXIND	0.14	0.12	0.12	0.15	0.13	0.10	0.17	0.12	0.18	0.11
MPTRXMAT	0.15	0.15	0.12	0.15	0.12	0.14	0.18	0.12	0.21	0.11
MPTRXUTL	0.17	0.12	0.17	0.17	0.10	0.11	0.24	0.18	0.15	0.18
MPTRXCOM	0.20	0.17	0.13	0.14	0.12	0.13	0.19	0.13	0.19	0.12
MPTRXHLC	0.14	0.13	0.15	0.11	0.10	0.09	0.16	0.13	0.15	0.14
MPTRXNCY	0.14	0.14	0.14	0.11	0.14	0.14	0.21	0.12	0.17	0.13
MPTRXENE	0.18	0.15	0.13	0.16	0.12	0.15	0.18	0.16	0.16	0.13
MPTRXYCY	0.18	0.16	0.16	0.17	0.14	0.17	0.23	0.15	0.22	0.22
US	0.16	0.12	0.11	0.16	0.12	0.15	0.16	0.11	0.15	0.14
IT	0.15	0.15	0.17	0.19	0.10	0.12	0.16	0.16	0.15	0.13
GR	0.16	0.16	0.16	0.19	0.12	0.15	0.21	0.17	0.18	0.12
DE	0.12	0.15	0.14	0.15	0.11	0.14	0.18	0.11	0.19	0.13
CN	0.12	0.11	0.12	0.12	0.10	0.09	0.17	0.10	0.13	0.12

Table 4.3: Mean relevance of the lags of the news sentiments of the European sectors and selected countries for the European companies by sector ranging from Jan. 1, 2005 to Dec. 31, 2014.

behavior of the average relevance of some sectors for all stocks is shown in Fig. 4.5. The analysis of the strength of the sectors in terms of the average absolute Lasso estimated coefficients coincides with the results for the US market. The news related to Financials and Energy seems to be important for all other sectors; see Table 4.4. Similarly to the US market, the strength of the financial news rises before the crisis and reaches a maximum in 2008; see Fig. 4.6 and Table 4.4. However, the pattern of the strength of the energy news is cyclical in contrast to the US market. This might be explained by the fact that, in contrast to the US, the European energy market is less independent and concerns over the prices spread out to the other sectors. Full detailed results of the analysis are reported in Appendix 4.

It can be concluded that the news sentiments related to several important sectors provide additional information about the future stock prices. For each market, several important sectors can be defined. The news spillover effects coming from these influential sectors seem to be at least as important as the direct effects. Moreover, the relevance of these spillover effects increases during periods characterized by general economic instability and/or financial market turbulence.

5 An empirical illustration

The results of the analysis presented in the previous section show that the sentiment data of several sectors have a significant information component, which can be used to improve the prediction of the assets' returns. This section illustrates the impact of the various sentiment series on the conditional mean and the conditional variance equations of ARMA-GARCH models. It is important to note that the purpose of this section is to give an insight into possible practical applications of the results presented earlier rather than draw general conclusions about sentiment augmented time series models.

Further on, a parsimonious ARMA(1,1)-GARCH(1,1) model with Gaussian innovations is

	FIN	TEC	IND	MAT	UTL	COM	HLC	NCY	ENE	YCY
asset _{t-1}	0.24	0.24	0.23	0.25	0.24	0.27	0.22	0.23	0.25	0.23
asset _{t-5}	0.23	0.23	0.23	0.22	0.22	0.23	0.23	0.23	0.22	0.22
asset _{t-22}	0.19	0.17	0.18	0.18	0.19	0.19	0.17	0.17	0.20	0.17
MPTRXFIN _{t-1}	0.77	0.65	0.98	0.96	0.85	0.61	0.77	0.75	0.84	0.68
MPTRXFIN _{t-5}	0.84	0.58	1.00	1.22	0.76	0.56	0.80	0.75	0.79	0.46
MPTRXFIN _{t-22}	0.61	0.52	0.88	0.83	0.59	0.63	0.65	0.67	0.67	0.39
MPTRXTEC _{t-1}	0.35	0.39	0.36	0.52	0.38	0.35	0.36	0.36	0.53	0.35
MPTRXTEC _{t-5}	0.30	0.36	0.28	0.46	0.43	0.30	0.31	0.34	0.46	0.18
MPTRXTEC _{t-22}	0.46	0.35	0.50	0.51	0.44	0.43	0.45	0.43	0.49	0.22
MPTRXIND _{t-1}	0.47	0.38	0.47	0.57	0.60	0.50	0.55	0.41	0.58	0.26
MPTRXIND _{t-5}	0.36	0.32	0.47	0.44	0.58	0.35	0.48	0.39	0.44	0.22
MPTRXIND _{t-22}	0.45	0.46	0.56	0.64	0.53	0.42	0.40	0.53	0.51	0.31
MPTRXMAT _{t-1}	0.32	0.28	0.29	0.46	0.35	0.25	0.36	0.28	0.42	0.22
MPTRXMAT _{t-5}	0.32	0.28	0.41	0.39	0.32	0.26	0.28	0.25	0.37	0.24
MPTRXMAT _{t-22}	0.34	0.32	0.32	0.37	0.36	0.33	0.33	0.31	0.32	0.19
MPTRXUTL _{t-1}	0.39	0.30	0.66	0.50	0.46	0.40	0.52	0.37	0.53	0.55
MPTRXUTL _{t-5}	0.45	0.37	0.40	0.37	0.37	0.29	0.39	0.30	0.50	0.22
MPTRXUTL _{t-22}	0.38	0.36	0.40	0.44	0.41	0.36	0.42	0.35	0.43	0.23
MPTRXCOM _{t-1}	0.44	0.40	0.46	0.49	0.42	0.38	0.41	0.34	0.48	0.30
MPTRXCOM _{t-5}	0.38	0.31	0.36	0.36	0.47	0.22	0.36	0.31	0.29	0.25
MPTRXCOM _{t-22}	0.33	0.37	0.39	0.38	0.43	0.37	0.43	0.33	0.39	0.21
MPTRXHLC _{t-1}	0.46	0.45	0.67	0.65	0.36	0.37	0.41	0.41	0.53	0.34
MPTRXHLC _{t-5}	0.40	0.39	0.49	0.64	0.41	0.37	0.40	0.41	0.43	0.26
MPTRXHLC _{t-22}	0.40	0.36	0.58	0.55	0.48	0.41	0.49	0.43	0.56	0.28
MPTRXNCY _{t-1}	0.47	0.33	0.68	0.46	0.42	0.33	0.52	0.35	0.46	0.25
MPTRXNCY _{t-5}	0.38	0.34	0.45	0.40	0.46	0.33	0.43	0.40	0.43	0.20
MPTRXNCY _{t-22}	0.34	0.34	0.49	0.47	0.35	0.37	0.47	0.36	0.40	0.20
MPTRXENE _{t-1}	0.87	0.77	1.12	0.84	0.83	0.65	0.77	0.71	0.87	0.62
MPTRXENE _{t-5}	0.57	0.63	0.86	0.80	0.69	0.53	0.72	0.71	1.25	0.38
MPTRXENE _{t-22}	0.68	0.65	0.88	0.87	0.83	0.57	0.87	0.54	0.89	0.49
MPTRXYCY _{t-1}	0.42	0.40	0.46	0.48	0.42	0.39	0.38	0.39	0.44	0.45
MPTRXYCY _{t-5}	0.36	0.33	0.36	0.36	0.41	0.30	0.46	0.30	0.50	0.21
MPTRXYCY _{t-22}	0.30	0.50	0.48	0.59	0.30	0.41	0.38	0.32	0.40	0.21
US _{t-1}	0.71	0.62	0.37	0.55	0.59	0.40	0.46	0.45	0.47	0.50
US _{t-5}	0.57	0.39	0.65	0.54	0.50	0.45	0.55	0.42	0.56	0.25
US _{t-22}	0.58	0.43	0.42	0.47	0.45	0.42	0.54	0.39	0.49	0.33
IT _{t-1}	0.44	0.33	0.41	0.55	0.46	0.52	0.46	0.40	0.46	0.26
IT _{t-5}	0.39	0.29	0.32	0.36	0.41	0.32	0.37	0.37	0.38	0.20
IT _{t-22}	0.38	0.37	0.68	0.49	0.41	0.34	0.38	0.39	0.44	0.20
DE _{t-1}	0.51	0.52	0.72	0.60	0.48	0.47	0.45	0.52	0.43	0.32
DE _{t-5}	0.44	0.37	0.50	0.52	0.46	0.43	0.54	0.43	0.53	0.27
DE _{t-22}	0.45	0.42	0.56	0.56	0.41	0.45	0.44	0.38	0.50	0.21
CN _{t-1}	0.38	0.38	0.40	0.45	0.32	0.36	0.46	0.32	0.38	0.43
CN _{t-5}	0.36	0.29	0.47	0.54	0.40	0.32	0.42	0.32	0.40	0.16
CN _{t-22}	0.47	0.38	0.47	0.44	0.38	0.37	0.40	0.33	0.34	0.26
GR _{t-1}	0.61	0.52	1.13	0.72	0.63	0.62	0.57	0.51	0.65	0.37
GR _{t-5}	0.38	0.41	0.52	0.58	0.54	0.42	0.53	0.40	0.67	0.38
GR _{t-22}	0.54	0.46	0.53	0.59	0.60	0.48	0.55	0.50	0.55	0.31

Table 4.4: Mean strength of the lags of the news sentiments on the European sectors and selected countries for the European companies by sector ranging from Jan. 1, 2005 to Dec. 31, 2014.

Fig. 4.5: The overall relevance of the news sentiments of the selected sectors and countries on the European stocks ranging from Jan. 1, 2005 to Dec. 31, 2014.

applied to the series of log returns of several firms. The choice of the model is due to the fact that lower order GARCH models are used in most applications. Moreover, this model has been selected according to the Akaike information criterion in the majority of cases. This benchmark model is compared to extensions in which the lag of the sentiment index is included in the model as an exogenous variable. The considered model has the following specification:

$$\begin{aligned} Y_t &= \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 \varepsilon_{t-1} + \gamma_1 S_{t-1} + e_t, \\ \text{Var}(e_t | \mathcal{F}_{t-1}) &= a_0 + a_1 \sigma_{t-1}^2 + b_1 e_{t-1}^2 + g_1 S_{t-1}, \end{aligned} \quad (5.1)$$

where $e_t = \sigma_t \varepsilon_t$, with ε_t being a sequence of iid random variables with mean 0 and variance 1, Y_t is the series of the log returns, S_{t-1} is the sentiment index of the pre-defined sector, $t = 1, \dots, T$, and $\alpha_0, \alpha_1, \beta_1, \gamma_1, a_0, a_1, b_1, g_1$ are the parameters to be estimated. For further details on time series models, we refer to Tsay (2005), Andersen et al. (2009), Francq and Zakoian (2011) and original work on the GARCH model by Bollerslev (1986).

The specification (5.1) results in three competing models. If $\gamma_1 = 0$ and $g_1 = 0$, the ARMA(1,1)-GARCH(1,1) model without exogenous sentiment variables is considered. If $g_1 = 0$, the sentiment variable is included in the conditional mean equation. If both γ_1 and g_1 need to be estimated, the sentiment index is included in the conditional mean and the conditional variance equations.

The analysis performed in the previous sections has shown that the sentiment spillover effects

Fig. 4.6: The overall strength of the 1 day lags of the news sentiments on the selected US sectors and countries for the US companies ranging from Jan. 1, 2005 to Dec. 31, 2014.

dominate the direct effect during periods of financial turbulence. It has been empirically shown that the sentiment data of the financial and energy sectors drive the market during certain periods. In order to validate this conclusion, we separately estimate the models with the firm-specific sector's sentiment and the sentiment data coming from the financial and energy sectors.

After estimating the models based on 200-day rolling windows and obtaining the one-day-ahead forecast series, the models are compared in terms of their predictive power. Their predictive power is expressed in terms of the mean squared prediction error (MSPE). The two-sided test of Diebold and Mariano (1995) (DM test) is applied to the squared errors to check whether the model provides a statistically significant improvement in comparison to the ARMA(1,1)-GARCH(1,1), which is always used as the benchmark.

Four companies, belonging to two industrial sectors (Basic Materials and Healthcare), are used for the illustrative purposes of this section: International Flavors & Fragrances Inc. (IFF), Williams Companies Inc. (WMB), Chevron Corporation (CVX), and Cigna Corporation (CI). The selection of the time intervals is motivated by Fig. 5.1. It is evident that the financial news has a strong effect on the companies related to Basic Materials in 2010 and is less important for the returns of this sector in 2008 and 2009. The energy news is of high importance for both Basic Materials and Healthcare at the beginning of the analysed time period (2007) and is less influential at the end of the period (2013).

	MSE 1.11.2008 - 1.11.2009	DM p -value	MSPE 1.11.2009 - 1.11.2010	DM p -value	MSPE 1.11.2010 - 1.11.2011	DM p -value
IFF (Basic Materials)						
$\gamma_1 = 0, g_1 = 0$	$4.18 \cdot 10^{-5}$		$5.89 \cdot 10^{-5}$		$3.06 \cdot 10^{-5}$	
$g_1 = 0, S = \text{MPTRXFIN}$	$4.96 \cdot 10^{-5}$	$3.76 \cdot 10^{-1}$	$9.70 \cdot 10^{-6}$	$8.05 \cdot 10^{-8}$	$4.29 \cdot 10^{-5}$	$1.51 \cdot 10^{-1}$
$S = \text{MPTRXFIN}$	$5.01 \cdot 10^{-5}$	$3.46 \cdot 10^{-1}$	$9.70 \cdot 10^{-6}$	$8.16 \cdot 10^{-8}$	$4.06 \cdot 10^{-5}$	$1.64 \cdot 10^{-1}$
$g_1 = 0, S = \text{MPTRXMAT}$	$8.94 \cdot 10^{-5}$	$6.00 \cdot 10^{-3}$	$1.47 \cdot 10^{-4}$	$3.12 \cdot 10^{-1}$	$6.01 \cdot 10^{-5}$	$1.02 \cdot 10^{-3}$
$S = \text{MPTRXMAT}$	$6.70 \cdot 10^{-5}$	$3.00 \cdot 10^{-3}$	$1.44 \cdot 10^{-4}$	$3.30 \cdot 10^{-1}$	$5.64 \cdot 10^{-5}$	$3.10 \cdot 10^{-3}$
WMB (Basic Materials)						
$\gamma_1 = 0, g_1 = 0$	$1.05 \cdot 10^{-4}$		$1.39 \cdot 10^{-4}$		$8.76 \cdot 10^{-5}$	
$g_1 = 0, S = \text{MPTRXFIN}$	$7.91 \cdot 10^{-5}$	$3.17 \cdot 10^{-1}$	$6.93 \cdot 10^{-5}$	$4.67 \cdot 10^{-3}$	$6.73 \cdot 10^{-5}$	$1.72 \cdot 10^{-1}$
$S = \text{MPTRXFIN}$	$7.53 \cdot 10^{-5}$	$2.46 \cdot 10^{-1}$	$6.94 \cdot 10^{-5}$	$4.78 \cdot 10^{-3}$	$6.75 \cdot 10^{-5}$	$1.75 \cdot 10^{-1}$
$g_1 = 0, S = \text{MPTRXMAT}$	$6.80 \cdot 10^{-5}$	$1.74 \cdot 10^{-1}$	$1.69 \cdot 10^{-4}$	$3.43 \cdot 10^{-1}$	$2.17 \cdot 10^{-5}$	$1.26 \cdot 10^{-9}$
$S = \text{MPTRXMAT}$	$7.75 \cdot 10^{-5}$	$3.18 \cdot 10^{-1}$	$1.83 \cdot 10^{-4}$	$1.93 \cdot 10^{-1}$	$2.14 \cdot 10^{-5}$	$1.23 \cdot 10^{-9}$

Table 5.1: MSPE of (5.1) (p -values of DM test compared to the model with $\gamma_1 = g_1 = 0$).

Fig. 5.1: The strength of 1 day lags of the news sentiments of the FIN and ENE sectors for the MAT and HLC US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

The mean squared prediction errors and the p -values of the two-sided DM test for the log returns of IFF and WMB are presented in Table 5.1. The comparison is always made with respect to the ARMA(1,1)-GARCH(1,1) model, i.e. model (5.1) with $g_1 = \gamma_1 = 0$. It is evident that including the exogenous variable of financial news in the conditional mean equation improves the predictive power of the ARMA-GARCH model from November 1, 2009 to November 1, 2010, which corresponds to the period of strong influence of the financial sector. An improvement is not observed during the year before and the year after the above mentioned period. The results for the conditional variance equation are not unambiguous and further research should be addressed to this question. In the majority of the analysed cases, exogenous news indices do not provide any further improvement of the predictive power of the model. Including sector related news (Basic Materials) in the model does not significantly improve the predictive power from November 1, 2009 to November 1, 2010, when the financial news was driving the market. The influence of the sector-specific news during other periods is more evident, but not always statistically significant.

Similar results are observed for CVX and CI, see Table 5.2. It is evident that including energy-related news in the conditional mean equation improves the predictive power of the ARMA-GARCH model in 2007 and is not relevant in 2014. The MSPE is not significantly reduced when the sentiment data are included in the conditional variance model. Moreover, sector-related sentiment data do not reduce the MSPE significantly in the period when the market was driven by the news coming from the Energy sector.

The results for the considered firms support the conclusion of the previous section and sug-

	MSPE 1.12.2006 - 1.12.2007	DM p -value	MSPE 1.12.2013 - 1.12.2014	DM p -value
CVX (Basic Materials)				
$\gamma_1 = 0, g_1 = 0$	$1.82 \cdot 10^{-4}$		$2.47 \cdot 10^{-5}$	
$g_1 = 0, S = \text{MPTRXENE}$	$1.13 \cdot 10^{-4}$	$2.96 \cdot 10^{-2}$	$2.73 \cdot 10^{-5}$	$5.21 \cdot 10^{-1}$
$S = \text{MPTRXENE}$	$1.11 \cdot 10^{-4}$	$2.21 \cdot 10^{-2}$	$2.85 \cdot 10^{-5}$	$3.47 \cdot 10^{-1}$
$g_1 = 0, S = \text{MPTRXHLC}$	$1.60 \cdot 10^{-4}$	$5.59 \cdot 10^{-1}$	$4.09 \cdot 10^{-5}$	$1.23 \cdot 10^{-2}$
$S = \text{MPTRXHLC}$	$2.19 \cdot 10^{-4}$	$3.70 \cdot 10^{-1}$	$2.37 \cdot 10^{-5}$	$7.99 \cdot 10^{-1}$
CI (Healthcare)				
$\gamma_1 = 0, g_1 = 0$	$9.58 \cdot 10^{-5}$		$1.21 \cdot 10^{-5}$	
$g_1 = 0, S = \text{MPTRXENE}$	$3.89 \cdot 10^{-5}$	$4.9 \cdot 10^{-3}$	$9.00 \cdot 10^{-6}$	$2.70 \cdot 10^{-1}$
$S = \text{MPTRXENE}$	$3.88 \cdot 10^{-5}$	$4.8 \cdot 10^{-3}$	$9.50 \cdot 10^{-6}$	$3.50 \cdot 10^{-1}$
$g_1 = 0, S = \text{MPTRXMAT}$	$1.60 \cdot 10^{-4}$	$3.9 \cdot 10^{-2}$	$4.09 \cdot 10^{-5}$	$2.29 \cdot 10^{-6}$
$S = \text{MPTRXMAT}$	$9.77 \cdot 10^{-5}$	$9.5 \cdot 10^{-1}$	$3.70 \cdot 10^{-6}$	$1.12 \cdot 10^{-5}$

Table 5.2: MSPE of (5.1) (p -values of DM test compared to the model with $\gamma_1 = g_1 = 0$).

gest that sentiment spillover effects dominate the direct effects in periods when the market is driven by the news coming from a different sector. Therefore, the time series models which are augmented by including the news of influential sectors might show better predictive power for returns. The periods when the sentiment indices of particular industries are more informative can be found by applying the procedure described in section 3. The illustrative examples discussed here support the results of section 4, which suggests the importance of the information contained in sentiment data for asset pricing theory.

Conclusion

The goal of this paper is to investigate the cross-industry patterns of the news and stock returns, and in particular to analyze how the news about one industry influences the stock returns in the other industries. For this purpose, the graphical Granger model has been applied to the Kalman smoothed sentiment data on 10 US and 10 non-US industries and on the excess returns of 78 US and 78 European companies. The sentiment data on several countries have been included in the analysis as control variables. The adaptive Lasso procedure has been applied to estimate the return-news networks. The network-based measures reflecting the relevance and the strength of each news source have been proposed and analyzed over a period of 10 years by employing a rolling window approach.

We found empirical evidence that the relevance of the news coming from different sectors shows a cyclical behavior and spreads evenly among the industries. Moreover, our results show that the strength of the influence of the news on some sectors grows just before periods of economic and financial instability and reaches a maximum during crises. Interesting patterns are observed in the causality of the financial and energy sentiments. These sentiments can be seen as the most influential, and the spillover effects from the sectors dominate the direct effects. Estimation results show that the overall connectedness among the stock returns and the news is stronger for the US market than for the European market. The importance of sentiment spillover effects has been empirically illustrated. We showed that the sentiment indices of specific industries can be successfully used to improve the predictive power of time series models for returns.

In future research we plan to relax the assumption of predefined lags in the graphical Granger model and to apply the same adaptive Lasso methodology to test the significance of arbitrary

5 *An empirical illustration*

lags. This could be especially interesting to show the persistence of the news coming from different sectors. Moreover, the study of the nonlinear or quantile cross-industry dependencies among the news could yield further insights into the analysis of the impact of direct sentiment effects as well as spillover sentiment effects on companies' excess returns.

Bibliography

- E. Akyıldırım, A. Altarovici, and C. Ekinici. Effects of firm-specific public announcements on market dynamics: Implications for high-frequency traders. *Handbook of High Frequency Trading*, page 305, 2015.
- D. E. Allen, M. J. McAleer, and A. K. Singh. Machine news and volatility: The dow jones industrial average and the trna real-time high-frequency sentiment series. In *The Handbook of High Frequency Trading*, pages 327–344. Academic Press San Diego, 2015.
- T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. V. Mikosch. *Handbook of financial time series*. Springer Science & Business Media, 2009.
- G. Appel. Become your own technical analyst: How to identify significant market turning points using the moving average convergence-divergence indicator or macd. *The Journal of Wealth Management*, 6(1):27–36, 2003.
- A. Arnold, Y. Liu, and N. Abe. Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 66–75. ACM, 2007.
- F. Audrino and L. Camponovo. Oracle properties and finite sample inference of the adaptive lasso for time series regression models. *arXiv preprint arXiv:1312.1473*, 2015.
- F. Audrino and S. D. Knaus. Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, pages 1–37, 2016.
- B. Becker. Starbucks’ bitter rewards – good for investors? <http://lipperalpha.financial.thomsonreuters.com/2016/05/starbucks-bitter-rewards-good-for-investors/>, 2016a. Accessed: 2010-09-30.
- B. Becker. Sentiment may have bottomed for volkswagen. <http://seekingalpha.com/article/3957529-sentiment-may-bottomed-volkswagen>, 2016b. Accessed: 2010-09-30.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- S. Borovkova. The role of news in commodity markets. *Available at SSRN 2587285*, 2015.
- S. Borovkova and A. Lammiman. The impact of news sentiment on energy futures returns. *Vrije Universiteit Amsterdam*, http://www.feweb.vu.nl/en/Images/paper%20borovkova_tcm97-204330.pdf, 2010.
- S. Borovkova and D. Mahakena. News, volatility and jumps: the case of natural gas futures. *Quantitative Finance*, 15(7):1217–1242, 2015.
- S. Borovkova, E. Garmaev, P. Lammers, and J. Rustige. Sensr: A news sentiment-based systemic risk indicator. *Available at SSRN*, 2016.
- M. Brenner and D. Galai. New financial instruments for hedge changes in volatility. *Financial Analysts Journal*, 45(4):61–65, 1989.

Bibliography

- C. T. Brownlees and R. F. Engle. Srisk: A conditional capital shortfall index for systemic risk measurement. *Department of Finance, New York University*, 2015.
- P. Bühlmann and S. Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.
- R. Cahan, J. Jussa, and Y. Luo. Breaking news: How to use news sentiment to pick stocks. *Macquarie US Equity Research*, 2009.
- F. X. Diebold and R. Mariano. Comparing forecast accuracy. *Journal of Business and*, 1995.
- X. Ding, Y. Zhang, T. Liu, and J. Duan. Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (ICJAI'15)*, pages 2327–2333, 2015.
- J. Durbin and S. Koopman. Time series analysis by space state methods. *Nova Iorque: Oxford University Press*, 2001.
- S. D. Erawan. *Essays on Behavioural Approach in Finance*. PhD thesis, University of St. Gallen, 2015.
- J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.
- L. Fang and J. Peress. Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052, 2009.
- C. Francq and J.-M. Zakoian. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2011.
- J. Friedman, T. Hastie, and R. Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.
- J. Friedman, T. Hastie, and R. Tibshirani. glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, 1, 2009.
- C. W. Granger. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control*, 2:329–352, 1980.
- M. Gustafsson, M. Hornquist, and A. Lombardi. Constructing and analyzing a large-scale gene-to-gene regulatory network lasso-constrained inference and biological validation. *IEEE/ACM Transactions on computational biology and bioinformatics*, 2(3):254–261, 2005.
- L. Jacob, G. Obozinski, and J.-P. Vert. Group lasso with overlap and graph lasso. In *Proceedings of the 26th annual international conference on machine learning*, pages 433–440. ACM, 2009.
- R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.
- M. D. Kirange, R. R. Deshmukh, H. M. Insaaf, M. Morawaka, W. Wanasinghe, A. Fernando, M. M. Mudugamuwa, and D. Dhammearatchi. Sentiment analysis of news headlines for stock price prediction. 2016.
- C. Li and H. Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, 2008.
- T. Loughran and B. McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

- T. Loughran and B. McDonald. Ipo first-day returns, offer price revisions, volatility, and form s-1 language. *Journal of Financial Economics*, 109(2):307–326, 2013.
- T. Loughran and B. McDonald. Measuring readability in financial disclosures. *The Journal of Finance*, 69(4):1643–1671, 2014.
- T. Loughran, B. McDonald, and H. Yun. A wolf in sheep’s clothing: The use of ethics-related terms in 10-k reports. *Journal of Business Ethics*, 89(1):39–49, 2009.
- A. C. Lozano, N. Abe, Y. Liu, and S. Rosset. Grouped graphical granger modeling for gene expression regulatory networks discovery. *Bioinformatics*, 25(12):i110–i118, 2009.
- A. Lugmayr and G. Gossen. Evaluation of methods and techniques for language based sentiment analysis for dax 30 stock exchange a first concept of a “lugo” sentiment indicator. *International SERIES on Information Systems and Management in Creative eMedia*, (1):69–76, 2013.
- R. Luss and A. d’Aspremont. Predicting abnormal returns from news using text classification. *Quantitative Finance*, 15(6):999–1012, 2015.
- P. K. Narayan and D. Bannigidadmath. Does financial news predict stock returns? new evidence from islamic and non-islamic stocks. *Pacific-Basin Finance Journal*, 2015.
- R. Nickerson. *Mathematical reasoning: Patterns, problems, conjectures, and proofs*. Taylor & Francis, 2011.
- J. Peng, P. Wang, N. Zhou, and J. Zhu. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*, 2012.
- J. Peress. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *The Journal of Finance*, 69(5):2007–2043, 2014. ISSN 1540-6261. doi: 10.1111/jofi.12179. URL <http://dx.doi.org/10.1111/jofi.12179>.
- R. L. Peterson. *Trading on Sentiment: The Power of Minds Over Markets*. John Wiley & Sons, 2016.
- M. Reilly, D. Posadas-Sanchez, L. Kettle, and P. Killeen. Making the trip worthwhile: do rats (*rattus norvegicus*) and pigeons (*columba livia*) forage prospectively. *Behavioural Processes (under review)*, 2011.
- J. Russel. Acircumplexmodelofafect. *Journal of Personalityand*, 1980.
- W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.
- T. Shimamura, S. Imoto, R. Yamaguchi, and S. Miyano. Weighted lasso in graphical gaussian modeling for large gene network estimation based on microarray data. *Genome Informatics*, 19:142–153, 2007.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.
- R. S. Tsay. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.

- M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.
- H. Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.
- H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Appendices

1 List of the companies

Name	TRBCEconomicSector
Bangkok Airways PCL	Industrials
Asset Acceptance Capital Corp	Industrials
Apple Inc	Technology
Abcam PLC	Healthcare
Abbott Laboratories	Healthcare
AES Corp	Utilities
Amgen Inc	Healthcare
Amazon.com Inc	Consumer Cyclical
Anadarko Petroleum Corp	Energy
Avery Dennison Corp	Industrials
American Water Works Company Inc	Utilities
American Express Co	Financials
Boeing Co	Industrials
Bank of America Corp	Financials
Baxter International Inc	Healthcare
BlackRock Inc	Financials
Ball Corp	Basic Materials
Citigroup Inc	Financials
Caterpillar Inc	Industrials
Cigna Corp	Financials
Colgate-Palmolive Co	Consumer Non-cyclical
CMS Energy Corp	Utilities
CVS Caremark Corp	Consumer Non-cyclical
Cisco Systems Inc	Technology
CenturyLink Inc	Telecommunication Services
Chevron Corp	Energy
Dow Chemical Co	Basic Materials
DTE Energy Co	Utilities
Devon Energy Corp	Energy
Ford Motor Co	Consumer Cyclical
Frontier Communications Corp	Telecommunication Services
Corning Inc	Technology
General Motors Co	Consumer Cyclical
Hasbro Inc	Consumer Cyclical
Harley-Davidson Inc	Consumer Cyclical
Hewlett-Packard Co	Technology
International Business Machines Corp	Technology
International Flavors & Fragrances Inc	Consumer Non-cyclical
Johnson & Johnson	Healthcare
JPMorgan Chase & Co	Financials
Coca-Cola Co	Consumer Non-cyclical
Eli Lilly and Co	Healthcare
Masco Corp	Consumer Cyclical
Mattel Inc	Consumer Cyclical
McDonald's Corp	Consumer Cyclical
MetLife Inc	Financials
Merck KGaA	Healthcare
Altria Group Inc	Consumer Non-cyclical
Merck & Co Inc	Healthcare
Morgan Stanley	Financials
Microsoft Corp	Technology
Nike Inc	Consumer Cyclical
National Oilwell Varco Inc	Energy
Eversource Energy	Utilities
Nucor Corp	Basic Materials
Oracle Corp	Technology
PG&E Corp	Utilities
PepsiCo Inc	Consumer Non-cyclical
Pfizer Inc	Healthcare
Procter & Gamble Co	Consumer Non-cyclical
Praxair Inc	Basic Materials
Sealed Air Corp	Basic Materials
Schlumberger NV	Energy
Southern Co	Utilities
AT&T Inc	Telecommunication Services
Target Corp	Consumer Cyclical
Textron Inc	Industrials
United Parcel Service Inc	Industrials
Vulcan Materials Co	Basic Materials
Verizon Communications Inc	Telecommunication Services
Wells Fargo & Co	Financials
Williams Companies Inc	Energy
Wal Mart Stores Inc	Consumer Cyclical
Xcel Energy Inc	Utilities
Exxon Mobil Corp	Energy
Yahoo! Inc	Technology
Cambridge Antibody Technology Group PLC	Healthcare
Bell Aliant Inc	Telecommunication Services

Table 3: The US companies used in the current studies.

Name	TRBCEconomicSector
Anglo American PLC	Basic Materials
Alcatel Lucent SA	Technology
Antofagasta PLC	Basic Materials
Anglo Pacific Group PLC	Basic Materials
Daisy Group PLC	Telecommunication Services
Electricite de France SA	Utilities
Eni SpA	Energy
Glencore PLC	Energy
Iberdrola SA	Utilities
Industria de Diseno Textil SA	Consumer Cyclical
Nestle SA	Consumer Non-cyclical
Rio Tinto PLC	Basic Materials
Roche Holding AG	Healthcare
SABMiller PLC	Consumer Non-cyclical
Swisscom AG	Telecommunication Services
STMicroelectronics NV	Technology
Sirius Minerals PLC	Basic Materials
Syngenta AG	Basic Materials
Telecity Group PLC	Technology
Telenor ASA	Telecommunication Services
United Utilities Group PLC	Utilities
Vitec Group PLC	Technology
Wolfson Microelectronics PLC	Technology
Abengoa Yield PLC	Utilities
ABB India Ltd	Industrials
Aditya Birla Minerals Ltd	Basic Materials
ArcelorMittal SA	Basic Materials
AstraZeneca PLC	Healthcare
Banco Bilbao Vizcaya Argentaria SA	Financials
GlaxoSmithKline PLC	Healthcare
Merck KGaA	Healthcare
Novartis	Healthcare
Banco Santander SA	Financials
Total Energy Services Inc	Energy
Anheuser-Busch Companies LLC	Consumer Non-cyclical
HSBC Holdings	Financials
Lloyds Banking Group	Financials
BNP Paribas	Financials
Allianz	Financials
UBS Group	Financials
Deutsche Bank	Financials
Logitech International	Technology
Infineon Technologies	Technology
SAP SE	Technology
Siemens	Industrials
Airbus	Industrials
Schneider Electric	Industrials
LINDE	Industrials
Vinci	Industrials
Glencore	Industrials
ThyssenKrupp	Basic Materials
BASF	Basic Materials
Anglo Pacific Group	Basic Materials
EOAN	Utilities
National Grid	Utilities
Enel SPA	Utilities
Engie SA	Utilities
Energie Baden-Wuerttemberg	Utilities
DTE Energy	Utilities
Orange	Telecommunication Services
Sanofi	Healthcare
Bayer	Healthcare
Bosch	Consumer Non-cyclical
Continental	Consumer Non-cyclical
Man SE	Consumer Non-cyclical
Peugeot	Consumer Non-cyclical
Volkswagen	Consumer Non-cyclical
Daimler	Consumer Non-cyclical
Sie de Saint-Gobain	Consumer Non-cyclical
BMW	Consumer Non-cyclical
Royal Dutch Shell	Energy
British Petroleum	Energy
Unilever	Consumer Non-cyclical
SabMiller	Consumer Non-cyclical
L'oreal	Consumer Non-cyclical
Moet Hennessy Louis Vuitton SE	Consumer Non-cyclical
Diageo PLC	Consumer Non-cyclical

Table 4: The European companies used in the current studies.

2 Overall relevance and strength of selected sectors and countries for US and European companies

Fig. 2: The overall relevance of the news sentiments of the selected sectors and countries on the US stocks ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 3: The overall strength of the news sentiments of the selected sectors and countries on the US stocks ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 4: The overall relevance of the news sentiments of the selected sectors and countries on the European stocks ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 5: The overall strength of the news sentiments of the selected sectors and countries on the European stocks ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

3 Strength by sector for US companies

Fig. 6: The strength of 1 day lags of the news sentiments of the YCY sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 7: The strength of 1 day lags of the news sentiments of the UTL sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 8: The strength of 1 day lags of the news sentiments of the TEC sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 9: The strength of 1 day lags of the news sentiments of the NCY sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 10: The strength of 1 day lags of the news sentiments of the MAT sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 11: The strength of 1 day lags of the news sentiments of the IND sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 12: The strength of 1 day lags of the news sentiments on the HLC sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 13: The strength of 1 day lags of the news sentiments of the FIN sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 14: The strength of 1 day lags of the news sentiments of the ENE sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 15: The strength of 1 day lags of the news sentiments of the COM sector for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 16: The strength of 1 day lags of the news sentiments of the US-related sentiment for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 17: The strength of 1 day lags of the news sentiments of the IT-related sentiment for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 18: The strength of 1 day lags of the news sentiments of the GR-related sentiment for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 19: The strength of 1 day lags of the news sentiments of the DE-related sentiment for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 20: The strength of 1 day lags of the news sentiments of the CN-related sentiment for the US companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

4 Strength by sector for European companies

Fig. 21: The strength of 1 day lags of the news sentiments of the YCY sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 22: The strength of 1 day lags of the news sentiments of the UTL sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 23: The strength of 1 day lags of the news sentiments of the TEC sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 24: The strength of 1 day lags of the news sentiments of the NCY sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 25: The strength of 1 day lags of the news sentiments of the MAT sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 26: The strength of 1 day lags of the news sentiments of the IND sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 27: The strength of 1 day lags of the news sentiments on the HLC sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 28: The strength of 1 day lags of the news sentiments of the FIN sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 29: The strength of 1 day lags of the news sentiments of the ENE sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 30: The strength of 1 day lags of the news sentiments of the COM sector for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 31: The strength of 1 day lags of the news sentiments of the US-related sentiment for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 32: The strength of 1 day lags of the news sentiments of the IT-related sentiment for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 33: The strength of 1 day lags of the news sentiments of the GR-related sentiment for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 34: The strength of 1 day lags of the news sentiments of the DE-related sentiment for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.

Fig. 35: The strength of 1 day lags of the news sentiments of the CN-related sentiment for the European companies ranging from the 1st of Jan., 2005 to the 31st of Dec., 2014.