School of Finance

University of St.Gallen

Do Newspaper Articles Predict Aggregate Stock Returns?

Manuel Ammann
Roman Frey
Michael Verhofen

Do Newspaper Articles Predict
Aggregate Stock Returns?

Manuel Ammann, Roman Frey, and Michael Verhofen[*]

University of St. Gallen

This version: 3rd December 2011

Abstract

We analyze whether newspaper content can predict aggregate future stock returns. Our study is based on articles published in the *Handelsblatt*, a leading German financial newspaper, from July 1989 to March 2011. We summarize newspaper content in a systematic way by constructing word-count indices for a large number of words. Word-count indices are instantly available and therefore potentially valuable financial indicators. Our main finding is that the predictive power of newspaper content has increased over time, particularly since 2000. We find that a cluster analysis approach increases the predictive power of newspaper articles substantially. To obtain optimal predictive power, we need at least seven clusters. Our analysis shows that newspaper content is a valuable predictor of future DAX returns in and out of sample.

Keywords: Word Count, Text Mining, Expected Returns,
            Tactical Asset Allocation
JEL classification: G10

---

[*]Manuel Ammann is Professor of Finance at the University of St. Gallen, Switzerland (Swiss Institute of Banking and Finance, Rosenbergstrasse 52, 9000 St. Gallen, Switzerland, manuel.ammann@unisg.ch), Roman Frey (corresponding author) is Research Assistant at the University of St. Gallen, Switzerland (Swiss Institute of Banking and Finance, Rosenbergstrasse 52, 9000 St. Gallen, Switzerland, roman.frey@unisg.ch) and Michael Verhofen is Lecturer of Finance at the University of St. Gallen, Switzerland and Portfolio Manager at Allianz Global Investors in Frankfurt, Germany (Allianz Global Investors, Mainzer Landstrasse 11-13, 60329 Frankfurt, Germany, verhofen@gmail.com).

1

1 Introduction

We analyze the predictive power of newspaper articles for stock market returns in Germany. Stock market prediction has attracted much attention in financial research. Popular predictors studied include economic and financial data such as past stock returns, macroeconomic variables, and time series of balance-sheet data. In addition to economic and financial data, daily newspapers also may contain relevant information for financial markets. We find that newspaper articles predict future DAX excess returns in and out of sample. Our results show that words such as "claim," "minus," and "black" predict stock returns. We find that the predictive power of newspaper content increased after 2000.

Text mining—similar to data mining—aims at finding patterns in unstructured text documents. Word-count indices are one technique for summarizing and quantifying text content. Word-count indices count the number of articles mentioning a specific word. A well-known word-count index is the "R-Word Index" constructed by *The Economist* (The Economist (1998)). The R-Word Index counts how many stories in a particular range of newspapers include the word "recession" within a certain period of time. The index can be regarded as an alternative indicator of economic activity and has accurately predicted the start and end of several recessions (The Economist (2002)). A similar indicator is the MarketPsych Fear Index. This index shows the relation between a moving average of the percentage of "fear" words in the U.S. financial news press and the Nasdaq 100 index (MarketPsych (2010)).

Word-count indices have major advantages. They are—unlike official and commonly used statistics—instantly available at any given point in time. Adverse time delays and lags can be avoided. Furthermore, word-count indices are potentially uncorrelated with other predicting variables, such as, for example, macroeconomic data. This makes them a valuable source of additional information that cannot be exploited by common predictors. In principle, increasing media coverage broadens the spectrum of structured

2

language to describe financial markets.

The goal of this paper is to analyze whether quantitative measures of newspaper language predict stock market returns in Germany. We particularly aim at assessing how including newspaper content—in addition to well-established predictor variables—impacts the performance of a market prediction model. We analyze the predictive power of both single words and clusters of words.

We make several contributions to the literature. First, we analyze the relation between information in newspaper articles and future stock market returns. Specifically, we conduct an in- and out-of-sample analysis to assess the predictive power of summarized newspaper content for the aggregated German stock market. We show whether newspaper articles provide valuable information that can be used for investment purposes. Second, instead of manually aggregating words into a few sentiment, interest, or relevance classes as is done in most other studies, we perform a cluster analysis and study the optimal level of fragmentation of quantitative measures of newspaper content for stock market prediction. Finally, we extend our study by analyzing the relation between newspaper articles and the development of Germany's real economy.

We find that newspaper articles predict German stock market returns in and out of sample. Summarized newspaper content is significantly correlated with and Granger-causes future stock returns. Our results show that this predictive power has increased in recent years. We find that using certain groupings of words—constructed by a cluster analysis—achieves the highest forecasting power for Germany's aggregated stock market. To obtain optimal predictive power, we need at least seven clusters. Our findings suggest that apparently random newspaper data can be structured and used to predict stock market changes.

The remainder of this paper proceeds as follows. Section 2 provides a brief literature review. Section 3 describes the dataset and the algorithm used to obtain the data. Section 4 discusses the different research designs

implemented in this study, and Section 5 presents the empirical in- and out-of-sample results. Section 6 discusses the robustness of our empirical findings. Section 7 concludes.

2 Literature Review

The body of literature analyzing the role the media plays in financial markets and the economy is growing rapidly. Here, we confine ourselves to briefly summarizing those studies that are most similar to our paper. Tetlock (2007) measures the interactions between the media and the stock market using content from a popular *Wall Street Journal* column. He finds that high media pessimism robustly predicts downward pressure on market prices and that unusually high or low pessimism forecasts high market trading volume. Tetlock, Saar-Tsechansky & Macskassey (2008) study a wide selction of newspaper articles in the United States and analyze the relation between individual firms' accounting earnings and stock returns. They find that stock prices under react to negative information and that the predictive power of negative words is largest for stories that focus on fundamentals.

Bollen, Mao & Zeng (2011) investigate whether public mood, as measured from tweets posted on twitter.com, is correlated with or even predictive of Dow Jones Industrial Average values. They find that the accuracy of Dow Jones predictions can be significantly improved by the inclusion of the public mood dimension. Tetlock (2011) shows that stock market returns generally respond less to old (stale) than to new information about firms. Griffin, Hirschey & Kelly (2011) find that in most developed countries, a firm's stock is more volatile on days on which there is public news about the firm. In emerging markets, however, volatility does not respond to either news or no-news days. Gross-Klussmann & Hautsch (2011) show that sentiment news items predict high-frequency returns, trading volume, and volatility on the London Stock Exchange. Carretta, Fiordelisi & Schwizer (2011) study the interactions between corporate governance news and the Italian stock market. They find that investors are sensitive to news stories

about profitable corporations. Fang & Peress (2009) find that the returns of stocks that have received high media coverage are lower than the returns of stocks receiving low media coverage.

Several studies analyze the general impact of news stories on financial markets. Engelberg & Parsons (2011) find that local press coverage significantly affects the daily trading volume of local retail investors. Das & Chen (2007) apply an algorithm to extract investor sentiment from stock message boards and find that tech-sector postings are related to stock returns. Other studies (e.g., Tetlock (2010), Schumaker & Chen (2009), Philpot & Johnson (2008), Bulkley & Herrerias (2006), Antweiler & Frank (2004), Doms & Morin (2004), Kloptchenko, Eklund, Karlsson, Back, Vanharanta & Visa (2004) , Coval & Shumway (2001)) confirm that news stories are associated with considerable response by the financial sector. For an introduction to text mining in general, the reader is referred to Weiss, Indurkhya, Zhang & Damerau (2004), Konchady (2006), Feldman & Sanger (2006), and Srivastava & Sahami (2009).

3   Data

This study uses the following datasets: newspaper stories (i.e., word-count indices), German stock market data, data on Germany's real economy, data on different control variables, and the risk-free interest rate.

The main dataset comprises newspaper articles. We obtain news stories from the *Handelsblatt*, a leading German financial newspaper. We summarize and quantify the newspaper data by constructing word-count indices. Word-count indices count the number of articles mentioning a specific word. The indices compiled in this study are based on articles published in the *Handelsblatt* from July 1989 through March 2011. Our study is based on a sample of 236 words. Table 1 in Appendix A.1 provides a list of all the words considered in this analysis, irrespective of any levels of significance found later in this paper. In addition the English word, we provide the original German word for which we actually searched. Our empirical results

are presented by the English word. We collect newspaper data in a monthly interval.

We construct the word-count indices as follows. First, we take a list of words frequently used in newspaper articles. We choose a wide range of words, i.e., words of every word class (e.g., nouns, verbs, adjectives, etc.) and category (e.g., positive, negative, different sentiments, etc.). This allows us to capture different market sentiments and conditions for several time periods. In a second step, we browse the online archive of the *Handelsblatt* and search for the occurrence of the individual words. Every article containing at least one instance of a specified word is counted as one hit. The output of the newspaper data collection described above is a matrix containing the number of occurrences of each word and time step, i.e., the values of the word-count indices.

For the aggregate stock market data, we use Germany's main stock market index, the DAX. We obtained the DAX price levels from July 1989 through March 2011 from Bloomberg. For the risk-free interest rate, we take the time series of monthly average of the three-month interest rates (code: SU0107) from the website of the Federal Reserve Bank of Germany (Bundesbank). Furthermore, as a proxy for Germany's economic activity we use data on industrial production in Germany (seasonally adjusted), retrieved from Bloomberg (ticker: GRIPI Index).

4 Research Design

This section presents the univariate, multivariate, and cluster analyses we conduct in the empirical section of the paper. We start with a univariate approach computing the linear correlations between future stock returns and word-count indices. Second, we implement a stepwise linear predictive regression. This approach regresses future stock returns—in excess of the risk-free rate—on a set of predictor variables. In addition to an in-sample regression analysis, we assess the out-of-sample performance. Finally, we discuss the theory of cluster analysis.

To improve statistical properties of the estimators, we apply a common technique in empirical finance (e.g., Tetlock et al. (2008)). We standardize the word-count indices by using z-scores. The z-score standardization is defined as

$$z_{i,t} = \frac{x_{i,t} - \mu_x}{\sigma_x},$$ (1)

where $i = 1, \ldots, N$ words and $t = \theta, \ldots, T$, while $\Delta t = 1$ month. $z_{i,t}$ denotes the standardized word-count index of word i at time t while $x_{i,t}$ denotes the number of occurrences of the same word. μ_x and σ_x are the mean and the standard deviation of x_i over the previous θ months. We choose $\theta = 24$ months.

4.1 Univariate Analysis

We provide a general overview of possible relations between the structured language content of the *Handelsblatt* and future DAX returns by conducting an univariate analysis. Our goal is to gain some initial insight into the development and the persistence of relations between newspaper stories and stock market behavior.

We compute the correlation coefficients, $\rho(z_{i,t}, r_{t+1})$, between the individual standardized word-count indices and future DAX excess returns, r_{t+1}. Furthermore, we determine the causal direction of the forecasting relationship by conducting a bivariate Granger causality test (Granger (1969)). If quantitative language measures partially cause stock market movements, then they, by definition, must precede these movements and therefore will be valuable for stock market prediction.

4.2 Multivariate Analysis: Stepwise Regression

There is a wide range of econometric tools for predicting asset returns. Since this study is based on a highly multivariate dataset, we apply a stepwise regression procedure according to Draper & Smith (1998). The stepwise selection method is a systematic forward procedure. It adds and removes

7

potential regressors from a multilinear model specification depending on their statistical significance in a regression estimation.

The stepwise method starts with an initial model containing the most correlated regressor—in this paper, the word-count indices, z_i—with the dependent variable. Then, it systematically checks the explanatory power of successively smaller or larger model specifications. With each added or removed word-count index the stepwise algorithm compares the p-values of the F-statistics based on the models as they existed before and after adding or removing an additional regressor. Given that a regressor is not considered in the model at a certain step, the null hypothesis is that the regressor would have a regression coefficient equal to zero if included in the estimation model. The regressor is eventually added to the model if there is statistical evidence to reject the null hypothesis. If a regressor is already in the model, however, it will be removed if there is no evidence to reject the null of a zero coefficient. Specifically, the stepwise algorithm is implemented as follows:

1. To determine the initial model, we compute the correlations of all possible predictor variables with the dependent variable. The independent variable with the highest correlation is taken as the first regressor.

2. Next, we systematically check whether there are any regressors not considered in Step 1 that have p-values lower than a previously fixed entrance tolerance. In this paper, this threshold is 2.5%. If there are such regressors, we include the one with the smallest p-value. We now repeat this step until there are no potential regressors with p-values lower than the entrance threshold and then proceed to Step 3.

3. After adding all regressors with significant predictive power to the model in a stepwise fashion, we check the model for any p-values higher than the exit tolerance. We fix the exit tolerance at 2.5% to match the entrance threshold. If there are variables in the model with p-values higher than the threshold, we eliminate the one with the highest p-

value and return to Step 2. Otherwise, the algorithm is stopped at this point.

The stepwise algorithm thus ends as soon as consideration of any further regressors would lead to a deterioration of the overall model fit.

We define the general stepwise regression equation described in the algorithm above as

$$r_{t+1} = \alpha + \sum_{j=1}^{J} \beta_j z_{j,t} + \gamma C_t + \varepsilon_t, \qquad (2)$$

where r_{t+1} denotes the excess return of the dependent variable at time $t+1$. α is the intercept of the regression and β_j denotes the exposures of r_{t+1} to the J word-count indices z_t at time t. γ reflects the vector of factor loadings for the control variables. C_t refers to the vector containing all control variables considered in the stepwise regression. ε_t denotes the zero-mean normally distributed residual.

In each stepwise regression, we include a set of control variables. The motivation for doing so is to evaluate whether newspaper articles predict future stock returns beyond established origins of predictability. Basically, we rely on standard variables found to have potential forecasting power in previous studies (e.g., Schrimpf, Schröder & Stehle (2007), Walkshaeusel & Lobe (2011)). If a control variable is a return itself, we take excess returns over DAX returns to avoid multicollinearity. In particular, we include the German Fama & French (1993) factors for the size (SMB) and the book-to-market ratio (HML). These are based on the S&P German Growth and Value indices. To control for possible autocorrelation in DAX log returns, we take a lagged DAX excess return, r_t. In addition, we control for influences from foreign stock markets by including a lagged MSCI Europe ex Germany, $MSCI^{EeG}$, return factor.

For macroeconomic influences we add the term spread, TS. Term spread is defined as the difference between the 10-year and the 1-year government bond yield. Additionally, we control for the development of Germany's

consumer price index. Since the index level is not stationary, we take a detrended log of the index, CPI. We compute the CPI trend component as a rolling average of the previous 12 months of the log CPI.

Finally, we control for past volatility in Germany's stock market. As a proxy for past volatility we include the first lag of the detrended squared DAX return residuals according to Tetlock et al. (2008). We demean the DAX log returns to obtain return residuals. After squaring the residuals, we subtract the past 36-month moving average of the squared DAX residuals. This procedure yields the control variable $vola$. The vector C_t containing all control variables we consider is given by $C_t = [SMB, HML, r_t, MSCI^{EeG}, TS, CPI, vola]$.

4.3 Out-of-Sample Analysis

In this subsection, we focus on the actual forecasting power of newspaper articles. In-sample analyses primarily assess the models' goodness of fit. Thus, the pure in-sample stepwise regression approach is prone to data mining. It might yield high R^2, giving the appearance of good model fit, but which, in fact, may be at least partially due to over-fitting. Therefore, we conduct an out-of-sample analysis to test the actual predicting power of word-count indices. In this paper, out-of-sample performance is defined as the ability of newspaper articles to predict the one-step-ahead development of DAX returns.

We assess the forecasting performance of newspaper stories with a rolling stepwise regression. Using rolling windows maximizes the number of possible forecasts we can produce. For the rolling regression we take a standard rolling window size of 60 months, $k = 60$. Additionally, we run the regression based on different rolling window sizes to check for any dependence on particular lengths of the time window. We stepwise regress the dependent variable on the standardized word-count indices

$$r_{t-k+2,t+1} = \alpha + \sum_{j=1}^{J} \beta_j z_{t-k+1,t}^j + \gamma C_{t-k+1,t}, \tag{3}$$

while $t = k, \ldots, T - 1$ observations. Based on the length of the time series of our sample, we obtain $T - k$ estimates of the regression coefficients. We denote them by α_l, $\beta_{j,l}$, and γ_l, while l is the rolling window index and $l = 1, \ldots, T - k$.

After computing the regression coefficients based on the rolling regression in Equation (3), we calculate the fitted one-step-ahead forecasts of the dependent variable:

$$\hat{r}_{t+1} = \alpha_l + \sum_{j=1}^{J} \beta_{j,l} z_{j,t} + \gamma_l C_t, \tag{4}$$

where $t = k, \ldots, T - 1$ and $l = 1, \ldots, T - k$.

We take two approaches to assess the accuracy of the forecasted values, $\hat{r}$. First, we plot the realized (observed) against the predicted returns and compute the best-fitting regression line. We run this test by regressing

$$r_t = v + w \cdot \hat{r}_t, \tag{5}$$

where v and w are the intercept and the slope, respectively, and estimated by ordinary least squares. If w is statistically different from zero, it can be concluded that the underlying model has statistically relevant forecasting power. We test w by an ordinary test statistic $t_w = w/\text{se}(w)$, where $\text{se}(w)$ denotes the standard error of w.

In our second approach we look at the difference between effectively realized values and the forecasted values. In particular, we compute a root mean squared error (RMSE) for each rolling window described above:

$$\text{RMSE}_h = \sqrt{\frac{\sum_{i=h}^{h+m-1}(\hat{r}_i - r_i)^2}{m}}, \tag{6}$$

where $h = 1, \ldots, T - k - m + 1$. m denotes the window size on which the RMSE computation is based. We take $m = 36$ months. The variable k again denotes the rolling window size of the rolling regression.

4.4 Cluster Analysis

The goal of a cluster analysis is to find useful classifications in multivariate data that are closely knit statistically and yet distinct from each other (Kettenring (2006)). It is a sort of learning process, meaning that we aim to discover hidden patterns in certain groupings recognizable in seemingly random data—in this paper, language in newspapers.

The empirical analysis of this paper builds on numerous word-count indices and is thus highly multivariate. In other empirical papers analyzing media content, e.g., Doms & Morin (2004), Tetlock (2007), and Tetlock et al. (2008), dimensionality is reduced by manually categorizing language. Tetlock (2007), as well as Tetlock et al. (2008), for instance, collapse their word-count matrix into two categories and build a positive and negative word category according to the Harvard-IV-4 psychosocial dictionary. Bollen et al. (2011) summarize Twitter statements in six proxies to measure human mood.

Our approach is different in that instead of categorizing words according to exogenous norms, we follow data directly. By running a cluster analysis algorithm, newspaper data are assigned to different groupings based on statistically similar patterns. Therefore, clusters are built and thus dimensionality reduced according to information comprised in newspaper articles. Accordingly, we believe cluster analysis to be an appropriate methodology for reducing dimensionality in structured newspaper language.

We build hierarchical clusters according to a linkage algorithm using Euclidean distances between the word-count indices. At each step, the cluster algorithm identifies the two clusters closest to each other and combines them into a new one (Allik & McCrae (2004)). We run this algorithm until all clusters—and thus all words of our dataset—are linked. Specifically, we implement a complete linkage method, also called furthest distance. This method considers the largest distance between objects in two clusters:

$$d(a,b) = \max(\text{dist}(s_{ai}, s_{bj})), \quad i \in (1, \ldots, q_a), j \in (1, \ldots, q_b). \tag{7}$$

a and b denote clusters, while s_{ai} is the ith object in cluster a. q denotes the number of objects in a cluster.

We use the output of the cluster analysis, i.e., the groupings of words, in a graphical and formal approach. In the graphical method, we show the results of the cluster analysis as a hierarchical tree diagram, also called a dendrogram. A dendrogram starts by taking single words. Then, the statistically closest two words are combined in one cluster. Intuitively, we would expect early joints to be words related in some way, e.g., by connotation, relevance, or interest.

With this formal methodology we determine the optimal number of clusters for explaining and forecasting future DAX excess returns. We again conduct an in- and out-of-sample analysis, but based on word-count clusters instead of individual words. In contrast, for a stepwise regression approach, it is the stepwise algorithm that—based on statistical significance—determines which words to include in the predicting analysis. In this subsection, we do not exclude any words but, instead, build clusters in a first step considering the entire sample, i.e., all word-count indices. Afterward, we run the regression analysis again based on the different clusters built from the word-count indices.

In the forecasting analysis we follow an approach very similar to that described for the out-of-sample analysis in Section 4.3. We first compute the predicted future DAX excess returns as defined in Equations (3) and (4). However, instead of the stepwise approach, we now instead use word-count clusters as explanatory variables. Second, to assess the quality of the predicted DAX excess returns, we regress the realized on the predicted DAX returns as denoted in Equation (5) and test the resulting slope coefficients for being statistically different from zero.

The goal of this approach is twofold. First, we run the regressions on a different number of clusters and check the individual goodness of fit of the regression—measured by the adjusted R^2. This allows us to determine the optimal level of fragmentation and thus the number of clusters that best

13

explain future stock returns. Second, we analyze the effective forecasting power of word-count clusters—compared to individual words as predicting variables—for future stock market developments.

5   Empirical Results

In this section, we present the empirical results of the univariate, multivariate, out-of-sample, and cluster analyses. We perform the empirical analysis for six subperiods: the entire sample (All, 1989/7–2011/3), first half of the sample (1st Half, 1989/7–1999/6), second half of the sample (2nd Half, 1999/7–2011/3), stock market boom during the 1990s (Boom, 1993/12–1998/8), final stage of the dotcom bubble including the post-crisis period (Dotcom, 1999/7–2004/3), and the subprime crisis (Subprime, 2006/1–2010/8).

These subperiods are chosen for their different characteristics—primarily the magnitude of average return and volatility—so as to assess whether the predictive power of newspaper articles depends on specific market sentiments. We investigate whether the predictive power is constant over time or if it depends on financial market development.

5.1   Univariate Analysis: Results

The results of the correlation and the Granger causality analysis provide a first understanding of possible relations and causalities between Germany's aggregate stock market index and newspaper stories.

Figure 1 shows an example of a word-count index, the "Recession Index." We plot the lagged monthly number of occurrences of the word "recession" in the *Handelsblatt* and the normalized DAX index level, denoted by $DAXn_{t+1}$. During the entire sample period, the correlation between the Recession Index, $x_{rec,t}$, and the DAX index level is $\rho(x_{rec,t}, DAXn_{t+1}) = 0.03$. According to The Economist (2002), one would expect the Recession Index to be negatively correlated with the aggregate stock market. Although the correlation is almost zero, of Figure 1 suggests a relationship between these

14

two variables.

Occurrence of the recession word soared during the subprime crisis, resulting in a structural break. We account for changing market conditions by computing the correlation for two subperiods individually. We find that the correlation coefficient based on a sample that ends at the beginning of the subprime period is -0.40. Taking into account the subprime crisis only results in a correlation of -0.71. By considering different market behavior our results thus confirm the negative relation between the occurrence of the word recession and the stock market development as claimed by The Economist (2002).

Table 2 presents the correlation coefficients between word-count indices and future DAX excess returns, $\rho(z_{i,t}, r_{t+1})$. For the entire sample period, we find that 25 words are correlated with future DAX returns at a significance level of 2.5%. To save space, we do not report correlations statistically insignificant at 2.5%. The majority (21 words) is positively related to the DAX while we find only four words with a negative correlation. Correlation coefficients range from -0.163 for the word "minus" to 0.205 for "claim [v]." Furthermore, we find that 60%, or 15 of 25, words are nouns.

Considering the subperiods, we find that during both the first half of our sample and the boom period only two words are significant, while there are 10 words for the second half. The results for the dotcom and the subprime periods paint a different picture. On the one hand, for the dotcom period we find positively related words only, in total 28. No words are negatively correlated. Again, the word "claim [v]" has the highest coefficient at 0.440. On the other hand, during the subprime crisis, all the significant words we find are negatively related to the DAX. The lowest correlation is for the word "worse" at -0.401. For all subperiods, approximately two thirds of the significant words are nouns.

Our findings show that the explanatory power of newspaper articles tends to increase in the second half of the sample. Possible reasons for this could include, first, that the first decade of 2000 has been affected by major turmoil

on financial markets. Volatile markets have attracted additional attention and thus might have increased the influence of newspaper content on stock markets. Second, there has been more detailed press coverage of financial news. That is, recent developments in the financial sector have an increasingly important position in the daily press.

The algebraic signs of the correlation coefficients mostly coincide with what we would expect from a word's connotation. That is, we expect a word with a positive connotation to be positively correlated with development of the DAX index returns. Examples during the entire sample period are "challenge [n]" with a correlation of 0.139 and "minus" with one of -0.163. Two words at variance with their connotations are "consumption" and "warning."

The results of the dotcom and the subprime periods suggest a shift in focus of general news coverage. During the dotcom and post-crisis period, we nearly always only find significant words with positive connotation, even though the DAX index dropped by more than two-thirds. When the number of articles containing positive words decreased, the DAX index fell, too. In other words, news coverage became *less positive* during the dotcom period. For the subprime crisis, however, we only find words with a negative connotation. As the number of articles containing negative words increased, the DAX index lost half its value. Therefore, news coverage was not less positive, but *increasingly negative*, during the subprime period.

To analyze the causal direction of the relation between newspaper stories and DAX returns we employ a bivariate Granger causality test. We find that word-count indices Granger-cause DAX returns. Our results show that the majority of words in our sample Granger-cause DAX returns at a 5% significance level. We also test the reverse causal direction and find that DAX returns Granger-cause only a third of all tested words at the same significance level. In both tests we allow for a maximum of five lags. We take this finding not as actual causality, but to mean that newspaper content has predictive power regarding future DAX movements.

5.2 Multivariate Analysis: Stepwise Regression Results

In this section, we present the results of the stepwise regression analysis. We run the stepwise regression according to the algorithm described in Section 4.2. To avoid multicollinearity among predictor variables, we check correlations between all control variables, finding them to be low to moderate.

5.2.1 Aggregate Stock Market

Table 3 presents the results of the stepwise regression of future DAX excess returns on word-count indices and control variables as defined in Equation (2). We find that during the entire sample period, five words—two positive and three negative—have a factor loading statistically significant for an entrance and exit threshold of 2.5%. For the various subsamples, we find newspaper articles to have stronger explanatory power during the second half of the sample period. During the boom period in the 1990s, no words are significant. We find three and six significant word-count indices for the dotcom and the subprime periods, respectively. In general, the multivariate analysis—including control variables—yields fewer words with significant explanatory power than the univariate analysis.

The main result of Table 3 is that newspaper articles continue to have significant explanatory power for the aggregate stock market after including a set of control variables. Our findings are in line with the results of the univariate analysis. Coefficients of words that are significant in both the univariate and multivariate analysis have the same algebraic signs. We find "claim [v]," "consumption," "minus," "private," and "worse" to have significant coefficients in both analyses within the same subperiod. Again, it is primarily nouns that have significant explanatory power.

Interestingly, for the second half we not only find more significant words but also larger coefficients. One possible interpretation is that the first decade of 2000 was characterized by strong market movements during economic turmoil. Consequently, the general news coverage was dense and might have increasingly affected aggregate stock markets. This interpreta-

17

tion is confirmed by results for the dotcom and subprime subperiods. For both periods we find considerably larger coefficients than for the first half of the sample or the boom period during 1990s.

Figure 2 is a heat map of words found to have significant explanatory power for future DAX development. We run the stepwise algorithm based on a rolling window of 60 months, i.e., for each window we run the algorithm once. Figure 2 shows the persistence of significant words during the sample period. First, we find that no word is consistently chosen by the algorithm. However, "claim [v]"—as already found in the univariate and multivariate analyses—is the word that attains significance most often. Second, during the first six years, i.e., from 1990 to 1996, negative words such as "black," "recession," "panic [n]," "collapse [n]," and "bear market" are often selected. At the beginning of the second half of the sample, we find that "claim [v]," "minus," "anger [n]," and "optimistic" are often significant.

To summarize, we find that newspaper stories explain stock market movements in a multivariate setting. We find that the explanatory power is not persistently loaded on specific words over time. Although there are significant coefficients on words during most of the rolling windows, which words are significant changes. Nonetheless, our findings suggest that during different time periods, words expressing certain sentiments dominate. Turmoil on financial markets seems to stimulate news coverage which enforces explanatory power of newspaper articles.

5.2.2 Real Economy

We next analyze the relation between newspaper content and Germany's real economy. We are interested in whether newspaper stories can explain changes in industrial production, which we take as a proxy for the real economy. Furthermore, we compare coefficients on words found to be significant for the real economy with those significant for the aggregate stock market.

Table 4 reports significant factor loadings on words selected by the stepwise regression approach for explaining changes in industrial production.

The entrance and exit threshold of the stepwise algorithm is 2.5%. During the entire sample period, there are only two words—"decline [n]" and "deflation"—with significant coefficients. Both coefficients are negative and comparatively small. Also, for the first and second halves of the sample we find only two significant factor loadings. Regarding the subperiods, there are nine words significant during the boom phase and non during the dotcom period. During the subprime period, there are eight significant word-count indices; six of them are negative.

Interestingly, it is words having an economic interest or background that tend to be significant, e.g., "deflation," "export [n]," "global economy," etc. Furthermore, during the boom period, the coefficient of the word "government" is -0.95, even though government is not necessarily associated with a negative connotation. A possible interpretation of this finding might have to do with expected interventions by the government. During the mid-1990s the economy was mainly prospering. Therefore, it might have been expected that government would intervene to cool things down, rather than act to stimulate economic development.

In contrast to the results obtained in the previous section based on the stock market, for the real economy we find most significant word-count indices during the boom period of the 1990's, whereas the stock market was most influenced by newspaper content during the first decade of 2000. Finally, we also find that factor loadings for the real economy are substantially smaller than for the stock market analysis.

5.3   Out-of-Sample Analysis: Results

In this section, we present the results of the out-of-sample analysis based on a stepwise regression approach. As all the previously reported results in this paper are in-sample based, the goal of this section is to discover whether individual words of newspaper stories not only explain but also predict stock markets.

In Figure 3 we plot the effectively realized, r, and the predicted, $\hat{r}$, DAX

excess returns. The predicted DAX excess returns are computed according to Equation (4) and based on the standard rolling window of 60 months. The regression line through the scatter plot of Figure 3 is computed by Equation (5). The main result of Figure 3 is that observed and predicted DAX excess returns are positively related during the entire sample period. The regression line in Figure 3 has a slope coefficient of $w = 0.128$ with a corresponding standard error of $\text{se}(w) = 0.081$.

To check for possible dependencies due to the length of the rolling window, we conduct the forecasting analysis for different window sizes. Table 5 reports the results of the out-of-sample regression analysis based on window sizes $k = \{54, 60, 66, 72, 78, 84, 90, 96\}$. The adjusted R^2 and the t-statistics, t_w, of the individual slope coefficients, w, both indicate that the predictive power of word-count indices tends to increase slightly with longer rolling windows. For 90- and 96-month windows we find that word-count indices have statistically significant predictive power for future DAX returns at a 5% level. We take a minimum rolling window size of 54 months to ensure a sufficient number of degrees of freedom in the rolling regression.

The results presented above show that the explanatory power of word-count indices is not constant over time. Therefore, we analyze the development of forecasting performance of newspaper content within the entire sample period. We take a constant rolling window size of 60 months.

Figure 4 plots the root mean squared errors (RMSE) of the predicted stock market returns computed by Equation (6). To compare the predictive power of words with current stock market behavior, Figure 4 additionally plots the development of the normalized DAX index level during the same time period. The shaded areas indicate the boom, dotcom, and subprime subperiods. The development of the RMSE—along with the DAX index level—reveals an interesting characteristic. First, the RMSE level varies over time, finding its trough at the beginning of 2011. Second, and most interesting, the predictive power of word-count indices increases over time. This is confirmed by both the generally falling RMSE and its linear trendline

20

as shown in Figure 4. This result is in line with the findings of the univariate and the in-sample analyses.

In general, the out-of-sample analysis reveals that words in newspapers can be used to predict future stock market movements. The predictive power is superior for longer rolling windows. The forecasting performance varies over time but is increasing during the second half of our sample period.

5.4 Cluster Analysis: Results

In this section, we present the results of the cluster analysis. The main purpose of this approach is to reduce the dimensionality of our highly multivariate word-count dataset. We first report the result of the graphical analysis based on a dendrogram, in which we aimed at finding structures in seemingly random newspaper data. We then present the results of the regression analysis on the optimal level of fragmentation of our dataset to explain and predict future changes in the DAX index.

Figure 5 is a dendrogram of the word-count indices. Initial words are selected by a stepwise regression based on an entrance and exit threshold of 10%. We find that 26 words are significant. We take these initial words to start the dendrogram.

As Figure 5 reveals, early joints in the dendrogram are mostly in line with the interest, meaning, and relevance they carry. For example, "long term" and "future" share a positive perspective of time, "threats [v]" and "falls" are closely linked to failure, "minus" and "shaky" are influenced by uncertainty. At higher levels of clustering, the structure shows patterns according to connotation: "supply [n]" and "resources" are added to the cluster of "long term" and "future," while "negative" joins the "threats [v]" and "falls" cluster. Ultimately, two main clusters of words—except for the term "deflation"—are formed. The top branch is dominated by strong negative words; the majority of words in the bottom carry a positive connotation. In general, the classification in Figure 5 reveals that language in newspaper stories comprises a structure and is not purely random.

5.4.1 Regression Analysis with Word Clusters

The purpose of this section is to identify the optimal level of fragmentation in clustering newspaper language to explain and predict future DAX returns. In the previous section, we predict future stock market movements based on individual words. In this section, we do not use individual words directly but combine them in different clusters. The question we want to answer now is how many clusters yield optimal forecasting accuracy.

We find that the optimal number of clusters for explaining future DAX returns in-sample is 8. We regress future DAX returns on 2 to 10 clusters of words and observe the quality of the regression. Interestingly, the explanatory power of newspaper content structured in a very low number of clusters remains very poor. Our results show that at least seven clusters are required to obtain reasonable explanatory power. This finding is particularly interesting in light on previous research in this area. Extant work—e.g., Tetlock (2007), Tetlock et al. (2008), Bollen et al. (2011), Doms & Morin (2004)—mostly classify media content in a very few groupings of words.

Next, we extend the cluster regression analysis by performing an out-of-sample or forecasting test. As defined in Equation (5), we regress realized, r_t, on predicted, $\hat{r}_t$, DAX returns. Predicted returns are again determined by a regression analysis based on 2 to 10 clusters. We find that newspaper articles have highest predictive power if combined in 7 clusters, which yields an adjusted R^2 of 10.37%. In line with the in-sample results, we again find that a very low number of clusters is not sufficient to predict the stock market.

The most interesting finding is that the forecasting performance of word-count clusters is significantly better than that of a stepwise regression approach. Compared to the stepwise out-of-sample results of Table 5, a cluster analysis obtains substantially higher R^2. Furthermore, the slope coefficient in Equation (5), w, of the prediction methodology based on 7 clusters is significant at a level of 1%.

In summary, our findings suggest two major conclusions. First, combining word-count indices in a reasonable number of clusters predicts the stock market better than do individual words. Second, to obtain optimal explanatory and predictive power, we need at least 7 clusters of words in the clustering algorithm.

6 Robustness Analysis

We now check the robustness of our results. As a dependent variable we replace the 1-month DAX return with the 3-month forward return. The question we focus on is how the predictive power of newspaper articles depends on the time horizon of the stock returns.

Table 6 reports the results of Equation (2) based on 3 months DAX forward excess returns. We take the 3-month average of the word-count indices as the independent variables. Table 6 shows that based on 3 months, more words are significant at an entrance and exit threshold of 2.5%. During the entire sample period, 14 words obtain a significant coefficient, eight of which are negative.

Thus our results for the DAX 3-month forward returns are consistent with previous findings. We find that the explanatory power of newspapers for forward returns increases over time. Furthermore, coefficients of significant word-count indices tend to be larger during the dotcom and the subprime periods. The majority of significant words are nouns. Although explanatory power is not loaded on the same words, our main findings are robust against different investment horizons.

7 Conclusion

In this paper, we analyze the predictive power of newspaper articles for German stock market returns. Based on newspaper articles published between July 1989 and March 2011 in the *Handelsblatt*, a leading German financial newspaper, we use word-count indices as a quantitative language measure.

Word-count indices are instantly available and therefore potentially valuable financial indicators. Our main finding is that newspaper content predicts future DAX returns, both in and out of sample.

Our first result is that word-count indices are significantly correlated with stock returns. Most of these words are also Granger causative of the DAX. Our second result is that newspaper articles explain future stock market movements beyond well-established predictor variables. In the out-of-sample analysis we find that the predictive power of quantitative newspaper language has increased over time, particularly since 2000.

The cluster analysis shows that language in newspaper articles is not random. We find clusters that are consistent with the words' meaning and relevance. Our analysis indicates that seven or more clusters are required to predict stock returns. Therefore, classifying words into a very low number of groups—as previous studies do—does not result in optimal forecasting power. We find that using clusters instead of single words substantially increases the predictive power of newspaper articles.

In summary, we find newspapers to provide additional information valuable for predicting future stock market developments, information that is not considered by common predictors. We therefore conclude that seemingly random newspaper data can be structured and used to predict future stock market returns in Germany.

References

Allik, J. & McCrae, R. R. (2004), 'Toward a Geography of Personality Traits: Patterns of Profiles across 36 Cultures', *Journal of Cross-Cultural Psychology* **35**(1), 13–28.

Antweiler, W. & Frank, M. (2004), 'Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards', *Journal of Finance* **59**(3), 1259–1294.

Bollen, J., Mao, H. & Zeng, X.-J. (2011), 'Twitter Mood Predicts the Stock Market', *Journal of Computational Science* **2**(1), 1–8.

Bulkley, G. & Herrerias, R. (2006), 'Does the Precision of News Affect Market Underreaction? Evidence from Returns following two Classes of Profit Warnings', *European Financial Management* **12**(5), 763–787.

Carretta, A., Fiordelisi, F. & Schwizer, P. (2011), 'The Impact of Corporate Governance Press News on Stock Market Returns', *European Financial Management* **17**(1), 100–119.

Coval, J. D. & Shumway, T. (2001), 'Is Sound Just Noise?', *Jounal of Finance* **56**(5), 1887–1910.

Das, S. R. & Chen, M. Y. (2007), 'Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web', *Management Science* **53**(9), 1375–1388.

Doms, M. & Morin, N. (2004), 'Consumer Sentiment, the Economy, and the News Media', *Federal Rserve Bank of San Francisco Working Paper* (9).

Draper, N. & Smith, H. (1998), *Applied Regression Analysis*, John Wiley & Sons, New York.

Engelberg, J. E. & Parsons, C. A. (2011), 'The Casual Impact of Media in Financial Markets', *Journal of Finance* **66**(1), 67–97.

Fama, E. & French, K. (1993), 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics* **33**, 3–57.

Fang, L. & Peress, J. (2009), 'Media Coverage and the Cross-section of Stock Returns', *Journal of Finance* **64**(5), 2023–2052.

Feldman, R. & Sanger, J. (2006), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press.

Granger, C. W. J. (1969), 'Investigating Casual Relations by Econometric Models and Cross Spectral Methods', *Econmetrica* **37**, 424–438.

Griffin, J. M., Hirschey, N. H. & Kelly, P. J. (2011), 'How Important Is the Financial Media in Global Markets?', *Review of Financial Studies* **24**(12), 3941–3992.

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* **27**(2), 321–340.

Kettenring, J. R. (2006), 'The Practice of Cluster Analysis', *Journal of Classification* **23**, 3–30.

Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanharanta, H. & Visa, A. (2004), 'Combining Data and Text Mining Techniques for Financial Reports', *Intelligent Systems in Accounting, Finance and Management* **12**, 29–41.

Konchady, M. (2006), *Text Mining Application Programming*, Thomson Learning, Boston, MA.

MarketPsych (2010), 'Investor fear'.
    URL: *http://www.marketpsych.com*

Philpot, J. & Johnson, D. T. (2008), 'Mutual Fund Performance and Fund Prospectus Clarity', *Journal of Financial Services Marketing* **11**, 211–216.

Schrimpf, A., Schröder, M. & Stehle, R. (2007), 'Cross-sectional Test of Conditional Asset Pricing Models: Evidence from the German Stock Market', *European Financial Management* **13**(5), 880–907.

Schumaker, R. P. & Chen, H. (2009), 'Textual Analysis of Stock Market Prediction using Breaking Financial News', *Association for Computing Machinery Transactions on Information Systems* **27**(2), Article 12.

Srivastava, A. & Sahami, M. (2009), *Text Mining: Classification, Clustering, and Applications*, Chapman and Hall / CRC.

Tetlock, P. C. (2007), 'Giving Content to Investor Sentiment: The Role of Media in the Stock Market', *Journal of Finance* **62**(3), 1139–1168.

Tetlock, P. C. (2010), 'Does Public Financial News Resolve Asymmetric Information?', *Review of Financial Studies* **23**(9), 3520–3557.

Tetlock, P. C. (2011), 'All the News That's Fit to Reprint: Do Investors React to Stale Information?', *Review of Financial Studies* **24**(5), 1481–1512.

Tetlock, P. C., Saar-Tsechansky, M. & Macskassey, S. (2008), 'More Than Words: Quantifying Language to Measure Firms' Fundamentals', *Journal of Finance* **63**, 1437–1467.

The Economist (1998), 'The Recession Index'. *The Economist*, December 10, 1998.

The Economist (2002), 'Words that can harm you'. *The Economist*, November 23, 2002.

Walkshaeusel, C. & Lobe, S. (2011), 'The Alternative Three-Factor Model: An Alternative beyond US Markets?', *European Financial Management* **00**(0), 1–38.

Weiss, S., Indurkhya, N., Zhang, T. & Damerau, F. (2004), *Text Mining: Predictive Methods for Analyzing Unstructured Information*, Springer.

27

A Appendix

A.1 List of all Words

English	German	English	German
accomplished	geschafft	concentration	Konzentration
affected	betroffen	conflict [n]	Konflikt
allowance	Wertberichtigungen	construction	Bau
all-time high	Höchststände	consumer	Verbraucher
anger	Ärger	consumption	Konsum
attempt [n]	Versuch	convinced	überzeugt
back	zurück	costs [n]	Kosten
bad	schlecht	crisis	Krise
bad	schlimm	decline [n]	Rückgang
bankrupt	Pleite	deep [adj.]	tief
bankruptcy	Insolvenz	deficiencies	Mangel
bankruptcy	Konkurs	deflation	Deflation
bankrupties	Pleiten	delayed	verzögert
bear market	Baisse	demand [n]	Nachfrage
benefit [v]	profitiert	depression	Depression
black	schwarz	disappointment	Enttäuschung
bonds	Renten	disappointments	Enttäuschungen
bonus	Bonus	disaster	Desaster
boom [n]	Boom	dismissal	Kündigung
boom [n]	Hochkonjunktur	diversification	Diversifikation
boom [n]	Konjunktur	dividend	Dividende
bull market	Hausse	doubt [n]	Zweifel
burden [v]	belasten	downswing [n]	Abschwung
burden [n]	Belastungen	duty	Pflicht
burdened	belastet	earning	Ertrag
calm	gelassen	emergency	Not
capital increase	Kapitalerhöhung	end [n]	Ende
capital increases	Kapitalerhöhungen	energy	Energie
challenge [n]	Herausforderung	error	Fehler
challenges	Herausforderungen	expansion	Expansion
chance	Chance	expectation	Erwartung
change [n]	Veränderung	expectations	Erwartungen
change [n]	Wandel	export [n]	Export
change [v]	Wechsel	fall [n]	Absturz
claim [v]	fordern	falling	fallende
claim [n]	Forderung	falls [v]	fällt
clear [adj.]	klar	falls [v]	sinkt
climax	Höhenflug	fear [n]	Angst
climax	Höhepunkt	fear [n]	Furcht
collapse [v]	stürzen	fears [n]	Ängste
collapsed	stürzte	fears [n]	Befürchtungen
commodities	Rohstoffe	fight [v]	kämpfen
competition	Wettbewerb	fluctuations	Schwankungen
concentrated	konzentriert	frustration	Frust

English	German	English	German
fun	Spass	new	neu
funds [n]	Fonds	new issues	Neuemissionen
fusion	Fusion	objectives	Ziele
future	Zukunft	oil	Öl
global	global	open [adj.]	offen
global economy	Weltwirtschaft	optimism	Optimismus
globalization	Globalisierung	optimistic	optimistisch
gold	Gold	outperform	übertroffen
good	gut	panic [n]	Panik
government	Regierung	perspective	Perspektive
growth	Wachstum	pessimism	Pessimismus
guarantee [n]	Garantie	pessimistic	pessimistisch
hard	hart	plan [n]	Plan
help	helfen	plus	plus
high	Hoch	positive	positiv
hope [n]	Hoffnung	prepare	gerüstet
impact [n]	Auswirkungen	prevent	verhindern
improved	verbessert	price drop	Preisverfall
improvement	Verbesserung	prices	Preise
inflation	Inflation	private	Private
information	Information	probability	Wahrscheinlichkeit
innovation	Innovation	problem	Problem
interest	Zinsen	problems	Probleme
interest rate increase	Zinserhöhung	profit [n]	Gewinn
interest rate reduction	Zinssenkung	progress	Entwicklung
inventory	Lager	prospects	Perspektiven
investigation	Untersuchung	protection	Schutz
investors	Investoren	purchase [n]	Kauf
joy	Freude	quality	Qualität
Keynes	Keynes	recession	Rezession
labor market	Arbeitsmarkt	record [n]	Rekord
laid off	entlassen	recovery	Aufschwung
layoff [n]	Kündigung	red	rot
long term	langfristig	reduced	reduziert
loses	verliert	redundancies	Entlassungen
loss	Verlust	reform [n]	Reform
lost	verlor	refuse	verweigert
machines	Maschinen	regulation	Regulierung
mass layoffs	Massenentlassungen	reorganization	Restrukturierung
mechanical engineering	Maschinenbau	rescue [v]	retten
media	Medien	research [n]	Forschung
medicine	Medizin	resources	Ressourcen
Merger	Übernahme	retracement	Einbruch
metal	Metall	return [n]	Rendite
minus	minus	rise [v]	steigt
missed	verfehlt	risk [n]	Risiko
missed	verfehlte	risky	riskant
monopoly	Monopol	safe [adj.]	sicher
negative	negativ	salaries	Gehälter

Table 1 – continued from previous page

English	German	English	German
salary	Gehalt	uncertainty	Unsicherheit
sales [n]	Verkauf	uncertainty	Verunsicherung
save	sparen	unclear	unklar
save	spart	unemployment	Arbeitslosigkeit
saved	gerettet	unload	entlasten
scenarios	Szenarien	verge	Rande
search [v]	suchen	warning	Warnung
search [v]	sucht	wave [n]	Welle
shaky	schwankend	weak	schwach
short [adj.]	knapp	weakness	Schwäche
short term	kurzfristig	welfare state	Sozialstaat
shortage	Engpass	worries [n]	Sorgen
shrank [v]	schrumpfte	worse	schlimmer
speculation	Spekulation	write offs	Abschreibungen
speculators	Spekulanten	young	jung
standstill	Stillstand		
start [n]	Anfang		
state [n]	Staat		
stock exchange	Börse		
stock market	Aktienmarkt		
stocks	Aktien		
strike [n]	Streik		
subsidies	Subventionen		
subvention	Subvention		
supply [n]	Angebot		
surplus	Überschuss		
surprise [n]	Überraschung		
survive	überleben		
tax	Steuern		
tax cut	Steuersenkungen		
technology	Technologie		
threat [v]	droht		
transport	Transport		
trend [n]	Trend		
turnover	Umsatz		

Table 1: This table reports all $N = 236$ words considered in this study. The words are sorted in alphabetical order and reported irrespective of any statistical significances. We systematically search all articles of the *Handelsblatt* for the occurrence of these words within our sample period. Since the *Handelsblatt* is a German newspaper, we looked for the German word in a first step and translated it into English word. For the English word we indicate the word class ([n] for noun, [v] for verb, [adj.] for adjective) in squared bracket if it is not apparent which class the word belongs to.

A.2 Tables and Figures
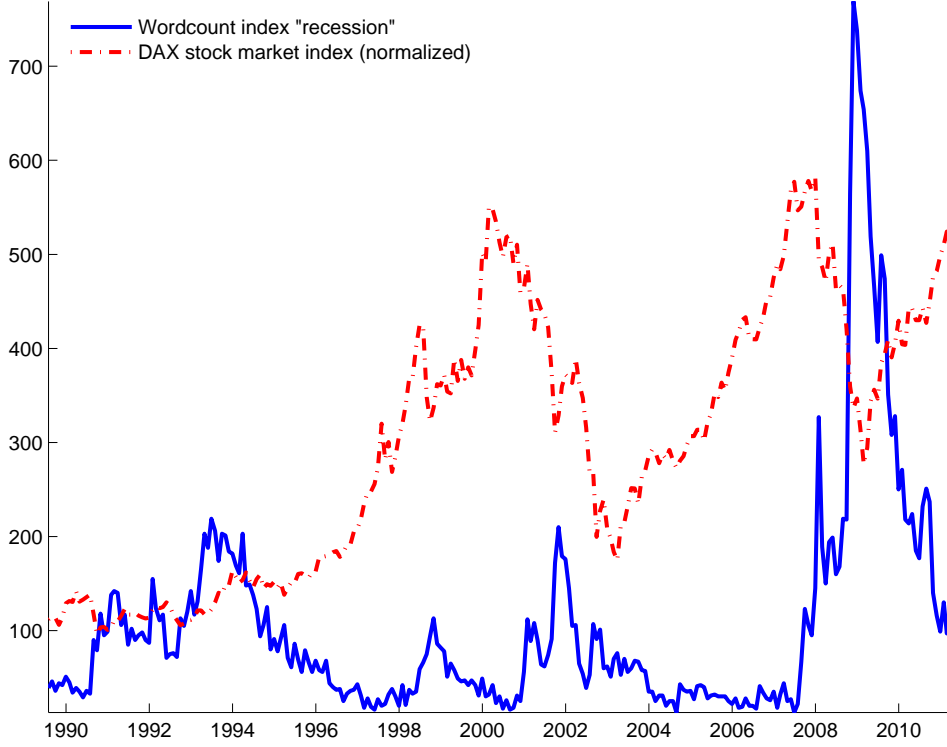
Figure 1: The Recession Index

This figure shows a time series plot of the monthly number of occurrences of the word "recession" in the *Handelsblatt* (solid line), and the normalized DAX index level (dot-dash line). During the entire sample period, i.e., July 1989 until March 2011, the correlation between the recession word-count, $x_{rec,t}$, and the DAX index level is $\rho(x_{rec,t}, \mathrm{DAXn}_{t+1}) = 0.03$. Although the correlation is close to zero, the plot suggests a relation between the recession word-count and the DAX index level. Occurrences of the word "recession" soared during the subprime crisis, indicating sort of a structural break. Allowing for changing market conditions, we compute the correlations for two subperiods. We find that the correlation based on a sample ending at the early beginning of the subprime period (January 2006) is -0.40. Considering the subprime crisis only (January 2006 until August 2010) results in a correlation of -0.71. By accounting for different market behavior our results confirm the negative relation between the recession word-count and the stock market development as claimed by The Economist (2002).

Table 2: Univariate Analysis—Correlations

The table shows correlation coefficients between word-count indices and future DAX excess returns, $\rho(z_{i,t}, r_{t+1})$, for the subperiods All, 1$^{\text{st}}$ Half (1H), 2$^{\text{nd}}$ Half (2H), Boom, Dotcom (Dotc.), and Subprime (Subp.). We report correlations statistically significant at a level of 2.5%. * indicates significance at 1%. For the English word we indicate the word class ([n] for noun, [v] for verb, [adj.] for adjective) in squared bracket if it is not apparent which class the word belongs to.

| | | All | 1H | 2H | Boom | Dotc. | Subp. |
| | | Jul89-Mar11 | Jul89-Jun99 | Jul99-Mar11 | Dec93-Aug98 | Jul99-Mar04 | Jan06-Aug10 |
English	German						
attempt [n]	Versuch	0.142				0.311	
bad	schlimm						−0.297
black	schwarz	0.160*	0.209				
bonds	Renten	0.168*					
boom [n]	Hochkonjunktur				0.319		
burden [n]	Belastungen						−0.304
capital increase	Kapitalerhöhung					0.306	−0.301
challenge [n]	Herausforderung	0.139				0.335	
chance	Chance					0.327	
change [n]	Wandel	0.142		0.192		0.350*	
claim [n]	Forderung						−0.317
claim [v]	fordern	0.205*		0.248*		0.440*	
clear [adj.]	klar					0.394*	
competition	Wettbewerb	0.150*		0.213		0.330	
concentrated	konzentriert					0.298	
consumption	Konsum	−0.137					
convinced	überzeugt	0.166*		0.206		0.360*	
disaster	Desaster						−0.308
downswing	Abschwung						−0.330*
expansion	Expansion	0.135		0.216			
export [n]	Export						−0.299
falling [adj.]	fallende						−0.322
fear [n]	Furcht					0.378*	
future	Zukunft	0.162*		0.206		0.320	
globalization	Globalisierung	0.185*				0.310	
guarantee [n]	Garantie					0.393*	−0.311
help [v]	helfen					0.316	
impact	Auswirkungen					0.312	
improvement	Verbesserung					0.326	
metal	Metall	0.138		0.190			
minus	minus	−0.163*					
negative	negativ				0.387*		
objectives	Ziele	0.160*					
oil	Öl						−0.315
open [adj.]	offen	0.144				0.344*	
pessimism	Pessimismus						−0.340
pessimistic	pessimistisch						−0.328
prepared	gerüstet					0.310	
private	privat	0.137		0.191			
probability	Wahrscheinlichkeit					0.366*	−0.303
protection	Schutz					0.337	
reform [n]	Reform	0.153*					
regulation	Regulierung	0.176*				0.313	
safe [adj.]	sicher					0.351*	
search [v]	suchen	0.138				0.382*	
shaky	schwankend	−0.141					
stock market	Aktienmarkt					0.317	
supply [n]	Angebot	0.181*		0.205			
trough	Tief						−0.297
uncertainty	Verunsicherung						−0.302
unclear	unklar					0.300	
unload	entlasten					0.341*	
verge [n]	Rande					0.354*	
warning	Warnung	0.141					
worse	schlimmer	−0.141		−0.226*			−0.401*
young	jung	0.160*	0.217				

Table 3: Stepwise Regression—Future DAX Excess Returns

The table shows the regression coefficients of a stepwise regression of the future DAX excess returns, r_{t+1}, on all standardized word-count indices, $z_{i,t}$. The stepwise algorithm is based on an entrance and exit threshold of 2.5%. The regression equation is $r_{t+1} = \alpha + \sum_{j=1}^{J} \beta_j z_{j,t} + \gamma C_t + \varepsilon_t$, where α is the intercept and β_j denotes the exposures of r_{t+1} to the J word-count indices, z_t, at time t. C_t includes the control variables. We run the regression for the periods All, 1st Half (1H), 2nd Half (2H), Boom, Dotcom (Dotc.), and Subprime (Subp.). For the English word we indicate the word class ([n] for noun, [v] for verb, [adj.] for adjective) in squared bracket if it is not apparent which class the word belongs to. All coefficients reported in this table are statistically significant at a level of 1%.

English	German	All Jul89-Mar11	1H Jul89-Jun99	2H Jul99-Mar11	Boom Dec93-Aug98	Dotc. Jul99-Mar04	Subp. Jan06-Aug10
		Dependent variable: r_{t+1} (DAX excess log return)					
α (intercept)		0.47	0.68	−0.28	1.88	−0.31	4.73
anger [n]	Ärger	−1.07		−2.09		−2.20	
challenges [n]	Herausforderungen						−3.79
claim [v]	fordern	2.14		2.29		3.76	
climax	Höhepunkt						−2.87
collapse [v]	stürzen		1.71				
consumption	Konsum	−1.00				−3.04	
fear [n]	Furcht						1.80
interest [n]	Zinsen	0.98					
merger	Übernahme			−3.42			
minus	minus	−1.47		−4.39			
negative	negativ			1.73			
oil	Öl		−1.41				
optimism	Optimismus			1.22			
plus	plus		−1.51				
private	privat			2.52			
profit [n]	Gewinn			2.60			
recovery	Aufschwung		−1.23				
saves	spart						4.83
supply [n]	Angebot		1.80				4.19
technology	Technologie			−1.82			
transport [n]	Transport			2.26			
worse	schlimmer						−2.81

Table 4: Stepwise Regression—Real Economy

The table shows the regression coefficients of a stepwise regression of future percentage changes in Industrial Production, r_{t+1}^{IP}, on all standardized word-count indices, $z_{i,t}$. The stepwise algorithm is based on an entrance and exit threshold of 2.5%. The regression equation is $r_{t+1}^{\text{IP}} = \alpha + \sum_{j=1}^{J} \beta_j z_{j,t} + \gamma C_t + \varepsilon_t$, where α is the intercept and β_j denotes the exposures of r_{t+1}^{IP} to the J word-count indices, z_t, at time t. C_t includes the control variables. We run the regression for the periods All, 1st Half (1H), 2nd Half (2H), Boom, Dotcom (Dotc.), and Subprime (Subp.). For the English word we indicate the word class ([n] for noun, [v] for verb, [adj.] for adjective) in squared bracket if it is not apparent which class the word belongs to. All coefficients reported in this table are statistically significant at a level of 1%.

		Dependent variable: r_{t+1}^{IP} (Relative changes in Industrial Prod.)					
		All	**1H**	**2H**	**Boom**	**Dotc.**	**Subp.**
		Jul89-Mar11	Jul89-Jun99	Jul99-Mar11	Dec93-Aug98	Jul99-Mar04	Jan06-Aug10
English	**German**						
α (intercept)		0.09	0.03	−0.40	−0.05	0.09	−0.24
bad	schlimm				0.52		1.58
boom [n]	Hochkonjunktur				−0.40		
consumption	Konsum			−0.39			
decline [n]	Rückgang	−0.26					
deflation	Deflation	−0.23		−0.31			
export [n]	Export						−0.42
falls	fällt						−0.87
frustration	Frust				0.45		
global economy	Weltwirtschaft		−0.33				
government	Regierung				−0.95		
guarantee [n]	Garantie				0.63		
laid off	entlassen						−0.76
new issues	Neuemissionen		0.42				
outperform	übertroffen						0.84
pessimism	Pessimismus						−0.60
price drop	Preisverfall						−0.99
record [n]	Rekord						−0.39
save	sparen				0.83		
tax cut	Steuersenkungen				−0.88		
threats [v]	droht				0.45		
trough	Tief				−0.33		

Table 5: Out-of-sample Analysis—Different Rolling Windows

The table presents the results of the out-of-sample analysis based on a rolling window stepwise regression approach as defined in Equations (3) and (4). We report the adjusted R^2, the regression coefficients v (intercept) and w (slope) with their t-statistics t_v and t_w, respectively. The results are indicated for different rolling window sizes denoted by k months. Specifically, we regress the realized, r_t, on the predicted, $\hat{r}_t$, DAX excess returns. The regression equation is defined as $r_t = v + w \cdot \hat{r}_t$. The scatter plot including the linear regression line for a window size of $k = 60$ months is shown in Figure 3. * and ** denote statistical significance at a level of 10% and 5%, respectively.

Window size (k)	adj. R^2	v	t_v	w	t_w
54	1.78%	0.259%	0.564	0.148*	1.913
60	1.27%	0.224%	0.471	0.128	1.580
66	1.25%	0.243%	0.500	0.151	1.554
72	1.21%	0.282%	0.569	0.134	1.501
78	1.82%	0.277%	0.547	0.175*	1.818
84	1.33%	0.279%	0.536	0.163	1.510
90	2.39%	0.092%	0.170	0.226**	2.000
96	2.33%	0.006%	0.011	0.261**	2.326

Table 6: Stepwise Regression—DAX 3M Forward Excess Returns

The table shows the regression coefficients of a stepwise regression of the DAX 3 months forward excess returns, r_{t+1}^{3M}, on the 3-month averages of all standardized word-count indices, $z_{i,t}^{3M}$. The stepwise algorithm is based on an entrance and exit threshold of 2.5%. The regression equation is $r_{t+1}^{3M} = \alpha + \sum_{j=1}^{J} \beta_j z_{j,t}^{3M} + \gamma C_t^{3M} + \varepsilon_t$, where α is the intercept and β_j denotes the exposures of r_{t+1}^{3M} to the J word-count indices z_t at time t. C_t^{3M} includes the control variables. We run the regression for the periods All, 1st Half (1H), 2nd Half (2H), Boom, Dotcom (Dotc.), and Subprime (Subp.). For the English word we indicate the word class ([n] for noun, [v] for verb, [adj.] for adjective) in squared bracket if it is not apparent which class the word belongs to. All coefficients in this table are statistically significant at a level of 1%.

		Dependent variable: r_{t+1}^{3M} (DAX 3M excess log return)					
		All	**1H**	**2H**	**Boom**	**Dotc.**	**Subp.**
		Jul89-	Jul89-	Jul99-	Dec93-	Jul99-	Jan06-
English	**German**	Mar11	Jun99	Mar11	Aug98	Mar04	Aug10
α (intercept)		4.00	−0.83	−0.13	5.83	0.43	−3.46
bankrupties [n]	Pleiten	−2.95	−4.02	−4.18		−7.38	
bonus	Bonus			−4.92	−6.37		
boom [n]	Hochkonjunktur						5.50
challenges [n]	Herausforderungen					7.33	
collapsed	stürzte		1.56				
commodities	Rohstoffe	−2.47					
consumer	Verbraucher					−5.97	
deflation	Deflation		1.52				
delayed	verzögert		−2.49				
depression	Depression						3.01
diversification	Diversifikation					−6.63	
downswing	Abschwung	2.64					
error	Fehler	−3.45		−4.22			
expectation	Erwartung		−3.43	6.01			
falls [v]	sinkt			8.02			
frustration	Frust	−2.20			−5.03		
fun	Spass						−8.23
inflation	Inflation	1.79					
innovation	Innovation	−1.28		−2.22			−4.83
interest [n]	Zinsen			−5.35			
investigation	Untersuchung			−4.07			
joy	Freude	3.74	4.23				
layoff	Kündigungen		3.61				
lost	verlor					12.28	
missed	verfehlte		1.92				
panic [n]	Panik				3.74		
problem	Problem	2.59					
salaries	Gehälter	−3.16				−7.50	
saved	gerettet						−7.60
speculators	Spekulanten					−4.93	
subsidies	Subventionen	−2.84					
surplus	Überschuss			8.33			
tax	Steuern	2.01					
trough	Tief	1.58					
wave [n]	Welle			−4.35			
write offs	Abschreibungen	−4.32		−4.36			

Figure 2: Heat Map of Words Explaining Future DAX Movements

The figure depicts a heat map of words possessing significant explanatory power for future DAX returns. We select words by a stepwise regression algorithm based on rolling windows. The regression is based on the standard window size of 60 months. The entrance and exit threshold of the algorithm is 2.5%. The x-axis indicates the rolling windows. The y-axis shows significant words. For each window, the stepwise regression is run once. If a particular word enters the regression by a significant factor loading, it is indicated by a dark spot in the heat map. To ensure legibility, we only report words that are selected in nine or more rolling windows.
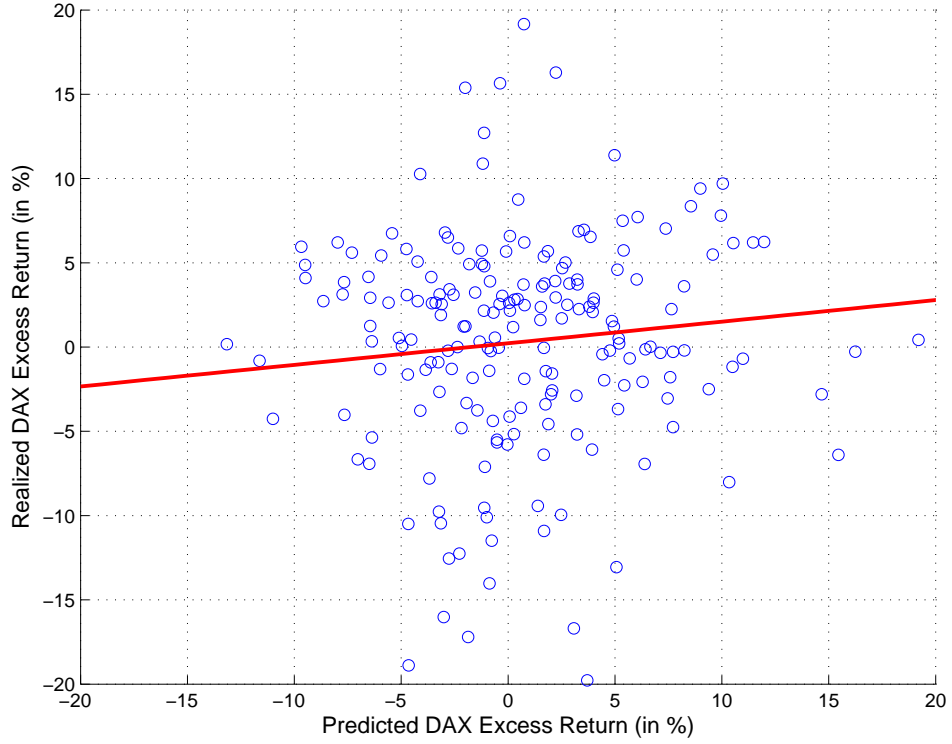
Figure 3: Realized vs. Predicted DAX Excess Returns

The figure depicts a scatter plot of realized, r, and predicted, $\hat{r}$, DAX excess log returns. The predicted DAX excess returns are based on a rolling window of $k = 60$ months. They are computed based on a stepwise regression according to the Equations (3) and (4). The regression line is defined according to the following regression equation: $r_t = v + w \cdot \hat{r}_t$. The estimated regression coefficients are $v = 0.224$ with a standard error of $se(v) = 0.476$ and $w = 0.128$ with a $se(w) = 0.081$. The adjusted R^2 of the regression is 1.27%.
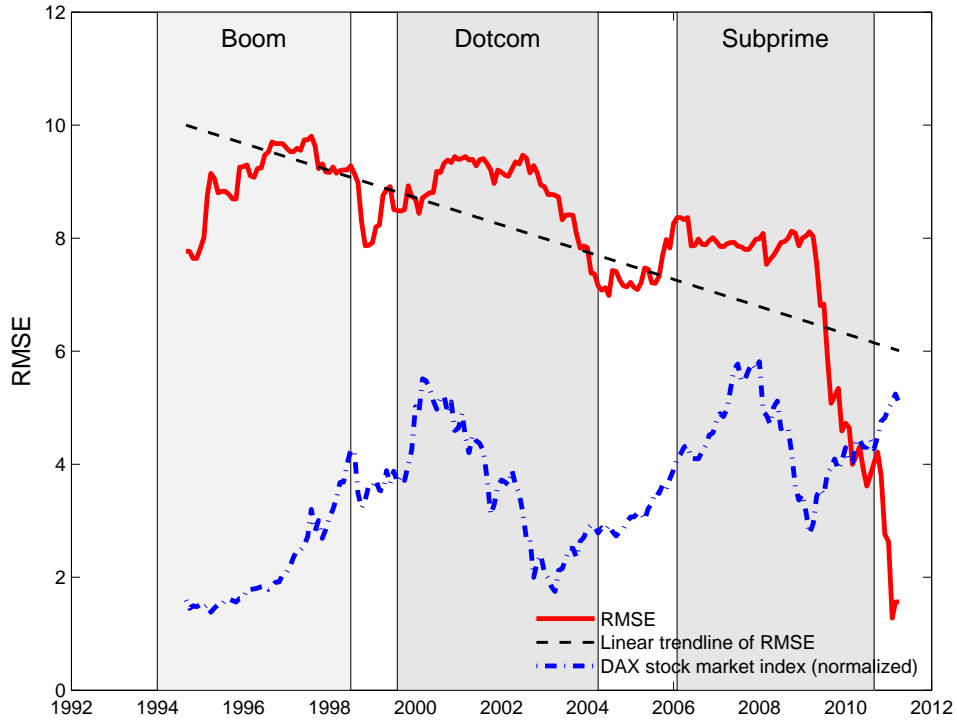
Figure 4: Realized vs. Predicted Returns—RMSE

The figure plots the root mean squared errors (RMSE) (solid line) for each window of the rolling window stepwise regression of Equations (3) and (4). By focusing on the difference between forecasted and realized DAX returns in different time windows, it measures the behavior of the forecasting performance over time. The dashed line depicts a linear trendline of the RMSE. To compare the development of the forecasting accuracy with current market situations, we additionally plot the normalized DAX index level (dot-dash line). The RMSE are computed according to Equation (6). The shaded areas indicate the time subperiods Boom (1993/12–1998/8), Dotcom (1999/7–2004/3) and Subprime (2006/1–2010/8).
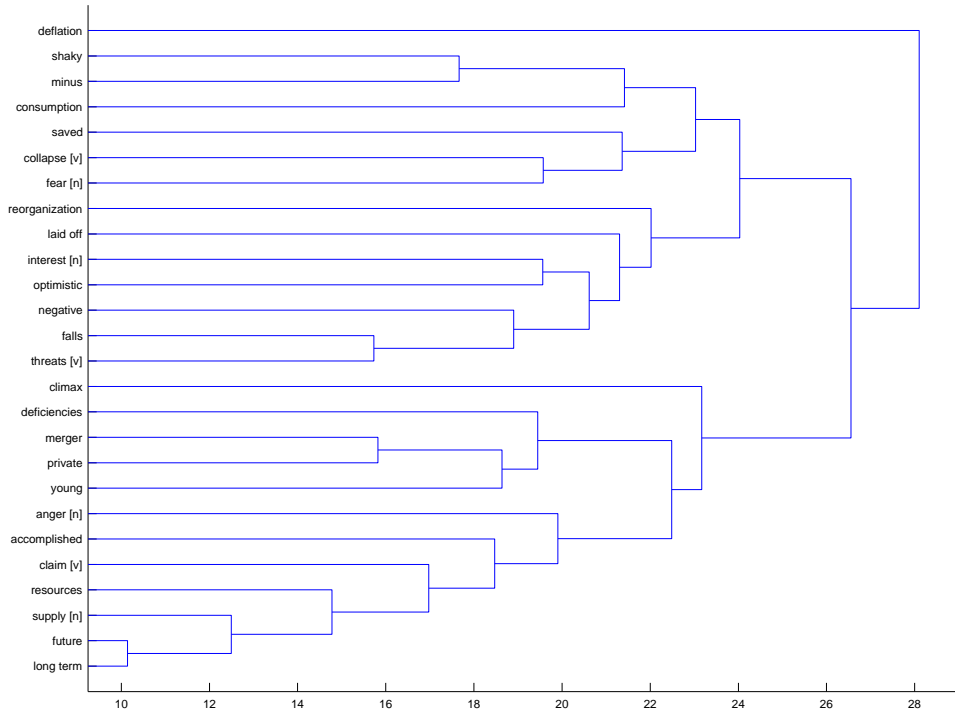
Figure 5: Dendrogram of Clusters of Word-count Indices

This figure shows a hierarchical cluster tree, i.e., a dendrogram, of all words found to be significant by the stepwise algorithm. The horizontal axis represents the Euclidean distance between the clusters. We select the entrance and exit threshold of the stepwise regression approach to be 10%. At each step of the tree, the clustering algorithm identifies the two clusters closest to each other and eventually combines them in a new cluster. This procedure is followed until all words and clusters are eventually linked, respectively.